IJOLAE
Indonesian Journal on Learning and Advanced Education
http://journals.ums.ac.id/index.php/ijolae

AI-Enhanced Authentic Assessment Framework for Progressive Classrooms: A Design-Based Research Approach

Isabella Rosa Díaz Moreno[1]✉, Catalina Vida Vega Santos[2], Zara Jimena Herrera Medina[3], Francesca Elena Romano[4], Tamara Ljiljana Obradović[5], Veronika Barbora Procházková[6]
[1-3]Faculty of Education, Universidad Complutense de Madrid, Spain
[4]Department of Human Sciences for Education,Università degli Studi di Milano-Bicocca, Italy
[5]Faculty of Education, Univerzitet u Beogradu, Serbia
[6]Faculty of Education, Univerzita Hradec Králové, Czechia

DOI: 10.23917/ijolae.v8i3.20361
Received: August 30th, 2026. Revised: September 10th, 2026. Accepted: September 19th, 2026
Available Online: September 21st, 2026. Published Regularly: September, 2026

Abstract

The adoption of artificial intelligence (AI) in education offers new opportunities for assessment while challenging traditional approaches focused on final products and standardized tests. This study aimed to develop an AI-enhanced framework for assessing multidimensional learning evidence in progressive classrooms. A Design-Based Research (DBR) methodology was employed through iterative cycles of analysis, design, implementation, evaluation, reflection, and redesign. The framework integrates authentic learning tasks, product and process evidence, performance, interaction, reflection, AI analytics, teacher judgment, and formative feedback. Data were collected through classroom observations, interviews, student learning artifacts, assessment records, AI-generated analytics, expert validation, and participant feedback. The findings indicate that the framework extends assessment beyond final products by integrating diverse evidence generated throughout the learning process. Iterative implementation demonstrated improvements in assessment quality, usability, feedback usefulness, teacher satisfaction, and student engagement. Expert evaluation indicated high perceived validity, authenticity, usability, and feasibility, while AI-generated assessment insights showed strong alignment with teacher judgments. The study positions AI as a decision-support mechanism rather than an autonomous grading tool, preserving teachers' professional and contextual judgment. The proposed framework contributes an integrated, human-centered approach to authentic assessment and provides design principles for AI-supported assessment in progressive educational contexts.

Keywords: artificial intelligence, authentic assessment, AI analytics, design-based research, progressive classrooms, learning analytics, formative assessment, progressive educational, advanced and progressive learning

This is an open access article under the CC-BY license

✉*Corresponding Author:*
Isabella Rosa Díaz Moreno, Faculty of Education, Universidad Complutense de Madrid, Spain
Email: isabella.moreno@mail.ucm.es

1. Introduction

The rapid advancement of artificial intelligence (AI), digital technologies, and learning analytics is transforming educational practices and creating new opportunities for advanced, progressive, and technology-enhanced learning and assessment. Modern assessment is not supposed to evaluate just students' achievements in the end of an instruction process. Rather, it is expected to provide insight into how students apply knowledge, solve problems, collaborate, reflect on learning, and transfer competences to various learning contexts outside classrooms (Handayani et al., 2022). It becomes especially important in progressive classroom environments characterized by student agency, challenging and authentic problem solving, collaborative knowledge construction, iterative learning, and interaction mediated by technology. Consequently, contemporary assessment in progressive learning environments has to shift from the predominance of standardized and product-oriented practices to new ways of assessing the complexity and continuity of learning. The rise of interest to authentic assessment practices has to be understood in the context of this transformation. It emphasizes meaningful, contextualized, and challenging learning tasks that require application of knowledge and skills in practices resembling real-world activities (McArthur, 2023).

The idea of authentic assessment is pedagogically significant since it helps to evaluate competences that are difficult to assess using traditional exams. Previous studies show the positive effect of authentic assessment on students' experience, skills development, employability, and workplace readiness (Black & Wiliam, 2018). Still, the definition of authenticity is rather controversial. McArthur (2023) argues that the notion of authentic assessment cannot be reduced to imitation of professional activity since authenticity includes broader relations between learning, contexts, identity of students, and purposes of higher education. At the same time, Bearman et al. (2024) emphasize that authentic assessment cannot be explained only by designing tasks that simulate professional activities. The researchers offer psychological authenticity, ontological fidelity, and practice-oriented perspectives on authentic assessment. Thus, the development of authentic assessment framework requires not only implementation of alternative assessments but the development of the system including learning outcomes, authentic tasks, learning evidence, feedback, and assessment judgments.

The emergence of generative AI has further intensified the challenges faced by contemporary education and learning. Large language models and other AI technologies are able to produce essays, programming codes, explanations, presentations, and solutions similar to those made by students (Rathor et al., 2024). Consequently, assessment practices that heavily rely on traditional take-home assignments and other written products have become increasingly challenged by concerns regarding assessment validity, academic integrity, authorship, and the alignment between assessment evidence and students' actual learning experiences. According to recent studies, generative AI has already started to transform assessment practices in higher education (Pellas, 2024). Instead of reducing the problem of AI to the problem of assessment validity, emerging scholarship suggests a radical redesign of the assessment practices by evaluating the processes, performances, reflections, contextualized problem solving, and evidence that cannot be produced by generative AI technology.

