ISSN 1798-4769 (print); ISSN 2053-0684 (online)
Journal of Language Teaching and Research, Vol. 17, No. 2, pp. 634-642, March 2026
DOI: https://doi.org/10.17507/jltr.1702.26

Exploring AI-Driven Written English Assessment: Toward Improved Assessment Quality and Learner Outcomes

Ahmed Alshehri
Department of English Language & Literature, College of Arts and Letters, University of Bisha, Saudi Arabia

Saddah Aldossary
English Department, College of Science and Humanities, Shaqra University, Saudi Arabia

Mohammad Jamshed*
English Department, College of Science & Humanities, Prince Sattam Bin Abdulaziz University, Saudi Arabia

Mohammad Rezaul Karim
English Department, College of Science & Humanities, Prince Sattam Bin Abdulaziz University, Saudi Arabia

***Abstract*—The use of AI-powered tools for assessing written English is revolutionizing conventional evaluation methods by improving accuracy, consistency, and teaching efficiency. However, there is limited research on how AI assessment tools affect the validity and fairness of language evaluation. This review synthesizes current research on the technological foundations, pedagogical impact, and ethical considerations of AI in this area. This review assesses the role of AI in education, highlighting AI-driven technologies such as automated writing evaluation and natural language processing that provide timely feedback, support personalized learning, and reduce instructor workload. The study employed the PRISMA method to investigate the impact of AI on written English assessment by analyzing empirical studies from Scopus, Web of Science, and Google Scholar. The findings revealed that AI tools significantly enhanced the evaluation of written English by providing reliable assessments, immediate feedback, fostering learner autonomy, and producing assessments consistent with those of human raters, thereby promoting self-regulated writing and increasing access to quality feedback. However, specific concerns about algorithmic bias, feedback clarity, and the diminishing role of human judgment in assessing creativity and discourse were also found. It was also found that educational automation negatively impacts linguistic diversity, critical thinking, data privacy, transparency, and equitable access to AI tools. The review suggests integrating AI capabilities with human experience to balance technological efficiency and pedagogical integrity in the use of AI tools. It recommends developing hybrid assessment frameworks that integrate AI analytics with human evaluation to enhance fairness, accuracy, and comprehension in written English assessment.**

***Index Terms*—contextual sensitivity, discourse competence, technological efficiency, pedagogical integrity, validity and fairness**

I. Introduction

In today's globalized world, adequate English proficiency is crucial, especially for EFL/ESL learners who take proficiency exams. Traditional assessments face criticism for being subjective and inefficient, requiring skilled language assessors and complicating the evaluation process (Poonpon et al., 2023). As educational systems increasingly integrate digital technologies, AI-driven assessment tools, such as automated essay scoring engines and natural language processing systems, have emerged as effective alternatives to traditional evaluation methods. These tools enhance the reliability, accuracy, and efficiency of writing assessments by swiftly analyzing linguistic features and content quality, which surpasses the speed of manual evaluations. Today, these technologies, especially AI, are transforming assessment methods in higher education by enabling innovative tools that can score tasks, provide detailed feedback, and monitor student progress continuously (Deepshikha, 2025; Tapalova & Zhiyenbayeva, 2022).

AI-based writing evaluation systems have become vital for English language instructors and learners alike (Alharbi, 2023; Wang, 2022; Shi & Aryadoust, 2024; Al-Obaydi & Pikhart, 2025; Lee, 2024). Grammarly, Criterion, ProWritingAid, and Write & Improve utilize NLP algorithms to provide real-time grammar, coherence, and stylistic feedback. AI-powered tools now assist with diagnostic evaluation, formative assessment, and personalized feedback, as well as automated correction. These technologies allow instructors to manage large-scale assessments while enhancing

* Corresponding Author.

students' writing with immediate, data-driven insights (Li et al., 2023). AI-powered academic feedback aligns with the global shift toward online instruction, where students benefit from algorithmic feedback to enhance language accuracy and pass standardized proficiency tests (Zhao et al., 2023). Thus, AI has become a vital pedagogical tool in writing education, shaping the measurement and development of writing skills.

The integration of AI-powered tools, especially massive language models like ChatGPT, has improved automated educational evaluation by offering context-aware feedback that goes beyond traditional scoring. Generative Artificial Intelligence (GenAI) is transforming higher education by influencing assessment practices and teaching methodologies, facilitating personalized learning experiences, and enhancing assessment efficiency (Bearman & Luckin, 2020; Chan & Hu, 2023; Tam, 2024; Hooda et al., 2022). These AI systems effectively identify error patterns, recommend specific improvements, and offer continuous writing support, ultimately fostering a more personalized learning environment. This AI-driven approach enhances education by focusing on data-informed practices, personalized learning, and essential 21st-century communication skills. It alleviates explicit performance anxiety associated with teacher assessments for EFL/ESL students by offering immediate, constructive feedback that increases motivation and academic performance, ultimately improving the overall educational experience.

A. Research Problem

Large language models, such as ChatGPT, enhance this by offering context-sensitive feedback, identifying learner errors, and promoting personalized writing improvements. This progression supports global educational goals focused on data-driven decisions and personalized learning, particularly aiding EFL and ESL learners. However, integrating AI raises concerns about validity and fairness, necessitating careful evaluation of the associated challenges and opportunities. This review paper analyzes the evolving AI-driven assessment landscape and its impact on quality and learner outcomes in various educational contexts.

