



Beyond Automation: A Conceptual Framework for AI in Educational Assessment

Rareş Fartuşnic

rares.c.fartusnic@gmail.com - corresponding author

Faculty of Social and Behavioural Science, University of Amsterdam, Netherlands

 <https://orcid.org/0009-0008-1009-633X>

Olimpius Istrate

European Baccalaureate Unit, OSG of the European Schools, Brussels, Belgium

 <https://orcid.org/0000-0002-1940-6284>

Ciprian Fartuşnic

Pedagogical Unit, OSG of the European Schools, Brussels, Belgium

 <https://orcid.org/0000-0003-4929-8453>

Abstract: This paper synthesizes the main applications of artificial intelligence in educational assessment through a systematic review following PRISMA guidelines. Our comprehensive analysis yielded 60 studies that revealed five key areas of AI-assisted educational assessment: assessment design, automatic grading, data analysis, performance prediction, and feedback provision. Based on identified patterns and implementation challenges, we propose a novel three-dimensional pedagogical framework for AI in educational assessment. Within this framework, we develop the Processual Assessment Integration Model (P-AI-M) to address the first dimension, distinguishing between assessment design/development and implementation/utilization phases. The complete framework integrates: (1) this processual dimension operationalized through P-AI-M; (2) a stakeholder dimension mapping the distinct roles and responsibilities of researchers, policy makers, school leaders, teachers, and students; and (3) a cognitive-taxonomic dimension aligning AI capabilities with revised Bloom's Taxonomy levels. The model is grounded in established educational theories including assessment for learning, constructive alignment, and sociocultural perspectives on evaluation. By addressing recurring gaps between technological capabilities and pedagogical integration, our multidimensional approach provides educators and researchers with a structured framework for understanding where AI can most effectively enhance assessment while preserving essential human

Received: 3.06.2025. Accepted and published: 29.07.2025

© Rareş Fartuşnic, Olimpius Istrate, Ciprian Fartuşnic, 2025. Published by the Institute for Education (Bucharest). This open access article is distributed under the terms of the Creative Commons Attribution Licence CC BY, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited:

Citation:

Fartuşnic, R., Istrate, O., & Fartuşnic, C. (2025). Beyond Automation: A Conceptual Framework for AI in Educational Assessment. *Journal of Digital Pedagogy*, 4(1) 83-102. Bucharest: Institute for Education. <https://doi.org/10.61071/JDP.2555>

expertise. The framework offers both theoretical grounding and practical guidance for implementing AI assessment tools in pedagogically sound, ethically responsible, and equitable ways across diverse educational contexts.

Keywords: artificial intelligence in education, educational assessment, item construction, automatic grading, data analysis, learning analytics, performance prediction, feedback, AI-supported assessment, Bloom's Taxonomy

1. Introduction

Assessment is a cornerstone of effective educational practice, providing insights into student learning that inform instructional decisions and practices, enabling a more comprehensive examination of effective pedagogies (Zou et al. 2024; Black & Wiliam, 2018; Earl, 2013; Hattie & Gan, 2011). When thoughtfully embedded within the educational process, especially through sustainable feedback, assessment transforms from a mere measurement tool into a dynamic catalyst that guides teaching strategies and empowers students to claim ownership of their learning journey (Boud & Soler, 2016; Carless, 2019). This approach aligns with contemporary educational theories that emphasize the formative potential of assessment to cultivate an active and autonomous learning (Panadero et al., 2018). The capacity for self-directed learning is an imperative acquisition for lifelong learners navigating increasingly complex informational ecosystems.

Artificial intelligence is revolutionising industries worldwide (Rashid & Kausik, 2024; Boulay et al., 2023) and it has emerged as a transformative force in the educational assessment area. It offers unprecedented opportunities to enhance evaluation practices, especially in understanding some of the currently considered "black boxes" of educational processes, such as the increased difficulty of developing and testing theories about learning mechanisms (Luckin et al., 2016; Zawacki-Richter et al., 2019). AI-powered assessment tools can analyse vast quantities of student data with remarkable speed and precision, enabling more dynamic, detailed, and personalized feedback than traditional methods allow, scaling current methods to substantially increased capacities, while also inspiring new ones (Cope & Kalantzis, 2019). New technologies can support important pedagogical aspects as identifying patterns in student performance that might escape human observation, potentially uncovering learning gaps and misconceptions that require targeted intervention (Kellogg et al., 2010), and providing specific support in offering meaningful and individualised feedback to our students (Hattie & Gan, 2011).

Additionally, AI systems can adapt assessment parameters in real-time based on student needs, creating more equitable evaluation experiences that accurately reflect individual capabilities (Holstein et al., 2019). The integration of AI in assessment also presents possibilities for expanding what can be measured in educational contexts. Beyond traditional knowledge recall, AI can facilitate complex competencies such as critical thinking, problem-solving, and creative expression through sophisticated analysis of student work processes and outputs (Chen et al., 2020; Hamada & Hassan, 2017). This capability aligns with contemporary educational goals that emphasize higher-order thinking skills and application of knowledge in authentic contexts.

Furthermore, AI can reduce administrative burdens on educators by automating routine assessment tasks, potentially allowing teachers to devote more attention to instructional design and meaningful student interactions (Larusson & White, 2014).

Thus, researching the intersection of AI and assessment is imperative given both the rapid technological advancement in this domain and the profound implications for educational practice (Holmes et al., 2019; Reich & Ito, 2017). There is an increasing need for guidance and support for educators and learners in the applicative realm of AI-supported assessment, given the growing complexity of assessment practices and the unprecedented development of technology (Boulay et al. 2023). The current study strives to overview the important areas and provide a novel three-dimensional pedagogical framework of practical orientation, embedded in theory, for AI use in educational assessment.

The study followed an analytical process guided by the PRISMA methodology (Gough et al., 2017; Page et al., 2021). A comprehensive search conducted across five major databases (Scopus, Web of Science, ERIC, IEEE Xplore, and Google Scholar) using the search string "(artificial intelligence OR AI OR machine learning) AND (education* assessment OR learning evaluation OR academic measurement)." Initial screening of 542 publications (2019-2024) yielded 155 relevant studies after applying inclusion criteria requiring empirical findings, peer-review, and explicit focus on AI applications in educational assessment. The final analysis included 60 studies after full-text evaluation, coding each study for AI application areas, theoretical foundations, implementation challenges, and reported outcomes.

Analysis of the selected literature revealed five promising areas of AI-assisted educational assessment: assessment design (23%), automatic grading (31%), data analysis (17%), performance prediction (14%), and feedback provision (15%). Key findings included: (1) a consistent bifurcation between pre-assessment and post-assessment AI applications; (2) recurring implementation gaps between technological capabilities and pedagogical integration; (3) varied stakeholder perspectives on AI assessment adoption; and (4) uneven distribution of AI applications across cognitive domains. These five identified areas of AI-assisted educational assessment—assessment design, automatic grading, data analysis, performance prediction, and feedback provision—directly informed our three-dimensional framework. The processual dimension emerged from the bifurcation between pre-assessment and assessment implementation activities, the stakeholder dimension addresses the varied perspectives on AI adoption across educational roles, and the cognitive-taxonomic dimension responds to the uneven distribution of AI applications across different levels of cognitive complexity.

This analysis revealed three distinct but interconnected dimensions that collectively addressed the challenges and opportunities of AI in educational assessment: a process-oriented dimension that distinguishes between different phases of assessment, a stakeholder-oriented dimension that clarifies roles and responsibilities, and a cognitive dimension that maps AI capabilities to established taxonomies of learning. We then iteratively refined this framework through theoretical validation, ensuring alignment with established educational principles including assessment for learning, constructive alignment, and sociocultural perspectives on evaluation. This methodological approach ensured that our framework is grounded in both empirical evidence from current research and theoretical perspectives that have shaped our understanding of effective assessment practices.

The remainder of this paper elaborates on our three-dimensional framework (Section 2), followed by in-depth exploration of assessment design and development (Section 3) and assessment implementation and utilization (Section 4), concluding with implications for practice and future research (Section 5).

This study uses existing literature to support, exemplify, and contextualize the proposed framework. Literature is drawn on to clarify key dimensions, highlight practical implications, and connect emerging AI capabilities to enduring pedagogical concerns. This constructive and integrative approach enables a theory-building orientation, aimed at supporting educators, researchers, and policy makers in navigating the complex and rapidly evolving intersection of AI and assessment. The framework does not claim comprehensive synthesis of all research in the field, nor does it assume within its scope to provide a critical analysis of the methodologies employed and findings re-reported by the evidence-based literature referenced across the article (a section considering limitations will still be discussed, however). Nonetheless, the relevance of this article is drawn from the need for a normative theoretical framework, orienting the practical applications and usage of AI tools in education. It offers a structured, theoretically-informed model to guide future development and implementation efforts, based on a descriptive overview of the current state of AI's integration in the field of education.