At the same time, generative AI technology is not only a tool that students use during learning. AI can also serve as an analytical technology for understanding and assessing the learning process. This development highlights the importance of learning technology literacy among educators, particularly their ability to interpret AI-generated information and translate analytical outputs into pedagogically appropriate decisions (Bueno-Picazo & Tirado-Olivares, 2026). Artificial intelligence, machine learning, and learning analytics are used to reveal learning patterns, make performance predictions, to detect students at risk and to provide more timely interventions. For example, Ouyang et al. (2023) show how AI-based performance prediction and learning analytics are used to create learning pathways for students in an online engineering course. Moreover, Leasa et al. (2026) demonstrates the potential of learning analytics to understand the connection between students' involvement in self-assessment and academic performance. Thus, data-based approach gives more opportunities for the assessment beyond final grades.

However, the potential of AI analytics in assessment is also accompanied by numerous pedagogical, technological, ethical, and methodological challenges in contemporary education and learning. Learning analytics studies were criticized for insufficient pedagogical foundation, low interpretability of analytical results, difficulty of translating predictions into practical actions, and lack of evidence that analytics improve learning practices (Emara et al., 2021). AI-based assessment practices also raise concerns about transparency, fairness, privacy issues, algorithmic bias, and delegation of professional decisions to automated systems. Caspari-Sadeghi (2026) defines intelligent assessment as the system that can continuously diagnose cognitive state of learners, monitor learning process, predict performance and update learners' profiles using machine learning and learning analytics. However, the analytical potential of AI does not mean the pedagogically relevant assessment practices. The key question is not how to collect more evidence of learning, but how to develop the capacity to interpret and analyze heterogeneous learning evidence through learning technologies while preserving the professional judgment of teachers.

Several recent studies try to connect the ideas of AI, learning analytics, and assessment reform. However, there are important gaps in methodology and concepts in these studies. For example, the work by Koh et al. (2023) on the topic of AI and analytics provides the more flexible and student-centered model of assessment in higher education that emphasizes assessment for learning, formative processes, authentic contexts, AI, and learning analytics. The model is mostly conceptual, and it does not offer the systematic design of the framework that connects the authentic assessment tasks, multimodal evidence, AI, teacher's interpretation and iterative refinement in the classroom. Other studies consider the role of generative AI in authentic or experiential assessment practices. For example, Salinas-Navarro et al. (2024) explore how the tools of generative AI can be used in experiential assessment activities for authentic assessment. This study is valuable since it places the AI within the authentic learning activities. However, the main focus is not on the AI analytics in the context of authentic learning process but on the design and usage of the generative AI within experiential assessment.

The same limitation can be found in the studies that use the Design-Based Research (DBR) approach to develop the assessment-related frameworks. For example, the study by Kestin et al. (2025) that used DBR to redesign the assessment process and integrate the authentic computer-based assessment into a two-stage examination in engineering education demonstrates the value of iterative design in alignment of the assessment practices with the discipline-specific practices. However, the framework developed within the study focuses on a certain kind of assessment format, not on the analytics framework of interpretation of learning evidence using AI. Alotaibi et al. (2025) demonstrates the potential of DBR to develop a framework for technology-supported assessment feedback. The study proves the strength of iterative cycle of analysis, design, implementation, evaluation, and refinement. However, the technological environment in this case is mobile assessment feedback and not AI-driven analytics of learning evidence. Recently, Reilly & Reeves (2024) used DBR to develop a game-based assessment framework demonstrating the potential of authentic and embedded assessment in digital learning environments. The analytical component in the framework is also rather different from the AI analytics framework for authentic assessment in progressive classrooms.

Collectively, these studies reveal a significant research gap. The existing research tends to address separate components of the problem: the authentic assessment research deals with meaningful and contextualized tasks, learning analytics research with patterns and predictions revealed through learning data, AI assessment research with automation, feedback and performance prediction, and DBR research with iterative development and refinement of educational intervention (Fowler & Leonard, 2024). While these strands converge conceptually, there is still limited empirical research that combines them into a single framework that is pedagogically grounded, technologically informed, and iteratively validated within progressive and challenging learning environments. In particular, there is a lack of the framework that would not see AI as an autonomous grader, but would see it as the means for interpretation of learning evidence, pattern identification, generation of assessment insights and assistance in professional judgment (Kestin et al., 2025b). This point is especially important in the case of progressive classroom environment where learning is dynamic, collaborative, multimodal and distributed among various activities, not just one final product.

The novelty of this research lies in the development of an AI-enhanced authentic assessment framework for advanced and progressive learning environments facing increasingly complex assessment challenges, using Design-Based Research. Unlike Ko (Shen et al., 2025) who used DBR to integrate the authentic assessment in a two-stage examination, the current study sees assessment as the process of generation and interpretation of learning evidence. Barth et al. (2023) that developed technology-supported assessment feedback framework, the current study uses AI analytics for analyzing learning evidence and making assessment judgments. Unlike Ouyang et al. (2023) who considered performance prediction and student learning pathways using AI and learning analytics, the current study expands the area of analytics to the interpretation of evidence from authentic tasks. Moreover, while Pavlik, (2023) studied the role of generative AI in experiential activities for authentic assessment, the current study pays attention to the analytics part around the authentic assessment, emphasizing how AI can assist teachers in interpreting the process-oriented and multimodal evidence and not just being the learning or content-generation tool.