B. Research Purpose Statement

This review synthesizes literature on AI-assisted evaluation in written English, emphasizing its effects on assessment methods, feedback quality, and learner engagement. It calls for empirical studies to explore the strengths and challenges of AI in writing assessment and outlines future research directions regarding its educational, ethical, and psychometric implications.

II. Literature Review

Artificial intelligence (AI) integration into EFL classrooms is gaining prominence because it can enhance assessment quality and learner outcomes. Recent studies show how AI impacts language development and assessment. Beyond basic corrections, artificial intelligence (AI) facilitates learning by promoting formative and summative evaluations. The long-term implications of AI in assessment include improving tools to enhance students' attitudes, motivation, and writing skills. AI focuses on increasing the validity and reliability of evaluations by predicting learning outcomes and personalizing assessments, while providing real-time monitoring and feedback that support immediate interventions to improve learner outcomes through tailored instruction (Biju et al., 2024).

Multiple studies have investigated the efficacy of AI-powered tools for assessing and evaluating written texts. For example, Deepshikha (2025) reviewed 77 studies on AI-driven grading and feedback systems, highlighting their role in addressing traditional grading inefficiencies and biases through machine learning and natural language processing. While these systems offered enhanced automation and personalized feedback, issues such as algorithmic bias, data privacy, and the need for human oversight persisted. The study advocates for a balanced integration of AI in education to uphold ethical standards and achieve equitable outcomes. Poonpon et al. (2023) examined the adoption of automated essay scoring (AES) for non-native speakers, focusing on language proficiency and cultural contexts. Using data from Khon Kaen University's English test and long short-term memory networks, the study revealed significant improvements in essay scoring, with results comparable to those of leading deep learning models. The findings hold immense pedagogical implications for enhancing educational assessments and fostering equitable opportunities for global language learners. Usher (2025) conducted a study involving 76 undergraduate students to assess the efficacy of AI chatbots compared to peer and instructor evaluations during a group project. Results indicated that AI chatbots provided higher grades and more comprehensive feedback than peers, although some responses were found to be irrelevant or contradictory.

In contrast, peer feedback, while less extensive, offered a more personalized touch. These findings suggest that combining chatbot assessments with traditional evaluation methods may enhance student learning by drawing on the strengths of both approaches. Chen (2025) explored how deep learning enhanced English language education by overcoming traditional assessment limitations, such as their slowness and subjectivity. A mixed methods approach combined perspectives from educators and students with performance metrics. The results revealed that AI improved accuracy, reduced grading time, and increased student engagement, while ethically addressing cultural bias and privacy concerns, underscoring deep learning's potential for fair and effective language learning experiences.

Similarly, Khasawneh (2024) studied the effects of AI-driven assessment and automated feedback on 529 undergraduates' foreign language proficiency in Jordan. Results showed that 53.65% of students valued timely feedback

for identifying strengths and weaknesses, while 37.57% had concerns about its impact on confidence. Notably, 90.12% felt personalized feedback helped with goal-setting, and 86.58% used automated feedback to correct language errors. The research underscores both the pros and cons of these systems, indicating that tailored feedback significantly improves the learning experience, as reflected in a mean motivational score of 4.59. Kanchana and Saha (2025) explored AI-driven rubrics designed for personalized and equitable assessments for engineering undergraduates learning English. Involving 64 first-year B. Tech students, the mixed-method study found that while the AI-generated rubrics were perceived as transparent and fair, there were concerns about the quality of feedback and AI decision-making. The study offers recommendations to improve rubrics for more inclusive, individualized evaluations in English instruction. Chen et al. (2024) examined the design and effects of AI-enabled assessment tools in language education. A systematic review and meta-analysis of 25 studies found that structural AI architecture is mainly used in upper primary classrooms for short interventions, showing a medium effect size and significant positive impacts on language learning outcomes. The study emphasizes that effective implementation of AI-enabled assessments can enhance language education and suggests further research into instructional designs. A study by Jin and Fan (2023) revealed that integrating AI technology into language assessment significantly improved task design, delivery, scoring, and feedback, while enhancing data collection and analysis for validation. It examined both the benefits and challenges for test-takers and proposed a conceptual model distinguishing assessment participation from technology perceptions. The analysis identified future research directions and strategies to increase test-taker engagement and promote equitable assessment practices. Khan et al. (2024) examined the effects of AI-assisted assessment on vocabulary knowledge and emotional resilience in 60 Afghan junior high students aged 13-14. Using a pretest-posttest design, results indicated significant improvements in vocabulary and resilience, although teacher support did not influence outcomes, unlike the positive trend seen with parental support. The research calls for collaboration between educators, parents, and policymakers to incorporate AI tools into education and enhance cognitive and emotional development.

In summary, AI-driven written English assessment can enhance both assessment quality and learner outcomes by improving writing accuracy, motivation, and self-regulation. These tools support personalized evaluation strategies, highlighting the importance of ongoing AI development and integration in educational practices.