The paper introduces a three-dimensional conceptual framework designed to guide the pedagogically grounded integration of artificial intelligence in educational assessment. This model, the Processual Assessment Integration Model (P-AI-M), brings together three intersecting dimensions:

- the process (addressing how assessment is designed, implemented, and utilized),
- the stakeholder (clarifying the roles and responsibilities of actors across the educational ecosystem), and
- the cognitive-taxonomic (mapping AI capabilities to levels of Bloom's revised taxonomy).

Together, these dimensions provide a normative structure for identifying, organizing, and interpreting the ways in which AI technologies can enhance assessment while maintaining pedagogical coherence and ethical integrity.

The P-AI-M framework did not emerge from a traditional systematic review based on inclusion/exclusion criteria or formal bibliometric validation. Instead, it is the product of a constructive, practice-informed synthesis that foregrounds relevance and real-world applicability. The model was built by analysing a wide range of current AI-supported assessment practices already in use – regardless of the level of validation or evidence-based endorsement they may currently hold. This deliberate choice reflects recognition that educators, institutions, and developers are already engaging in meaningful experimentation and instrumentation. By examining what is actively being used, piloted, or proposed in educational settings, the model aims to reflect the lived complexity of practice while offering a structured lens through which those practices can be critically examined, organized, and improved.

2. A Pedagogical Multidimensional Framework for AI in Assessment

When discussing artificial intelligence in educational evaluation, it is essential to establish a clear conceptual framework that situates technological capabilities within sound pedagogical principles. To this end, we propose a **three-dimensional framework – the Processual Assessment Integration Model (P-AI-M)** – that captures the multifaceted nature of AI integration in assessment. This model comprises: (1) a processual dimension, which distinguishes between the design/ development and implementation/ utilization phases of assessment; (2) a stakeholder dimension, which clarifies the distinct roles and responsibilities of re-searchers, policy makers, school leaders, teachers, and students; and (3) a cognitive-taxonomic dimension, which maps AI-supported assessment applications to levels of revised Bloom’s Taxonomy. Together, these three dimensions offer a comprehensive structure for conceptualizing how AI can be meaningfully and ethically embedded into assessment practices. In the sections that follow, we elaborate on each dimension, discussing both established practices and emerging possibilities.

2.1. A focus on the process

Our systematic review revealed a clear pattern in how AI applications in educational assessment naturally cluster into two distinct phases of the assessment lifecycle. The first phase encompasses preparatory activities before student engagement, while the second involves the execution and utilization of assessment data during and after student participation. This bifurcation emerged consistently across the literature, with AI tools and approaches aligning distinctly with either (1) Assessment design and development or (2) Assessment implementation and utilization. This process-oriented conceptualization directly responds to calls in the literature for more integrated assessment approaches that view evaluation as a continuous cycle rather than isolated events (Boud & Soler, 2016; Carless, 2019). Each facet contains specific components that articulate how AI can enhance assessment while maintaining pedagogical primacy.

The first facet – *assessment design and development* – encompasses the preparatory activities that occur before students engage with assessment tasks. Here, AI can assist in strategic planning by analysing curriculum standards and learning outcomes to suggest appropriate assessment approaches. It can support item construction through automated generation of questions, problems, and tasks aligned with specific learning outcomes and offer specific suggestions for personalised tasks. Additionally, AI can enhance quality assurance by identifying potential biases, predicting item difficulty, and ensuring alignment with educational standards (Boulay et al., 2023).

The second facet – *assessment implementation and utilization* – focuses on the activities that occur during and after student engagement with assessment tasks. This includes administration and adaptive delivery, where AI can personalize assessment experiences based on real-time performance. It extends to response analysis and scoring, where AI can evaluate complex student outputs such as essays, projects, and problem-solving approaches. Furthermore, it encompasses interpretation and feedback systems that translate assessment results into actionable insights for both educators and learners. Finally, it addresses learning enhancement, where assessment data drives personalized learning pathways and addresses underperformance through targeted remediation strategies, inspired by the assessment as/for learning paradigm (Schellekens et al., 2021).

Table 1

The two-facets AI-supported assessment framework (on the processual dimension)

Facet 1:	Facet 2:
Assessment design and development	Assessment implementation and utilization
<ul style="list-style-type: none"> • Strategic planning of assessment • Item construction and validation • Assessment assembly and quality assurance 	<ul style="list-style-type: none"> • Administration and adaptive delivery • Scoring and response analysis • Interpretation and feedback • Learning enhancement

This dual-facet framework provides a comprehensive structure for examining the most prominent aspects identified across the research literature: AI for automated assessment and personalized feedback. These themes appear consistently across different studies and contexts, making them the central focus of research in AI for educational assessment. Additionally, our framework incorporates emerging applications and critical considerations to provide a comprehensive understanding of the field, addressed in the next chapters of this article.

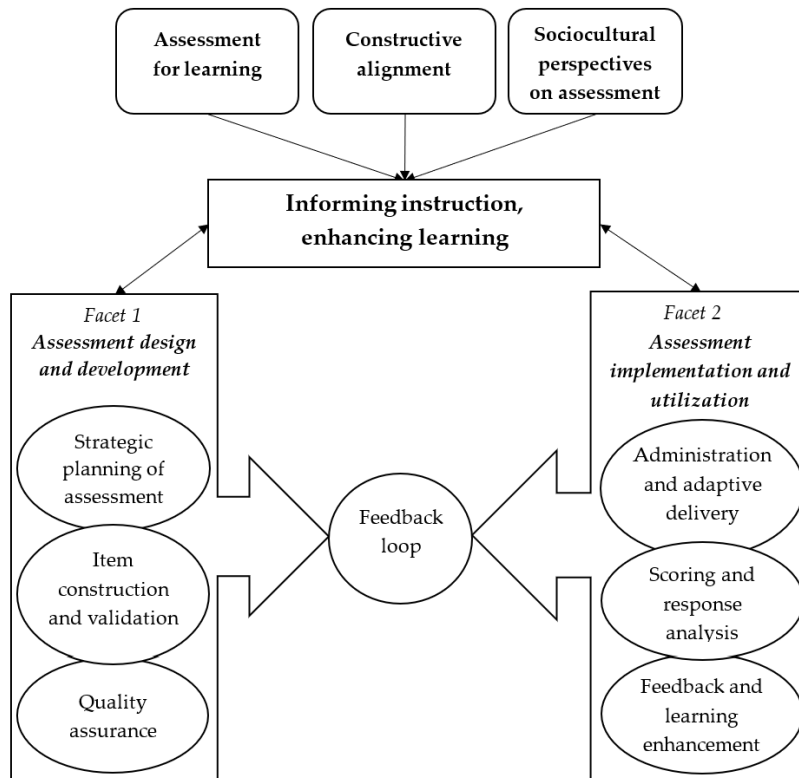
Table 1 denotes an integrated approach acknowledging assessment not as an isolated evaluative event but as an integral component of the educational process that both informs instruction and enhances learning. The merits of this approach are related to the comprehensive understanding of the institutional assessment providing a measure of a student's performance and status at a particular moment in time and intended for a specific use (i.e. judgment of ability, advancement, placement, etc).

Our process-focused framework draws on three key theoretical perspectives:

1. *Assessment for learning*: The processual dimension prioritizes formative functions of assessment, emphasizing how evaluation data can guide instruction and support student learning (Black & Wiliam, 2018; Schellekens et al., 2021). AI amplifies this approach by providing more detailed, timely, relevant and actionable information to support the dynamic features of learning. This perspective goes beyond the common understanding that assessments carry mainly a specific “stake” or consequence for students, offering educators an opportunity to exert influence and control through attainment pressure (Elmore, 2019)
2. *Constructive alignment*: Following Biggs' (2014) principles, the framework ensures that assessment methods are aligned from the curriculum perspective – in particular in relation with the learning outcomes foreseen and instructional learning activities recommended. AI tools enhance this alignment by offering sophisticated means to assess complex learning outcomes, ensure a comprehensive assessment and avoid gaps in coverage.
3. *Sociocultural dimension of assessment*: The framework recognizes assessment as a culturally situated practice (Gipps, 1999) and emphasizes the importance of considering diverse learner backgrounds and needs. AI applications must be designed and implemented with cultural responsiveness in mind, for example by controlling the questions' difficulty in relation to diverse students' cultural backgrounds. In this way, we could better acknowledge how current assessment systems have a deep embedded social and cultural purpose and not always serve as useful information about the development of learners' capabilities (Elmore, 2019).

These theoretical foundations ensure that technological innovations serve pedagogical purposes rather than allowing technology to drive assessment practices. This way, our *Processual Assessment Integration Model (P-AI-M)* emerges as a novel prescriptive framework for the AI use in education assessment. This top-down model uses the three theoretical key perspectives described above as a system of normative orientation for the processes underlying both facets. At the same time, the model concludes in an iterative feedback loop, informing instruction and enhancing learning (see Figure 1).

