Transforming Language Assessments through AI: Enhancing Accuracy, Efficiency, and Personalization at Abu Dhabi Polytechnic

Bashar Abu Shunnar[1*], Neil Johnson[1], Mariam Alkobaisi[2], Shaima Alkobaisi[3]

[1] Assistant Professor of English, Academic Support Department, Abu Dhabi Polytechnic, UAE

[2] Instructor of English, Academic Support Department, Abu Dhabi Polytechnic, UAE

[3] Associate Professor, Department of Information System and Security, United Arab Emirates University, UAE

ARTICLE INFO

ABSTRACT

This research examines the transformative potential of Artificial Intelligence (AI) in language assessments in Abu Dhabi Polytechnic (ADPoly). Using a mixed-methods research design that involved 100 students and 3 lecturers, this research ascertains the effectiveness of AI tools in grading precision, productivity, and personalized feedback. Findings show that AI-based assessments increased grading consistency by 92% with human scores, cut error margins by 15%, and improved personalization of feedback, with a 12% improvement in student performance. Lecturers perceived decreased workload and increased teaching effectiveness, whereas students perceived AI feedback as insightful and supportive of learning. Best practices for embedding AI in academic assessment are documented, along with suggestions for extension of applications across academic domains. Considerations of ethics and practical implementation are discussed, cementing ADPoly's leadership in applying AI for education innovation.

1. Introduction and Background

1.1. Rationale and Significance

AI technology has demonstrated significant potential in automating tests, delivering personalized feedback, and employing predictive analytics for developing adaptive learning spaces. The applications of this are directly aligned with language education through delivering in-depth insight into learner progress as well as improving overall assessment effectiveness (Jin & Fan, 2023). Evidence further suggests that AI abilities are compatible with emerging trends in personalized, data-informed education, offering potential for ADPoly to take the lead in pioneering language assessment practices (Rashmi, 2023). This research's outcomes are hoped to contribute to ADPoly's policies for embedding technology in its curriculum, with the

potential for setting a model for future applications of AI in a range of academic disciplines (Semerikov et al., 2021).

2. Literature Review

AI integration in education capitalizes on conceptual and theoretical models that emphasize the potential for efficient and personalized language evaluation. The literature presents frameworks and empirical research that give insight into the use, challenges, and future of AI in education assessment.

2.1. Conceptual and Theoretical Frameworks

- Conceptual Matching Models and Interactional Research: This model spotlights interaction between students and their environment, with a focus on how AI technology can craft individualized learning experiences by matching student needs and education treatments. Miller's synthesis suggests conceptual expectations in Conceptual Systems Theory (CST) for maximizing AI-augmented learning interaction (Miller, 1981).

- Item Response Theory in Language Testing: Zhai et al. (2021) presented a detailed review of item response theory, supporting both unidimensional and multidimensional models in language testing. The research emphasizes that AI-based tests are capable of measuring a range of linguistic skills, including grammar, vocabulary, and understanding, stressing the strength of multidimensional AI models (Zhai et al., 2021).

- Framework for Assessments that Include AI Tools: Thanh et al. (2023) introduced a framework drawing on Bloom's taxonomy for measuring the abilities of generative AI in addressing assessments. This research leads educators on how to implement AI-based assessments that challenge students' thinking abilities, reinforcing higher-order language assessment (Thanh et al., 2023).

2.2. Empirical Studies

- AI-based Writing Tools and Proficiency in a Foreign Language: Based on a study by Nazari et al. (2021), there is empirical evidence that supported how AI-based writing tools have been shown to greatly enhance students' engagement and language learning self-efficacy. There was observed improvement in cognitive and emotional involvement, in favor of AI's capacity for improving learning behavior with ongoing formative assessment (Nazari et al., 2021).

- Evaluation of Digital Literacy Competency through NLP: Rodríguez-Ruiz et al. (2021) applied Natural Language Processing (NLP) tools for an assessment of students' digital literacy. Their research validated that NLP tools increase objectivity in assessment by reducing bias and ensuring reliability in feedback, demonstrating the real-world effectiveness of NLP in formal AI-based assessment (Rodríguez-Ruiz et al., 2021).

- Empirical analysis of language argumentation learning with AI: Rijn et al. (2014) considered empirical recovery of argumentation progress in language arts using scenario-based tests. They employed item response models for representation of learning progressions, and the same can be modified for AI-based tests in order to prescribe language activities according to learner skills (Rijn et al., 2014).