III. Methodology

This systematic review adhered to PRISMA guidelines to investigate the impact of Artificial Intelligence (AI) on written English assessment. A thorough search was conducted via Scopus, Web of Science, and Google Scholar. Only empirical or experimental studies on AI-assisted writing evaluation in ESL/EFL contexts were included. Although there are few empirical studies, highlighting this research is crucial for establishing a robust evidence base regarding AI's effect on writing assessment practices.

A. Inclusion Criteria

1. Empirical or experimental studies examining AI-assisted assessment of written English in ESL/EFL contexts.
2. Peer-reviewed articles published in reputable international journals between 2020 and 2025.
3. Studies available in full text and written in English.
4. Research utilizing quantitative/qualitative/mixed-methods to assess the impact of AI on writing accuracy, feedback, and overall performance.

B. Exclusion Criteria

1. Theoretical or conceptual papers without empirical data.
2. Studies focusing on speech, reading, or other language skills rather than writing.
3. Non-peer-reviewed papers, conference abstracts, or book chapters.
4. Articles not accessible in full text or not written in English.
5. The search string used in both databases was: (“Artificial Intelligence” OR “AI” OR “Automated Writing Evaluation” OR “AWE”) AND (“English writing” OR “writing assessment” OR “academic writing”) AND (“feedback” OR “evaluation” OR “language learning”).

This search identified 232 Web of Science and 187 Scopus studies. Eight studies were chosen for final analysis after applying the inclusion and exclusion criteria. The following significant works were incorporated: Zhao et al. (2023), Jiang et al. (2023), Li et al. (2023), Jamshed et al. (2024, 2025), Cheng et al. (2024), Al Ghaithi and Behforouz (2025), Mahdi and Alkhateeb (2025), and Rahman et al. (2023). All selected studies explored AI in writing assessment, focusing on feedback production, learner engagement, evaluation reliability, and pedagogical adaptation.

The adopted review model for this study was thematic analysis, which facilitated the identification and synthesis of themes within the selected literature. Two researchers independently reviewed all eligible articles, coding them into categories such as pedagogical effectiveness, feedback quality, ethical considerations, and learner autonomy. Discrepancies in coding were resolved by consensus to maintain inter-rater reliability. Despite the limited number of studies, this systematic review aims to provide a comprehensive synthesis of the empirical evidence on the role of AI in written English assessment. The scarcity of experimental research indicates that this field is still emerging, highlighting

the pressing need for additional empirical studies that evaluate not only AI's technological capabilities but also its implications for pedagogy, ethics, and psychometrics in language assessment.

IV. Results

This systematic review synthesized eight empirical studies published from 2022 to 2025 that examined the integration of Artificial Intelligence (AI) into written English assessment. The findings revealed that AI had a multidimensional impact on writing assessment, affecting scoring reliability, feedback quality, learner engagement, pedagogical alignment, and ethical considerations. These findings are organized into six interrelated thematic domains that highlight both the pedagogical potential and the challenges associated with AI-assisted evaluation.

A. Enhancing Scoring Reliability and Objectivity

AI-based writing evaluation tools demonstrated greater reliability and objectivity compared to conventional human scoring methods. For example, Jiang et al. (2023) evaluated AI models, including GPT-4. They found a strong correlation (over 0.90) between AI scores and human evaluations of grammar, syntax, and lexical accuracy, underscoring the reliability of large language models in this field. Zhao et al. (2023) further investigated automated scoring for picture-cued writing tasks and found that AI scoring decreased rater subjectivity and improved consistency across essays. Their research showed that using AI in evaluations enhanced assessment turnaround time without compromising accuracy. Collectively, these findings highlight AI's potential to significantly enhance large-scale writing assessment frameworks through greater consistency and efficiency. Table 1 summarizes significant reliability findings and shows how AI systems enhance assessment across several educational settings.

Table 1
AI Models and Their Contribution to Scoring Reliability

Study	AI System / Model	Context	Key Findings
Jiang et al. (2023)	GPT-4, GPT-3.5, Baidu Cloud, iFLYTEK	Chinese L2 writing	GPT-4 achieved 0.90 reliability with human raters; substantial grammatical precision
Zhao et al. (2023)	AI-Assisted Scoring (Picture-Cued)	EFL learners	Enhanced inter-rater reliability and reduced time bias
Li et al. (2023)	AWE + Peer Review	University writing courses	Balanced objectivity with pedagogical relevance; higher fairness

These findings indicate that AI reduces human error and enhances evaluation consistency, though factors such as argument structure and tone still require human interpretation for appropriate judgment.

B. Validity and Construct Representation in Writing Evaluation

The reviewed studies emphasize the limitations of AI tools in assessments, particularly regarding construct validity. Cheng et al. (2024) noted that AI assessment models often focus on surface linguistic accuracy, overlooking critical writing elements such as creativity, coherence, and discourse organization, where AI's abilities are insufficient. Conversely, Mahdi and Alkhateeb (2025) introduced an AI-driven rubric designed to enhance validity through criterion-based evaluations that correspond with communicative competence and rhetorical effectiveness. Their findings indicated that AI could mimic human judgment when applied to context-rich data; however, reliance on uniform rubrics may lead to standardized writing that hampers originality. While AI contributes to increased reliability, it may compromise validity, especially concerning creativity and higher-order thinking. To address these issues, a hybrid assessment model is recommended that integrates AI's quantitative assessment with human evaluation of qualitative factors.