The processual dimension of our framework is directly informed by three theoretical perspectives identified in our literature review as essential for effective AI-enhanced assessment. Assessment for learning theory (Black & Wiliam, 2018) emphasizes the formative potential of evaluation, which our dual-facet approach captures by connecting design decisions with implementation outcomes. Constructive alignment principles (Biggs, 2014) necessitate viewing assessment as part of an integrated educational system, reflected in our framework's emphasis on linking assessment processes to curriculum and instruction. The sociocultural dimension of assessment (Gibbs, 1999) highlights how evaluation practices are culturally situated, which our process model acknowledges by emphasizing contextual adaptation in both design and implementation phases. Together, these theoretical foundations necessitate a process-oriented approach that views assessment not as a static event but as a dynamic component of the educational journey, with AI enhancing different phases while preserving educational integrity.

Figure 1*The illustration of the Processual Assessment Integration Model (P-AI-M)*

Current studies and evidence will be subsequently utilized in Chapter 3 for illustrating instantiations of the two facets in practice, highlighting current capabilities and instrumentations that could justify the convergence towards the overarching processual dimension.

2.2. A focus on the roles

Our systematic review consistently highlighted implementation challenges stemming from misalignments between different educational actors in AI assessment integration. Successful implementations, by contrast, demonstrated clear role definition and cross-stakeholder collaboration (Tsai et al., 2019; Holstein et al., 2019). This finding aligns with distributed cognition theory and socio-technical systems perspectives, which emphasize how technological implementation requires coordination across multiple actors in complex educational settings (Buckingham Shum et al., 2019).

Drawing on these insights, we developed a stakeholder dimension that maps the ecosystem of AI assessment implementation by identifying five key stakeholder groups—researchers, policy makers, school leaders, teachers, and students—each with distinct responsibilities, concerns, and areas of expertise essential to successful AI integration. For each stakeholder, the table presents three dimensions: primary focus areas that demand their attention, key responsibilities that fall within their purview, and implementation considerations that should guide their decision-making.

Table 2*Stakeholder roles in educational AI assessment*

Stakeholder	Primary focus areas	Key responsibilities	Implementation considerations
Researchers	Validity evidence; Ethical frameworks; Longitudinal impacts	Developing assessment models; Investigating bias; Validating effectiveness	Balance innovation with methodological rigor; Consider educational theory alongside technical capabilities
Policy makers	Governance structures; Equity safeguards; System integration	Creating regulatory frameworks; Ensuring fair implementation; Resource allocation	Develop policies based on evidence; Balance innovation with protection; Support capacity building
School leaders	Professional development; Institutional adoption; Data systems	Leading organizational change; Building assessment culture; Resource management	Prioritize pedagogical purpose; Create collaborative implementation teams; Develop assessment literacy
Teachers	Pedagogical integration; Feedback utilization; Student support	Designing assessments; Interpreting results; Guiding learning	Maintain professional judgment; Focus on formative potential; Connect assessment to instruction
Students	Feedback engagement; Self-regulation; Learning pathways	Self-assessment; Setting goals; Using feedback	Develop assessment literacy; Maintain agency in process; Connect assessment to personal growth

The stakeholder matrix emerged from our analysis of implementation studies that revealed how disconnects between different educational actors often undermined AI assessment initiatives. For example, many studies documented researcher-developed tools that failed to gain traction because they were designed without adequate teacher input (Knox et al., 2019; Selwyn, 2019), or policy frameworks that lacked sufficient guidance for classroom-level implementation (Reich & Ito, 2017). By explicitly mapping the complementary roles and responsibilities across the educational ecosystem, our framework addresses this recurring gap in the literature while acknowledging that successful AI integration depends not just on technological sophistication but on thoughtful human oversight distributed across organizational levels. This dimension prevents technological determinism by distributing accountability and agency across the educational community, a concern frequently raised in our reviewed literature (Holmes et al., 2019).

This framework recognizes that **successful AI assessment integration depends not only on technological capabilities but also on thoughtful human oversight across organizational levels**. By clarifying these complementary roles, the table provides a roadmap for collaborative implementation that balances innovation with ethical considerations, technical capabilities with pedagogical needs, and system-level changes with individual learning experiences. Understanding these varied perspectives is essential for developing AI assessment systems that are not only technically sound but also educationally valuable and equitably implemented.

2.3. A focus on the evaluative tasks

While the stakeholder matrix establishes who should be involved in AI assessment integration and their respective responsibilities, it is equally important to understand what specific assessment functions AI can support within established pedagogical frameworks. Instructional design and, consequently, the educational assessment have long been guided by Bloom's Taxonomy, which provides a hierarchical classification of cognitive processes from basic recall to complex creation. This taxonomy serves as an ideal structure for mapping AI capabilities to specific assessment needs across different levels of cognitive complexity. Therefore, a third and last cognitive-taxonomic dimension is presented.

Our review revealed that current AI assessment tools tend to cluster either at lower cognitive levels (focused on knowledge retrieval and basic understanding) or at advanced levels (supporting complex analysis and creation), with

limited integration across the full spectrum of cognitive processes. This finding highlighted the need for a comprehensive framework that maps AI capabilities to the entire range of cognitive operations. The following table addresses this gap by illustrating how artificial intelligence can complement human skills at each taxonomic level while supporting assessment functions linked to specific learning outcomes. This perspective is deeply rooted in Vygotsky's sociocultural theory of learning, particularly the concept of scaffolding within the zone of proximal development, where AI tools can provide the appropriate level of support needed for learners to progress to higher cognitive levels (Vygotsky, 1997). This theoretical lens informed our understanding of how AI can serve not just as an evaluation tool but as a supportive mechanism that facilitates cognitive development across multiple dimensions.

Furthermore, this approach reflects contemporary assessment theories emphasizing the need for multidimensional evaluation methods that capture the full spectrum of learning outcomes embedded in various national education systems' curricula. Often defined as general and specific competences, these learning outcomes cover a complex array of knowledge, skills and attitudes, defined both at a specific subject level, but also at a wider level. Our approach is therefore relevant to map not only the individual performance of a student at one subject, but also how the learners/graduate competence profiles are met.

By connecting stakeholder roles with taxonomically-organized assessment applications, we establish a comprehensive model that bridges organizational considerations with classroom-level pedagogical practices to ensure AI assessment tools are both theoretically grounded and practically applicable. The framework presented in Table 3 integrates a revised Bloom's Taxonomy with emerging AI capabilities to provide a comprehensive view of how human skills and artificial intelligence can complement each other in educational assessment. It is built based on the Oregon State University Ecampus resource for the teaching staff (the first three columns of the table), to which we added a fourth column dedicated to AI-supported assessment.

Table 3

AI assessment applications mapped to Bloom's taxonomy framework

Level	Distinctive human skills*	How GenAI can supplement learning*	How AI can support assessment/ evaluative tasks**
CREATE	Engage in both creative and cognitive processes that leverage human lived experiences, social-emotional interactions, intuition, reflection, and judgment to formulate original solutions	Support brainstorming processes; suggest a range of alternatives; enumerate potential drawbacks and advantages; describe successful real-world cases; create a tangible deliverable based on human inputs	Assess originality by comparing student work against corpus knowledge/ state of the art; provide constructive feedback on creative products; generate alternative solutions for comparison; evaluate alignment with provided criteria/ rubrics; support portfolio-based assessment
EVALUATE	Engage in metacognitive reflection; holistically appraise ethical consequences of other courses of action; identify significance or situate within a full historical or disciplinary context	Identify pros and cons of various courses of action; develop and check against evaluation rubrics	Analyse evaluation justifications for depth and coherence; provide multi-perspective feedback on ethical reasoning; assess the comprehensiveness of critical reviews; benchmark evaluations against expert examples; generate counterarguments to test robustness of student evaluations
ANALYZE	Critically think and reason within the cognitive and affective domains; justify analysis in depth and with clarity	Compare and contrast data, infer trends and themes in a narrowly - defined context; compute; predict; interpret and relate to real-world problems, decisions, and choices	Identify logical gaps in student analyses; provide immediate feedback on analytical processes; suggest additional analytical perspectives; evaluate the quality of evidence used in arguments; create customized analytical challenges based on student performance patterns

APPLY	Operate, implement, conduct, execute, experiment, and test in the real world; apply human creativity and imagination to idea and solution development	Make use of a process, model, or method to solve a quantitative or qualitative inquiry; assist students in determining where they went wrong while solving a problem	Imagine hypothetical problem situations that the student must solve using particular knowledge and skills; simulate real-world environments for authentic assessment (also possible within VR settings); automate assessment of procedural knowledge through step tracking; provide scaffolded support during application tasks; generate variations of practice scenarios with adaptive difficulty
UNDERSTAND	Contextualize answers within emotional, moral, or ethical considerations; select relevant information; explain significance	Accurately describe a concept in different words; recognize a related example; translate to another language	Automatically grade essays based on given (general) criteria; give knowledge-based hints to support demonstration of understanding; provide ongoing contextual support and supplementary questions (while taking the test) to gauge focused/ deep understanding; analyse concept maps for comprehension assessment; identify misconceptions in student explanations; personalize assessment based on learning pathways
REMEMBER	Recall information in situations where technology is not readily accessible	Retrieve factual information; list possible answers; define a term; construct a basic chronology or timeline	Indicate the most significant information to be included in a test/ exam; organise the flow of knowledge-based items in adaptive tests; generate personalized retrieval practice exercises; implement spaced repetition testing algorithms; create customized flashcards based on individual learning gaps; track knowledge retention over time

* Bloom's Taxonomy Revisited – by Oregon State University Ecampus, 2024 – based on MAGE framework (Zaphir & al, 2024).