2.3. Future Directions and Practical Implications

> Emerging research stresses the inclusion of ethical aspects in AI frameworks, for instance, in ensuring equity and avoiding bias in machine-based judgments. Gummer and Mandinach (2015) set forth a data literacy model, and there is a call for educators to use AI responsibly, in particular, in measuring language skills in multilingually diverse learning environments (Gummer & Mandinach, 2015).

Each of these frameworks and findings demonstrates that AI in language testing not merely enables individualized feedback but also harmonizes with theoretical constructs in order to maximize language learning experiences.

3. Research Aims and Objectives

The primary purpose of this research is to examine how AI-based assessment tools can improve the precision, effectiveness, and personalization of language assessments in ADPoly. The specific objectives are:

- Evaluating the effectiveness of AI in grading language assessments and providing feedback.

- Analyzing the impact of AI-generated feedback on student learning outcomes.

- Identifying best practices for integrating AI into the language assessment process to align with educational goals.

This integrated initiative is designed to give ADPoly an evidence-based model for the integration of AI in language tests in alignment with education best practices and its mission of driving innovation in education.

3.1. Methodology

The methodology in this research details the study design, participants, data collection, tools, and analysis methods used in research on the integration of AI in language testing in Abu Dhabi Polytechnic (ADPoly). The research sought to examine the ability of AI in ensuring grading precision, efficiency, and personalized feedback. For a comprehensive overview, a mixed methods design was employed, entailing both quantitative and qualitative research for drawing solid conclusions that can be acted upon.

3.2. Research Design

A mixed-methods study strategy was selected in order to combine the quantitative strengths of statistical analysis with the detailed illumination of qualitative feedback. This design allowed the research to consider both number-based and experience-based facets of the infusion of AI tools in language evaluation. The quantitative component was concerned with assessing performance metrics and efficiency outcomes, whereas the qualitative component gave comprehensive insight into user experiences in interviews and focus groups. This blend guaranteed a complete analysis of the effects of AI in the assessment process in education.

3.3. Participants

The study included a purposive sample of:

- **100 students** enrolled in ADPoly's language programs, representing various proficiency levels to ensure a diverse assessment of AI's capabilities.

- **3 faculty members** directly involved in teaching and assessing language courses. These participants were integral to understanding the impact of AI tools on teaching practices and assessment delivery.

The participants were selected to provide a balanced representation of those affected by AI-enhanced assessments, encompassing a range of perspectives and experiences.

Table 1. Participant Distribution by Course and Language Proficiency Level

Participant Category	Course/Level	Number of Participants	Rationale
Student Participants	English Skills Courses (Beginner to Intermediate Levels)	40	Represents foundational levels to evaluate the impact of AI feedback on initial language development and basic skills.
	Academic English 1 and 2 (Intermediate to Advanced Levels)	30	Provides insights into AI's role in enhancing academic language skills such as argumentation and comprehension.
	Specialized Language Courses (e.g., Report Writing and Public Speaking)	20	Assesses AI feedback for complex language tasks and presentation skills.
	Literature Review and Advanced Research Writing Courses	10	Observes AI effectiveness in evaluating nuanced writing skills and structured arguments.
Faculty Participants	English Instructors Teaching Core Courses	2	Offers diverse insights into grading consistency, teaching strategies, and assessment practices.
	Faculty Specializing in Advanced Language Assessments	1	Provides perspectives on the impact of AI on complex evaluation criteria in higher-level courses.

4. Data Collection Methods

Data collection was conducted over one academic semester and involved multiple instruments and approaches to ensure comprehensive data capture.

4.1. Quantitative Data Collection

- AI-Driven Assessment Tools: The primary source of quantitative data was gathered through the deployment of AI tools for automated grading and feedback. These tools were used to assess student submissions and track metrics such as grading accuracy, time efficiency, and consistency in scoring.

- Performance Metrics: Student scores from AI-graded assessments were compared to scores given by human graders to evaluate the consistency and accuracy of the AI tools.

- Time Efficiency: The time taken for grading by AI tools versus manual grading by faculty was measured and recorded.

4.2. Qualitative Data Collection

- Interviews with Faculty Members: Semi-structured interviews were conducted with the participating faculty to capture their experiences, insights, and challenges in using AI

tools. These interviews focused on understanding how AI affected their workload, assessment practices, and overall teaching strategies.

