# Investigate How AI Algorithms Can Be Used to Automate English Language Proficiency Assessments

Zein Bassam Bani Younes[1]

[1] Assistant Professor, Department of English Language and Literature, Jadara University, Jordan

Correspondence: Zein Bassam Bani Younes, Assistant Professor, Department of English Language and Literature, Jadara University, Jordan.

Received: November 3, 2024 Accepted: February 14, 2025 Online Published: April 17, 2025

doi:10.5430/wjel.v15n5p285 URL: https://doi.org/10.5430/wjel.v15n5p285

**Abstract**

This study explores the integration of artificial intelligence (AI) tools in English language learning and assessment among university students, focusing on its impact on the accuracy, efficiency, and comprehensiveness of evaluating speaking, listening, reading, and writing skills. Utilizing a quantitative research design with an online questionnaire, data were collected from students actively using AI in their academic pursuits. Adopting a cross-sectional survey methodology, the study investigates how AI-driven assessments can enhance learning by providing instant feedback, streamlining evaluation processes, and potentially reducing the burden on educators. The findings suggest that AI offers promising opportunities to support language acquisition through automated scoring systems and personalized learning experiences tailored to individual needs. However, concerns persist regarding the reliability, fairness, and accuracy of AI-generated assessments, raising the need for standardized frameworks to ensure validity and minimize biases. The study also highlights the necessity of addressing technical challenges, such as system errors and user adaptability, to optimize AI's effectiveness in educational settings. Furthermore, successful AI implementation in language assessment requires comprehensive training programs to familiarize students and educators with the technology, fostering confidence and competence in its use. By expanding the knowledge of AI's role in education, this study underscores the importance of making informed, data-driven decisions regarding AI adoption in academic environments to maximize its benefits while mitigating potential risks.

**Keywords:** AI, language learning, assessment, educational technology, university students

## 1. Introduction

### *1.1 Introduce the Problem*

Artificial Intelligence (AI) has long been a subject of human fascination and inquiry, evolving from mythological narratives to a formally recognized field of research. The field was formally initiated at the Dartmouth workshop in 1956, setting the foundation for AI research in subsequent decades (Kaplan & Haenlein, 2019). From early mechanical computing devices to modern neural networks, AI has developed significantly, particularly in the realm of language processing and automated assessment. The evolution of AI and its application in education, particularly in language proficiency assessment, presents both opportunities and challenges that warrant further exploration.

The primary problem under study is the impact of AI on language learning and assessment. AI-driven tools, including Natural Language Processing (NLP) and Automated Speech Recognition (ASR), have transformed the landscape of language proficiency testing. While these advancements enhance efficiency, accessibility, and personalization in language learning, they also raise concerns regarding accuracy, fairness, and ethical considerations. This study aims to evaluate the historical trajectory, technological advancements, and future implications of AI in language proficiency assessment.

### *1.2 Explore the Importance of the Problem*

The role of AI in education, particularly in language proficiency assessment, has gained significant traction due to its ability to provide scalable and personalized learning solutions. The increasing reliance on AI-driven assessment methods necessitates a deeper understanding of their reliability, effectiveness, and ethical implications (Pedro et al., 2019). AI applications such as IBM Watson, Deep Blue, and expert systems have demonstrated high computational capabilities (Crevier, 1993; LeCun et al., 2015; Rababah, 2025), highlighting AI's potential to revolutionize education.

However, the integration of AI in language assessment is not without challenges. Questions arise regarding the fairness of AI-driven evaluation methods, the ability to accurately measure human linguistic capabilities, and the broader impact on pedagogical practices. Furthermore, AI's influence on educational accessibility aligns with the objectives of the fourth Sustainable Development Goal, emphasizing equitable education for all. This study seeks to address these concerns by exploring the transformative effects of AI on language proficiency assessment.

### *1.3 Describe Relevant Scholarship*

The development of AI in language proficiency assessment has evolved over multiple phases:

Early Beginnings (1960s-1980s): The advent of Computer-Assisted Language Learning (CALL) in the 1960s and 1970s marked the initial steps toward automated language assessment. These early systems primarily utilized simple grammar and vocabulary exercises, often relying on multiple-choice formats (Chapelle, 2001).

Advancements in NLP (1990s): The 1990s saw significant progress in NLP, enabling more sophisticated assessment tools. A notable example is the Educational Testing Service's e-rater, which evaluates written essays based on grammar, syntax, and coherence (Burstein et al., 1998).

Introduction of Automated Speech Recognition (2000s): The early 2000s witnessed the emergence of ASR technologies, which facilitated automated evaluation of spoken language proficiency. ETS's SpeechRater exemplifies these innovations by analyzing pronunciation, fluency, and intonation (Zechner et al., 2009).

Integration of AI and Machine Learning (2010s): The 2010s saw the incorporation of AI and ML into language assessment, significantly enhancing accuracy and adaptability. AI-driven platforms such as Duolingo's English Test leverage machine learning algorithms to tailor assessments to users' proficiency levels, promoting personalized learning experiences (Settles et al., 2013).

Recent literature also highlights the broader role of AI in education, emphasizing personalized learning, adaptive teaching methods, and data-driven decision-making (Chen et al., 2020; Sayed et al., 2024). Researchers such as Roll and Wylie (2016) argue that AI not only supports traditional teaching practices but also redefines educational methodologies by integrating intelligent learning assistants and diagnostic tools. However, existing studies indicate that educators' perspectives on AI integration remain underrepresented (Zawi-Richter et al., 2019). This study aims to bridge this gap by analyzing AI's impact on teaching methodologies and pedagogical frameworks.

### *1.4 State Hypotheses and Their Correspondence to Research Design*

This study hypothesizes that AI-driven language proficiency assessment enhances learning outcomes through personalized and adaptive feedback mechanisms. However, it also examines potential drawbacks, including biases in AI-driven evaluations, ethical concerns, and challenges in standardization. The research adopts a qualitative literature review approach, synthesizing existing studies on AI's role in language assessment while identifying key trends, technologies, and theoretical frameworks.

By systematically reviewing advancements in AI and NLP, this study aims to contribute to ongoing discussions on the effectiveness, fairness, and future implications of AI-driven language proficiency assessment. Furthermore, it seeks to inform educators, policymakers, and developers about best practices for integrating

## 2. Literature Review

The advancements in Artificial Intelligence (AI) have significantly influenced language proficiency assessment, particularly in evaluating spoken English skills. Automated scoring systems have gained prominence over the past few decades, offering increased accuracy and efficiency. This section reviews key studies that explore the implementation, effectiveness, and challenges of AI-driven language assessments.

Evanini, Hauck, and Hakuta (2017) examine the use of automated scoring technology in evaluating K–12 students' spoken responses on English Language Proficiency (ELP) tests. With various states' rising adoption of computer-based assessments, the reliability of automated scoring remains a central concern. While these technologies offer substantial advantages, including efficiency and scalability, they also present challenges in evaluating diverse linguistic skills (Evanini et al., 2017). The study emphasizes the necessity of determining whether automated scoring can accurately assess students' exam-oriented skills and proposes recommendations for integrating AI-driven evaluation in ELP assessments (Evanini et al., 2017).

Similarly, the IEEE Transactions on Professional Communication (2010) research highlights AI's role in aviation-related language assessments. The International Civil Aviation Organization (ICAO) directive mandates that pilots and air traffic controllers demonstrate sufficient English proficiency for international communications. The Versant Aviation English Test, a fully automated listening and speaking assessment, was developed to address this requirement. Findings suggest that automated scoring aligns closely with human evaluations ($r = 0.94$), indicating its reliability in high-stakes scenarios (IEEE Transactions on Professional Communication, 2010).