** Assessment extension – a proposal for the taxonomy's AI-based assessment dimension; it mostly refers to digital exams.

Organized hierarchically from foundational remembering to complex creation, the table delineates three essential dimensions. The second column identifies distinctive human skills that remain vital in an AI-augmented world, highlighting the irreplaceable aspects of human cognition at each taxonomic level. The third column outlines how generative AI can supplement student learning processes, serving as a tool that enhances rather than replaces human thinking (Oregon State University Ecampus, 2024).

In the last column, we propose a structured list of examples of how AI can support assessment and/or evaluative tasks, providing concrete applications for educational measurement from basic knowledge verification to sophisticated creative evaluation. Moreover, across all distinctive human skills, AI could provide immediate thorough feedback during various stages of assessment, supporting real-time amelioration. **The cognitive-taxonomic dimension** responds directly to recurring calls in the literature for assessment approaches that capture the full spectrum of learning outcomes embedded in contemporary curricula (Cope & Kalantzis, 2019; Chen et al., 2020). By mapping AI capabilities to Bloom's Taxonomy, our framework offers educators a structured approach to understanding where artificial intelligence can most effectively enhance assessment while preserving human expertise in areas requiring judgment, creativity, and ethical reasoning. This dimension addresses the tension identified in our review between technological capabilities and pedagogical needs by illustrating how AI and human assessment approaches can complement rather than replace each other, with each playing distinct roles across different cognitive domains. It further responds to concerns about AI potentially narrowing assessment focus to easily measurable outcomes by explicitly mapping how AI can support evaluation across the full range of cognitive processes, from basic recall to complex creation.

Together, these dimensions offer educators a structured approach to understanding how AI tools can most effectively strengthen assessment practices, while preserving the central role of uniquely human capabilities in learning and evaluation. Subsequently, an overview of the current research landscape will be explored in relation to the present framework, and connections to literature will serve as an illustration of practical options for each element comprised by the processual, stakeholder and taxonomical dimensions.

2.4. Integration of the Three Dimensions

The three dimensions of our P-AI-M framework—processual, stakeholder-oriented, and cognitive-taxonomic—emerged not as isolated components but as an integrated response to the multifaceted challenges of AI integration in educational assessment identified in our systematic review. Each dimension addresses distinct but interconnected aspects of assessment: the 'what' (processual), the 'who' (stakeholder), and the 'how' (cognitive-taxonomic). Together, they form a comprehensive framework that bridges theoretical foundations with practical applications.

The integration of these dimensions creates a dynamic model that reflects the complexity of educational assessment while offering practical guidance for implementation. For example, the processual dimension's distinction between assessment design and implementation is directly linked to the stakeholder dimension's differentiation of roles, with researchers and policy makers typically more involved in design aspects while teachers and students engage more with implementation. Similarly, the cognitive-taxonomic dimension intersects with both the processual dimension (different assessment phases may target different cognitive levels) and the stakeholder dimension (various stakeholders may prioritize different cognitive aspects of assessment).

Our review of the literature highlighted that existing approaches to AI in educational assessment often address only one of these dimensions in isolation, leading to implementation challenges and limited adoption. By integrating these three perspectives, our framework provides a more holistic approach that acknowledges the multidimensional nature of assessment and the complex interplay between processes, stakeholders, and cognitive domains in effective AI integration.

3. Assessment Design and Development

Having established our comprehensive three-dimensional framework—encompassing processual, stakeholder, and cognitive-taxonomic dimensions—we now delve deeper into the first facet of the processual dimension: assessment design and development. While the stakeholder roles (from researchers to students) and cognitive levels (across Bloom's Taxonomy) remain integral considerations throughout, this section examines how AI can enhance the preparatory phases of assessment, from strategic planning and item construction to assembly and quality assurance, providing practical applications that bridge theory with implementation.

3.1. Strategic Planning of Assessment

Artificial intelligence can significantly enhance how educators and institutions conceptualize and develop comprehensive evaluation strategies, moving beyond just automating the grading of completed assessments. This application focuses on using AI to shape the entire assessment framework from conceptualization to implementation.

In test purpose determination and specification, the application of AI in analysing curriculum content effectively identifies key assessment targets (Owan et al., 2023). Large language models demonstrate significant utility in test specification by processing extensive course-related text data, identifying essential themes and concepts, and generating appropriate test item recommendations. AI systems contribute to assessment design through several modalities: identification of critical knowledge and competencies from curriculum materials, recommendation of assessment frameworks that align with established learning objectives, thus ensuring a calibration of the assessment to the class content.

Another important aspect in assessment planning and calibration is the comprehensive coverage of the specified areas of the assessment domain, thus providing content validity. LLMs can analyse content areas and instructional objectives in the cognitive domain, and therefore, they can help educators design assessment methods exhaustively representing

it (Owan et al., 2023). This can be achieved by ensuring appropriate weighting of topics based on curriculum emphasis, align assessment difficulty with intended learning outcomes and include different item types for complementary skill assessment.

AI can also help obtain other validation evidence by analysing response processes of students (processing validity), internal structure of assessments (internal validity), as well as relationships to other relevant variables (criterion validity), all being important for the quality and accuracy of the assessment results, as well as the subsequent confidence in the derived learning goals (Kaldaras et al., 2024).

3.2. Item Construction and Validation

3.2.1. Item Construction

Item construction and validation represent two interconnected areas where AI technologies are making substantial contributions to educational assessment. These applications leverage natural language processing, machine learning, and large language models to enhance both the creation and evaluation of assessment items.

AI has significantly enhanced the process of building assessment items, particularly in high-stakes exams. Traditional methods of item generation often rely on human expertise, which can be time-consuming and prone to bias. AI technologies, such as Natural Language Processing (NLP) and Machine Learning (ML), have introduced automated item generation (AIG) as a viable alternative (Circi et al., 2023). This option allows for increased item production capabilities, especially useful for large-scale assessment. Allowing various types of input, ranging from templates to strong theories, its accessibility creates considerable opportunities for educators. Organized in multiple stages, the automatic item generation process can prepare and understand text for subsequent item production, as well determining usefulness of the generated items (Tan et al., 2024).

Beyond powerful capacities of production, LLMs underlying AIG also have the ability of generating highly structured items with unambiguous prompts and solutions. Exceeding frequently practiced traditional memory-recall questions, AI systems can provide multiple-choice items (with stem, correct answer and possible distractors), essay prompts, short-answer questions, and problem-based items, with the main advantage of increasing accessibility of a diverse plethora of assessment options corresponding to a complementary ability evaluation. There is practical evidence of questions constructed through AI, considered clear and relevant with regard to the assessment subject (Owan et al., 2023).

The power and speed with which these tools can build proposals for assessment items make AI-assisted assessment creation an increasingly valuable option in the perspective of scaling assessment education. AI systems also provide, beyond versatility and accessibility, unprecedented possibilities in large-scale testing, by offering exponentially more resources, as well as by combining AI technologies and psychometric methods to produce substantial joint developments in reliability and validity of assessment.

3.2.2. Item Validation

Research evidence indicates the potential use of neural networks in estimating both person and item parameters from Item Response Theory (IRT), such as item difficulty and item discrimination (the ability of an item to distinguish between increments of respondents' ability level) (Zhang & Chen, 2024). Moreover, AI systems can offer meaningful explanations of parameters across different cognitive domains as well as performance prediction (Su et al., 2021). AI's ability to estimate item parameters at incredibly high precision also allows for applications in more complex methods of cognitive diagnostics (Li et al., 2022).

AI derived options can also provide necessary complementarity for traditional methods that can be difficult to implement in practice, due to their sensitivity to many assumptions. In certain situations, AI systems could be preferred to classical Factor Analysis in studying the item-factor relationship, by being more accurate and adaptable (Milano et al., 2024).