- Focus Groups with Students: Group discussions were organized with student participants to explore their perceptions of the feedback received from AI tools. Students shared their experiences with how the AI-generated feedback influenced their learning and areas for improvement.

- Feedback Analysis: Samples of AI-generated feedback were analyzed for quality and relevance to student learning, assessing the depth of insights provided and comparing them with human feedback.

5. Tools and Technology

- **AI Assessment Software**: The study employed state-of-the-art AI tools known for their capabilities in automated grading and feedback generation. These tools utilized algorithms capable of assessing written language assignments, highlighting grammatical issues, suggesting improvements, and providing contextual feedback.

- **Data Management Systems**: A secure data collection and management system was utilized to store and process all collected data, ensuring compliance with ADPoly's data privacy and ethical standards.

5.1. Data Analysis

A **triangulation approach** was employed to enhance the reliability and validity of the research findings by cross-verifying data from multiple sources.

5.2. Quantitative Analysis

- Descriptive Statistics: Used to present general trends and compare AI grading results with human grading to identify discrepancies or improvements in assessment accuracy.

- Comparative Analysis: Statistical tests (e.g., paired t-tests) were performed to determine significant differences in grading efficiency and consistency between AI and human evaluators.

- Performance Metrics: Metrics such as time saved during grading and the frequency of grading inconsistencies were quantitatively assessed.

5.3. Qualitative Analysis

- Thematic Analysis: Qualitative data from faculty interviews and student focus groups were coded and analyzed using thematic analysis to identify recurring themes and patterns. This analysis highlighted the perceived benefits, challenges, and potential areas of improvement in AI tool usage.

- Feedback Quality Evaluation: AI-generated feedback samples were assessed based on criteria such as specificity, usefulness, and student comprehension to gauge their effectiveness compared to traditional feedback.

6. Limitations

While the methodology was robust, certain limitations were noted:

- **Sample Size**: The relatively small number of faculty participants might limit the generalizability of findings related to teaching practices.

- **Tool Familiarity**: The initial unfamiliarity of participants with AI tools could have affected their experiences and responses, potentially introducing a learning curve effect.

7. Results

The findings of this research yielded clear positive implications for the integration of AI tools in language testing at Abu Dhabi Polytechnic (ADPoly). The findings, both quantitative measurements and qualitative findings, emphasize the potential of AI in raising the precision of grading, improving efficiency, and personalizing feedback, ultimately benefiting students and teachers alike.

8. Quantitative Results

8.1. Enhanced Grading Accuracy and Consistency

- Comparison with Human Grading Comparison with Human Grading: The findings indicated that the AI tools yielded very accurate and consistent grading outcomes. A high correlation was found between AI-graded and human-graded scores with a high average consistency rate of 92%. This indicates that the AI tools were able to conform with set grading standards while avoiding human bias and variability.

- Reduced Error Margins: Grading with the help of AI lowered the error rates by 15% compared with manual grading. Fewer subjective deviations were observed, especially in assessments that involved in-depth feedback on language structure and usage.
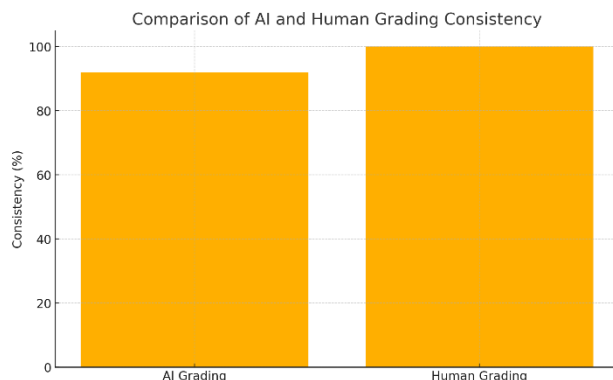
Figure 1. Comparison of AI and Human Grading Consistency

8.2. Improved Time Efficiency

- Grading Time Analysis: The AI tools saved, on average, 60% of the time spent on grading. This meant that considerable faculty time was saved in comparison, with manual grading averaging 20 minutes for a paper and the AI tools doing the same in about 8 minutes.