Studies analyzing computer-based assessments in the United States also reveal an increasing reliance on AI for evaluating young learners' English proficiency. Downey, Suzuki, and Van Moere (2010) explore both the advantages and limitations of automated English language testing for children. Their study dissects the components of an automated scoring system, providing insights into its design, implementation, and performance.

In a broader review of language assessment technologies, Bahari (2021) analyzed 286 articles published between 2002 and 2018 to identify key trends in AI-assisted language testing. The study highlights the shift toward more interactive and adaptive assessment models that cater to individual learner differences. These advancements reflect a growing emphasis on evaluating language proficiency holistically rather than isolating specific linguistic skills.

Another notable contribution comes from De Damp, Van der Walt, and Niesler (2009), who investigate AI's capability to assess spoken English by analyzing various speech characteristics, including pronunciation, fluency, and repetition patterns. Their findings indicate a strong correlation between machine-generated scores and human ratings, reinforcing the potential of AI-driven evaluation systems.

Bamdev et al. (2023) emphasize the need for automated English proficiency scores to align with human comprehension, particularly in academic and professional settings. Their research applies machine learning techniques to evaluate spoken responses based on five linguistic dimensions: fluency, pronunciation clarity, content coherence, grammatical accuracy, and vocabulary use. The study finds that regression models outperform other scoring methods, offering a reliable approach to assessing speech quality. By identifying key linguistic markers, their analysis bridges the gap between automated evaluation and human judgment.

Machine learning has also been employed to enhance automated scoring accuracy in non-native English speakers. Johnson, Kang, and Ghanem (2015) introduce a computational model that incorporates suprasegmental speech features, such as stress, rhythm, and intonation, to improve spoken language assessment. Their study, which analyzes 120 monologues from the Cambridge ESOL Common English Examinations, achieves a Pearson correlation of 0.718 with human proficiency scores, demonstrating strong alignment with expert evaluations.

Collier and Huang (2020) examine the Texas English Language Proficiency Assessment System (TELPAS), a standardized test used annually to assess English learners (ELs) in K–12 education. Their study underscores TELPAS's role in informing EL classification, educational accountability, and instructional accommodations. The researchers highlight recent structural modifications aimed at enhancing the reliability and effectiveness of AI-assisted scoring within the TELPAS framework.

Beyond language proficiency assessments, AI has also been utilized to detect plagiarism in spoken responses. Hauck, Wolf, and Mislevy (2016) integrate text similarity analysis with automated speech scoring to distinguish between plagiarized and original spoken responses. Their system, tested on a large dataset from an operational English proficiency exam, achieves an F1-score of 0.706, suggesting that AI can improve the validity of both human and automated assessment methods.

Van Moere and Downey (2016) further explore the evolution of automated scoring technologies, emphasizing their widespread acceptance in language assessment. While AI does not replicate human evaluators' cognitive processes, it efficiently identifies, categorizes, and weighs linguistic features in both speech and writing, producing scores comparable to human raters (Shermis, 2014; Van Moere, 2012).

Vajjala et al. (2018) investigate the linguistic features essential for AI-driven essay scoring across multiple datasets, such as TOEFL11SUBSET and FCE. Their predictive modeling, incorporating vocabulary usage, syntactic complexity, and error analysis, demonstrates an accuracy rate of 73% for proficiency classification and a correlation coefficient 0.8 in regression models. The study highlights the influence of a speaker's native language on their English writing patterns and emphasizes the necessity of linguistic diversity in training AI models.

Gayed et al. (2022) explore AI KAKU, an AI-driven web application designed to support English as a Foreign Language (EFL) learners in collaborative writing tasks. Their findings suggest that AI-powered tools like AI KAKU offer structured assistance beyond conventional word processors, potentially improving students' writing performance. The study calls for further research into AI-assisted learning environments, showcasing the expanding role of AI in language acquisition.

In the context of the Fourth Industrial Revolution, AI has become an integral part of daily life, extending beyond education into various domains. AI-powered assistants such as Alexa and robotic household devices illustrate how intelligent systems enhance everyday tasks (Kuddus, 2022). Within education, AI-driven platforms provide personalized learning experiences, tailoring instruction to individual student needs (Kuddus, 2022). However, concerns persist among educators regarding the implications of AI in language learning, highlighting the need for ongoing research into its pedagogical effectiveness and ethical considerations.

## 3. Methodology

### *3.1 Introduction*

This chapter outlines the methodology employed to investigate the influence of AI tools on English language learning and assessment among university students. Using a quantitative approach, this study utilizes an online survey to gather data from university students to obtain comprehensive information that aligns with the research objectives.

### *3.2 Research Design*

Using a cross-sectional survey design allows for capturing a wide range of participant responses at a single point in time. This design provides a broad perspective on how AI tools are integrated into language learning practices.

### *3.3 Population and Sample*

The study focuses on university students actively involved in learning English as their second language. Eligible participants must meet specific criteria:

- Currently enrolled in a university
- Actively using AI-based tools for language learning

- Willing to participate voluntarily

Convenience sampling will be utilized to recruit participants, with a target sample size of at least 50 students to ensure robust statistical analysis.

*3.4 Data Collection Instrument*

Since the intended study population has Internet access, the preferred method of data collection will be an online survey, which is easily accessible by providing a web link. These are both closed-ended and open-ended questions, where it is possible to define numerical criteria and receive more qualitative data.

*3.5 Questionnaire Structure*

The questionnaire is structured into the following sections:

3.5.1 Demographic Information

- Age
- Gender
- The academic characteristics include the university affiliation and students' academic level.
- Native language

3.5.2 Language Learning Background

- Duration spent on English language learning
- Level of English presently possessed by an individual
- Frequency and duration used for the AI tool

3.5.3 Usage Patterns of AI Tools

- Candidate’s name and the type of tools they choose, such as grammar checkers and automated essay scoring.
- Frequency of use: How and when is the AI tool used, concerning the subject and its classes (e.g., for assignments, independent learning)

3.5.4 Perceived Effectiveness

- Likert-scale questions that address observed changes in writing, speaking, listening, and reading abilities.
- Establishment of the most effective and least effective features of an AI tool

3.5.5 Usability and User Experience

- Ease of use
- User Interface Satisfaction
- Technical challenges encountered

3.5.6 Impact on Learning Autonomy:

- Acquisition of autonomy and motivation to learn. Helped format the written assignments and edit the written material.

3.5.7 Open-Ended Feedback:

- Detailed feedback on AI tools
- Suggestions for improvement
- Additional comments

*3.6 Pilot Testing*

Before full implementation, a pilot test involving a small group of university students (10-15 participants) will refine the questionnaire based on feedback received regarding clarity and relevance.

*3.7 Data Collection Procedure*

The online questionnaire will be circulated through various channels such as university email lists, social media platforms (e.g., Facebook, Twitter, LinkedIn), and educational forums to reach university students. Participants are assured of the study's purpose, voluntary participation, and confidentiality before providing electronic consent.

*3.8 Ethical Considerations*

The study adheres to ethical guidelines, safeguarding participants' rights by:

- Informed Consent: Ensuring participants have detailed information about the study before consenting.
- Confidentiality: Maintaining participant anonymity and confidential handling of data.