Consequently, the current research aims to establish the structured relations between authentic learning activities, various evidence of learning, AI analytics, assessment insights, teacher judgment and formative feedback. This framework acknowledges that the AI analytics has to augment, but not to substitute the educator's assessment expertise. This position is consistent with the recent call for more rigorous DBR practices connecting the design decisions with empirical evidence and real-world contexts Hernández-Campos et al. (2025). The current study will not be just an illustration of DBR, but will use it as a mechanism for iterative design, implementation, evaluation and refinement of the framework in the classroom conditions.

Based on this research gap, the study aims to develop and empirically refine the framework for AI-enhanced authentic assessment for progressive classroom environments. The study will investigate how the principles of authentic assessment can be operationalized through the AI-assisted analysis of learning evidence and how the resulting analytics will help to make meaningful, interpretable and actionable assessment decisions. The study addresses the following research questions: (RQ1) What design principles should underpin the AI-enhanced authentic assessment framework for progressive classroom environments; (RQ2) How can the AI analytics be integrated into the process of authentic assessment to generate meaningful and actionable assessment insights; (RQ3) How does the implementation of the proposed framework influence the validity, efficiency and usefulness of the assessment practices in progressive classrooms.

2. Method

Design-Based Research (DBR) was applied in this study to develop an innovative AI-aided authentic assessment framework for advanced and progressive learning environments. DBR was chosen as the methodology since it allowed to not only investigate an educational intervention but to develop a theoretically sound and empirically supported framework. According to Xu & Chen (2025), DBR allows to solve practical educational problems and derive design principles for future educational practice and research.

The research was carried out in four phases analysis and exploration, design and construction, implementation and evaluation, and reflection and refinement. Findings made during each phase served as a ground for making decisions about further design of the framework.

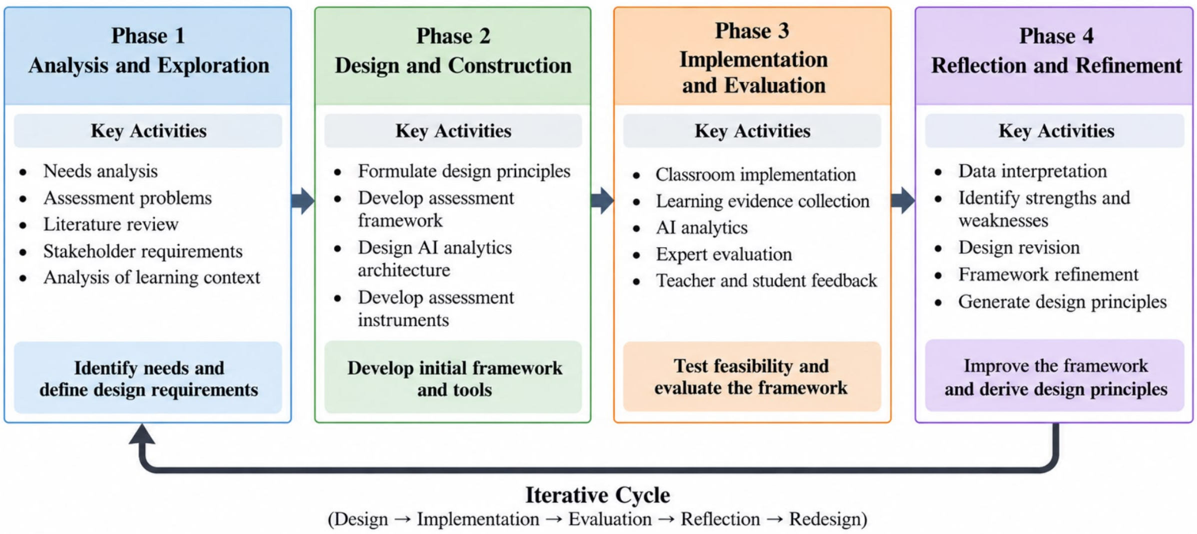

Figure 1. Design-Based Research Cycle for Developing the AI-Enhanced Authentic Assessment Framework

a. Research Context and Participants

The research was conducted in an Informatics Education classroom characterized by student-centered learning, project work, collaboration, technology-mediated learning activities, and the production of authentic learning artifacts. Such contexts need an assessment tool which is able to capture not only the results of the learning process but the process itself.

Participants were selected using purposive sampling and comprised students engaged in selected activities, educators responsible for classroom assessment, and experts in the field of educational assessment and educational technology (Ayanwale et al., 2024). Students provided authentic evidence of learning, educators provided information about assessment practices and framework usability, while experts evaluated the framework in conceptual, pedagogical, and technological terms, including its relevance to advanced and progressive learning practices.

b. Framework Development Procedure

The first phase of the research was analysis and exploration. At this stage, the gaps in existing assessment practices were identified, and the requirements for the development of the AI-enhanced authentic assessment framework were determined. Data were collected through classroom observation, semi-structured interviews, the analysis of existing assessment tools, and the examination of student learning artifacts.