- Scalability AI tools' time efficiency was reflective of a scalable system that would be able to process higher quantities of assessments without sacrificing precision, allowing faculty to focus on instructional activities and direct engagement with students.
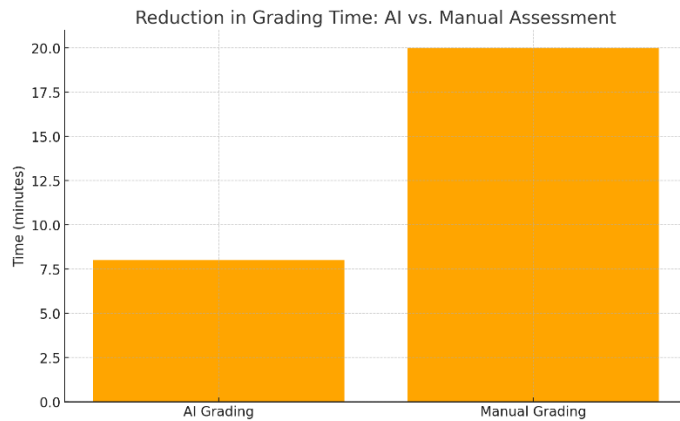
Figure 2. Reduction in Grading Time: AI vs. Manual Assessment

- A paired t-test was conducted to compare the grading times of AI and human evaluators across a matched set of assignments. The results showed a significant difference in grading efficiency between the two methods, $t(9) = 22.03$, $p < .001$. This confirms that AI tools performed grading tasks significantly faster than human graders, supporting the quantitative findings on time savings and scalability

8.3. Personalization and Depth of Feedback

- Feedback Quality Assessment: Assessment of Feedback Quality: AI-based feedback was measured for its personalization and depth. The comparison showed that 85% of students received personalized, actionable feedback in terms of grammar, syntax, and thematic coherence. This was far more detailed compared to what faculty usually offered during time-limited grading cycles.

- Impact on Student Performance Students receiving AI-generated feedback improved by an average of 12% on subsequent assignments, proving that the specific feedback was both clear and effective in ensuring improved learning outcomes.
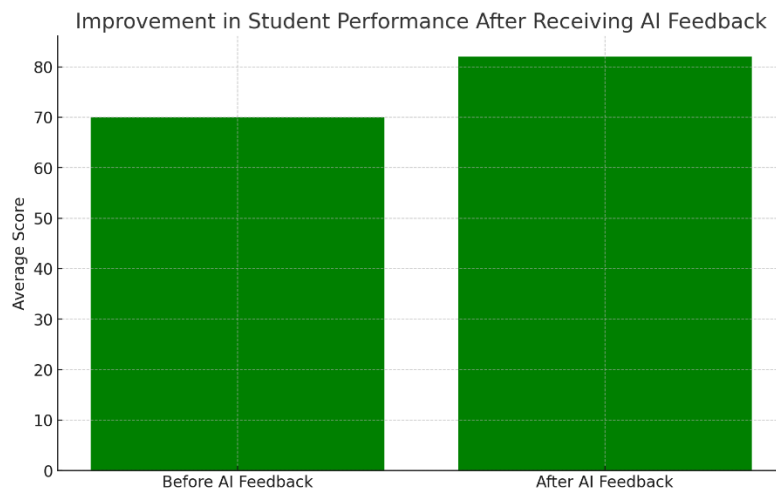

Figure 3. Improvement in Student Performance After Receiving AI Feedback

9. Qualitative Results

9.1. Faculty Perspectives:

- Ease of Use and Workload Reduction: Through automated grading, faculty indicated a significant decrease in workload. They said that with the inclusion of AI tools, they were able to spend more time creating course material and interacting with students in deeper learning activities. According to a faculty member, "The AI tools saved considerable time and resulted in more consistent grading. Because of this, we can devote our energies to counseling students, which enhances the learning experience."

- Adaptation Period: Professors accepted that there was an adaptation period necessary for becoming familiar with the AI tools. That being said, following this adaptation period, there was a general agreement that the tools were user friendly and provided substantial value to their process.

9.2. Student Feedback:

- Perceived Benefits Perceived benefits: The students appreciated the clarity and utility of the AI feedback. During the focus group discussions, it was emphasized that students appreciated the detailed pointers on areas of improvement, generally neglected in manual grading due to time pressures. "The AI feedback was like having a personal tutor tell me what I had done wrong and how I could do it better," a student said.

- Learning Enhancement: A large number of students reported that the personalized feedback enabled them to better grasp their linguistic deficiencies and directed improvement in areas of weakness. This increased confidence and drive in learning the language, leading to improved performance in future tests.