- Voluntary Participation: Allowing participants to withdraw without consequences.
- Data Security: Ensuring secure storage accessible only by authorized research team members.

Data Analysis: To explore relationships and patterns within data, both quantitative methods (descriptive and inferential statistics) and qualitative methods (thematic analysis) will be employed.

*3.9 Delimitations*

- The study focuses solely on university students learning English as a second language.
- Limited only to AI-based tools for language learning.
- Geographically restricted to universities reachable online.
- The study concentrates on current experiences rather than long-term effects.
- Only including participants experienced with AI-based language learning tools.

*3.10 Conclusion*

- This chapter specifically discusses the research methods used to examine factors affecting the effectiveness of AI-based tools in learning English among university students. Using an online survey, the study seeks to obtain details that can help answer the research questions and provide information on the performance and ease of AI instruments in language learning.

**4. Data Analysis**

*4.1 Introduction*

This chapter provides a detailed examination of the responses collected from the participants to investigate the survey results concerning the integration of AI in language proficiency assessment. The purpose is to identify opportunities for improving language assessment through the use of AI technologies in speaking proficiency, adaptive reading, and writing. Consequently, to add a few notes to the benefits and concerns regarding integrating AI-driven assessment tools into educational settings, the chapter compares respondents' demographic profiles and their views on the capabilities of the AI system.

*4.2 Demographic Information*

The impression was randomly created, which is why the number of respondents was considerable, and they were of different ages and educational levels. As for the demography, a quarter of the respondents were below 25 years old, which points to the high engagement of young people in the educational technology market. Another 38% of respondents were 30 years old and above, securing input from people with more experience. Also, 20% of the participants were 20 years of age, bringing new ideas to the process, while 18% were 26 to 30 years old, thus providing a reasonable balance. The gender distribution was equal, with fifty percent of the participants being male while the other fifty percent were female. Concerning education, the respondents had diverse backgrounds; nevertheless, many reported they had postgraduate or PhD degrees, proving high academic activity. Other proclaimed native languages were Arabic, Sindhi, and Urdu to confirm the multiethnicity of the participants.

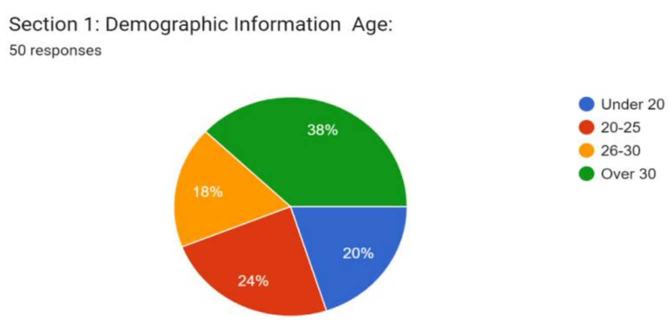


Figure 1. A chart showing the age demographics

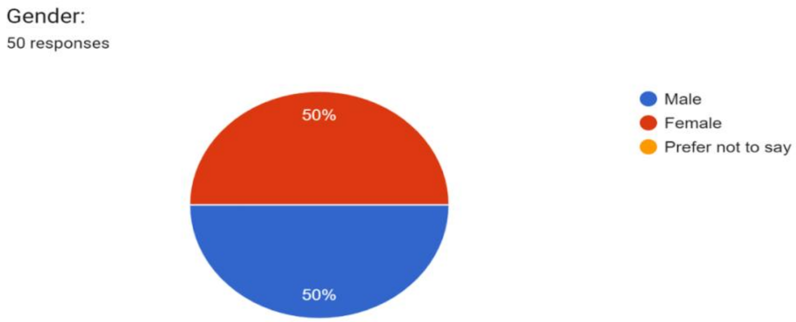


Figure 2. A chart showing the gender demographics

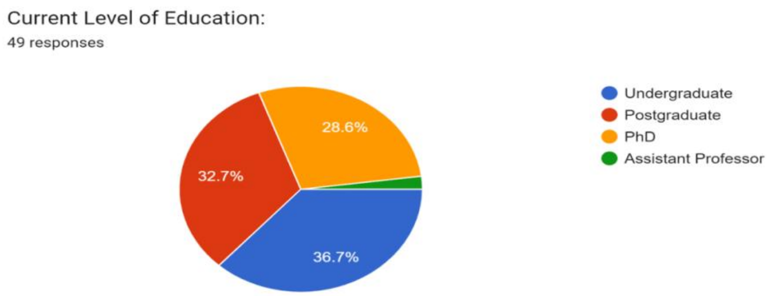


Figure 3. A chart showing the Educational Level of respondents

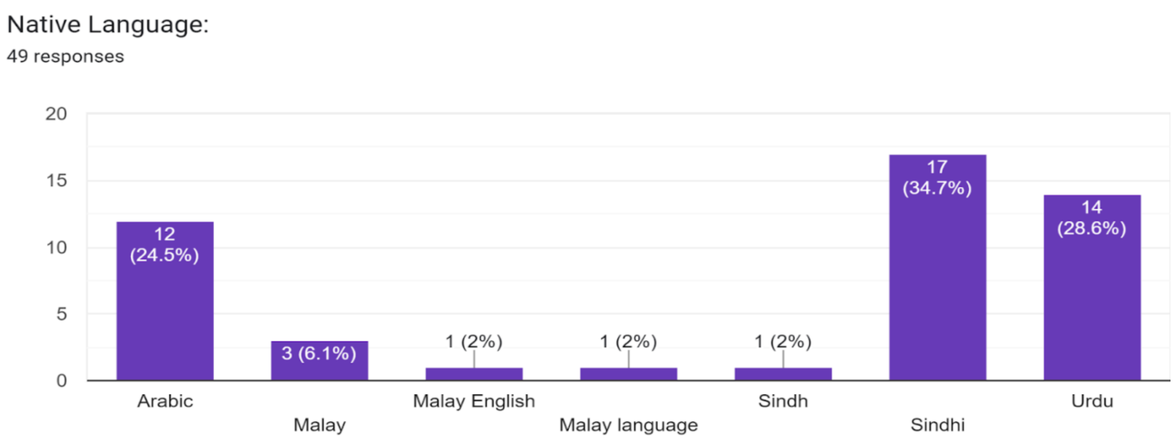


Figure 4. A chart showing the native language of the respondents

*5.3 Analysis of Pie Charts*

5.3.1 Speaking Proficiency Assessment

Concerning the most effective AI method for evaluating speaking fluency, 40% of respondents chose speech recognition, 30% chose pronunciation analysis, 24% perceived value in using both approaches, and 6% were unclear. These results indicate stakeholders' preferences and thus illustrate the plurality of views on using AI tools to select speaking proficiency in educational environments.

