

Learning Assessment in an AI World: A New Paradigm

Thomaz Edson Veloso da Silva¹

Center of Excellence in Social Technologies (NEES)
Federal University of Alagoas (UFAL)
Maceió, Alagoas, Brazil
Educometrika LTDA
Fortaleza, Ceará, Brasil
thomaz.veloso@nees.ufal.br
thomaz@educometrika.com.br

May 15, 2025

Abstract

This paper explores the emergence of a new paradigm in learning assessment grounded in Thomas Kuhn's theory of scientific revolutions. Traditional, summative assessment methods increasingly fail to capture the complexity of 21st-century learning, such as creativity, collaboration, and the need for real-time feedback. Anomalies accumulate, signaling a crisis in the current educational paradigm. In response, we argue that integrating generative Artificial Intelligence (AI) in education constitutes a Kuhnian paradigm shift. This shift redefines how learning is evaluated, enabling personalized, formative, and continuous assessment practices embedded in authentic learning experiences. We outline the characteristics of this new paradigm, discuss its ethical and equity implications, and propose a research agenda to support its realization. The goal is not merely to enhance existing tools but to reconceptualize the foundations of assessment in an AI-driven world.

¹This work has been supported by the Brazil's Ministry of Education (MEC) under the number TED 1970.

1 Introduction

Artificial Intelligence (AI) integration into educational systems has grown steadily over the past decades, transforming instructional practices, learning environments, and assessment methods. However, conventional educational assessment relies heavily on standardized, summative testing to evaluate student recall and procedural knowledge (HILLMAYR et al., 2020). While this model provided administrative scalability during the industrial era of schooling, it has struggled to capture the full range of competencies required in contemporary, knowledge-based societies, including creativity, critical thinking, collaboration, and emotional intelligence (SHUTE, 2008).

This paper contends that recent advances in generative AI represent a technological enhancement of assessment tools and a catalyst for a paradigmatic transformation in how learning is assessed. In line with Thomas Kuhn's theory of scientific revolutions, we argue that the accumulation of anomalies in the current assessment paradigm, such as its failure to provide timely feedback, address diverse learning trajectories, and assess higher-order competencies, signals the emergence of a crisis that necessitates a fundamental shift (KUHN, 1962). This new paradigm centers on continuous, personalized, and embedded assessment models enabled by generative AI.

Rather than positioning AI as an external add-on to traditional assessments, we envision a redefinition of assessment, where evaluation is seamlessly interwoven with learning experiences. Generative AI tools offer the potential for real-time, context-sensitive, and adaptive feedback loops that empower both learners and educators. At the same time, this transformation raises profound questions regarding equity, transparency, data privacy, and the pedagogical values underpinning algorithmic decision-making.

This paper seeks to contribute to the design of a more responsive and human-centered educational paradigm by critically examining the philosophical, technical, and ethical dimensions of AI-driven assessment. We begin by situating this transformation within a historical trajectory of AI in education, identifying the constraints of current models, and articulating a vision for the future of learning assessment grounded in educational theory and technological possibility.

2 From Automation to Adaptation

The field of Artificial Intelligence in Education (AIED) has undergone a substantial evolution since its origins in the 1970s. Early efforts were marked by rule-based Intelligent Tutoring Systems (ITS), mirrored traditional pedagogical models by delivering structured content and scripted feedback aligned with predetermined instructional sequences (CORBETT; KOEDINGER; ANDERSON, 1997). These systems reflected a mechanistic view of learning and assessment, emphasizing correctness and efficiency over depth and nuance.

The turn of the millennium brought significant advances in machine learning and educational data mining, catalyzing the development of adaptive learning environments (KOEDINGER et al., 2013). These systems began incorporating learner interaction data to adjust instructional content and pacing, signaling a shift toward more individualized educational pathways. However, assessment practices often remained anchored in performance metrics such as task completion and test scores, reinforcing a narrow conception of learning as observable output.

More recent breakthroughs in deep learning and generative AI have opened new pos-

sibilities for the world and learning assessment. Emerging systems can now interpret unstructured learner inputs, such as open-ended writing, spoken dialogue, and creative problem-solving, and provide real-time, formative feedback. These capabilities depart from static measurement models, introducing the potential for assessments sensitive to learners' cognitive processes, affective states, and social interactions.

Importantly, this historical trajectory reveals technological evolution and the epistemological assumptions underlying assessment. **While early systems aimed to automate traditional models of instruction and testing, contemporary AI invites us to reconsider what constitutes evidence of learning and how it should be interpreted.**

The next section explores how the COVID-19 pandemic further accelerated this shift, exposing the limitations of conventional models and foregrounding the urgency of innovation.