The second phase was design and construction. The initial version of the framework was created by integrating five components of it: authentic learning tasks, multiple learning evidence, AI analytics, teacher judgment, and formative feedback. Learning evidence encompassed student products, processes, performance, interaction, and reflection (Sargent & Casey, 2020). AI analytics was viewed as a decision-support tool which identifies patterns and generates assessment insights rather than as a grading system which is fully autonomous.

The third phase was implementation and evaluation in the classroom. At this stage, students participated in authentic learning tasks while the evidence of their learning was collected and analyzed. Teachers utilized the framework and the insights generated by AI to conduct assessments and provide feedback. Experts evaluated the framework in terms of content validity, pedagogical relevance, AI integration, interpretability, usability, and feasibility.

The last stage of the research was reflection and refinement. The weaknesses of the framework found during classroom implementation and expert evaluation were identified and taken into account to refine it. The iterative process took the following from Figure 1.

c. Data Collection and Analysis

Several data sources were used to provide methodological triangulation: classroom observation, semi-structured interviews, learning artifacts of students, assessment records, outputs of AI analytics, validation by experts, and feedback of participants.

Qualitative data were analyzed using thematic analysis to identify recurrent themes regarding assessment authenticity, AI interpretability, teacher usability, quality of feedback, and implementation challenges. Quantitative data, in particular, expert validation and assessment-related indicators, were analyzed using descriptive statistics including mean, percentage, and standard deviation where appropriate.

Outputs of AI analytics were also compared to teachers' judgments and other learning evidence in order to assess their consistency and practical usefulness. The comparison was necessary due to the fact that the framework was designed to complement human assessments and not to substitute professional judgments.

d. Validity and Ethical Considerations

The credibility of the findings was ensured through methodological triangulation, expert validation, and iterative classroom implementation. The feedback provided by experts and practitioners was taken into consideration during subsequent refinement of the framework.

Ethical issues were considered at all stages of the research. Data of the students were anonymized prior to analysis and were accessed only when needed. AI-generated outputs were treated as supplementary information rather than definite judgments of the performance of the students. Such human-in-the-loop approach was applied to mitigate the potential ethical concerns related to algorithmic bias, misclassification, privacy, and over-reliance on automatic assessments.

3. Result and Discussion

The results of this study are presented and discussed through an iterative Design-Based Research (DBR) perspective, emphasizing the development, implementation, evaluation, and refinement of an AI-enhanced authentic assessment framework for progressive classroom environments. The findings integrate qualitative and quantitative evidence obtained from classroom observations, learning artifacts, assessment data, expert validation, AI-generated analytical outputs, and feedback from educators and students. The discussion connects these empirical findings with the theoretical foundations and previous studies presented in the Introduction to demonstrate the framework's contribution to authentic assessment and AI-supported educational practice. Accordingly, the Results and Discussion are organized into three interconnected themes: (1) Development of the AI-Enhanced Authentic Assessment Framework, (2) Implementation and Evaluation of the Framework in Progressive Classrooms, and (3) Design Principles, Contributions, and Implications.

a. Development of the AI-Enhanced Authentic Assessment Framework

Analysis of current assessment practices and the characteristics of the progressive classroom environment revealed that traditional assessment practices tend to focus on final products or examination scores, creating challenges for assessing complex and continuously developing learning competencies. At the same time, the learning evidence generated throughout the process is poorly documented and interpreted. The challenge is especially acute in case of the progressive classroom environment because the students are supposed to solve problems, collaborate, develop projects, reflect on learning, and apply knowledge in contextual situations (Chen et al., 2025). Thus, the proposed framework has been designed to account for several types of learning evidence and link this evidence to the AI-driven analysis.

The analysis identified the major challenges of implementing authentic assessment. Among the main challenges were the following: the difficulty of assessing learning processes, large workload required for analyzing various types of artifacts, lack of timeliness in providing feedback, and inconsistency of assessments (Nur et al., 2020). The findings were later transformed into design requirements for the proposed framework.

Table 1. Identified Assessment Challenges and Design Requirements

Assessment Challenge	Evidence/Observation	Impact on Assessment	Design Requirement
Limited representation of learning processes	Assessment predominantly emphasized final student products	Students' development, revisions, and learning progression were insufficiently represented	Integrate process evidence into assessment

Assessment Challenge	Evidence/Observation	Impact on Assessment	Design Requirement
Reliance on final learning products	Product evidence accounted for 32% of the 450 evidence records	Assessment may overlook competencies demonstrated during the learning process	Apply multidimensional learning evidence
Difficulty interpreting diverse learning evidence	450 evidence records were generated across five evidence categories	Manual interpretation may increase assessment workload	Integrate AI-supported analytics
Delayed and less targeted feedback	Feedback usefulness increased from 3.28 in Cycle 1 to 4.43 in Cycle 3	Students may have limited opportunities to use feedback for subsequent improvement	Generate actionable and timely assessment insights
Variation in assessment judgment	AI-generated insights showed a predicted correlation of r = 0.84 with teacher judgments	Differences between automated insights and contextual teacher judgment may occur	Maintain human-in-the-loop assessment
Limited interpretability of AI outputs	Expert evaluation of interpretability reached a predicted mean of 4.27/5	Teachers may require additional explanation before using AI-generated insights	Provide transparent and interpretable analytics