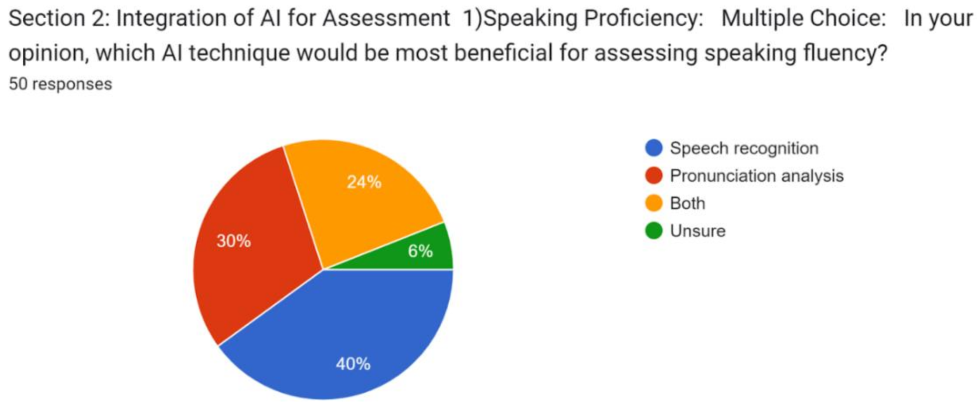


Figure 5. Responses on which AI tool is more beneficial for assessing speaking fluency

5.3.2 Listening Comprehension Assessment

Learners' talker evaluation, whereby their capability to comprehend spoken language at different speeds and accents was evaluated, drew different responses. On the Likert-type scale concerning the general effectiveness of AI, 38% said it was very effective, while the remaining 34% described it as extremely effective. 26% stated it as moderate, while only 2 % strongly agreed with the effectiveness of the same. This evidence suggests substantial agreement on the ability of AI to accurately estimate the levels of listening comprehension irrespective of the language environment.

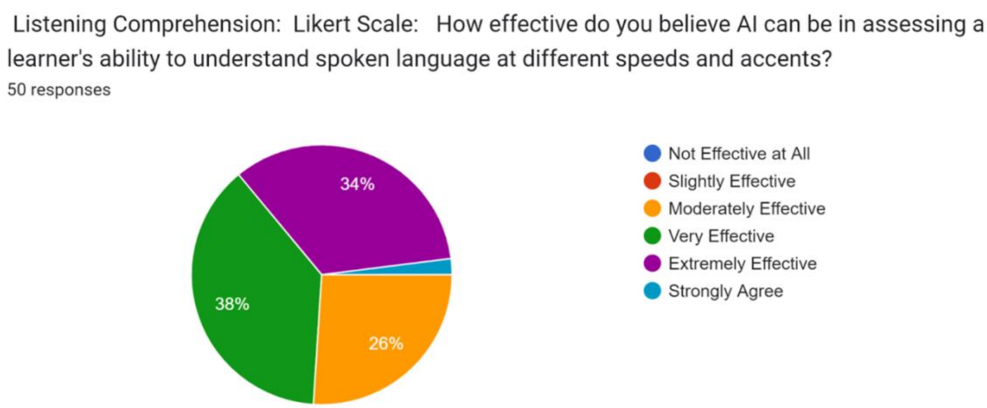


Figure 6. Responses about the effectiveness of AI for assessing a learner's ability to understand spoken language

### *5.4 Integration of AI for Assessment*

#### 5.4.1 Adaptive Reading Assessment

Some topics covered in the survey included, but were not limited to, the respondents' perception of what functionalities of AI could improve on the adaptive reading assessments. Used findings show that 28% of respondents stated text summarization analysis is vital, and 38% underlined the importance of vocabulary recognition tools. Besides, 32% of the respondents suggested it would be most useful if the two components were implemented together. Qualitative responses stressed that AI can help quicken the work of summarizing texts, detecting terms necessary for better understanding, and adapting the text's complexity to learning advancement.

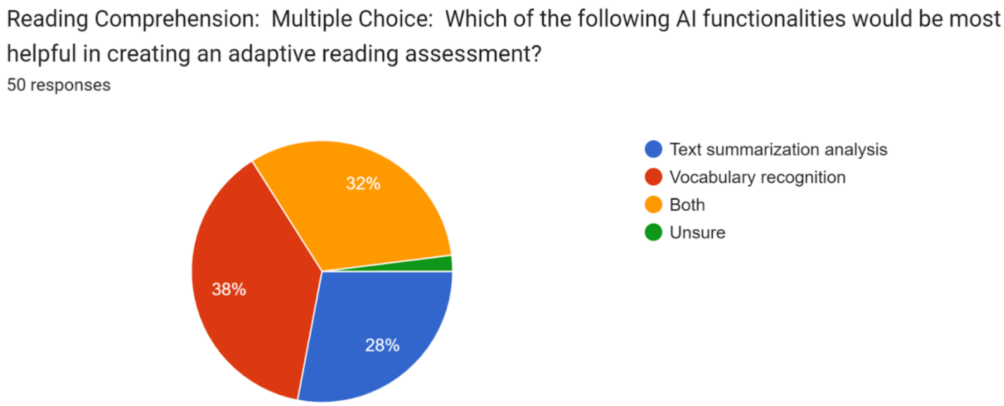


Figure 7. Responses about the suitability of AI functions for creating adaptive reading assessment

5.4.2 Reading Assessment Adjustment

Regarding the flexibility in managing the reading assessments based on the learners' performance, 26% stated moderate flexibility, 32% stated complete flexibility, and 40% foresaw enhanced flexibility. These findings reveal a positive link between the participants' perception of AI adaptability in a learning context and the overall effectiveness of future learning needs to be met.

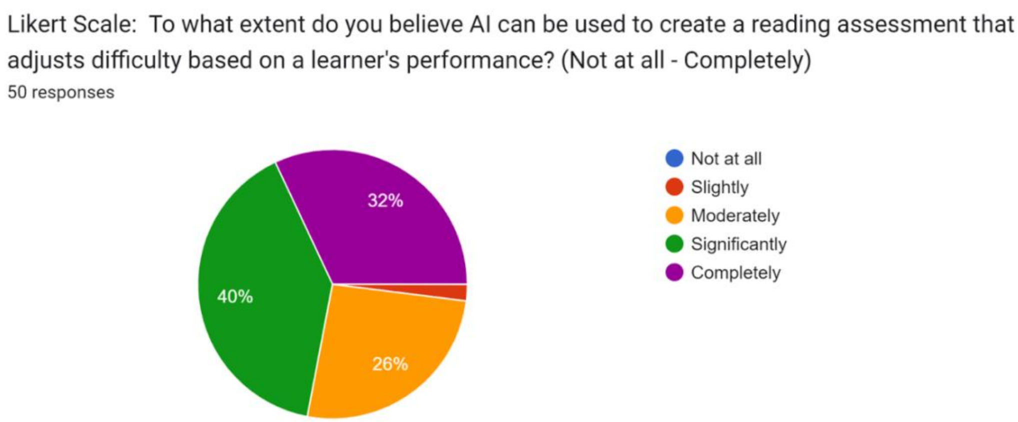


Figure 8. The extent of AI to create a reading assessment that adjusts difficulty based on the learner's performance

5.4.3 Writing Assessment

Interview responses shed light on the organizers' opinions about using AI to assess writing skills. Some of the features highlighted as particularly useful were grammar and mechanics checking (22%), checking the structure of the sentence and the utilized words (40%), as well as the check for plagiarism (10%); only 28% pointed out the importance of the combination of all these characteristics. Narrative responses emphasized two aspects concerning grammar and syntax check, analysis of context relevance and coherence in writing, as well as identification of cases of plagiarism. These functionalities help the learners enhance their vocabulary usage in various forms of writing, including emails and academic essays.