3 COVID-19 as a Catalyst for Educational AI

The global COVID-19 pandemic acted as a profound disruption to traditional educational models, exposing systemic fragilities and forcing rapid innovation across all dimensions of learning. Among the most significant outcomes was the accelerated adoption of digital and AI-enhanced tools to maintain continuity in teaching and assessment under conditions of remote or hybrid instruction (HILLMAYR et al., 2020).

This abrupt transition laid bare the inadequacies of summative, high-stakes assessments, which often depend on in-person supervision and standardized delivery. In contrast, the urgent demand for flexible, scalable, and inclusive assessment approaches has brought renewed attention to the capabilities of AI systems. Generative AI, in particular, demonstrated its value in supporting formative assessment by providing real-time feedback, adapting to diverse learner contexts, and sustaining engagement through conversational and multimodal interfaces (ISOTANI et al., 2023).

Moreover, the pandemic underscored longstanding equity gaps in access to quality educational resources and technologies, emphasizing the need for inclusive strategies to ensure that underserved communities are not left behind in the digital transformation of education. In this context, AI tools must support innovation and respond to ethical imperatives: ensuring fair, transparent, and inclusive deployment that does not exacerbate digital divides (ISOTANI et al., 2023). Thus, the crisis catalyzed a rethinking of assessment in its technological affordances and societal responsibilities.

These lessons persist as we transition to a post-pandemic era. The convergence of necessity and technological possibility during COVID-19 has irreversibly shifted expectations for what educational assessment can and should accomplish.

In the next section, we articulate a vision for AI-driven assessment that responds to these demands with coherence, intentionality, and ethical foresight.

4 A Vision for the Future: New Paradigm for AI-driven Learning Assessment?

As we move beyond emergency responses to educational disruption, there is a growing consensus that returning to traditional assessment practices is neither desirable nor sufficient. The transformative potential of generative AI invites us to rethink the epistemological,

pedagogical, and technological foundations of how learning is evaluated. This section outlines a future-oriented paradigm of AI-driven assessment that emphasizes personalization, continuity, contextual relevance, and learner empowerment.

4.1 From Episodic Testing to Embedded Evaluation

In contrast to conventional assessments that are episodic, summative, and detached from learning processes, AI-driven assessment allows evaluation to occur seamlessly within ongoing educational activities. AI systems can provide continuous formative feedback through real-time analysis of learner inputs across modalities (text, speech, interaction data). This enables educators to make timely pedagogical decisions while empowering learners to take ownership of their development (SHUTE, 2008).

Moreover, generative AI enables the assessment of complex cognitive skills such as critical thinking, creativity, collaboration, and emotional intelligence, areas traditionally challenging to measure through conventional tests. Immersive technologies such as augmented reality (AR) and virtual reality (VR) can enhance assessment capabilities by creating authentic, interactive environments where learners demonstrate competencies through applied tasks and real-world scenarios. Thus, AI-driven assessment represents a paradigm shift, as presented in Figure 1, making evaluation a seamless, integral part of the learning experience rather than a separate, periodic event.

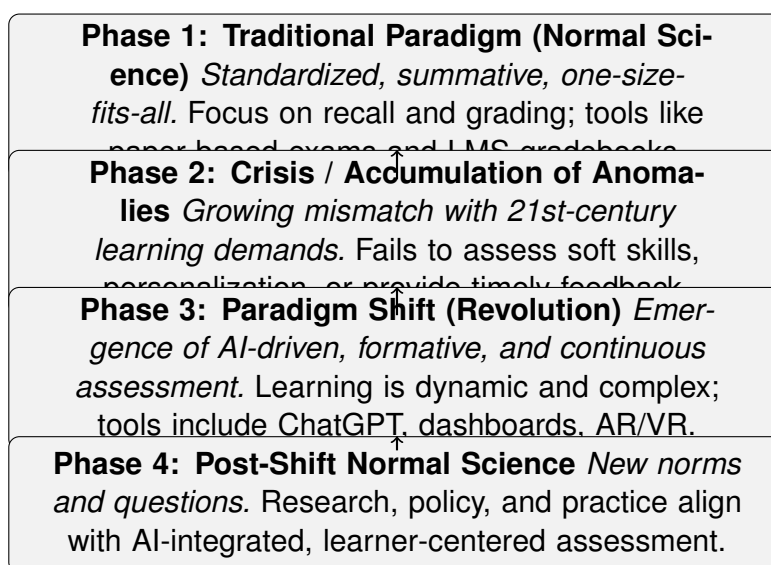


Figure 1: Representation of a Kuhnian paradigm shift in learning assessment, illustrating the transition from traditional summative evaluation toward AI-driven, personalized, and continuous assessment practices in response to the inadequacy of conventional models in addressing 21st-century learning demands.