3.3. Assessment Assembly and Quality Assurance

Once individual items have been constructed and validated, they must be assembled into coherent assessment instruments. AI supports this process through sophisticated assembly algorithms and quality control mechanisms.

As iterated earlier, AI supports the strategic selection of items to create cohesive assessments. It can interpret relevant materials and texts, identify redundant or overlapping items, ensure appropriate difficulty distribution and maintain internal consistency, thus providing support in identifying the appropriate items and ensuring convergent validity (Owan et al., 2023).

In terms of validity for the resulting assessment instrument, AI systems provide different support possibilities. Content validity measures using AI are developed for paralleling human evaluation, adding another layer of verification for the complete coverage of the assessment factors' domain by the selected items of the (Milano et al., 2025). Also, powerful options for validating the internal structure in large-scale assessment that use AI are available (Urban & Bauer, 2021). Specifically, algorithms of deep learning provide new performance possibilities in learning how the structural organization of items reflect the latent traits intended for assessment. Spectacular new developments attack some of the hitherto considered "black-boxes" of assessment challenges, namely the response processes. Research using biometrics and machine-learning enables insight in strategies and specific behaviours (Yaneva et al., 2022). These systems can also analyse student behaviour during assessments, offering a more comprehensive understanding of their cognitive and affective states (Liu et al., 2024). For instance, AI-powered platforms can track engagement levels, detect emotional states, and provide real-time interventions to support student well-being (Saputra et al., 2024; Apetorgbor et al., 2024).

These are just a few of the overwhelming tools and methods enabled by AI systems, dedicated for evaluating assessment quality.

AI also allows for important features with respect to the superordinate principle of cultural sensitivity and correction for bias and threats to fairness. AI systems can identify multiple sources of potential bias in assessment materials by analysing the content of test items and signalling specific issues (Owan et al., 2023). These tools can highlight cultural or linguistic barriers that may disadvantage certain students. Also, they can identify stereotypical representations in item contexts. Moreover, the AI systems have the potential to detect interference with item clarity leading to inaccuracies in interpretation by analysing language usage that may create construct-irrelevant variance and noise. Also, it can check assumptions about assessment content that relies on prior knowledge, not covered by the curriculum. AI-powered tools like Finetune (*Generate* and *Catalog*) have been used to align educational content with various standards and frameworks, reducing subjectivity and improving precision (Bolender et al., 2024).

4. Assessment Implementation and Utilization

Building upon our exploration of assessment design and development, we now turn to the second facet of the processual dimension within our three-dimensional framework: assessment implementation and utilization. This section examines how AI transforms the execution, analysis, and educational application of assessments while still acknowledging the distinct roles of various stakeholders and the cognitive processes being evaluated. The implementation phase represents the critical juncture where theoretical design principles meet practical classroom realities, creating opportunities for more responsive, equitable, and pedagogically sound assessment practices.

4.1. Administration and Adaptive Delivery

The second dimension of our framework addresses how AI can enhance the administration, scoring, interpretation, and educational use of assessments. This dimension focuses on the functional aspects of assessment as an integral component of the teaching and learning process.

4.1.1. Testing Environment Monitoring

AI systems can enhance assessment integrity through proctoring capabilities that detect suspicious behaviours, authentication methods to verify test-taker identity, environmental monitoring to ensure testing conditions, anomaly detection in response patterns or testing behaviours (Young, 2024).

4.1.2. Adaptive Assessment Delivery

One of the most innovative contingencies of AI usage in assessment implementation refers to the adaptive testing systems used to adjust the difficulty of questions based on a student's performance, providing a more accurate measure of their abilities (Msayer et al., 2024; Khlaif et al., 2024). Some of the most important advantages refer to precision in dynamically matching item difficulty to estimated student's ability to maintain optimal mental challenge, reduced testing time and test items while preserving measurement precision, and customized assessment pathways based on response patterns.

4.2. Response Analysis and Scoring

Traditional assessment methods often involve manual scoring, which can be time-consuming and prone to human error, or formulaic evaluation, excessively template-driven evaluation struggling with generalization and context.

Answering problems related to resource optimality AI-based systems offer automated grading and real-time feedback, enabling educators to quickly identify areas where students may need additional support. Recently, AI-powered tools started being successfully used in essay scoring, demonstrating high accuracy and consistency (Hartmann, 2023; Mahamuni et al., 2024). Recent studies emphasize the utility of AI tools for automatic grading of assessments, showcasing how AI can streamline the evaluation process. This includes the development of algorithms capable of effectively assessing student performance on various types of assessments, such as essays, quizzes, and practical assignments (Hrich et al., 2024).

For example, AI systems specialised on automated essay scoring (AES) evaluate written essays on dimensions including content relevance and accuracy, organization and coherence, grammar and mechanics, vocabulary usage and style, argument development. Major systems include e-rater (ETS), Intelligent Essay Assessor (Pearson), and features in platforms like Turnitin.

Apart from resource efficiency, the use of AI could also deliver serious contributions to the problem of lack of flexibility and rigid use of templates and answer keys instead of contextual meaning and understanding of the evaluation criteria. For instance, Richardson and Clesham (2021) describe how Pearson's Intelligent Essay Assessor uses latent semantic analysis (LSA), which uses words as vectors projected in a multidimensional semantic space. Words and sentences can be compared on similarity of meaning based on the contexts of the words used (relation to the rest of the text), allowing AI to evaluate meaning rather than merely matching keywords. Evidence of reliability comes in support for their practical use (Richardson & Clesham, 2021). Also, the possibility of evaluating scoring process transparency brings potential interest for educators, due to the capacity of grasping an understanding over the decisions made by the system by interrogating the "reasoning" behind the scoring process (Kaldaras et al., 2024).

4.3. Interpretation and Feedback Systems

AI enhances the interpretation of assessment results by offering comprehensive data analysis, personalized feedback, predictive analytics, and valuable insights for educators. This capability allows for a more responsive educational environment that can quickly adapt to students' needs.

4.3.1. Comprehensive Data Analysis

AI systems can process large volumes of assessment data quickly and efficiently, including both quantitative data (like scores and completion times) and qualitative data (like student-written responses). Machine learning algorithms can identify patterns, trends, and anomalies in student performance that may not be evident through traditional assessment methods.

AI systems excel at identifying patterns in large datasets that might be invisible to human analysts. As Owan et al. (2023) explain, AI can:

- Analyse student response patterns across multiple assessments to identify conceptual misunderstandings.
- Detect correlations between performance on different question types or topics.
- Identify anomalies that may indicate unusual learning challenges or strengths.
- Recognize trends in performance over time that suggest learning progression or regression.

4.3.2. Performance Prediction

AI can use historical assessment data to predict future performance. For example, it can evaluate how students' scores in formative assessments might relate to their expected scores in summative assessments. This predictive analysis can guide teachers in implementing early interventions for students who may be at risk of falling behind (Owan et al., 2023).

Key predictive capabilities include:

- Early identification of students at risk of academic difficulties;
- Projection of learning trajectories based on current performance;
- Estimation of knowledge acquisition rates over time;
- Prediction of performance on future assessments based on current patterns.

4.3.3. Multi-Level Feedback Systems in Learning Enhancement

AI-powered systems can generate detailed reports automatically, providing administrators with actionable insights into student performance (Bolender et al., 2024; Apetorgbor et al., 2024). These reports can include system-wide performance analytics, achievement gap identification, curriculum effectiveness indicators, and resource allocation recommendations, among many others.

AI-based assessment systems enable the identification of learning gaps, allowing educators to tailor instruction to individual needs based on indices of class-level performance patterns, topic-specific mastery levels, instructional effectiveness indicators, and even explicitly recommended teaching interventions.

The role of AI is also recognized in providing real-time feedback for students (Owan et al., 2023). AI can provide immediate feedback based on their assessment performance:

- Specific knowledge and skill gaps;
- Personalized learning recommendations;
- Progress tracking against learning goals;
- Metacognitive guidance on learning strategies.

Teachers and educators could guide intelligent tutoring systems and use collected data for monitoring how performance improvement recommendations are implemented and how effective they are.

5. Concluding Thoughts

The integration of artificial intelligence into educational assessment could represent a major transformation in how we conceptualize, design, and implement evaluative practices. Our systematic review reveals that AI's potential extends far beyond mere automation of existing processes – it offers unprecedented opportunities to reconceptualize assessment as an integral, dynamic component of the educational journey rather than an isolated evaluative event.