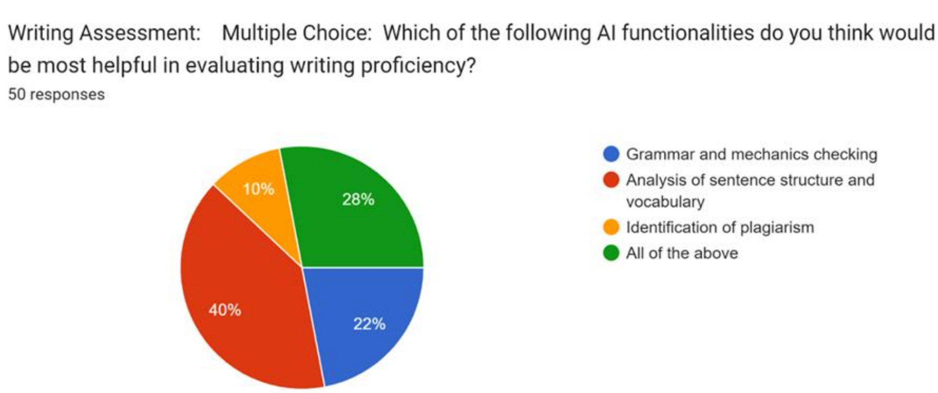


Figure 9. AI's functionalities that can be most helpful in evaluating writing proficiency

### *5.5 Challenges and Considerations*

#### 5.5.1 Limitations of Current AI Tools

Some of the challenges that were discovered about the currently available AI-based language proficiency assessment tools are as follows: As many as 50 percent of respondents mentioned the impossibility of the AI to decode slang or idiomatic expressions. Also, the fears associated with bias owing to the training data (24%) and the difficulties in evaluating the higher-order thinking skills (24%) show that there is more to be desired regarding assigning technologies driven by artificial intelligence.

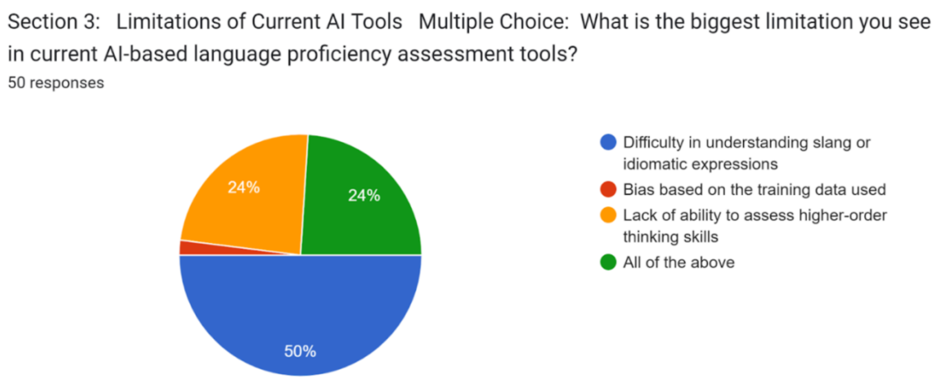


Figure 10. Limitations of AI tools

### *5.6 Holistic AI Assessment System*

To a slightly lesser extent, respondents highlighted the need for AI algorithms to be explainable (38%), feeding them with big and diverse data (24%), and receiving detailed feedback on the peculiarities of their strong and weak points (16%). Most (22%) understood that all these elements must be incorporated to improve the credibility and accuracy of using AI in evaluation processes in education. Also, responses to the level of importance of AI assessments for learners according to their native languages varied from minor to very important, based on the learners. This raises the emergent concern of learners and learning approaches that respect language differences.

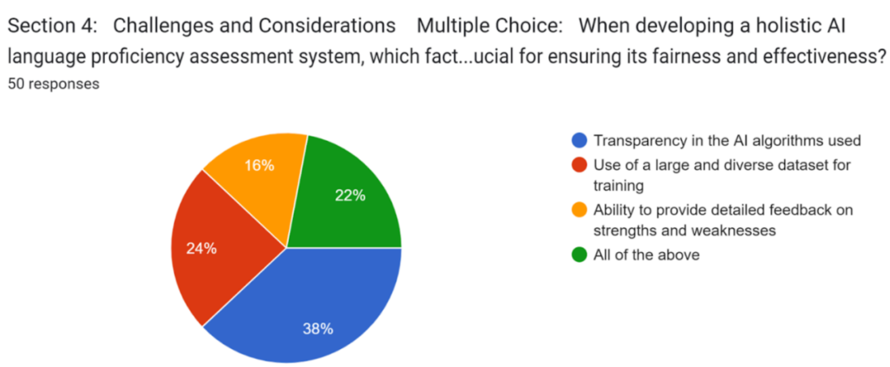


Figure 11. The factor that is most crucial for ensuring its fairness and effectiveness when developing an AI language assessment system

From the chart below, one gets an understanding of the respondents' perception of the suitability of adjusting AI language assessments according to the native language of the learners. While 38% of the respondents considered it "Very Important," 30% chose "Important," which also pointed out that respondents highly value personal assessments. Consequently, 26% of the respondents deemed it 'Moderately Important,' while only 6% felt it was 'Slightly Important.' None of the participants found it 'Not Important.' This establishes research consensus on the importance of customization in AI language assessment to fit all respondents' language profiles.

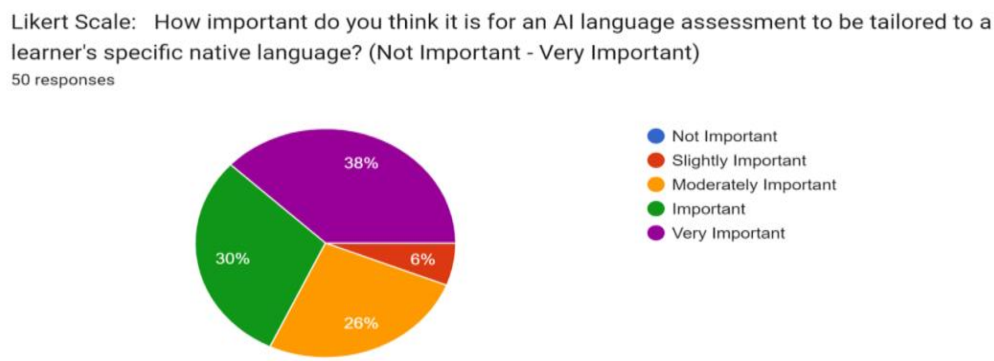


Figure 12. Importance of AI language assessment to be tailored according to the native language

*5.7 Analysis of Manual Response*

Table 1. Provide an example of how an AI-powered listening assessment could be designed to evaluate a learner's ability to follow complex instructions or conversations

| **Respond Number:** | **How AI Can Evaluate Learners' Spoken Language Proficiency** |
|---|---|
| **1.** | AI-powered assessments can be integrated into language learning platforms to track a learner's progress over time, identify areas of improvement, and tailor learning experiences to individual needs. |
| **2.** | AI can generate comprehensive reports on a learner's spoken language proficiency, including scores for clarity, conciseness, fluency, and coherence. This provides actionable feedback for targeted skill development. |
| **3.** | AI-driven assessments can evaluate the appropriateness of speech for different contexts, assessing if a learner can adjust their language register and formality level. |
| **4.** | Semantic analysis in AI can evaluate how well a learner organizes and presents ideas verbally, assessing the clarity of thought and communication of complex concepts. |
| **5.** | By analyzing prosodic features like intonation, rhythm, and stress patterns, AI can assess the naturalness and fluency of speech and capture the nuances of spoken language proficiency. |
| **6.** | Collaborative AI tools can facilitate peer-to-peer evaluations, with AI providing objective assessments and constructive feedback. |
| **7.** | AI can use speech recognition to transcribe spoken words into text, with NLP algorithms analyzing grammar, vocabulary usage, and coherence. |
| **8.** | AI can transcribe spoken language into text, identifying speech patterns, vocabulary usage, and grammatical errors. |
| **9.** | AI can transcribe spoken language into text, identifying speech patterns, vocabulary usage, and grammatical errors. |

