

Assessing Language Competence in Adaptive AI Learning

Karimova Diyora Abduvaxidovna

Tashkent University of Information Technologies

named after Muhammad al-Khwarazmi

Foreign Language Department

diyora.karimova.87@mail.ru

Abstract. *This article examines outlines an integrated model and foundational principles for the responsible implementation of adaptive artificial intelligence (AI) in higher education language proficiency assessment. It addresses critical ethical and practical challenges, focusing on fairness, accountability, validity, and reliability. The text warns against "metric dominance" and unintended biases caused by continuous algorithmic adjustments. To counter this, it proposes a dual-control architecture where pedagogical principles and human oversight govern automated processes.*

Key words: *Adaptive artificial intelligence, language proficiency assessment, higher education, fairness, accountability, validity, hybrid assessment models, dual-control architecture.*

Introduction

A fifth fundamental principle concerns fairness and accountability. Since adaptive artificial intelligence modifies instructional conditions according to individual learner characteristics, equity depends on whether these adaptations broaden educational opportunities or, conversely, limit them without explicit justification. For example, when a system continuously allocates simplified materials to students identified as low-performing, it may inadvertently deprive them of exposure to more sophisticated linguistic structures that are essential for further development. To avoid such outcomes, adaptive mechanisms should incorporate opportunities for increased challenge and guarantee that every learner engages with core activities necessary for achieving programme outcomes.[1]

Accountability is equally important. Adaptive decisions should be systematically documented, and instructors must retain the authority to override automated recommendations whenever pedagogical considerations require it. This principle is particularly relevant in higher education assessment, where decisions related to grades, progression, or certification must be transparent, explainable, and open to scrutiny. Consequently, the educator's role evolves from that of a passive observer into that of a designer, supervisor, and evaluator of adaptive learning pathways.[2]

Based on these considerations, the assessment of students' language proficiency should rely on a clearly structured system of criteria applicable across diverse adaptive trajectories. Such criteria should support both formative purposes, aimed at promoting learning, and summative purposes, intended to evaluate achievement. At the same time, they must capture the multidimensional nature of language competence while remaining practical and measurable.[3]

An effective framework may be organised around the four major language skills and supplemented by common evaluative dimensions, including task fulfilment, linguistic accuracy and range, coherence and cohesion, and strategic competence.[4] Receptive skills additionally require indicators related to the depth of comprehension and inferential understanding, whereas productive skills call for attention to register appropriateness and the effectiveness of interaction. Importantly, all criteria

should be expressed through explicit performance descriptors to ensure a consistent interpretation by both human assessors and AI-supported scoring systems.[5]

Methodology

Validity represents the overarching criterion guiding assessment quality. Evaluation procedures should measure the intended construct itself rather than unrelated proxies. In adaptive contexts, construct-irrelevant variation may increase because students complete different tasks. To preserve validity, the item bank should be developed according to a detailed blueprint ensuring that all adaptive pathways cover equivalent construct domains despite variations in difficulty level. Similar to adaptive testing models, comparability is achieved through calibrated tasks and shared proficiency scales.[6] However, language assessment also requires communicative authenticity and evidence derived from discourse-level performance. Accordingly, validity evidence should include analyses of content alignment, investigations of response processes to confirm that tasks elicit the targeted language behaviours, and correlations with external indicators such as teacher evaluations or established proficiency benchmarks.[7] Reliability, traditionally understood as the consistency of measurement, requires reinterpretation within adaptive systems.

Consistency should not imply identical tasks for all learners; rather, it should ensure that individuals with comparable levels of competence receive similar evaluative outcomes. Reliability may be enhanced by collecting evidence from multiple observations across varied tasks, implementing rubric-based assessment for productive activities, and employing moderation practices. In contexts involving automated scoring, reliability additionally depends on maintaining algorithmic stability and preventing model drift, as updates to scoring systems may alter the meaning of results over time.[8] Therefore, institutions should implement version control for scoring algorithms and conduct regular recalibration using benchmark performances. Methodological traditions in Russian assessment research likewise emphasise that reliability derives from the entire assessment process, including assessor preparation and testing conditions, rather than from instruments alone. Within AI-enhanced environments, this understanding extends to the governance and documentation of algorithmic modifications.[9]

Result and discussion

Objectivity and comparability require that assessment criteria be applied consistently. In evaluating speaking and writing, complete reliance on automated procedures remains problematic in higher education because discourse appropriateness and pragmatic effectiveness are highly context dependent. A more balanced approach involves hybrid assessment models.[10] In such models, AI systems generate preliminary analytical indicators, such as measures of lexical diversity, fluency, or recurring error patterns, while human evaluators apply holistic rubrics to judge communicative effectiveness. The primary challenge is to avoid metric dominance, whereby easily quantifiable indicators overshadow the intended construct. Research in Uzbek pedagogy similarly advocates criterion-referenced assessment that prioritises communicative outcomes while preserving the instructor's methodological oversight. Hybrid approaches also provide valuable opportunities for professional development, as AI-generated feedback can facilitate assessor calibration and encourage students' self-assessment when linked directly to established descriptors.[11]