Our article explored current developments in AI use for enhancing learning, focussing on education assessment, as one of the most important and promising areas of application. The model proposed assumes that despite the current limitations in rigorously understanding how brains work, the theoretical advances are offering a promising perspective (Churchland and Abbot, 2016). AI tools still have documented limitations from the technological perspective and the areas explored in our normative model accurately follow these limitations. However, we highlight the limitations of our model when considering its real-life applications. The lack of enabling factors (for example, a systemic policy support, an enhanced accessibility, adequate training of teachers), as well as specific contexts (for example, rigid educational cultures and resistance to change, limited funding, insufficient cooperation between key stakeholders) could reduce

the practical value of this model, despite its normative nature. Also, at list at this stage, AI systems are **tools with no inherent moral agency**, trained on available data with its correspondent bias and error. Thus, uses of AI always pose the risk of circular reproduction of error, and their applications are always bound to the quality of its training data.

The *Processual Assessment Integration Model (P-AI-M)* proposed in this paper provides a comprehensive framework that bridges theoretical foundations with practical applications across three critical dimensions that work in synergy:

1. **The processual dimension:** Our dual-facet approach to assessment processes—distinguishing between design/development and implementation/utilization—creates a dynamic framework that recognizes assessment as a continuous cycle rather than a static event. This processual understanding demonstrates how AI can enhance different phases of assessment while maintaining pedagogical integrity. The iterative feedback loops embedded in this dimension ensure that assessment continuously informs instruction and enhances learning in a virtuous cycle.
2. **The stakeholder dimension:** By explicitly mapping the roles, responsibilities, and concerns of researchers, policy makers, school leaders, teachers, and students, our framework acknowledges that successful AI integration depends on coordinated efforts across the educational ecosystem. This multi-stakeholder approach ensures that AI implementation balances innovation with ethical considerations, technical capabilities with pedagogical needs, and system-level changes with individual learning experiences. The stakeholder dimension prevents technological determinism by distributing accountability and agency across the educational community.
3. **The cognitive-taxonomic dimension:** By aligning AI capabilities with Bloom's Taxonomy levels, our framework provides educators with a structured approach to understanding where AI can most effectively enhance assessment while preserving human expertise. This dimension illustrates how artificial intelligence can complement human skills across the spectrum from basic recall to complex creation, maintaining the centrality of uniquely human capabilities in areas requiring judgment, creativity, and ethical reasoning.

The integration of these three dimensions creates a **comprehensive model** that provides both theoretical grounding and practical guidance. When implemented together, they address persistent challenges in educational assessment: resource constraints, scalability limitations, and the need for more nuanced, timely feedback, while preserving the human elements essential to meaningful education.

From a **theoretical perspective**, the P-AI-M framework is grounded in established educational principles including assessment for learning, constructive alignment, and the sociocultural dimensions of assessment. This theoretical grounding ensures that technological innovations serve educational objectives rather than directing or constraining pedagogical practices. The model acknowledges that assessment is not merely a measurement tool but a catalyst that informs teaching strategies and empowers students in their learning journey.

The **practical implications** of our framework are substantial. In assessment design and development, AI offers capabilities ranging from strategic planning and item construction to quality assurance and bias detection. In implementation and utilization, AI enhances administration through adaptive delivery, sophisticated response analysis, comprehensive interpretation of results, and personalized feedback mechanisms. These applications collectively address persistent challenges in educational assessment: resource constraints, scalability limitations, and the need for more nuanced, timely feedback.

We acknowledge both the promise and limitations of our framework. Despite the current technical limitations in AI, the normative model we propose accurately identifies areas where AI can enhance educational assessment while remaining cognizant of its boundaries. However, successful implementation requires enabling factors including systemic policy support, enhanced accessibility, adequate teacher training, and collaborative stakeholder engagement. Moreover, the ethical implications of AI in assessment demand ongoing vigilance regarding data privacy, algorithmic transparency, and equity considerations.

The P-AI-M framework ultimately represents a **balanced approach to AI integration** that neither uncritically embraces technological determinism nor dismisses innovation due to resistance to change. Instead, it offers a structured pathway for thoughtful implementation that enhances assessment practices while preserving educational values. By situating AI capabilities within sound pedagogical principles, the framework helps educators and institutions to meaningfully employ assessment innovation while ensuring that technology serves as a tool for educational empowerment rather than a mechanism that reduces learning to data points.

Important challenges remain in determining how these technologies can be implemented ethically and effectively across diverse educational contexts, ensuring they enhance rather than undermine pedagogical objectives (Knox et al.,

2019; Selwyn, 2019). Vigilance in understanding algorithmic processes and sensitivity to error or inequity is imperative for ensuring principles of fairness, inclusivity and equality of chance (Kizilcec & Lee, 2020). Important research also addresses whether AI assessments produce valid and reliable evaluations across different student populations, addressing concerns about methodological integrity and accuracy of measurement (Saadati, 2023). Additionally, research should explore how educators can best integrate AI assessment tools into their practice while maintaining their professional judgment and pedagogical authority (Tsai et al., 2019). There is also focus on the broader ethical dimensions of AI in assessment, including data privacy, transparency of algorithmic decision-making, and student agency (Buckingham Shum et al., 2019). As educational systems increasingly adopt these technologies, evidence-based guidelines are essential to ensure that AI serves as a tool for educational empowerment rather than a mechanism that narrows curriculum or reduces students to data points or becomes blindly governed by technocentric systems (Selwyn, 2019). By systematically examining both the affordances and limitations of AI in assessment, researchers can contribute vital knowledge that informs policy and practice, ultimately ensuring that technological innovation serves genuine educational advancement (Holmes et al., 2019).

As AI continues to evolve, this framework provides both theoretical grounding and practical guidance for implementing assessment tools in pedagogically sound, ethically responsible, and equitably distributed ways – ensuring that technological advancement serves genuine educational enhancement rather than narrowing curricular focus or diminishing the rich complexity of human learning.

Bibliography

- Apetorgbor, M., Narad, S., Akpabio, E., Dhawale, C., Konto, A., & Verma, P. (2024). Leveraging Artificial Intelligence for Effective Assessment and Evaluation in Education: A Comprehensive Review. *2nd DMIHER International Conference on Artificial Intelligence in Healthcare, Education and Industry (IDICAIEI)*, Wardha, India, 2024, pp. 1-6, <https://doi.org/10.1109/IDICAIEI61867.2024.10842940>
- Biggs, J. (2014). Constructive Alignment in University Teaching. *HERDSA Review of Higher Education*, 1, 5-22.
- Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy & Practice*, 25(6), 551–575. <https://doi.org/10.1080/0969594X.2018.1441807>
- Bolender et al. (2024). Generative AI in K12: Analytics from Early Adoption. *Journal of Measurement and Evaluation in Education and Psychology*, 15(Special Issue), 361-377. <https://doi.org/10.21031/epod.1539710>
- Boud, D., & Soler, R. (2016). Sustainable assessment revisited. *Assessment & Evaluation in Higher Education*, 41(3), 400-413. <https://doi.org/10.1080/02602938.2015.1018133>
- Buckingham Shum, S. J., Ferguson, R., & Martinez-Maldonado, R. (2019). Human-centred learning analytics. *Journal of Learning Analytics*, 6(2), 1-9. <https://doi.org/10.18608/jla.2019.62.1>
- Benedict du Boulay (ed), Antonija Mitrovic (ed), Kalina Yacef (ed). 2023. Handbook of Artificial Intelligence in Education. Edward Elgar Publishing. Cheltenham. <https://doi.org/10.4337/9781800375413>
- Bulut, O., Beiting-Parrish, M., Casabianca, J.M., Slater, S.C., Jiao, H., Song, D., Ormerod, C., Fabyi, D.G., Ivan, R., Walsh, C., Rios, O., Wilson, J., Yildirim-Erbasli, S.N., Wongvorachan, T., Liu, J.X., Tan, B., & Morilova, P. (2024). The Rise of Artificial Intelligence in Educational Measurement: Opportunities and Ethical Challenges. <https://doi.org/10.48550/arXiv.2406.18900>
- Carless, D. (2019). Feedback loops and the longer-term: Towards feedback spirals. *Assessment & Evaluation in Higher Education*, 44(5), 705-714. <https://doi.org/10.1080/02602938.2018.1531108>
- 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. <https://doi.org/10.1016/j.caeai.2020.100002>
- Churchland, A. K., Abbot, L. F. (2016). Conceptual and technical advances define a key moment for theoretical neuroscience. *Natural Neuroscience* 2016; 19:348–349. <https://doi.org/10.1038/nn.4255>
- Circi, R., Hicks, J., & Sikali, E. (2023). Automatic item generation: foundations and machine learning-based approaches for assessments. *Frontiers in Education*, 8. <https://doi.org/10.3389/feduc.2023.858273>
- Cope, B., & Kalantzis, M. (2019). Education 2.0: Artificial intelligence and the end of the test. *IEEE Transactions on Learning Technologies*, 13(3), 528-543. <http://dx.doi.org/10.1163/25902539-00102009>