Table 2. Describe how AI could be used to evaluate a learner's ability to write clear, concise, and grammatically correct sentences in a specific context (e.g., an email or a formal essay)

| **Example** | **Description** |
|---|---|
| Curated Audio Clips | Careful selection of audio clips covering various difficulties and topics, like daily activities, academic subjects, and professional interactions. |
| Question Variety | Incorporate multiple question types such as MCQs, SAQs, sequencing tasks, and true/false statements for comprehensive assessment. |
| Thematic Focus | Use audio clips that vary in complexity and themes, including everyday scenarios, academic topics, and professional contexts. |
| Detail Comprehension | MCQs to evaluate understanding of key details in the audio. |
| Summarization | SAQs require learners to summarize parts of the conversation. |
| Logical Arrangement | Sequencing tasks to arrange steps or events in the correct order. |
| Context Understanding | True/false statements to assess factual understanding and context comprehension. |
| Complex Dialogues | Present dialogues between multiple speakers on specific topics (e.g., business meetings and medical consultations) for assessment. |
| Real-World Scenarios | Include audio clips from real-world scenarios, such as technical discussions, to test comprehension and inference skills. |

Table 3. Describes how AI Can Evaluate learners ' written Language Proficiency

| **Responds:** | **How AI Can Evaluate Learners' Written Language Proficiency** |
|---|---|
| Automated Scoring Systems | AI assigns scores based on clarity, conciseness, grammar, and coherence, providing an objective evaluation of writing proficiency. |
| Tailored Writing Prompts | AI generates writing prompts for emails or essays and assesses responses for effective communication, format adherence, and grammatical accuracy. |
| Real Time Feedback | AI tools offer immediate feedback as learners write, highlighting areas for clearer, more concise sentences and enabling continuous refinement. |
| Semantic Analysis | AI evaluates the organization and presentation of ideas in emails or essays, ensuring content relevance, grammatical correctness, and coherence. |
| Contextual Language Use | AI assesses tone and formality in different contexts, guiding learners on appropriate language registers for emails and essays. |
| Adaptive Learning Platforms | AI personalizes learning based on individual writing strengths and weaknesses, providing tailored exercises and feedback. |
| Comparative Analysis | AI compares learners' writing to proficient samples, offering benchmarks and guidance for improvement relative to standards. |
| Grammar and Syntax Analysis | AI-powered tools identify and correct grammatical errors, punctuation mistakes, and sentence structures, enhancing writing clarity and correctness. |
| Contextual Analysis | AI algorithms ensure writing relevance, coherence, and logical connection of sentences, assessing overall paragraph structure. |
| Collaborative Writing Tools | AI integrates with writing tools for peer review, providing structured feedback and comparative analysis to foster collaborative learning and skill improvement. |

Table 4. Describe any practical challenges you foresee in implementing a large-scale AI-driven automated language proficiency assessment system

| **Respond:** | **Concerns** |
|---|---|
| **1** | Lack of transparency and accountability in AI scoring. |
| **2** | Potential bias due to non-diverse training data. |
| **3** | Difficulty in adapting to evolving language trends. |
| **4** | Questions about the interpretability of AI-generated scores. |
| **5** | Inaccuracy in detecting nuanced grammatical errors. |
| **6** | Underdeveloped in providing correct and accurate information. |
| **7** | Struggles with understanding idiomatic expressions and cultural nuances. |
| **8** | Inaccurate assessment of spoken language fluency with diverse accents. |
| **9** | Challenges in evaluating complex writing tasks like essays. |
| **10** | Limited in assessing the idea organization and coherence in writing. |

Table 5. Practical Challenges

| **Respond:** | **Practical Challenge** |
|---|---|
| **1** | Continuous system maintenance and updates to ensure relevance. |
| **2** | Securing sufficient financial resources for development and operation. |
| **3** | Addressing regulatory and ethical considerations on data privacy and security. |
| **4** | Providing adequate training and support for educators and administrators. |
| **5** | Gaining user acceptance and trust in AI-driven assessments. |
| **6** | Ensuring integration with existing educational frameworks and technologies. |
| **7** | Adapting the system for scalability across diverse educational settings. |
| **8** | Handling data privacy and security for large volumes of personal data. |
| **9** | Regularly updating AI models to adapt to evolving language trends. |
| **10** | Addressing bias and fairness in AI algorithms for assessments. |

### *5.8 Conclusion*

This chapter consolidates findings by surveying the literature on AI use in language proficiency assessment, providing valuable evidence and insights into the perceived state of affairs and opinions of the stakeholders – educators and learners. AI's applicability is highlighted regarding its employment possibilities for improving educational assessments in terms of accuracy, individual approaches, and speed. However, it also accepts major problems that need to be solved, like cultural sensitivity, eliminating prejudice, and the openness of the assessment procedures. Such conclusions enrich the current debate on the augmentation of AI solutions in the learning process and provide guidance for the development of further work that explores the possibilities of AI in furthering the processes of language learning and assessment.

## 6. Discussion

### *6.1 Introduction*

Considering the results of the literature analysis and the methodology used, this chapter discusses the possibilities of the transition to using AI-based tools to improve the assessment of English language proficiency. Thus, based on the research participants' input and prior research findings, the study will offer a detailed account of AI's opportunities, constraints, and implementation issues in improving language acquisition, especially regarding addressing speaking, listening, reading, and writing skills.

*6.2 Key Findings and Their Implications*

6.2.1 Advancements in Accuracy and Efficiency

Consequently, the literature also keeps emphasizing the increase in efficiency and accuracy of language assessments by using AI. Evanini et al. (2017) provide an example of how automated scoring helps in increasing efficiency in K–12 ELP assessments; the Versant Aviation English Test is an example illustrating how AI rating correlates with reference human rating (IEEE Transactions on Professional Communication, 2010). Such developments also stress AI’s ability to provide accurate determinations, reduce time consumption in the assessment processes, and lessen the workload on the human analyzers.

6.2.2 Implications

Enhanced Efficiency: AI analyzes scoring data, which is the computer process, and makes it easier for scoring devices to give timely feedback to users and assessors.

Consistency: AI being scored ensures the correct marks are met each time, guaranteeing that human error and bias do not substantially affect the grading result.

Scalability: Any new technologies delivered for educators' presence must allow large group assessments and operations, and be usable for many situations.

*6.3 Challenges in Validity and Reliability*

Nevertheless, controversy exists as to the credibility of AI assessments. Literature also stressed the need for cross-checking the automated scoring techniques when establishing test arrangements (Evanini et al., 2017), besides bias in training data that may influence the results (Hauck, Wolf, & Mislevy, 2016).

*6.4 Implications*

Ensuring Validity: Good examples of AI assessments are the ones that measure the right skills users intend to possess.

Bias Mitigation: One of the most important tasks in training data is dealing with bias, which can be tricky. Partly, just selecting the right data can solve the issue of bias, and then fairness and accuracy can be maintained.

Complex Skill Evaluation: Acquiring not only the evaluation of easy basic technical skills, but also a demanding AI mind to struggle through difficult psychological problems.

