AI IN ENGLISH LANGUAGE TEACHING AND RESEARCH

Artificial intelligence (AI) has rapidly changed English language teaching (ELT) and English language research (ELR). With the ongoing evolution of AI, educators and educational researchers are now confronted with growing challenges to rethink their traditional pedagogical techniques, assessment methods, and processes of knowledge generation. This book, "AI in English Language Teaching and Research," is a collection of scholarly viewpoints on how AI may both improve and challenge language education. This book offers a wide-ranging and critical assessment of the roles of AI in defining the future of English language education and addressing practical applications as well as pedagogical, ethical, and societal consequences. It is a source of inspiration and a practical guide for educators and future educators.

CO-WRITERS
Collaborative Writing of Indonesian English Teachers

IKAPI
IKATAN PENERBIT INDONESIA
Anggota IKAPI 199/JTE/2020

RIZQUNA
cv_rizqunaa@gmail.com
081294856272
Jl. KS Tubun Gang Camar RT 05/04
Karangsalam Kidul, Kedungbanteng, Banyumas

AI IN ENGLISH LANGUAGE TEACHING AND RESEARCH

RIZQUNA

Editor:
Novalita Fransisca Tungka

AI IN ENGLISH LANGUAGE TEACHING AND RESEARCH

Nurul Hasanah Fajaria, Lu'luil Maknun, Muhamad Hasbi, Neneng Islamiah, Hendrikus Male, Eka Sustri Harida, Melli Kusmaningrum, Siti Mariam, Catur Kepirianto, Areta Puspa, Teguh Sulistyo, Eka Rizki Amalia, Ni Luh Putu Ning Septyarini Putri Astawa, Bernadin Maria Noenoek Februati, Hidayah Nor, Linda Septiyana, Fitria Nur Hamidah, Zahratun Nufus, Dini Deswarni, Nurohman, Zurriat Nyndia Rahmawati, Nine Febrie Novitasari, Annisa Khairir Rizki, Ana Utami Fatoni, Praptika Septi Femilia, Ahmad Ridho Rojabi, Lisa Rakhmanina, Maslihatul Bisriyah, Isry Laila Syathroh, Ida Ayu Oka Purnami.

Table of Contents

Chapter 2

AI-Based Assessment in English Language Teaching

Lu'luil Maknun
Universitas Mercu Buana Yogyakarta
lulu@mercubuana-yogya.ac.id

1. INTRODUCTION

Artificial intelligence (AI) has increasingly changed the way language learning is tested and taught in schools. The introduction of AI-based assessment tools in English language education, such as automated writing assessment systems, AI-generated feedback platforms, and speech recognition technologies, has made it easier and more consistent to assess how well students are developing their language skills. These systems use natural language processing (NLP) and machine learning algorithms to analyze linguistic features and provide automated feedback. This allows teachers to review many student submissions more quickly. These technologies are especially important in English as a Foreign Language (EFL) settings, where teachers often struggle to provide students with timely and personalized feedback (Hussein et al., 2019; Liu, 2024; Maknun et al., 2024; Mualim et al., 2025). Furthermore, AI-assisted assessment systems have shown promise in enhancing evaluative accuracy and

providing immediate feedback that fosters students' language development and engagement (Dong, 2024; Zheng et al., 2024).

One of the most widely discussed applications of AI in language assessment is automated writing evaluation (AWE). These systems can automatically score essays, detect grammatical errors, and provide revision suggestions to learners. Recent studies indicate that AI-based scoring can achieve considerable levels of accuracy and reliability, particularly when supported by clearly defined rubrics and linguistic features. In some cases, AI-generated scores have shown substantial alignment with human ratings, suggesting that AI can serve as a complementary tool to assist teachers in writing assessment (Mizumoto & Eguchi, 2024; Taşçı, 2025; Maknun et al., 2024). Nevertheless, AI systems tend to focus more effectively on surface-level linguistic features such as grammar, vocabulary, and syntax, while struggling to evaluate deeper aspects of writing such as coherence, argumentation, and rhetorical quality (Setiyowati & Ardaniah, 2024; Margana et al, 2026). These limitations indicate that AI-based assessment cannot yet fully replace human evaluators but may instead function best as part of a hybrid assessment model combining machine efficiency with human judgment.