C. Impact of AI Feedback on Writing Development

A key finding from recent reviews suggests that AI-assisted feedback markedly improves learners' linguistic and metacognitive development. Research by Jamshed et al. (2024) demonstrates that feedback generated by ChatGPT significantly enhances English learners' grammatical accuracy, sentence complexity, and lexical diversity, resulting in statistically significant improvements compared to traditional teacher comments. Furthermore, Jamshed et al. (2025) highlight that AI-driven corrective feedback delivered via WhatsApp encourages learners to make iterative revisions. The immediate, accessible nature of mobile feedback promotes self-regulated learning, particularly in environments with limited teacher presence, thereby increasing learner satisfaction through timely responses and clear grammatical explanations. An overview of the pedagogical outcomes of various AI feedback platforms, underscoring their contribution to learner development, is presented in Table 2.

TABLE 2
EMPIRICAL FINDINGS ON AI FEEDBACK AND WRITING IMPROVEMENT

Study	Feedback Platform	Learning Outcome	Pedagogical Insight
Jamshed et al. (2024)	ChatGPT	Enhanced grammar, fluency, and lexical accuracy	AI feedback fosters self-editing and reflection
Jamshed et al. (2025)	WhatsApp (AI-driven feedback)	Improved accuracy and revision quality	Mobile tools sustain continuous engagement
Al Ghaithi & Behforouz (2025)	AI Writing Evaluation System	Reduced anxiety; higher confidence	AI perceived as non-judgmental support

D. Pedagogical Alignment and Learner Engagement

AI-driven writing evaluation systems are reshaping classroom dynamics and enhancing student engagement. Research by Li et al. (2023) emphasizes the advantages of integrating Automated Writing Evaluation (AWE) with peer review in large university writing classes, thereby promoting collaborative learning and self-assessment, which, in turn, deepen analytical awareness and improve revision practices. Al Ghaithi and Behforouz (2025) further demonstrate that AI tools boost motivation by providing consistent and immediate feedback. Students regard AI systems as "available tutors," resulting in more frequent writing attempts and fostering a culture of iterative practice. However, Abdul Rahman et al. (2023) warn of the potential for "feedback dependency," in which students may lean too heavily on machine validation, thereby compromising their critical thinking and classroom engagement. Overall, the research indicates that AI can enhance student engagement when used in conjunction with guided instruction, with a focus on AI as a facilitator of reflective learning rather than a mere correctness evaluator.

E. Psychological and Cognitive Dimensions of AI-Based Assessment

AI assessments influence learners' psychology significantly beyond mere skill enhancement. Research by Al Ghaithi and Behforouz (2025) reveals that continuous AI feedback can reduce writing anxiety among less skilled learners, thereby boosting their confidence and encouraging them to engage more freely with language. However, Li et al. (2023) suggest that advanced learners may experience digital fatigue and stress due to frequent corrective feedback. Cheng et al. (2024) emphasize that effective AI integration hinges on appropriate cognitive pacing and literacy concerning feedback. They advocate training learners to evaluate AI suggestions critically rather than accepting them unquestioningly. Teachers play a vital role in this process, supporting students to maintain their independence while benefiting from AI-assisted evaluations. These findings highlight that psychological outcomes vary based on learner profile, task complexity, and the pedagogical context in which AI is deployed as illustrated in Table 3.

TABLE 3
PSYCHOLOGICAL AND COGNITIVE EFFECTS OF AI-ASSISTED WRITING ASSESSMENT

Study	Psychological Finding	Pedagogical Implication
Al Ghaithi and Behforouz (2025)	Reduced writing anxiety; increased risk-taking	AI feedback can scaffold confidence for weaker learners
Li et al. (2023)	Increased technostress among advanced learners	Over-correction may lead to performance anxiety
Cheng et al. (2024)	Need for feedback, literacy, and reflective interpretation	Training required to promote critical engagement

F. Ethical, Transparency, and Data Privacy Concerns

AI-based assessments face significant ethical and methodological challenges, particularly regarding algorithmic bias, which often reflects Western linguistic norms and disadvantages writers from diverse backgrounds, thereby undermining evaluation fairness. Transparency and data security are critical issues; Mahdi and Alkhateeb (2025) describe AI scoring systems as "black-box" mechanisms lacking transparent decision-making processes. They recommend implementing explainable AI frameworks and ethical rubrics to enhance accountability. Additionally, Zhao et al. (2023) and Li et al. (2023) emphasize that numerous commercial AI platforms frequently retain and reuse student data, raising serious concerns regarding consent and intellectual property rights.