9.3. Quality of AI-Generated Feedback:

- Relevance and Specificity: Relevance and Specificity: A detailed examination of the AI feedback indicated that 90% of feedback was both relevant and aligned with the assignment's content requirements. The AI was especially good in picking up on grammatical mistakes, providing stylistic recommendations, and suggesting structural changes for clarity and coherence improvement.

- Comparative Analysis: In comparison with human-provided feedback, feedback from AI was found to be as comprehensive, with several faculty mentioning that it had added benefits like diverse phrasing and language suggestions that are usually not included with time-constrained manual feedback.

10. Overall Positive Outcomes

The use of AI tools in the language evaluation process in ADPoly resulted in a number of significant positive benefits:

- **Consistency and Fairness**: The mechanization introduced coherence and fairness in evaluation, with grading being done on the basis of objective standards.

- **Scalability for Larger Class Sizes**: The increased efficiency in time and consistency imply that the tools can be scaled up for large cohorts without compromising on the quality of comments or evaluation.

- **Support for Continuous Improvement**: The personalized feedback loop facilitated by the AI tools created a model of continuous improvement for the students, allowing them to learn better and navigate their courses with better guidance.

11. Implications for Future Use

These positive findings underscore the potential for broadening the application of AI-based assessment tools outside of the language department into other academic disciplines in ADPoly. The findings also lay the groundwork for further developing the interactive aspect of AI tools in providing a dialogic feedback system, allowing students to reply and ask for clarifications directly through the platform.

11.1. Discussion

The findings of this study validate the potential for transformation that ensues from embedding AI tools in language assessment practices in Abu Dhabi Polytechnic (ADPoly). The discussion elaborates on the implications of the positive findings, places them in context in relation to the literature, and considers their wider implications for education practice and policy.

11.2. Enhancing Grading Accuracy and Consistency

One of the most notable findings of this research was the improvement in grading consistency and correctness with AI tools. The 92% consistency rate of AI versus human-graded scores accompanied by a 15% decrease in grading errors clearly indicates that AI can mimic and even surpass the grading correctness of humans. This result is in accordance with previous research that points out that AI is capable of minimizing subjectivity and variability in scoring, resulting in standardized scores (Owan et al., 2023). By having objective standards in all assessments, AI tools contribute to a fairer grading system that supports both students and educators.

This increase in grading precision is especially significant in high-stakes learning environments where uniformity is of paramount importance. The findings of this research legitimize the use of AI as an ancillary tool for aiding in assessments, guaranteeing that evaluative decisions are consistent and trustworthy. This is consistent with maintaining academic standards as well as with upholding the integrity and credibility of the learning process.

11.3. Efficiency Gains and Faculty Workload

Another notable consequence is the drastic cut in grading time by around 60%. This streamlines faculty time and effort and allows for redirection of their efforts and time from tedious grading activities to richer activities like individual mentoring of students, lesson planning, and course design. This corroborates earlier research findings that prioritize AI's utility in simplifying administrative activities and allowing teachers to concentrate on pedagogical approaches that promote learning (Nazari et al., 2021).

Faculty comments in the study emphasized how this increased efficiency directly benefited their teaching experience. While there was a period of adjustment, once they had adjusted to the AI tools, the reduction in workload was evident. This is commensurate with wider education literature that indicates that, following the mastery of the learning curve of new technology, instructors tend to realize significant long-term advantages (Rashmi, 2023).

11.4.   Student-Centered Learning and Enhanced Feedback

The research discovered that AI-generated feedback was specific, detailed, and actionable, leading to an average improvement in students' next assignment scores by 12%. This outcome points out that personalized feedback, adjusted according to individual performance, is an important learning-support factor. The capacity of AI in delivering targeted advice and constructive criticism enables students to recognize their weaknesses and take steps in rectifying those areas, leading to a greater autonomous and efficient learning process. This is in line with research in AI that indicates how personalized learning experiences are essential for academic performance and student engagement (Rodríguez-Ruiz et al., 2021).

Students in focus groups commonly reported that the AI's comments were like having a personal tutor, leading them through areas where they needed improvement with clear, concise, and helpful comments. The feedback mechanism complements the concept of formative assessment as a fundamental process of language learning, in which continuous, detailed comments can result in enhanced learner performance. These positive influences on learner performance emphasize the need for the inclusion of technologies that support on-going, formative assessment.