*6.5 Integration of AI Across Language Skills*

Our findings reveal AI's effectiveness in assessing diverse language skills:

Speaking Proficiency: AI language tools, e.g., voice recognition and pronunciation analysis, give learners much information on the ability of fluency and clarity of the speech (de Wet, Van der Walt, & Niesler, 2009).

Listening Comprehension: Artificial intelligence can assess comprehension in people with diverse accents and at different paces, which is crucial in learning the spoken language.

Reading and Writing: AI applications can boost various tasks, such as text summarization, vocabulary recognition, and grammar checking; thus, they can help create individualized learning and real-time feedback.

*6.6 Implications*

Holistic Skill Development: AI learning tools significantly contribute to the development of language skills through individualized and adaptive learning methods.

Adaptive Learning: AI techniques adapt the tests according to the specific needs of the learners; hence, they help learners to learn language effectively.

Enhanced Feedback: The AI tools give feedback immediately, which improves the raw skills and ensures that the students are taken care of in terms of grammar and coherence.

*6.7 Practical Challenges and Recommendations*

Technological and Implementation Barriers

6.7.1 Implications

6.7.1.1 Continuous Improvement: AI system updates and regular maintenance are extremely important for the effectiveness and security of the AI system.

6.7.1.2 Data Security: Protecting the participants' data is a necessary measure to observe ethical standards and gain the trust of the involved parties.

6.7.1.3 Technological Advancements: It is indispensable to develop conversational AI capable of grasping the complete canvas of linguistic paradigms and cultural contexts.

*6.8 User Acceptance and Training*

Successful integration of AI tools hinges on user acceptance and adequate training for educators and students, addressing misconceptions and enhancing engagement.

*6.9 Implications:*

Educator Training: Helping teachers become AI proficient will improve teaching effectiveness.

Student Engagement: Educating students about AI's benefits contributes to their acceptance and engagement with AI-driven assessment.

## 7. Conclusion

Rehumanizing AI technology's application in language proficiency can greatly contribute to the fact that these instruments tend to be more precise and offer more efficient and personalized learning experiences.

Although the validity issues, the usability of the technology, and the fact that it is far from being what would make the deployment of such systems difficult, addressing the above-listed concerns is of pivotal importance for the realization of the high potential that these systems boast. Through the assistance of AI, the areas that benefit from the use of the technology are emphasized, and its limitations are taken into account, which is to the great advantage of the educational institutions that can bring fundamental advancements to the language assessment processes, matching the skills of the eligible students to address their difficulties.

The study at hand represents AI, which offers excellent opportunities for teachers and students to take advantage of in the education sector, and yet encourages continuous learning. It is also the first step in exploring the limits of AI's application to language learning. Besides this, research and development activities ought to be carried out in diverse academic settings that involve human and machine collaboration, as material objects would.

Moreover, integrating artificial intelligence (AI) within language proficiency assessments represents a watershed moment in educational practices. It offers new pathways toward more accurate assessment results, operational efficiencies, and personalized learning. By synthesizing major findings and associating key insights with them, this paper painstakingly traces the multifarious sides of some AI-driven tools regarding the four critical skills: speaking, listening, reading, and writing.

*7.1 Summary of Findings*

Through a comprehensive literature review and rigorous survey analysis, several pivotal insights have surfaced:

Improved Accuracy and Efficiency: AI-enabled technologies have significantly improved scoring accuracy and consistency for various language skills over the years. Well-designed automated systems streamline the time-consuming aspects of assessments, which have timely feedback that is important to teachers and students alike (Evanini et al., 2017; IEEE Transactions on Professional Communication, 2010).

Concerns About Validity and Reliability: Although it still results in time saved, validity and reliability concerns persist since AI-powered assessments have a rough chance of data bias. It is essential to tackle these issues so that assessments can consider the fairness and non-biased issue of tests and evaluate language capabilities (Hauck, Wolf & Mislevy, 2016).

Integrated Skills: The most advanced AI conducts spoken language fluency evaluations, listening comprehension assessments, and essay writing proficiency checks, which produce personalized feedback to provide individual learning paths. This flexibility boosts academic performance in meeting the needs of a wide range of learners, both native and non-English speaking (de Wet, Van der Walt, & Niesler, 2009).

*7.2 Practical Implications*

There are significant challenges to the practical use of AI in language assessment and legal and training for a broad user base with long-term needs entwined with short-term constraints due to economics and other real-world factors. Overcoming these obstacles necessitates long-term investments in digital innovation, strong data governance systems, and focused educational interventions to impart confidence and proficiency in educators and learners.

*7.3 Future Directions*

In the future, we will need to use AI more effectively in language education (and indeed elsewhere), and work needs to be undertaken to place some of these key areas at its heart, hopefully allowing it to gather speed over time.

AI Algorithm Refinement: Upgrading such algorithms to address better the complex nuances of languages, idiomatic expressions, and cultural contexts will showcase accurate assessments, leading to improved agreement reliability.

Ensuring Bias Mitigation and Transparency: Maintaining fairness in educational outcomes is necessary to work towards unbiased AI systems that provide transparency with minimal artifacts in assessment processes.

Longitudinal Impact Studies: Long-term research of AI-driven assessments' lasting effects on proficiency and achievement will be key to understanding broader impacts.

Pedagogical Integration: Understanding how AI tools can support existing teaching methods (e.g., personalized and adaptive models) will drive further value to both educators and learners.

### *7.4 Conclusion*

This way of thinking demonstrates that AI has the power to change language proficiency assessment. In summary, using AI in assessment to ensure accurate and efficient results with increased personalization will help create a more inclusive learning environment within schools. But to reap these rewards, they need overall data security measures and substantial initiatives to tackle tech complexities, address biases, and improve user experience through structured training and supplemental support.

As AI becomes more of a reality in language education, the possibilities seem limitless for furthering educational equity and excellence. With the constant conceptual innovation and evidence-informed practices of AI-enabled assessment stakeholders, they can work together to create a future where, through more relevant learning experiences and appropriate assessments, learners can excel at their fullest potential in an ever-globalizing society.

**Acknowledgments**

I would like to express my gratitude to Jadara University and its students for their support and participation in this study. Their engagement was invaluable in facilitating the research process.

**Author's Contributions**

Dr. Zein Bani Younes was responsible for the entire study.

**Funding**

Not applicable.

**Competing Interests**

Not applicable.

**Informed Consent**

Obtained.

**Ethics Approval**

The Publication Ethics Committee of the Sciedu Press.

The journal's policies adhere to the Core Practices established by the Committee on Publication Ethics (COPE).

**Provenance and Peer Review**

Not commissioned; externally double-blind peer-reviewed.

**Data Availability Statement**

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

**Data Sharing Statement**

No additional data are available.



**References**

Bahari, A. (2021). Computer-assisted language proficiency assessment tools and strategies. *Open Learning: The Journal of Open, Distance and e-Learning, 36*(1), 61-87. https://doi.org/10.1080/02680513.2020.1790320

Bamdev, P., Grover, M. S., Singla, Y. K., Vafaee, P., Hama, M., & Shah, R. R. (2023). Automated speech scoring system under the lens: Evaluating and interpreting the linguistic cues for language proficiency. *International Journal of Artificial Intelligence in Education, 33*(1), 119-154. https://doi.org/10.1007/s40593-022-00293-w

Benkő, A., & Lányi, C. S. (2007). Developing intelligent emotional agents. In Proceedings of the Regional Conference on Embedded and Ambient Systems (pp. 75-80).