Despite the growing adoption of AI in language assessment, several unresolved issues remain. Concerns regarding the reliability and validity of automated scoring, potential bias in algorithmic decision-making, and ethical implications related to academic integrity have sparked significant debate among researchers and educators. Some studies also highlight that AI-generated feedback may be inaccurate or misleading, potentially negatively influencing students' learning when used without critical evaluation (Steiss et al., 2024; Chowdhury et al., 2024). Consequently, integrating AI into English language assessment requires careful consideration of pedagogical, methodological, and ethical dimensions. This chapter, therefore, explores the role of AI-based assessment in English language teaching by examining key areas such as automated writing feedback, AI-supported speaking assessment, AI-generated feedback

systems, reliability and validity, and the ethical challenges associated with AI use in language assessment.

2. DISCUSSION

The use of artificial intelligence (AI) in English language assessments has changed how teachers assess students' English-speaking and writing skills and how well they are learning. AI technologies enable automatic analysis of language data, allowing teachers to deliver faster, more consistent, and more scalable assessments across different language skills. AI-based assessment tools are becoming more common in English language teaching (ELT) to support writing evaluation, speaking assessment, and feedback. These changes could make assessments more efficient and keep students more interested, but they also raise concerns about reliability, validity, and ethical use. The next sections discuss four important aspects of using AI in ELT assessments: automated writing feedback, AI-supported speaking assessment, AI-generated feedback systems with reliability and validity issues, and ethical problems in AI-assisted language assessment.

2.1 Automated Writing Feedback

Automated writing feedback, commonly known as Automated Writing Evaluation (AWE), is one of the most prevalent applications of AI-driven assessment in English language education. AWE systems use natural language processing and machine learning algorithms to analyze students' written work and provide instant feedback on grammar, vocabulary, spelling, and sentence structure. Teachers can grade more quickly in large classrooms thanks to tools such as automated essay-scoring systems and AI-powered writing assistants, which can review many student submissions in just a few seconds. This technological capability is especially useful in EFL settings, where teachers often have a lot of work to do and not enough time to give students detailed feedback on their writing (Hussein et al., 2019).

Studies show that automated writing systems can support formative assessment by helping students identify mistakes in their writing and make better revisions. Students who get quick feedback from AI-based tools often change their work more often, which can help them write better over time. Even though AWE systems offer these benefits, they still can't fully assess higher-order writing skills such as argument development, coherence, creativity, and rhetorical effectiveness. AI systems may not fully understand a piece's communicative intent or the quality of its ideas, as these aspects of writing require complex human judgment and context interpretation. Consequently, numerous scholars advocate for the utilization of AI-driven writing feedback as an adjunct rather than a substitute for teacher assessment (Maknun & Nurnadiah, 2022; Mizumoto & Eguchi, 2024). In this mixed approach, AI helps teachers by providing feedback and identifying mistakes, while teachers focus on more important tasks, such as writing quality and critical thinking.

2.2 AI-Supported Speaking Assessment

AI-assisted speaking assessment is another new way that AI is being used in language tests. AI systems can now analyze learners' spoken language by considering factors such as pronunciation accuracy, fluency, intonation, and speaking rate (Maknun, 2020). Better speech recognition and machine learning technologies have made this possible. AI-powered platforms can listen to what students say, write it down, and then use preset scoring algorithms to look at language features. Many standardized language tests and digital learning platforms have used this technology to automatically score and give feedback on speaking tasks.

There are several educational benefits to using AI to support speaking tests. First, it lets students practice speaking more often and get feedback right away without needing a teacher to be there all the time. Second, AI-based speaking tools can provide objective, consistent feedback by using the same scoring system for all students. This consistency can help make human scoring less subjective and make the assessment process more efficient. AI

systems can also track how well students speak over time, allowing teachers and students to see how far they've come and what they need to work on (Maknun, 2024; Zheng et al., 2024).

Nonetheless, AI-assisted speaking evaluation encounters numerous challenges. Speech recognition systems might not correctly understand non-native accents, background noise, or the different ways people speak that are common in EFL settings. Because of this, automated scores might not always show how well learners can communicate. Additionally, speaking proficiency encompasses not only pronunciation and fluency but also pragmatic competence, interactional skills, and contextual appropriateness, factors that continue to pose challenges for AI systems in terms of accurate evaluation. Therefore, as with writing tests, AI-based speaking tests should be combined with teacher judgment to provide a more complete picture of how well students can communicate.

2.3 AI-Generated Feedback Systems, Reliability, and Validity

AI-generated feedback systems are designed to provide students with quick feedback on how effectively they write and talk. These systems analyze what students say or write, identify language patterns, and establish standards to recommend ways to improve. This type of feedback can support formative assessment in language education by prompting students to revise their work and reflect on how they use language. AI's ability to deliver rapid feedback is especially helpful in large online classes, where teachers can't always answer every question right away.