- Earl, L. M. (2013). *Assessment as learning: Using classroom assessment to maximize student learning* (2nd ed.). Corwin Press.
- Elmore, R. F. (2019). The Future of Learning and the Future of Assessment. *ECNU Review of Education*, 2(3), 328-341. <https://doi.org/10.1177/2096531119878962>
- Fernández-Sánchez, A., Lorenzo-Castiñeiras, J.J. and Sánchez-Bello, A. (2025), Navigating the Future of Pedagogy: The Integration of AI Tools in Developing Educational Assessment Rubrics. *Eur J Educ*, 60: e12826. <https://doi.org/10.1111/ejed.12826>
- Gibbs, G. (1999). Using Assessment Strategically to Change the Way Students Learn. In S. Brown, & A. Glasner (Eds.), *Assessment Matters in Higher Education: Choosing and Using Diverse Approaches* (pp. 1250-1263). The Society for Research into Higher Education & Open University Press.
- González-Calatayud, V., Prendes-Espinosa, P., & Roig-Vila, R. (2021). Artificial Intelligence for Student Assessment: A Systematic Review. *Applied Sciences*, 11(12), 5467. <https://doi.org/10.3390/app11125467>
- Gough, D., Oliver, S., & Thomas, J. (2017). Introducing systematic reviews. In: Gough, D and Oliver, S and Thomas, J, (eds.) *An Introduction to Systematic Reviews*. (pp. 1-16). Sage Publications: London, UK. https://books.google.be/books/about/An_Introduction_to_Systematic_Reviews.html?id=ucf0AwAAQBAJ
- Hamada, M., & Hassan, M. (2017). An interactive learning environment for information and communication theory. *Eurasia Journal of Mathematics, Science and Technology Education*, 13(1), 35-59. <https://doi.org/10.12973/eurasia.2017.00603a>
- Hartmann, P. (2023). Digitization of High-Stakes Exams. Empirical Insights and Design Recommendations for the Digital Execution and Scoring of Exams. (Dissertation). Göttingen Graduate School of Social Sciences: Göttingen, 2023. <https://doi.org/10.53846/goediss-10080>
- Hattie, J., & Gan, M. (2011). Instruction based on feedback. In R. E. Mayer & P. A. Alexander (Eds.), *Handbook of research on learning and instruction* (pp. 249–271). Routledge.
- Holstein, K., McLaren, B. M., & Aleven, V. (2019). Co-designing a real-time classroom orchestration tool to support teacher-AI complementarity. *Journal of Learning Analytics*, 6(2), 27-52. <https://doi.org/10.18608/jla.2019.62.3>
- Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial intelligence in education: Promises and implications for teaching and learning*. Center for Curriculum Redesign.
- Hrich, N., Azekri, M., Elhaddouchi, C., & Khaldi, M. (2024). Enhancing Educational Assessments: Score Regeneration through Post-Item Analysis with Artificial Intelligence. *International Journal of Information and Education Technology*, 14(10), 1414–1420. <https://doi.org/10.18178/ijiet.2024.14.10.2172>
- Hrich, N., Azekri, M., & Khaldi, M. (2024). Artificial Intelligence Item Analysis Tool for Educational Assessment: Case of Large-Scale Competitive Exams. *International Journal of Information and Education Technology*, 14 (6), 822-827. <https://doi.org/10.18178/ijiet.2024.14.6.2107>
- Kaldaras, L., Akazeze, H. O., & Reckase, M. D. (2024). Developing valid assessments in the era of generative artificial intelligence. *Front. Educ.* 9:1399377. <https://doi.org/10.3389/feduc.2024.1399377>
- Karan, B., & Angadi, G. R. (2024). Potential Risks of Artificial Intelligence Integration into School Education: A Systematic Review. *Bulletin of Science, Technology & Society*, 43(3-4), 67-85. <https://doi.org/10.1177/02704676231224705> (Original work published 2023)
- Kellogg, R. T., Whiteford, A. P., & Quinlan, T. (2010). Does Automated Feedback Help Students Learn to Write? *Journal of Educational Computing Research*, 42(2), 173–196. <https://doi.org/10.2190/ec.42.2.c>
- Khlaif, Z., Odeh, A., & Bsharat, T. R. (2024). Generative AI-Powered Adaptive Assessment. In M. Sanmugam, D. Lim, N. Mohd Barkhaya, W. Wan Yahaya, & Z. Khlaif (Eds.), *Power of Persuasive Educational Technologies in Enhancing Learning* (pp. 157-176). IGI Global Scientific Publishing. <https://doi.org/10.4018/979-8-3693-6397-3.ch007>
- Kılınc, S. (2024). Comprehensive AI assessment framework: Enhancing educational evaluation with ethical AI integration. *Journal of Educational Technology and Online Learning*, 7(4 - ICETOL 2024 Special Issue), 521-540. <https://doi.org/10.31681/jetol.1492695>
- Kizilcec, R. F., & Lee, H. (2020). Algorithmic fairness in education. In W. Holmes & K. Porayska-Pomsta (Eds.), *Ethics in artificial intelligence in education* (pp. 71-87). Routledge. <http://dx.doi.org/10.4324/9780429329067-10>

- Knox, J., Wang, Y., & Gallagher, M. (2019). *Artificial intelligence and inclusive education: Speculative futures and emerging practices*. Springer. <https://doi.org/10.1007/978-981-13-8161-4>
- Larsson, J. A., & White, B. (2014). *Learning analytics: From research to practice*. Springer. <https://doi.org/10.1007/978-1-4614-3305-7>
- Li, G., Hu, Y., Junkai Shuai, Yang, T., Zhang, Y., Dai, S., & Xiong, N. (2022). NeuralNCD: A Neural Network Cognitive Diagnosis Model Based on Multi-Dimensional Features. *Applied Sciences*, 12(19), 9806–9806. <https://doi.org/10.3390/app12199806>
- Liu, Z., Guo, R., Jiao, X., Gao, X., Oh, H., Xing, W. (2024). How AI Assisted K-12 Computer Science Education: A Systematic Review. *2024 ASEE Annual Conference & Exposition, Portland, Oregon*. <https://doi.org/10.18260/1-2--47532>
- Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence unleashed: An argument for AI in education*. Pearson.
- Ma, X. (2024). Artificial intelligence-driven education evaluation and scoring: Comparative exploration of machine learning algorithms. *Journal of Intelligent Systems*, 33(1), 20230319. <https://doi.org/10.1515/jisys-2023-0319>
- Mahamuni, A. J., Parminder, & Tonpe, S. S. (2024). Enhancing Educational Assessment with Artificial Intelligence: Challenges and Opportunities. *2024 International Conference on Knowledge Engineering and Communication Systems (ICKECS)*, 112, 1–5. <https://doi.org/10.1109/ickecs61492.2024.10616620>
- Milano, N., Casella, M., Esposito, R., & Marocco, D. (2024). Exploring the Potential of Variational Autoencoders for Modeling Nonlinear Relationships in Psychological Data. *Behavioral Sciences*, 14(7), 527–527. <https://doi.org/10.3390/bs14070527>
- Milano, N., Ponticorvo, M., & Marocco, D. (2025). *Comparing Human Expertise and Large Language Models Embeddings in Content Validity Assessment of Personality Tests*. ArXiv.org. <https://arxiv.org/abs/2503.12080>
- Oregon State University Ecampus (2024). Bloom's Taxonomy Revisited, based on MAGE framework. <https://ecampus.oregonstate.edu/faculty/artificial-intelligence-tools/blooms-taxonomy-revisited/>
- Owan, V. J., Abang, K. B., Idika, D. O., Etta, E. O., & Bassey, B. A. (2023). Exploring the potential of artificial intelligence tools in educational measurement and assessment. *Eurasia Journal of Mathematics, Science and Technology Education*, 19(8), em2307. <https://doi.org/10.29333/ejmste/13428>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., & al. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews *BMJ* 2021; 372 :n71, <https://doi.org/10.1136/bmj.n71>
- Panadero, E., Andrade, H., & Brookhart, S. (2018). Fusing self-regulated learning and formative assessment: A roadmap of where we are, how we got here, and where we are going. *Australian Educational Researcher*, 45(1), 13-31. <https://doi.org/10.1007/s13384-018-0258-y>
- Rashid, A. B. & Kausik, A. K. (2024). AI revolutionizing industries worldwide: A comprehensive overview of its diverse applications. *Hybrid Advances*, 7, 100277, <https://doi.org/10.1016/j.hybadv.2024.100277>
- Reich, J., & Ito, M. (2017). *From good intentions to real outcomes: Equity by design in learning technologies*. Digital Media and Learning Research Hub.
- Richardson, M., & Clesham, R. (2021). Rise of the machines? The evolving role of AI technologies in high-stakes assessment. *London Review of Education*, 19(1). <https://doi.org/10.14324/lre.19.1.09>
- Saadati, S. M. (2023). The Need for More Attention to The Validity and Reliability of AI-Generated Exercise Programs. *Health Nexus*, 1(1), 77-81. <https://doi.org/10.61838/kman.hn.1.1.12>
- Sahin, A., Thompson, N., & Ercikan, K. (2024). Opportunities and challenges of AI in educational assessment. *Journal of Measurement and Evaluation in Education and Psychology*, 15 (Special issue), 260-262. <https://doi.org/10.21031/epod.1607441>
- Sahito, Z. H., Sahito, F. Z., & Imran, M. (2024). The Role of Artificial Intelligence (AI) in Personalized Learning: A Case Study in K-12 Education. *Global Educational Studies Review*, IX(III), 153-163. [https://doi.org/10.31703/gesr.2024\(IX-III\).15](https://doi.org/10.31703/gesr.2024(IX-III).15)
- Saputra, I., Kurniawan, A., Yanita, M., Putri, E. Y., & Mahniza, M. (2024). The Evolution of Educational Assessment: How Artificial Intelligence is Shaping the Trends and Future of Learning Evaluation. *The Indonesian Journal of Computer Science (IJCS)* 13(6). <https://doi.org/10.33022/ijcs.v13i6.4465>