TABLE 4
ETHICAL AND METHODOLOGICAL CHALLENGES IN AI-BASED WRITING EVALUATION

Study	Challenge Identified	Proposed Solution
Cheng et al. (2024)	Algorithmic bias	Use culturally diverse training datasets
Mahdi and Alkhateeb (2025)	Scoring opacity	Develop explainable AI rubrics
Zhao et al. (2023)	Data storage and privacy issues	Enforce anonymization and consent policies
Li et al. (2023)	Ownership of learner-generated texts	Define institutional data governance
Rahman et al. (2023)	Limited construct validity	Combine AI analytics with human interpretation

Ethical accountability is crucial to the effective integration of artificial intelligence (AI) into writing assessments, ensuring credibility through transparent algorithms and secure data management. AI assessments enhance accuracy and efficiency, provide immediate feedback, aid continuous learning, and reduce teacher workload while engaging students. However, issues such as construct validity and algorithmic transparency underscore the need for hybrid models that integrate human and AI assessments. Future research should focus on developing AI tools that are explainable,

culturally adaptive, and ethically responsible, ensuring fairness and upholding the pedagogical integrity of written English evaluations.

V. Discussion and Analysis

The findings of this review indicate that Artificial Intelligence (AI) is transforming the principles, processes, and ethics of written English assessment. By synthesizing insights from eight empirical studies conducted between 2022 and 2025, the discussion situates these results within established theoretical frameworks and pedagogical paradigms, including construct validity, assessment for learning, feedback literacy, and AI ethics. The introduction of AI enhances efficiency and precision in writing assessments; however, it also raises complex questions about validity, fairness, learner autonomy, and ethical accountability.

A. Reliability and Construct Validity: Balancing Objectivity With Depth

The reviewed literature indicates that AI-based writing evaluation systems are reliable and objective, adhering to the psychometric principle of construct validity, which emphasizes measurement consistency for assessment credibility. AI systems mitigate rater bias and offer standardized evaluations, enhancing fairness across diverse learner populations. However, concerns raised by Cheng et al. (2024) and Mahdi and Alkhateeb (2025) caution against excessively prioritizing linguistic accuracy, which may result in "construct underrepresentation." This is relevant because AI struggles to assess cognitive and rhetorical aspects of writing, including argumentation, coherence, and creativity. Thus, while AI can enhance reliability, it also presents challenges to authentic validity—where the assessment's capacity to reflect a construct's complexity is critical. This theoretical tension aligns with Weigle's (2002) argument that practical writing assessment should integrate both analytic and holistic dimensions to maintain educational significance. Current research suggests that the most effective solutions are hybrid systems that combine AI's algorithmic strengths in evaluating lower-order writing elements with human assessment for higher-order discourse analysis.

B. AI Feedback and Assessment for Learning: A Paradigm Shift

AI feedback systems are transforming assessment from summative evaluations to formative interactions, as highlighted in research by Jamshed et al. (2024, 2025). These systems use platforms such as ChatGPT and WhatsApp to provide iterative feedback that enhances linguistic accuracy and self-reflection among learners. This approach aligns with the Assessment for Learning (AfL) principles proposed by Black and Wiliam (2009), viewing feedback as a continuous, dialogic process. AI feedback also fosters learner agency, encouraging self-directed correction and experimentation, as noted by Al Ghaithi and Behforouz (2025). However, concerns raised by Rahman et al. (2023) suggest that over-reliance on AI feedback may lead to "surface compliance," in which students prioritize error correction over deeper understanding and rhetorical skills. This aligns with Carless and Boud's (2018) feedback literacy model, which emphasizes students' ability to engage critically with feedback. Thus, to enhance the efficacy of AI feedback, teachers must moderate it, ensuring it is framed in a way that promotes reflective awareness rather than a mechanical response.

C. Pedagogical Integration and Learner Engagement

AI plays a dual role as both an evaluative instrument and a teaching facilitator in educational environments, leading to impactful changes in classroom dynamics. Research by Li et al. (2023) and Al Ghaithi and Behforouz (2025) shows that incorporating AI into peer review processes fosters collaborative and reflective learning experiences. Students show greater engagement in self-evaluation and use machine-generated feedback to assess their revisions, thereby improving their understanding of linguistic and stylistic choices. This aligns with Vygotsky's socio-constructivist theory, which emphasizes learning in social contexts. Nonetheless, concerns persist that AI could supplant critical human interactions in education. Over-reliance on AI feedback may jeopardize the interactive components of writing instruction, which are vital for effective communication pedagogy. Therefore, it is crucial to use AI as a supporting tool in a human-centered educational framework, where teachers play a key role in contextualizing AI feedback, encouraging student reflection, promoting equitable access to technology, and ensuring that AI outputs align with educational objective.

D. Cognitive and Affective Dimensions of AI-Based Assessment

AI integration significantly affects learners' cognition and emotions. Research by Al Ghaithi and Behforouz (2025) indicates that consistent AI feedback alleviates students' writing anxiety. Conversely, Li et al. (2023) note that advanced learners experience greater cognitive fatigue from excessive digital corrections. These observations align with Sweller's (2011) Cognitive Load Theory, which posits that overwhelming cognitive demands can impede learning. To counteract this, AI systems should deliver adaptive feedback that varies in complexity and frequency according to the learner's proficiency. Instructors are encouraged to implement self-regulated learning strategies (Zimmerman, 2002) to help students manage feedback, prioritize revisions, and sustain motivation. When applied thoughtfully, AI can foster cognitive engagement while protecting emotional well-being.