4.2 Personalization Beyond Adaptivity

While adaptive learning platforms have historically offered content adjustments based on learner performance, generative AI enables a more holistic personalization of the assessment. These systems can evaluate correctness, reasoning processes, creativity, and affective responses. By constructing dynamic learner profiles, AI can tailor the difficulty

and modality of assessment tasks, aligning them with individual goals and cultural contexts (LUCKIN et al., 2016).

Zhao (ZHAO, 2025) argues that educational systems must move away from the traditional "grammar of schooling," characterized by uniform curriculum and standardized assessments, toward models that nurture individual strengths and promote value creation. This perspective reinforces the imperative for AI-driven assessments to move beyond measuring conformity and instead support learners' unique trajectories, talents, and aspirations.

In this sense, AI-driven assessments will be highly personalized, adapting dynamically to individual learners' needs, preferences, and learning contexts. Traditional one-size-fits-all standardized tests fail to accommodate individual variability and often inadequately measure deeper competencies and skills. In contrast, AI systems will leverage comprehensive learner profiles, capturing nuanced cognitive, behavioral, and emotional indicators to deliver tailored assessments that genuinely reflect individual learning journeys.

It is also important to notice that transparency and explainability will become integral components of AI-driven assessment, building trust and understanding among learners and educators. AI assessment methods will evaluate performance and clearly articulate the rationale behind feedback and recommendations, empowering learners to take ownership of their educational development.

4.3 Authenticity and Multimodality

Future assessments must reflect the complexity of real-world learning demands. Generative AI facilitates the creation of immersive, authentic tasks using technologies such as virtual reality (VR) and augmented reality (AR). Learners may be assessed through simulations, design challenges, or collaborative problem-solving contexts in which cognitive and social skills are demonstrated organically. These multimodal environments not only enrich assessment validity but also motivate deeper engagement (KOEDINGER et al., 2013).

Conversely, Zhao's (ZHAO, 2025) emphasis on problem-oriented pedagogy and developing competencies that matter in real life resonates with this vision. When thoughtfully designed, AI-driven assessment can align with Zhao's call to empower students as active agents capable of solving complex, meaningful problems that extend beyond the classroom.

4.4 Ethical Imperatives and Design Principles

A commitment to ethical design must temper AI's power to reshape assessment practices. From the outset, transparency, explainability, and bias mitigation must be embedded in algorithmic systems. Furthermore, equity must remain a foundational concern, particularly in ensuring that technological innovations serve rather than marginalize underserved communities. Achieving this balance requires interdisciplinary collaboration among educators, technologists, ethicists, and policymakers.

Together, these shifts suggest not merely a reform of assessment practices but a paradigmatic realignment of what it means to understand and support learning. The following section illustrates practical scenarios that exemplify this emerging vision.

5 Challenges and Ethical Considerations

As AI-driven assessment systems become increasingly embedded in educational infrastructures, they bring transformative potential and complex ethical and practical challenges. Without careful design and governance, these technologies risk reproducing and amplifying existing educational inequities and undermining public trust.

AI systems trained on historical data risk encoding and perpetuating biases related to gender, race, socioeconomic status, and language. If not addressed, these biases may result in systematically inaccurate or unfair evaluations for marginalized groups. Fairness in AI assessment demands proactive auditing, diverse training datasets, and participatory design practices that include the voices of affected communities (VELOSO et al., 2023).

One of the key ethical concerns in AI-based assessment is the opacity of algorithmic decision-making. Learners and educators must understand how assessment outcomes are generated to ensure accountability and foster trust. Explainable AI (XAI) frameworks should, therefore, be prioritized in the development of educational technologies (LUCKIN et al., 2016).

Continuous and multimodal assessment models rely on collecting and analyzing vast amounts of learner data. This raises significant concerns regarding privacy, consent, and surveillance. Robust data governance policies, including clear terms of use, data minimization principles, and privacy-preserving architectures, are essential to protecting learner rights (SHUTE, 2008).

The digital divide continues to pose a critical barrier to the equitable implementation of AI-enhanced assessment. Students in underserved or low-connectivity contexts may lack access to the necessary hardware, infrastructure, or support systems. Any move toward AI-driven assessment must be accompanied by targeted policies and investments to ensure inclusive access and opportunity (ISOTANI et al., 2023).

In sum, the deployment of AI in assessment must be guided by technical feasibility and a principled commitment to human dignity, justice, and democratic accountability. These challenges are not peripheral but central to the future of trustworthy and transformative educational technology.