- Schellekens, L. H., Bok, H. G. J., Jong, L. H., Schaaf, M. F., Kremer, W. D. J., & Vleuten, C. P. M. (2021). A scoping review on the notions of Assessment as Learning (AaL), Assessment for Learning (AfL), and Assessment of Learning (AoL). *Studies in Educational Evaluation*, 71, 101094. <https://doi.org/10.1016/j.stueduc.2021.101094>
- Selwyn, N. (2019). What's the problem with learning analytics? *Journal of Learning Analytics*, 6(3), 11-19. <https://doi.org/10.18608/jla.2019.63.3>
- Su, Y., Cheng, Z., Luo, P., Wu, J., Zhang, L., Liu, Q., & Wang, S. (2021). Time-and-Concept Enhanced Deep Multidimensional Item Response Theory for interpretable Knowledge Tracing. *Knowledge-Based Systems*, 218, 106819. <https://doi.org/10.1016/j.knosys.2021.106819>
- Tan, B., Nour Armoush, Mazzullo, E., Okan Bulut, & Gierl, M. J. (2024). *A Review of Automatic Item Generation Techniques Leveraging Large Language Models*. <https://doi.org/10.35542/osf.io/6d8tj>
- Tsai, Y. S., Poquet, O., Gašević, D., Dawson, S., & Pardo, A. (2019). Complexity leadership in learning analytics: Drivers, challenges and opportunities. *British Journal of Educational Technology*, 50(6), 2839-2854. <https://doi.org/10.1111/bjet.12846>
- Urban, C. J., & Bauer, D. J. (2021). *Deep Learning-Based Estimation and Goodness-of-Fit for Large-Scale Confirmatory Item Factor Analysis*. <https://doi.org/10.48550/arXiv.2109.09500>
- Vygotsky, L. S. (1997). *Educational psychology*. Jamaica Hills, NY: Saint Lucie Press. (Original work published 1926).
- Yaneva, V., Clauser, B. E., Morales, A., & Paniagua, M. (2022). Assessing the validity of test scores using response process data from an eye-tracking study: a new approach. *Advances in Health Sciences Education*. <https://doi.org/10.1007/s10459-022-10107-9>
- Yilmaz, K., & Deniz, K. Z. (2024). Natural language processing and machine learning applications for assessment and evaluation in education: opportunities and new approaches. *Journal of Measurement and Evaluation in Education and Psychology*, 15(4), 421-445. <https://doi.org/10.21031/epod.1551568>
- Young, L. (2024). Innovations in online exam proctoring for secure assessments. *Turnitin*. <https://www.turnitin.com/blog/secure-assessment-innovations-in-online-exam-proctoring?>
- Zaphir, L., Lodge, J. M., Lisec, J., McGrath, D., & Khosravi, H. (2024). How critically can an AI think? A framework for evaluating the quality of thinking of generative artificial intelligence. The University of Queensland, Australia. <https://doi.org/10.48550/arXiv.2406.14769>
- Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education—where are the educators? *International Journal of Educational Technology in Higher Education*, 16(1), 1-27. <https://doi.org/10.1186/s41239-019-0171-0>
- Zhang, L., & Chen, P. (2024). A neural network paradigm for modeling psychometric data and estimating IRT model parameters: Cross estimation network. *Behavior Research Methods*, 56(7), 7026–7058. <https://doi.org/10.3758/s13428-024-02406-3>
- Zou, Yuxue & Yuan, Minhui & Mo, Lan & Mustakim, Siti. (2024). Enhancing Teaching and Learning through Assessment Strategies: A Practical Guide. *International Journal of Academic Research in Business and Social Sciences*, 14. <https://doi.org/10.6007/IJARBS/v14-i7/21928>

Abbreviations

The following abbreviations are used in this manuscript:

P-AI-M	Processual Assessment Integration Model
PRISMA	Preferred Reporting Items for Systematic reviews and Meta-Analyses

Author Information and Declarations

Author Contributions: Conceptualization, R.F., O.I. and C.F.; methodology, R.F., O.I. and C.F.; validation, R.F., O.I. and C.F.; formal analysis, R.F., O.I. and C.F.; resources, R.F., O.I. and C.F.; data curation, R.F., O.I. and C.F.; writing, R.F., O.I. and C.F.; review and editing, O.I.; visualization, R.F.; supervision, C.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable. Please contact the authors for further information regarding the processing of data (PRISMA phases).

Acknowledgments: During the preparation of this study, the authors used Elicit Plus and Claude 3.7 Sonnet to find additional sources of information, and to improve coherence and alignment of the information retrieved. All AI suggestions have been selected and re-processed by the authors, to whom all ideas and intellectual input in this text belong. The authors have reviewed and edited the AI outputs and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest

Author Biographies

Rares Fartușnic is specializing in psychometry and psychological methods, currently in his third year of Psychology studies at the University of Amsterdam, Faculty of Social and Behavioural Science. He is an active member of the Assessment and Individual Differences Lab (AID-Lab) at the University of Bucharest, Faculty of Psychology and Education Sciences, where he contributes to various research activities, including data collection, coding, analysis, and academic writing. His academic focus centres on understanding individual differences through rigorous methodological approaches, with his work at AID-Lab reflecting his growing interest in empirical psychological research. Mr. Fartușnic is currently focused on advancing his knowledge in psychological assessment and research methodology, underpinned by applications in data science.

Olimpiu Istrate, PhD in education sciences, brings over 20 years of experience in the areas of digital pedagogy, education programmes evaluation, and teacher training. Since 2004, he has taught specialized courses at the University of Bucharest's Faculty of Psychology and Education Sciences, Teacher Training Department – at both bachelor's and master's levels – including *Computer Assisted Instruction*, *Virtual Learning Environments*, *New Media in Education and Training*, *AI in education*. Currently, Dr. Istrate works as an e-assessment specialist on secondment to the Office of the Secretary General of the European Schools in Brussels. His notable achievements include co-founding Academia Online (2003), founding Elearning.Romania (2005), serving as regional education manager at Intel (2007-2010), founding iTech.ro (2010), coordinating eTwinning nationally (2011-2012), working as senior education officer at the International Federation of RCRC in Geneva (2013-2016), and co-founding digitaledu.ro (2015). Dr. Istrate has authored several books, chapters, and articles in the field of digital pedagogy, including his influential 2022 paper "Digital Pedagogy. Definition and Conceptual Area", published in the Journal of Digital Pedagogy. His latest assignments include collaboration with UNESCO IBE, the European Commission, the World Bank, and various national working groups for innovation in education.

Ciprian Fartușnic, PhD in education sciences, has over 25 years of experience in education research, policy analysis, curriculum, program and project assessment and teacher training. Former director of the Institute of Education Sciences in Romania (2014-2018), Dr. Ciprian Fartușnic is currently working as Head of the Pedagogical Development Unit, Office of the Secretary General of the European Schools. He held the position of an associated professor at the University of Bucharest, University of Iasi, and National School of Political Sciences and Public Administration - Bucharest, teaching courses at both bachelor's and master's level, including Didactics of Social Sciences, Theories of Learning, Education Policies. Dr. Ciprian Fartușnic co-ordinated the competence based curriculum reform and was involved in the design and delivery of a national continuous training program for teachers to support this reform, with an important online component. He is a member of national and international research teams on various topics. He worked as an expert for UNICEF, the European Commission, UNESCO (International Bureau of Education) and the World Bank. Dr. Ciprian Fartușnic was involved in the development of the Romanian Digital Competence Framework and he is a member of the EPAN (Council of Europe), REFERNET (CEDEFOP), International Research Network on Equity in Youth Education and Training (IRNEYET), and the Learning and Teaching, Bologna Follow-Up Group (European Commission).