E. Ethical, Transparency, and Data Governance Challenges

The ethical dimensions of AI-based assessment are crucial alongside their pedagogical implications. Research by Cheng et al. (2024), Mahdi and Alkhateeb (2025), and Zhao et al. (2023) emphasizes the need for transparency and accountability in these systems. The "black-box" nature of AI scoring, which produces results without clear reasoning, undermines the transparency principle highlighted by the International Language Testing Association (ILTA, 2018). Additionally, algorithmic bias poses a significant issue: models trained primarily on Western English data may disadvantage non-native language features, risking cultural inequity in assessments. This points to the necessity for bias-aware dataset design and inclusive algorithm calibration, as noted by Jiang et al. (2023) and Cheng et al. (2024). The ethical landscape is further complicated by data privacy concerns, with Zhao et al. (2023) indicating that many commercial AI platforms collect and reuse learner data without explicit consent. To uphold ethical integrity, institutions should implement strong data governance frameworks to protect user autonomy, ownership, and confidentiality. The multidimensional implications of these aspects are further illustrated in Table 5, which summarizes the theoretical, pedagogical, and ethical interpretations from the discussed studies.

TABLE 5
THEORETICAL AND PEDAGOGICAL IMPLICATIONS OF AI IN WRITING ASSESSMENT

Dimension	Empirical Evidence	Theoretical Interpretation	Pedagogical / Ethical Implications
Reliability & Validity	AI scoring shows high consistency but limited construct coverage (Jiang et al., 2023; Cheng et al., 2024)	Supports psychometric reliability but challenges construct validity (Messick, 1995)	Requires hybrid human–AI models for balanced evaluation
Feedback & Learning	Immediate AI feedback enhances self-editing and motivation (Jamshed et al., 2024, 2025)	Aligns with AfL and feedback literacy frameworks (Black & Wiliam, 2009; Carless & Boud, 2018)	Teachers must mediate AI feedback to promote reflective learning.
Learner Engagement	AI–peer integration increases participation and awareness (Li et al., 2023)	Embodies socio-constructivist interaction (Vygotsky, 1978)	Position AI as a scaffold, not an evaluator; retain dialogic pedagogy
Cognitive-Affective Impact	Reduced anxiety but potential technostress (Al Ghaithi & Behforouz, 2025; Li et al., 2023)	Explained by Cognitive Load Theory (Sweller, 2011)	Adopt adaptive pacing and promote self-regulated learning
Ethics & Governance	Algorithmic bias, opacity, and privacy risks (Mahdi & Alkhateeb, 2025; Zhao et al., 2023)	Reflects ethical AI and fairness principles (Floridi et al., 2018)	Implement explainable AI and institutional data protection policies

F. Toward a Human–AI Synergy in Writing Assessment

The integration of empirical and theoretical knowledge indicates that artificial intelligence (AI) should not operate as an independent evaluator within writing assessments but rather collaborate with human evaluators. This partnership capitalizes on AI's computational accuracy to enhance human judgment, while educators provide ethical oversight, contextual knowledge, and pedagogical empathy. Such a cooperative approach aligns with Bachman's Assessment Use Argument (AUA), which emphasizes that assessment validity is based on practical utility, fairness, and the consequences that result. In the context of written English assessment, AI increases efficiency and scalability, while human evaluators maintain authenticity and uphold ethical standards. This partnership embodies the human-centered AI paradigm, which promotes accountability, inclusivity, and educational benefits. The primary challenge is to establish this synergy through focused teacher training, the development of explainable AI systems, and the implementation of equitable assessment policies that ensure both technological efficiency and adherence to humanistic principles.

VI. CONCLUSION

The synthesis of recent empirical studies revealed that AI tools significantly enhanced the evaluation of written English by improving reliability, providing immediate feedback, and increasing learner autonomy. Eight studies from 2022 to 2025 demonstrated that AI technologies, such as large language models and automated writing evaluation (AWE), provided consistent assessments comparable to those of human raters. These tools support self-regulated writing aligned with formative assessment, democratizing access to quality feedback. However, challenges related to construct validity, interpretive transparency, and ethical governance persist. AI evaluations often focus on surface-level attributes like grammar while neglecting deeper writing elements, and algorithmic scoring can introduce biases that compromise fairness. The findings advocate a human-AI collaborative model in which computational efficiency complements human judgment. Instructors play a vital role in mediating AI feedback and fostering reflective learner engagement. For AI to serve ethically and effectively, its implementation must ensure accountability and inclusivity. Future research should investigate the long-term impacts of AI feedback on writing through longitudinal, cross-linguistic studies, alongside advancements in explainable AI (XAI) and culturally adaptive datasets to enhance fairness. Ultimately, AI should be seen as a collaborator that enhances human expertise and supports transformative learning in writing assessment.

ACKNOWLEDGEMENTS

The authors extend their appreciation to Prince Sattam bin Abdulaziz University for funding this research work through the project number (2025/01/36905).