Fairness as an assessment principle demands sensitivity to learners' linguistic backgrounds, accessibility needs, and technological circumstances. Adaptive environments should not disadvantage students because of factors unrelated to language ability, such as reduced typing speed caused by limitations of digital devices. In listening assessment, exposure to diverse accents should be deliberately incorporated rather than occurring by chance, while reading tasks should minimise topic familiarity bias through careful text selection.[12] Fairness can be investigated through analyses of differential group performance and qualitative reviews examining cultural neutrality. International frameworks emphasise that fairness does not entail identical treatment for all learners; instead, it requires equal opportunities to demonstrate competence according to shared interpretive standards.

Within university settings, this principle necessitates transparent policies concerning permitted resources, testing conditions, and the use of AI tools during instructional activities as opposed to formal assessment.[13]

Finally, this study proposes an integrated model for the responsible implementation of adaptive AI. The model is based on a dual-control architecture in which pedagogical principles define the evidence required, while algorithms optimise learning processes within those predefined boundaries. Educators establish competency frameworks, approve item-bank specifications, and validate assessment rubrics. AI systems personalise practice activities and collect evidence of learning, whereas final evaluative judgments emerge from triangulating multiple sources of information. Adaptive records should therefore be combined with standardised performance tasks administered at predetermined checkpoints.[14]

Such checkpoints are essential because, despite the richness of continuous adaptive evidence, it may be influenced by variations in learning conditions that are difficult to control. Standardised activities—including oral interviews, supervised writing tasks, or proctored reading and listening examinations—can function as interpretive anchors, while adaptive records provide valuable developmental insights that contextualise learner progress.[15]

Conclusion

the ethical and effective application of adaptive artificial intelligence in language education depends upon maintaining alignment with pedagogical objectives, ensuring transparency, securing representative data, supporting formative feedback processes, and upholding fairness through accountability. On this basis, language assessment criteria should demonstrate construct validity, promote reliability, ensure objectivity through clearly documented procedures and rubrics, and guarantee fairness across diverse learners and adaptive pathways. The principal contribution of this approach lies in linking personalised adaptive learning with a stable criteria-based framework and an evidence-informed assessment model. In doing so, it ensures that individualised learning trajectories do not undermine the comparability, credibility, and interpretability of language proficiency judgments in higher education.

References

- [1] J. Jalolov, *Methodology of Foreign Language Teaching: A Textbook*. Tashkent: O‘qituvchi, 2012.
- [2] O‘. Hoshimov and I. Yoqubov, *Methodology of English Language Teaching: A Study Guide*. Tashkent: Sharq, 2003.
- [3] V. S. Avanesov, *Theory and Practice of Pedagogical Measurements*. Moscow: Pedagogika, 2002.
- [4] I. L. Bim, *Methodology of Teaching Foreign Languages as a Science and Problems of the School Textbook*. Moscow: Russkiy yazyk, 1977.
- [5] L. F. Bachman and A. S. Palmer, *Language Assessment in Practice: Developing Language Assessments and Justifying Their Use in the Real World*. Oxford: Oxford University Press, 2010.
- [6] C. A. Chapelle, M. K. Enright, and J. M. Jamieson, *Building a Validity Argument for the Test of English as a Foreign Language*. New York: Routledge, 2008.
- [7] W. Holmes, M. Bialik, and C. Fadel, *Artificial Intelligence in Education: Promises and Implications for Teaching and Learning*. Boston: Center for Curriculum Redesign, 2019.
- [8] H. D. Brown, *Principles of Language Learning and Teaching*, 6th ed. White Plains, NY: Pearson Education, 2014.
- [9] J. C. Richards and T. S. Rodgers, *Approaches and Methods in Language Teaching*, 3rd ed. Cambridge: Cambridge University Press, 2014.
- [10] D. Nunan, *Language Teaching Methodology: A Textbook for Teachers*. New York: Prentice Hall, 1991.
- [11] L. Bachman, *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press, 1990.

- [12] G. Fulcher and F. Davidson, *Language Testing and Assessment*. London: Routledge, 2007.
- [13] M. McNamara, *Language Testing*. Oxford: Oxford University Press, 2000.
- [14] R. Ellis, *The Study of Second Language Acquisition*, 2nd ed. Oxford: Oxford University Press, 2008.
- [15] S. Thornbury, *How to Teach Speaking*. London: Longman, 2005.