Even with these improvements, issues about the trustworthiness and validity of AI-generated feedback remain a key problem in research on language assessment. Reliability is the degree to which the outcomes of an evaluation are consistent across various settings or contexts (Maknun, 2024). Validity is the degree to which a test accurately measures what it is supposed to measure (Maknun, 2024). AI systems rely on training data and mathematical models that may not completely represent the diverse linguistic backgrounds of language learners. This is why automated scoring systems may

produce inconsistent scores or fail to recognize certain language features. Some systems, for example, may care more about grammar than how well they communicate or how good the content is (Liu, 2024).

Another issue is how easy it is to understand AI algorithms. Many AI-based assessment systems are like "black boxes," making it impossible for people to understand how scores are generated. Because of this lack of openness, teachers and students may struggle to determine whether the evaluation process is fair and reliable. Researchers emphasize the need to integrate AI-generated feedback with human oversight to ensure that assessment judgments remain conducive to learning and aligned with educational objectives (Steiss et al., 2024).

2.4 Ethical problems in AI-assisted language assessment.

There are also serious moral questions arising from the growing use of AI in language assessment. One main issue is that AI algorithms might not be fair or may exhibit bias. The results of automated assessments might be influenced by the quality and representativeness of the databases used to train AI systems. If the training data does not cover a wide range of linguistic backgrounds, AI systems could unintentionally make it difficult for some groups of learners. For instance, voice recognition algorithms might not operate as well for those who speak with a strong regional or non-native accent. This could lead to unjust judgments.

Another ethical problem is academic honesty and the risk that AI tools could be used for malicious purposes. As AI writing assistants and generative AI technologies become more prevalent, students may rely too much on automated tools to write (Hasbi, 2024). This could make it hard to distinguish between work done by a computer and work done by a real student. Teachers have a harder time ensuring that assessments accurately reflect how effectively kids can speak and write in English in this setting. So, teachers need to find ways to assess that encourage kids to use AI ethically and help them improve their language skills.

When employing AI for testing, you should also think about how safe and private your data is. Many AI systems maintain a lot of student information, including written texts, speech recordings, and performance records. If you do not take appropriate steps to protect your data, private student information could be misused or made public (Hasbi et al., 2026). Because of this, schools and tech companies that build AI-based assessment systems need to ensure they comply with ethical and data protection rules. In the end, it's necessary to address these moral difficulties so that AI technologies are used safely and effectively in English language instruction.

3. FUTURE DIRECTION

A number of educators and researchers are talking about what the expanding use of AI in language assessment will mean for teaching and learning English in the long term. AI-based evaluation has many benefits, such as being faster, easier to use, and providing immediate feedback. These elements can support formative assessment by providing students with regular feedback on their language abilities and making it easy for teachers to see how well their students are doing. AI technologies can help teachers perform their duties more effectively and give students more opportunities to practice and review, especially in large EFL classes. As a result, AI-assisted evaluation is often seen as a promising way to meet the growing need for personalized, data-driven learning spaces (Zawacki-Richter et al., 2019).

Some people, on the other hand, warn that using AI to grade too much could hurt education in ways you did not mean to. One major issue is that automated systems would focus on things that can be measured, such as the perfection of grammar or the diversity of vocabulary, rather than more crucial qualities of language use, like creativity, critical thinking, and the desire to communicate. AI systems may also provide feedback that disregards context or the subtleties of pedagogy, thereby misguiding students when used without human assistance. Consequently, numerous experts assert

that AI should be viewed not as a replacement for human assessment but as a supplementary tool that augments educators' professional judgment and pedagogical expertise (Luckin et al., 2016).

Teachers and schools need to be careful and balanced when incorporating AI technology into language instruction. They should consider both the benefits and drawbacks of AI-based evaluation. Teachers need to know how to evaluate AI-generated input critically and apply it in ways that support their teaching (Hasbi et al., 2024). People who build AI-based testing systems should also keep working to make algorithms more open, fair, and compatible with many languages. AI will work with people in the future to improve language tests, make them fairer, and make them more valuable.

Author

Dr. Lu'luil Maknun, M.Pd., serves as an assistant professor in the English Education Department at Universitas Mercu Buana Yogyakarta, Indonesia. She earned her bachelor's, master's, and doctoral degrees from Universitas Negeri Yogyakarta. Her academic and research work focuses on English language teaching, language assessment, and the integration of technology in EFL learning environments. She has published papers in a reputable international journal indexed in Scopus. She has a strong interest in exploring technology-enhanced instruction and assessment to support effective and innovative English learning. Her contributions encompass teaching, research, and community engagement, all aimed at enhancing the quality of English education in Indonesia.