REFERENCES

[1] Alharbi, W. (2023). AI in the Foreign Language Classroom: A Pedagogical Overview of Automated Writing Assistance Tools. *Education Research International*, *2023*, 1–15. https://doi.org/10.1155/2023/4253331

[2] Al-Obaydi, L. H., & Pikhart, M. (2025). Correction: AI partner versus human partner: comparing AI-based peer assessment with human-generated peer assessment in examining writing skills. *Language Testing in Asia*, *15*(1). https://doi.org/10.1186/s40468-025-00383-8

[3] Bachman, L. F. (2024). What Is the Construct? The Dialectic of Abilities and Contexts in Defining Constructs in Language Assessment 1. In *The Writings of Lyle F. Bachman* (pp. 220-250). Routledge.

[4] 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). https://doi.org/10.1186/s40468-024-00322-z

[5] Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability (formerly: Journal of personnel evaluation in education)*, *21*(1), 5-31. https://doi.org/10.1007/s11092-008-9068-5

[6] Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment & Evaluation in Higher Education*, *43*(8), 1315–1325.

[7] 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*(1). https://doi.org/10.1186/s41239-023-00411-8

[8] Chan, M., & Luckin, R. (2020). Preparing university assessment for a world with AI: Tasks for human intelligence. In *Re-imagining university assessment in a digital world* (pp. 49-63). Cham: Springer International Publishing.

[9] Chen, A., Zhang, Y., Jia, J., Liang, M., Cha, Y., & Lim, C. P. (2024). A systematic review and meta-analysis of AI-enabled assessment in language learning: Design, implementation, and effectiveness. *Journal of Computer Assisted Learning*, *41*(1). https://doi.org/10.1111/jcal.13064

[10] Cheng, Y. (2025). Transforming English language education with AI-driven deep learning models for scalable adaptive and inclusive assessment. *International Journal of Information and Communication Technology*, *26*(12), 15–31. https://doi.org/10.1504/ijict.2025.146163

[11] Deepshikha, D. (2025). A comprehensive review of AI-powered grading and tailored feedback in universities. *Discover Artificial Intelligence*, *5*(1). https://doi.org/10.1007/s44163-025-00517-0

[12] Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ... & Vayena, E. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and machines*, *28*(4), 689-707.

[13] Ghaithi, A. A., & Behforouz, B. (2025). Effects of interaction with AI-assisted writing evaluation on EFL students' writing performance. *Knowledge Management & E-Learning*, 206–224. https://doi.org/10.34105/j.kmel.2025.17.009

[14] Hooda, M., Rana, C., Dahiya, O., Rizwan, A., & Hossain, M. S. (2022). Artificial intelligence for assessment and feedback to enhance student success in higher education. *Mathematical Problems in Engineering*, *2022*, 1–19. https://doi.org/10.1155/2022/5215722

[15] Jamshed, M., Fatimah Albedah, Mohd Sajid Ansari, & Sameena Banu. (2025). Assessing the Efficacy of AI-Driven Corrective Feedback via WhatsApp Application to Improve ESL Learners' Writing Skills: An Experimental Study. *International Journal of Interactive Mobile Technologies* (iJIM), *19*(07), 190–205. https://doi.org/10.3991/ijim.v19i07.52709

[16] Jamshed, M., Manjur Ahmed, A. S. M., Sarfaraj, M., & Warda, W. U. (2024). The Impact of ChatGPT on English Language Learners' Writing Skills: An Assessment of AI Feedback on Mobile. *International Journal of Interactive Mobile Technologies*, *18*(19), 18-36. https://doi.org/10.3991/ijim.v18i19.50361

[17] Jiang, Z., Xu, Z., Pan, Z., He, J., & Xie, K. (2023). Exploring the role of artificial intelligence in facilitating assessment of writing performance in second language learning. *Languages*, *8*(4), 247.

[18] Jin, Y., & Fan, J. (2023). Test-Taker Engagement in AI Technology-Mediated Language Assessment. *Language Assessment Quarterly*, *20*(4–5), 488–500. https://doi.org/10.1080/15434303.2023.2291731

[19] Kanchana, S., & Saha, P. R. (2025). Evolving Student Assessment: AI-Driven Rubrics for Personalized and Equitable English Language Learning. *Journal of Engineering Education Transformations*, *38*(2), 584–590. https://doi.org/10.16920/jeet/2025/v38is2/25072

[20] Khan, M. A., Kurbonova, O., Abdullaev, D., Radie, A. H., & Basim, N. (2024). Is AI-assisted assessment liable to evaluate young learners? Parents support, teacher support, immunity, and resilience are in focus in testing vocabulary learning. *Language Testing in Asia*, *14*(1). https://doi.org/10.1186/s40468-024-00324-x

[21] Khasawneh, D. M. A. S. K. (2024). Improving the learning of language proficiency at tertiary education level through AI-driven assessment models and automated feedback systems. *Migration Letters*, *21*(2), 712-726.

[22] Lee, O. (2024). Examining AI-Based Accuracy Assessment in L2 learners' writing. *Journal of Pan-Pacific Association of Applied Linguistics*, *28*(2), 39–55. https://doi.org/10.25256/paal.28.2.3

[23] Li, W. Y., Kau, K., & Shiung, Y. J. (2023). Pedagogic exploration into adapting automated writing evaluation and peer review integrated feedback into large-sized university writing classes. *SAGE Open*, *13*(4), 1-18. https://doi.org/10.1177/215824402312090

[24] Liao, X., Zhang, X., Wang, Z., & Luo, H. (2024). Design and implementation of an AI-enabled visual report tool as formative assessment to promote learning achievement and self-regulated learning: An experimental study. *British Journal of Educational Technology*, *55*(3), 1253-1276.