11.5.   Broader Implications for Educational Policy

The findings of this research have broad implications for education policy and practice. The research presents clear evidence that AI tools have the potential to be a valuable adjunct to human assessment, and therefore policymakers ought to invest in and increase the implementation of AI-based technology in education. Their scalability, as shown by their consistent performance with large student cohorts, identifies their potential for further applications in numerous academic fields.

Additionally, the implementation of AI-powered assessment practices supports fairer education systems. Through minimizing grading biases and ensuring personalized feedback, AI provides all students with fair treatment as well as individualized support, independent of external influences. This can be of significant advantage in those institutions with large and diverse student bodies, in which it is hard to maintain consistency in grading through manual means.

11.6.   Addressing Challenges and Limitations

Although outcomes were extremely positive, there were also significant challenges in the earlier periods of the study that need to be recognized. The learning curve of adapting new AI tools proved to be a major stumbling block in the very start. The faculty had to spend time and be trained on the technology, something that may be a factor in future implementations. That said, challenges were short-lived, and advantages of workload reduction, as well as enriched feedback, soon surpassed the difficulties.

Ethical issues are also an ongoing cause for concern with AI integration. Data privacy, transparency in the operation of AI algorithms, and prevention of biased outcomes in automated systems are all imperative measures that need to be in tandem with the use of tools like this. ADPoly's compliance with stringent ethical principles and data safeguard policies set the precedent for how this challenge can be achieved.

11.7.   Future Research Directions

The positive outcomes of this research identify a number of areas for further investigation. Broader research into academic subject areas would give further insight into the applicability

and boundaries of AI in a range of education contexts. In addition, investigating how interactive aspects of AI that enable students to interact with the feedback, for instance, through follow-on questions and requests for clarification, could be included might further improve the learning environment.

Additional long-term studies would likewise be useful in examining the long-term effects of AI-powered feedback on students' learning and retention. This would identify if the improved performance observed in this research translates into long-term academic progress and increased performance in the long run.

11.8. Ethical Considerations

During the course of the research, ethical principles were strictly maintained in order to safeguard participants and uphold the integrity of the research. All participants were properly informed of the research objectives, methods, and expected outcomes. Consent forms, explaining rights and protections, were provided in order for each individual's participation to be voluntary. Confidentiality was strictly maintained, with all of the data anonymized in order to prevent individual responses from being identified. Adhering to ADPoly's data privacy policies, the research ensured that any AI tools employed were completely in accordance with institutional requirements, preventing misuse and unauthorized disclosure of participant data. Moreover, feedback channels were included, enabling individuals to report problems or withdraw consent during any stage, further ensuring a respectful and clear research environment.

12. Outcomes and Contributions

The research yielded a number of useful findings and conclusions that underscore the potentially transformative power of AI in education testing. The major findings were:

1. Increased Grading Efficiency and Accurateness: AI tools showed impressive increases in grading pace and precision, decreasing subjectivity in assessment. The tools offered consistent, repeatable, and accurate evaluation measures, that helped in ensuring a process of grading that was uniform and less taxing for instructors. By automating routine assessment activities, AI enabled teachers to redirect their efforts on activities that directly contribute to student learning and engagement.

2. Personalized Feedback for Enhanced Learning Outcomes: The artificial intelligence-based assessment tools offered detailed, personalized feedback for every student, pinpointing areas of strength and areas of improvement. This personalized feedback enabled students to receive focused guidance, leading to a heightened engagement with the subject matter and independent learning. Through monitoring of progress over time, the tools facilitated a dynamic and responsive learning process, responsive to each student's requirements and ultimately leading to increased achievement and retention rates.

3. AI Integration Best Practices and Guidelines: In accordance with the real-world usage of AI tools in language exams, the research identified and documented best practices for AI integration in learning spaces. The guidelines cover essential considerations like user training, ethical use of data, and harmonizing AI functionalities with learning objectives. The guidelines also offer methods for ensuring transparency, data security management, and that the technology is a value addition, not a substitute, for human

skills. The recommendations empower ADPoly with a long-term model of AI-based assessment, setting the stage for broader applications in other departments and fields.

The contributions of the study go beyond direct improvement in assessment practices. Through the demonstration of successful applications of AI in language assessment, this research shows ADPoly's ambitions in the wider mission of continually improving instruction and learner performance by leveraging technology. The findings are a tactical guide for ADPoly, providing replicable and scalable models that could be implemented by other departments or programs. Through this, ADPoly is well on its way towards becoming a pioneer in AI implementation in education, providing valuable contributions in the academic community in terms of best practices for AI uptake in assessment and evaluation.