References

Dong, J. Q. 2024. A Research Agenda for Digital Transformation: Multidisciplinary Perspectives. Edward Elgar Publishing.

Hasbi, M., Islamiah, N., Sulistyo, T., Marhaban, S., Sari, A. L., Nor, H. (2025). *Deep Learning: A Handbook for English Language Teachers*. Rizquna.

Hasbi, M., Alamsyah, A., Faozan, A., Astawa, N. L. P. N. S. P., Fauzi, A. R. (2024). *Useful AI Tools For English Teachers*. Rizquna. http://e-repository.perpus.iainsalatiga.ac.id/21414

Hasbi, M. (2024). The ways QuillBot enhances English academic writing. In *Useful AI Tools for English Teachers* (pp. 29-40). Rizquna. http://e-repository.perpus.iainsalatiga.ac.id/21413/

Hussein, M. A., Hassan, H, & Nassef, M. 2019. Automated language essay scoring systems: a literature review. *PeerJ Computer Science*. *5.* https://doi.org/10.7717/peerj-cs.208

Liu, W. (2024). A systematic review of automated writing evaluation feedback: Validity, effects, and students' engagement. *Language Teaching Research Quarterly*, 45, 86–105. https://doi.org/10.32038/ltrq.2024.45.05

Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence unleashed: An argument for AI in education*. Pearson Education.

Maknun, L. (2020). The implementation of Orai as artificial intelligence for digital native students in English-speaking learning. *Indonesia Technology-Enhanced Language Learning* (pp. 131-138). https://itell.or.id/conference/index.php/itell/itell20/paper/view/96

Maknun, L., & Nurnadiah, S. (2022). Students' perception of plagiarism in thesis writing. *Jurnal Inovasi penelitian, 3*(5), 6401-6412. https://doi.org/10.47492/jip.v3i5.2091

Maknun, L. (2024). Criteria of a good test. In B. S. H. Hutauruk (Eds.). *An Introduction of English Language Testing* (pp. 105–123). Future Science.

Maknun, L., Zamzani, Z, & Jamilah, J. (2024). Unveiling Indonesian EFL teachers' perceptions and challenges of technology-based assessment as and for learning. *International Journal of Language Testing, 14*(1), 82–98. https://doi.org/10.22034/IJLT.2023.400628.1260

Maknun, L., Zamzani, Z., & Jamilah, J. (2024). Leveraging Technology-Based Afl and Aal within The Framework of English Differentiated Instruction in Indonesia. *Teaching English with Technology, 24* (2), 47–70. https://doi.org/10.56297/vaca6841/LRDX3699/VSCH7944

Maknun, L. (2026). The instructional design of educational technology. In A. P. D. Devi (Eds.), *Teaching university-level English Course* (pp. 29–42). Rizquna.

Mualim, M., Margana, Widyantoro, A., & Maknun, L. (2025). Fostering EFL Pre-Service Teachers' TPACK through Inquiry-Based, Technology-Saturated, and Flipped Instructional Model, International *Journal of Modern Education and Computer Science (IJMECS), 17*(4), 45-57. https://doi.org/10.5815/ijmecs.2025.04.03

Mizumoto, A., & Eguchi, M. (2024). Exploring the potential of using an AI language model for automated essay scoring. *Research methods in Applied Linguistics*, 2(2). https://doi.org/10.1016/j.rmal.2023.100050

Setiyowati, R., & Ardaniah, V. (2024). A systematic review of AI-based and teacher-based writing assessment. *English Language and Literature International Conference Proceedings*. https://jurnal.unimus.ac.id/index.php/ELLIC/article/view/18463

Steiss, J., Tate, T., & Warschauer, M. (2024). Generative Artificial Intelligence in Educational Contexts: A Systematic Review Of Opportunities, Challenges, And Ethical Implications. *International Research Journal of Adavnced Engineering and Technology (IRJAET, 2*(10), 102-111. https://aimjournals.com/index.php/irjaet

Taşçı, S. (2025). Human and AI scoring of EFL writing: The influence of rubrics and genre on reliability. *Journal of Education and New Approaches*, *8*(2), 191-210. https://doi.org/10.52974/jena.1785369

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education-where are the educators?. *International Journal of Educational Technology in Higher Education*, *16*(39). https://doi.org/10.1186/s41239-019-0171-0

Zheng, Y., Warschauer, M., & Farkas, G. (2024). Artificial intelligence for language learning: Entering a new era. *Language Learning & Technology*, *28*(1), 1–4. https://doi.org/10.64152/10125/73569