[25] Mahdi, H. S., & Alkhateeb, A. (2025). Revolutionising essay evaluation: a cutting-edge rubric for AI-assisted writing. *International Journal of Computer-Assisted Language Learning and Teaching (IJCALLT)*, *15*(1), 1-19.
[26] Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, *50*(9), 741–749. https://doi.org/10.1037/0003-066x.50.9.741
[27] Poonpon, K., Manorom, P., & Chansanam, W. (2023). Exploring effective methods for automated essay scoring of non-native speakers. *Contemporary Educational Technology*, *15*(4), ep475. https://doi.org/10.30935/cedtech/13740
[28] Rahman, N. a. A., Zulkornain, L. H., Mat, A. C., & Kustati, M. (2023). Assessing Writing Abilities using AI-Powered Writing Evaluations. *Journal of ASIAN Behavioural Studies*, *8*(24), 1–17. https://doi.org/10.21834/jabs.v8i24.420
[29] Ranalli, J. (2018). Automated written corrective feedback: how well can students make use of it? *Computer Assisted Language Learning*, *31*(7), 653–674. https://doi.org/10.1080/09588221.2018.1428994
[30] Shi, H., & Aryadoust, V. (2024). A systematic review of AI-based automated written feedback research. *ReCALL*, *36*(2), 187–209. https://doi.org/10.1017/s0958344023000265
[31] Sweller, J. (2011). Cognitive load theory. In *Psychology of learning and motivation* (Vol. 55, pp. 37-76). Academic Press.
[32] Tam, A. C. F. (2024). Interacting with ChatGPT for internal feedback and factors affecting feedback quality. *Assessment & Evaluation in Higher Education*, *50*(2), 219–235. https://doi.org/10.1080/02602938.2024.2374485
[33] Tapalova, O., & Zhiyenbayeva, N. (2022). Artificial Intelligence in Education: AIED for Personalised Learning Pathways. *The Electronic Journal of e-Learning*, *20*(5), 639–653. https://doi.org/10.34190/ejel.20.5.2597
[34] 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
[35] Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes* (Vol. 86). Harvard University Press.
[36] Wang, Z. (2020). Computer-assisted EFL writing and evaluations based on artificial intelligence: a case from a college reading and writing course. *Library Hi Tech*, *40*(1), 80–97. https://doi.org/10.1108/lht-05-2020-0113
[37] Weigle, S. C. (2002). *Assessing writing*. Cambridge University Press.
[38] Zhao, R., Zhuang, Y., Zou, D., Xie, Q., & Yu, P. L. H. (2022). AI-assisted automated scoring of picture-cued writing tasks for language assessment. *Education and Information Technologies*, *28*(6), 7031–7063. https://doi.org/10.1007/s10639-022-11473-y
[39] Zimmerman, B. J. (2002). Becoming a Self-Regulated Learner: An Overview. *Theory into Practice*, *41*(2), 64–70. https://doi.org/10.1207/s15430421tip4102_2

Ahmed Alshehri is currently working as an Assistant Professor at the Department of English Language & Literature, College of Arts and Letters, University of Bisha, Saudi Arabia. He earned the Master of Arts in Linguistics from Essex and the Doctor of Philosophy in Applied Linguistics from the University of York. He teaches EFL classes. His area of interest includes applied linguistics and he has made significant contributions to this field over many years. He publishes in renowned Scopus-indexed journals.

Saddah Aldossary is currently working as an Assistant Professor in the Department of English at the College of Science and Humanities, Shaqra University, Saudi Arabia. He earned his Master's degree in Applied Linguistics and his Ph.D. in TESOL. His areas of interest include educational research, particularly TESOL and TEFL. He is passionate about teaching English by integrating traditional and innovative methods through blended and e-learning. He also focuses on second language acquisition, applied linguistics, the use of social media for professional development, and enhancing student learning experiences. Dr. Aldossary is actively publishing in reputable Scopus-indexed journals.

Mohammad Jamshed got his Ph.D. degree in 2018 from Aligarh Muslim University, Aligarh, and U.P. (INDIA). His areas of interest include postcolonial literature, travel writings, comparative studies, and EFL pedagogy in Saudi Arabia and larger Arab world. He has presented papers at conferences and published around three dozen articles in Scopus-indexed journals of repute both in literature and ESL/EFL pedagogy and research. He is currently working as an Associate Professor, at the department of English, College of Science & Humanities, Prince Sattam Bin Abdulaziz University, Al Kharj, 11942, Kingdom of Saudi Arabia.

Mohammad Rezaul Karim is currently working as an Associate Professor of English in the College of Science and Humanities, Prince Sattam bin Abdulaziz University, KSA. He holds a Ph.D. from Gauhati University, India. He has been teaching English language to the undergraduate students for the last 7 years. He has presented papers at both national and international conferences, published more than 25 research articles and papers in various Scopus and WOS indexed journals, and authored two books. His main area of interest is English language and comparative literature. He is available at karimrezaul318@gmail.com