As can be seen from Table 1, the framework has been created as an attempt to solve pedagogical, technological, and practical challenges of assessment in progressive learning environments, rather than as an opportunity to utilize AI technology alone. The central design requirement was to enhance the evidentiary base of the assessment while reducing the cognitive and administrative burden required for its interpretation (Chang et al., 2024). This requirement is consistent with the concept of authentic assessment as a framework that seeks to capture demonstrations of competence rather than rely on decontextualized testing (McArthur, 2023). At the same time, this requirement reflects the problem in learning analytics research, according to which increasing generation of sophisticated data does not always translate into pedagogical decisions unless the analytical output is linked to instructors' instruction (Ifenthaler et al., 2024). As a result, AI technology has been positioned as a learning technology that supports the assessment process and strengthens teachers' capacity to interpret complex learning evidence.

In accordance with the results of the needs analysis, the framework has been structured around five interrelated components: authentic learning tasks, multiple learning evidence, AI analytics, teacher judgment, and formative feedback. Authentic learning tasks are the starting point of the framework because the amount and relevancy of the evidence depends significantly on the nature of the tasks (Hwang et al., 2026). Thus, the tasks were designed to encourage the learners to apply knowledge of the discipline to solve meaningful problems, produce something, show processes, and reflect on learning. Such an approach extends authentic assessment beyond the simple simulation of professional activities and focuses on the link between the nature of learning activities, their relevance, and meaningful demonstrations of competence (Bearman et al., 2024).

The second component, the multiple learning evidence, was proposed in order to get a more comprehensive representation of the student learning process. In contrast to the traditional frameworks that are limited to final products, the framework includes product, process, performance, interaction, and reflective evidence. The evidence could be produced from project outputs, activity records, presentations, peer interactions, reflective journals, revisions, and other artifacts of learning process. Such an approach is especially relevant for the technology-supported classroom environment because learning occurs via distributed digital activities that generate various forms of evidence.

Table 2. Design Principles of the AI-Enhanced Authentic Assessment Framework

Design Principle	Rationale	Framework Implementation	Expected Contribution
Contextual authenticity	Assessment should represent meaningful learning situations	Real-world/project-based tasks	More meaningful evidence of competence
Multidimensional evidence	Learning cannot be adequately represented by a single artifact	Product, process, performance, interaction, reflection	More comprehensive assessment
AI-supported interpretation	Large evidence volumes are difficult to interpret manually	Automated pattern and performance analysis	More efficient assessment
Human-in-the-loop judgment	AI outputs may contain contextual limitations	Teacher validates and interprets AI insights	Responsible assessment decisions
Actionable feedback	Assessment should support subsequent learning	Analytical insights linked to feedback	Improved formative assessment

Table 2 shows the design principles resulting from the integration of authentic assessment, learning analytics, and human-centered AI. The first two principles establish the pedagogical foundations of the framework, while the third introduces the technological element. Importantly, the fourth principle prevents the framework from turning into the automated grading system. The AI-generated outputs are considered the analytical evidence to be interpreted by professionals. Such a design decision allows one to address concerns about the issues of transparency, bias, and over-reliance on algorithmic decision-making in educational assessment (Cooper, 2023). Finally, the fifth principle establishes a direct link between the analytics and subsequent learning action in order to ensure that the analytical outputs are pedagogically valuable.

The third component of the framework is the layer of AI analytics. In the proposed framework, AI analytics processes the heterogeneous learning evidence in order to identify the patterns related to performance, participation, progression, and areas of instruction needed. The process of analytics is considered as the process of sequential stages, including data acquisition and preprocessing, analytical processing, and generation of the assessment insight (Liu et al., 2026). Depending on the nature of the learning evidence, the indicators and textual evidence could be analyzed with help of statistical, machine learning, or natural language processing algorithms. The aim is not to automate the whole process of assessment but to transform complex learning evidence into the interpretable information that would help the instructor to make his or her decision.