6 Research Agenda: What's Needed to Achieve this Paradigm

Realizing the full potential of AI-driven assessment requires a coordinated and multidisciplinary research agenda that addresses foundational, technological, ethical, and contextual dimensions. The paradigm shift outlined in this paper is not self-executing; it demands sustained inquiry, experimentation, and institutional support to transition from vision to practice.

Future research should prioritize the development of AI models capable of assessing higher-order cognitive, social, and emotional skills. These models must go beyond pattern recognition and content retrieval to evaluate reasoning, metacognition, collaboration, and creativity. Building reliable and valid algorithms for these competencies will require iterative design, testing, and validation cycles across diverse learning contexts.

To foster trust and usability, research must focus on creating explainable AI (XAI) systems that clearly articulate the rationale behind their feedback and recommendations. Human-centered design methodologies should ensure educators and learners can interact meaningfully with AI systems and interpret their outputs.

Robust policy frameworks are essential to guide ethical data use in AI assessment. Research should contribute to the development of international standards for data privacy, consent, algorithmic accountability, and fairness. These frameworks must be adaptable to different regulatory and cultural contexts while upholding universal human rights principles.

Finally, more applied research is needed on the real-world implementation of AI assessment tools, especially in low-resource and underserved settings. This includes examining infrastructural constraints, teacher readiness, student perceptions, and the long-term impacts of AI-enhanced assessment on learning equity. Pilot projects and participatory action research can provide actionable insights for scaling responsible innovations (ISOTANI et al., 2023; VELOSO et al., 2023).

Together, these research priorities form a roadmap for operationalizing a new assessment paradigm that is not only technologically sophisticated but also pedagogically sound, ethically grounded, and socially just.

7 Final Remarks

This paper has argued that we are on the cusp of a paradigm shift in educational assessment, driven by the capabilities and implications of generative Artificial Intelligence. Drawing on Kuhn's framework of scientific revolutions, we have positioned this transformation not as a linear technological progression but as a fundamental rethinking of what constitutes valid, ethical, and meaningful evaluation in the 21st century.

Traditional episodic, standardized, and summative assessment models are increasingly misaligned with contemporary learning goals that emphasize creativity, collaboration, and critical thinking. Generative AI offers the means to embed continuous, personalized, and authentic assessment into the learning flow. Yet, these affordances have significant challenges: algorithmic bias, privacy concerns, lack of transparency, and unequal access.

To navigate this transformation responsibly, we proposed a comprehensive research agenda focused on advancing assessment algorithms, ensuring interpretability, building ethical data governance frameworks, and addressing equity through implementation science. Ultimately, the goal is to improve assessment efficiency and align it more closely with pedagogical values and social justice imperatives.

Following Kuhn's perspective, this transition does not represent a mere technological upgrade but a scientific revolution in educational thought: **a necessary shift in the fundamental assumptions that guide how and why we assess learning**. This requires reimagining assessment not as a gatekeeping mechanism but as an ongoing dialogue between learners, educators, and intelligent systems that is dynamic, inclusive, and profoundly human-centered.

References

CORBETT, A. T.; KOEDINGER, K. R.; ANDERSON, J. R. Intelligent tutoring systems. *Handbook of human-computer interaction*, Elsevier, v. 5, p. 849–874, 1997.

HILLMAYR, D. et al. The potential of digital tools to enhance mathematics and science learning in secondary schools: A context-specific meta-analysis. *Computers & Education*, v. 153, p. 103897, 2020.

ISOTANI, S. et al. Aied unplugged: Leapfrogging the digital divide to reach the underserved. In: *Artificial Intelligence in Education*. [S.l.]: Springer, 2023. p. 253–271.

KOEDINGER, K. R. et al. New potentials for data-driven intelligent tutoring system development and optimization. *AI Magazine*, AAAI Press, v. 34, n. 3, p. 27–41, 2013.

KUHN, T. S. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press, 1962.

LUCKIN, R. et al. Intelligence unleashed: An argument for ai in education. *Pearson Education*, 2016.

SHUTE, V. J. Focus on formative feedback. *Review of Educational Research*, v. 78, n. 1, p. 153–189, 2008.

VELOSO, T. E. et al. Its unplugged: Leapfrogging the digital divide for teaching numeracy skills in underserved populations. In: SPRINGER. *Towards the Future of AI-augmented Human Tutoring in Math Learning 2023 - Proceedings of the Workshop on International Conference of Artificial Intelligence in Education co-located with The 24th International Conference on Artificial Intelligence in Education (AIED 2023)*. [S.l.], 2023.

ZHAO, Y. Artificial intelligence and education: End the grammar of schooling. *ECNU Review of Education*, v. 8, n. 1, p. 3–20, 2025. Disponível em: <<https://doi.org/10.1177/20965311241265124>>.