Berlinski, D. (2000). *The advent of the algorithm*. Harcourt Books.

Burstein, J., Chodorow, M., & Leacock, C. (1998). Automated essay scoring: A cross-disciplinary perspective. *ETS Research Report*.

Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics, 1*(1), 1-47. https://doi.org/10.1093/applin/1.1.1

Chalmers, D. J. (1993). Toward a theory of consciousness. Indiana University.

Chapelle, C. A. (2001). *Computer applications in second language acquisition: Foundations for teaching, testing, and research*. Cambridge University Press.

Chassignol, M., Khoroshavin, A., Klimova, A., & Bilyatdinova, A. (2018). Artificial intelligence trends in education: A narrative overview. *Procedia Computer Science, 136*, 16-24.

Chen, L., Chen, P., & Lin, Z. (2020). Artificial intelligence in education: A review. *IEEE Access, 8*, 75264-75278.

Chen, X., Xie, H., Zou, D., & Hwang, G. J. (2020). Application and theory gaps during the rise of artificial intelligence in education. *Computers and Education: Artificial Intelligence, 1*, 100002.

Collier, J. K., & Huang, B. (2020). Test review: Texas English Language Proficiency Assessment System (TELPAS). *Language Assessment Quarterly, 17*(2), 221-230. https://doi.org/10.1080/15434303.2020.1717499

Crevier, D. (1993). AI: The tumultuous search for artificial intelligence. BasicBooks.

de Wet, F., Van der Walt, C., & Niesler, T. R. (2009). Automatic assessment of oral language proficiency and listening comprehension. *Speech Communication, 51*(10), 864-874. https://doi.org/10.1016/j.specom.2009.04.001

Downey, R., Suzuki, M., & Van Moere, A. (2010). High-stakes English-language assessments for aviation professionals: Supporting the use of a fully automated test of spoken-language proficiency. *IEEE Transactions on Professional Communication*, *53*(1), 18-32.

Evanini, K., Hauck, M. C., & Hakuta, K. (2017). Approaches to automated scoring of speaking for K–12 English language proficiency assessments. *ETS Research Report Series, 2017*(1), 1-11. https://doi.org/10.1002/ets2.12147

Gayed, J. M., Carlon, M. K. J., Oriola, A. M., & Cross, J. S. (2022). Exploring an AI-based writing assistant's impact on English language learners. *Computers and Education: Artificial Intelligence, 3*, 100055. https://doi.org/10.1016/j.caeai.2022.100055

Hannaford, P. (2000). Alan Walsh 1916–1998. *Historical Records of Australian Science, 13*(2), 179-206.

Hauck, M. C., Wolf, M. K., & Mislevy, R. (2016). Creating a next-generation system of K–12 English learner language proficiency assessments. *ETS Research Report Series, 2016*(1), 1-10. https://doi.org/10.1002/ets2.12147

Henaff, M., Bruna, J., & LeCun, Y. (2015). Deep convolutional networks on graph-structured data. *arXiv preprint* arXiv:1506.05163.

Hwang, G. J., Xie, H., Wah, B. W., & Gašević, D. (2020). Vision, challenges, roles and research issues of artificial intelligence in education. *Computers and Education: Artificial Intelligence*, *1*, 100001.

Johnson, D. O., Kang, O., & Ghanem, R. (2015). Language proficiency ratings: Human vs. machine. *Pronunciation in Second Language Learning and Teaching Proceedings, 7*(1), 1-12.

Jurafsky, D., & Martin, J. H. (2009). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Prentice-Hall.

Kaplan, A., & Haenlein, M. (2019). Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons, 62*(1), 15-25. https://doi.org/10.1016/j.bushor.2018.08.004

Krashen, S. D. (1982). *Principles and Practice in Second Language Acquisition*. Pergamon Press.

Kuddus, K. (2022). Artificial intelligence in language learning: Practices and prospects. In *Advanced Analytics and Deep Learning Models* (pp. 1-17). *Springer*. https://doi.org/10.1007/978-3-030-83267-5

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436-444. https://doi.org/10.1038/nature14539

Lopez, A. (2020). AI in language assessment: Current trends and future directions. *Journal of Educational Technology*.

McCorduck, P. (2004). *Machines who think* (2nd ed.). A.K. Peters, Ltd.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational Measurement* (3rd ed., pp. 13-103). American Council on Education and Macmillan.

Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for integrating technology in teacher knowledge. *Teachers College Record, 108*(6), 1017-1054. https://doi.org/10.1111/j.1467-9620.2006.00684.x

Müller, V. C., & Bostrom, N. (2016). Future progress in artificial intelligence: A survey of expert opinion. In V. C. Müller (Ed.), *Fundamental issues of artificial intelligence* (pp. 555-572). Springer.

Neisser, U. (1967). *Cognitive Psychology*. Prentice-Hall.

Newquist, H. P. (1994). *The Brain Makers: Genius, ego, and greed in the quest for machines that think*. Macmillan/SAMS.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. Morgan Kaufmann.

Pedro, F., Subosa, M., Rivas, A., & Valverde, P. (2019). *Artificial intelligence in education: Challenges and opportunities for sustainable development.*

Piaget, J. (1972). *The Psychology of the Child*. Basic Books.

Rababah, L. (2025). An Experimental Study of the Effectiveness of Role-play in Improving Fluency in Jordanian EFL Students' Speaking Skills. *World Journal of English Language, 15*(4), 30.

Rogers, E. M. (2003). *Diffusion of Innovations* (5th ed.). Free Press.

Roll, I., & Wylie, R. (2016). Evolution and revolution in artificial intelligence in education. *International Journal of Artificial Intelligence in Education*, *26*(3), 582-599. https://doi.org/10.1007/s40593-016-0110-3

Russell, S. J., & Norvig, P. (2021). *Artificial intelligence: A modern approach* (4th ed.). Pearson.

Sayed, B. T., Bani Younes, Z. B., Alkhayyat, A. et al. To be with artificial intelligence in oral test or not to be: a probe into the traces of success in speaking skill, psychological well-being, autonomy, and academic buoyancy. *Lang Test Asia*, *14*(49). https://doi.org/10.1186/s40468-024-00321-0

Settles, B., Brust, C., Gustafson, E., Hagiwara, M., & Madnani, N. (2013). Duolingo English Test: An AI-powered language proficiency assessment tool. *Duolingo Research Report*.

Vajjala, S. (2018). Automated assessment of non-native learner essays: Investigating the role of linguistic features. *International Journal of Artificial Intelligence in Education, 28*(1), 79-105. https://doi.org/10.1007/s40593-017-0154-7

Van Moere, A., & Downey, R. (2016). Technology and artificial intelligence in language assessment. In D. Tsagari & J. Banerjee (Eds.), *Handbook of Second Language Assessment* (pp. 341-358). *De Gruyter Mouton*. https://doi.org/10.1515/9781614513827-021

Wikipedia contributors. (2024, June 16). *History of artificial intelligence*. Wikipedia. Retrieved from https://en.wikipedia.org/wiki/History_of_artificial_intelligence#:~:text=The%201956%20Dartmouth%20workshop%20was,considered%20the%20birth%20of%20AI

Zawacki-Richter, O., Mar ń, 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

Zechner, K., Higgins, D., Xi, X., & Williamson, D. (2009). Automatic scoring of non-native spontaneous speech in tests of spoken English. *Speech Communication, 51*(10), 883-895.