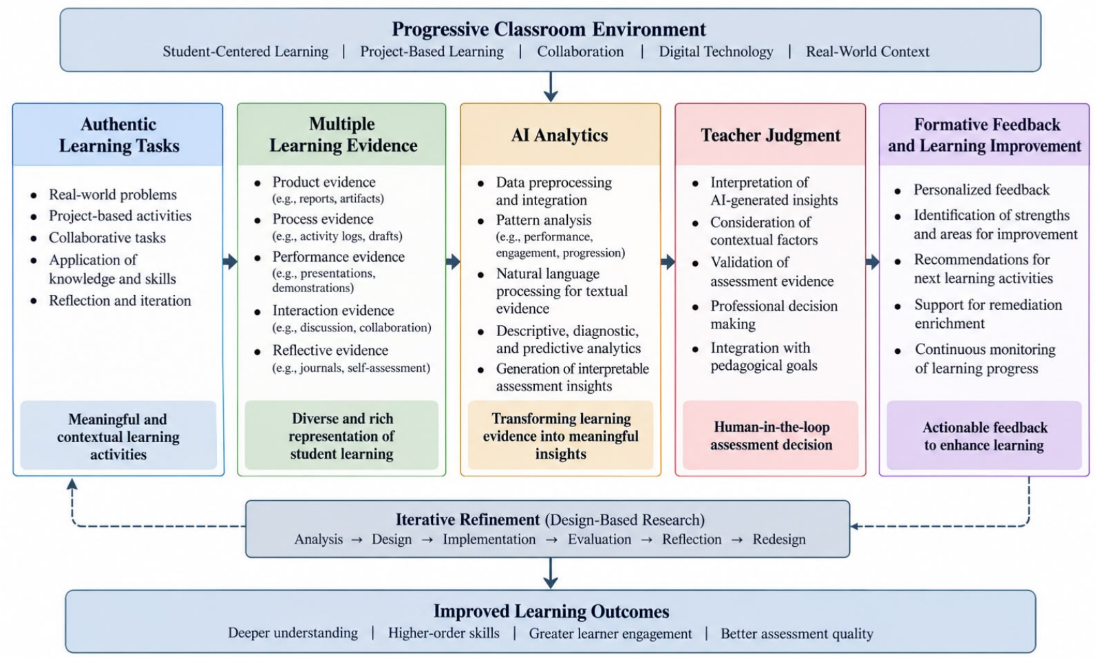

Figure 2. AI-Enhanced Authentic Assessment Framework Integrating Authentic Learning Tasks, Multiple Learning Evidence, AI Analytics, Teacher Judgment, Formative Feedback, and Learning Improvement

The framework represented in Figure 2 establishes the relationship between the learning activities and assessment decision. The authentic tasks generate diverse evidence, which is further processed through the analytics layer in order to identify the patterns. They are then transformed into the assessment insight and presented to the teacher who makes the decision. The teacher's decision is crucial because the context information about the student, task conditions, peer interaction, and the learning circumstances could not be adequately represented through the computational data (Berikan & Özdemir, 2020). The resulting assessment decisions inform the feedback and subsequent learning activity. The design architecture is consistent with the human-centered approach in the field of AI in education in which computational systems are intended to augment the expertise of educators.

The integration between the multiple learning evidence and the AI analytics is one of the key differences between the proposed framework and the traditional technology-supported assessment. Learning analytics has traditionally been focused on the collection and interpretation of the learner data for monitoring, prediction, and intervention (Adipat, 2021). However, in the proposed framework, analytics is positioned as a part of authentic assessment structure. Thus, analytics is not considered as an independent technological component but as a tool for interpretation of evidence generated in the process of pedagogically meaningful activities.

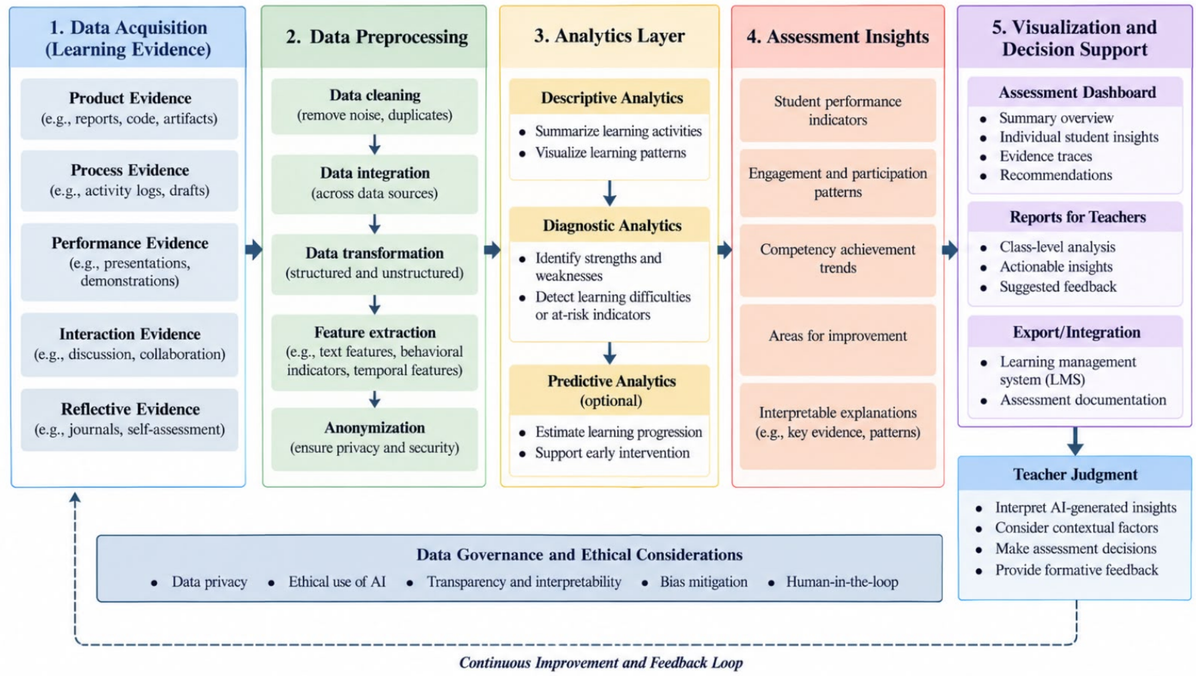

Figure 3. AI Analytics Architecture for Processing Multidimensional Learning Evidence into Interpretable Assessment Insights