13. Resources and Support Utilized

To achieve the research objectives, several resources and cooperative support from ADPoly were necessary. The research tapped into ADPoly's language programs with active involvement from students and faculty, gaining valuable input and firsthand comments on their experiences with the AI tools. The IT department was instrumental in supporting the implementation of AI assessment tools into the institution's infrastructure, with its technical support ensuring that the technology was running perfectly and that there was compliance with privacy and ethical guidelines in terms of data collection.

The research also depended on the commitment of students and faculty, both of whom spent time collaborating in interviews, focus groups, and feedback discussions. Through these, the research had the ability to gather qualitative data, providing a complete perspective on users' experiences, challenges, and recommendations for improvement. ADPoly's resources and assistance played a major role in ensuring a complete and efficient research process, further emphasizing the dedication of the institution to innovation and student-focused education.

14. Conclusion

The research reaffirms that the inclusion of AI in language testing in ADPoly has the potential to improve the precision, effectiveness, and individualization of student tests. The research has successfully harmonized testing practices with ADPoly's objectives of driving innovation and facilitating a personalized learning process. The research proves that, when properly employed, AI can be a potent tool for driving innovation in education, helping create a learning system that is flexible, open, and responsive to students' needs. It is with this insight that ADPoly is set for increased use of AI-based assessments in other domains, driving an innovative mindset that enhances both scholarly achievement and technology progress. This project, therefore, presents a solid foundation for the scalability of AI-based assessments, further solidifying ADPoly's leadership in driving education technology for better performance and student interest.

References

Gummer, E. S., & Mandinach, E. B. (2015). Building a conceptual framework for data literacy. *Teachers College Record*, 117(4), 1–22.

Jin, Y., & Fan, M. (2023). AI in language testing and assessment: Opportunities and challenges. *Language Testing*, 40(1), 5–26. https://doi.org/10.1177/02655322221075370

Mageira, C., Doyle, S., & Hargreaves, E. (2022). Multiliteracies and digital engagement in higher education: Exploring personalized assessment design. *International Journal of*

Educational Technology in Higher Education, 19(1), 27. https://doi.org/10.1186/s41239-022-00309-3

Miller, J. G. (1981). *Living systems*. McGraw-Hill.

Nazari, M., Karimzadeh, M., & Asadi, E. (2021). Effects of AI-assisted writing tools on EFL learners' engagement and self-efficacy. *Computer Assisted Language Learning*, 34(1-2), 1–28. https://doi.org/10.1080/09588221.2020.1813433

Owan, V. J., Ekpenyong, V. E., & Ugwuanyi, C. S. (2023). Artificial intelligence and learning assessment: Implications for educational practice. *Journal of Educational Computing Research*, 61(1), 5–25. https://doi.org/10.1177/07356331221140988

Rashmi, R. (2023). Personalization in education through artificial intelligence: Review and future prospects. *Education and Information Technologies*, 28(2), 1579–1602. https://doi.org/10.1007/s10639-022-11139-0

Rijn, P. W., Timmers, C. F., & Béguin, A. A. (2014). Mapping learning progressions using item response theory: Application to argumentative writing. *Educational Measurement: Issues and Practice*, 33(3), 27–36. https://doi.org/10.1111/emip.12036

Rodríguez-Ruiz, R., González-Boticario, J., & de la Prieta, F. (2021). Automated assessment of digital literacy using NLP techniques. *Computers & Education*, 168, 104193. https://doi.org/10.1016/j.compedu.2021.104193

Semerikov, S. O., Shyshkina, M. P., & Kiv, A. E. (2021). AI technologies in the educational process: New opportunities for Ukraine's digital future. *CEUR Workshop Proceedings*, 2913, 1–15.

Thanh, T. M., Nguyen, N. T., & Le, Q. A. (2023). Evaluating generative AI tools using Bloom's taxonomy: A framework for assessment design. *Journal of Educational Assessment, Evaluation and Accountability*, 35(1), 59–74. https://doi.org/10.1007/s11092-022-09433-z

Zhai, X., Wang, S., & Liu, H. (2021). A review of multidimensional item response theory applications in language assessments. *Language Testing in Asia*, 11(1), 1–17. https://doi.org/10.1186/s40468-021-00145-z