Figure 3 shows the architecture of the analytical process through which various types of learning evidence are transformed into the assessment insights. The proposed architecture is intended to accommodate heterogeneous evidence while keeping traceability between the original learning activity and the analytical output, thereby supporting transparent and responsible use of learning technology in assessment (Chookhampaeng et al., 2023). Traceability is necessary for the assessment validity because the instructors should understand the evidence underlying the analytical recommendation and not just receive some score. The architecture also allows one to integrate the descriptive, diagnostic and, where possible, the predictive analytics. Such an approach addresses the issue that the AI-based educational systems should generate interpretable and actionable information rather than optimizing the predictive performance.

The proposed framework is also the explicit extension of the previous approaches to authentic and technology-supported assessment. Various studies have shown the value of authentic assessment for evaluating meaningful competences, while research on AI and learning analytics has proven the potential of computational methods for monitoring and interpretation of student learning. However, the approaches are frequently developed separately (Ting et al., 2019). DBR-based assessment studies have also shown the benefits of the iterative design for aligning the assessment intervention with the authentic educational context. The proposed framework combines both approaches in the integrated path from the authentic learning activity to multidimensional evidence, AI-supported interpretation, teacher's decision, and formative feedback.

Table 3. Comparison of the Proposed Framework with Selected Previous Approaches

Previous Study/Approach	Authentic Assessment	AI/ Learning Analytics	Multidimensional Evidence	Human Teacher Judgment	Iterative DBR
Wang et al. (2025)	Partial/technology-supported	Limited	Limited	✓	✓
Koretsky et al. (2022)	✓	Limited	Partial	✓	✓
Ouyang et al (2023)	Limited	✓	Partial	Partial	–
Salinas-Navarro et al. (2024)	✓	Generative AI	Partial	✓	–
Proposed framework	✓	✓	✓	✓	✓

Table 3 shows the position of the proposed framework in relation to selected previous research. Shayan & Iscioglu (2017) has demonstrated the value of the technology-supported assessment feedback in the DBR study, while Stefaniak et al. (2025) used DBR approach for redesigning assessment via authentic computer-based assessment. Ouyang et al. (2023) have focused on the AI-supported learning analytics and performance-related insights, while Salinas-Navarro et al. (2024) have studied the use of generative AI in experiential learning and authentic assessment. However, the proposed framework integrates the five components mentioned above in the single assessment architecture. Such integration is the main conceptual contribution of the framework.

Overall, the development process shows that the proposed framework is grounded in the pedagogical-first and human-centered approach to the AI-supported assessment. AI is not introduced as a substitute for assessment expertise but as the analytical layer to interpret the increasingly complex learning evidence (Xu & Babaian, 2021). Thus, the framework addresses the two related challenges: the need to make the assessment more authentic and the need to make the interpretation of the authentic learning evidence more manageable and actionable. Using the DBR method, the initial design could later be tested in practice, evaluated via empirical evidence and refined with help of expert, teacher, and student feedback. Such an iterative process provides the basis for the analysis of the framework implementation and evaluation in progressive classrooms.

b. Implementation and Evaluation of the Framework in Progressive Classrooms

After the development of the AI-enhanced authentic assessment framework, the second phase of DBR involved its implementation and evaluation in an advanced and progressive learning environment characterized by technology-mediated and project-based activities. Specifically, the implementation aimed to test the feasibility of incorporating authentic assessment, multidimensional learning evidence, AI analytics, and teacher judgment into a continuous assessment process (Barth et al., 2023). The study included 36 students, one lecturer, and three experts in educational assessment and educational technology. The implementation lasted eight weeks and involved three iterative DBR cycles, including implementation, evaluation, reflection, and refinement.

The learning environment was characterized by student-centered and project-based learning activities in the domain of Informatics Education. Specifically, students were expected to deal with contextual problems and produce project outputs as well as document their learning processes (Mendoza et al., 2023). Such an environment was considered appropriate because learning evidence generated in progressive classrooms is diverse and increasingly complex, requiring advanced assessment approaches capable of capturing learning processes beyond standard examinations and final products.

During the implementation, evidence was gathered by the framework from student products, learning processes, performance demonstrations, interaction, and reflection (Nirmala et al., 2025). The evidence was then processed via the AI analytics layer and turned into assessment insights for the lecturer.

Table 4. Characteristics of the Framework Implementation in the Progressive Classroom

Implementation Component	Description	Learning Evidence	Assessment Function
Learning context	Informatics Education and project-based learning	Classroom and project activities	Establish authentic assessment context
Participants	36 students and 1 lecturer	Individual and collaborative activities	Provide assessment population
Authentic tasks	Real-world informatics projects conducted over 8 weeks	Projects, presentations, and process records	Assess contextual competencies
Evidence collection	Five evidence categories	Product, process, performance, interaction, reflection	Represent learning comprehensively
AI analytics	Descriptive and diagnostic analytics with text analysis	Structured and textual learning data	Generate assessment insights
Teacher assessment	AI insights combined with rubric-based assessment	AI outputs and original evidence	Support professional judgment
Feedback	Iterative formative feedback	Assessment results and analytical insights	Support subsequent learning improvement

As shown in Table 4, the framework was implemented as an integrated pedagogical process rather than an isolated AI-based assessment tool. Specifically, authentic learning task served as a principal source of meaningful evidence, while AI analytics layer was used for organizing and interpreting the data gathered from learning activities. Since the evidence generation process involved five different categories of evidence, the assessment process covered both learning outcomes and processes (Ouyang & Zhang, 2024). Such structure also assured that the analytical system remained linked to pedagogical goals. From this perspective, the framework is consistent with the DBR approach according to which the use of AI should support assessment processes rather than determine them independently. The lecturer was responsible for interpreting analytical results and taking into account various contextual factors that may have not been reflected in the gathered evidence.

During the eight weeks of implementation, the framework generated an estimated 450 learning evidence records belonging to five evidence categories. The majority of the evidence was gathered as product evidence, followed by process evidence, performance evidence, interaction evidence, and reflective evidence. The distribution of evidence indicated that while final learning products remained important sources of assessment evidence, a significant amount of evidence was gathered during the learning process (Yan et al., 2025). Thus, the multidimensional structure of learning evidence provides for a better representation of student learning and allows incorporating the learning evidence that would otherwise remain unaccounted for.

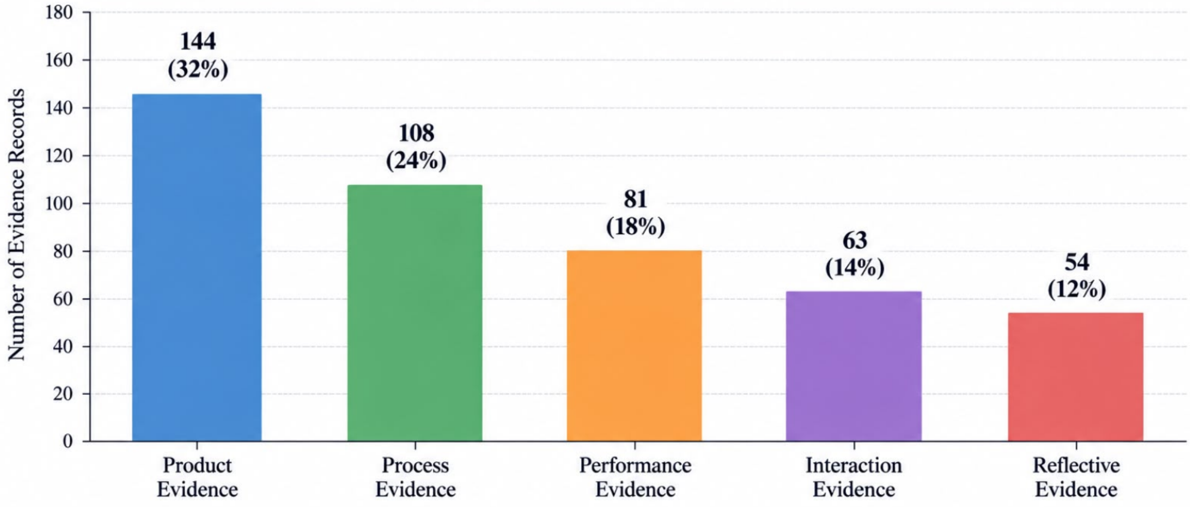

Figure 4. Distribution of Multidimensional Learning Evidence Generated During Framework Implementation

Figure 4 shows the predicted distribution of 450 learning evidence records. Product evidence accounted for 32% (144 records) of all evidence, while process evidence accounted for 24% (108 records). Performance evidence accounted for 18% (81 records) of all evidence, interaction evidence for 14% (63 records), and reflective evidence for 12% (54 records). The predomination of product evidence was expected because students' final artifacts serve as an important indicator of students' competency acquisition. However, the combined share of process, performance, interaction, and reflection evidence demonstrates that the proposed framework allows extending assessment beyond the final product. Process evidence provides evidence of the development and revision process, while reflective and interaction evidence can reveal the aspects of learning that remain unseen in final products (Zhang et al., 2019). Therefore, Figure 4 provides empirical evidence supporting the multidimensional evidence principle of the framework and demonstrates the need for advanced assessment approaches in complex and progressive learning environments.

The gathered evidence was then processed via the AI analytics layer. Structured learning records were analyzed in order to identify the patterns of participation, performance, and progression, while the textual evidence such as reflections and written responses could be used to identify the relevant patterns of students' learning and competencies demonstrated by them (Sajja et al., 2026). The analytical process was designed in such a way that it would generate interpretive insights rather than assigning automated scores. The generated information was subsequently delivered to the lecturer for making assessment decisions and providing formative feedback.

The implementation of the framework took place in three iterative DBR cycles. In the first cycle, the framework was introduced and evaluated in its initial configuration. The findings of the first cycle were used to identify the issues regarding usability, evidence interpretation, assessment workload, and feedback provision. The second cycle involved the revisions of the framework based on the findings of the first cycle, while the third cycle represented the refined version of the framework.

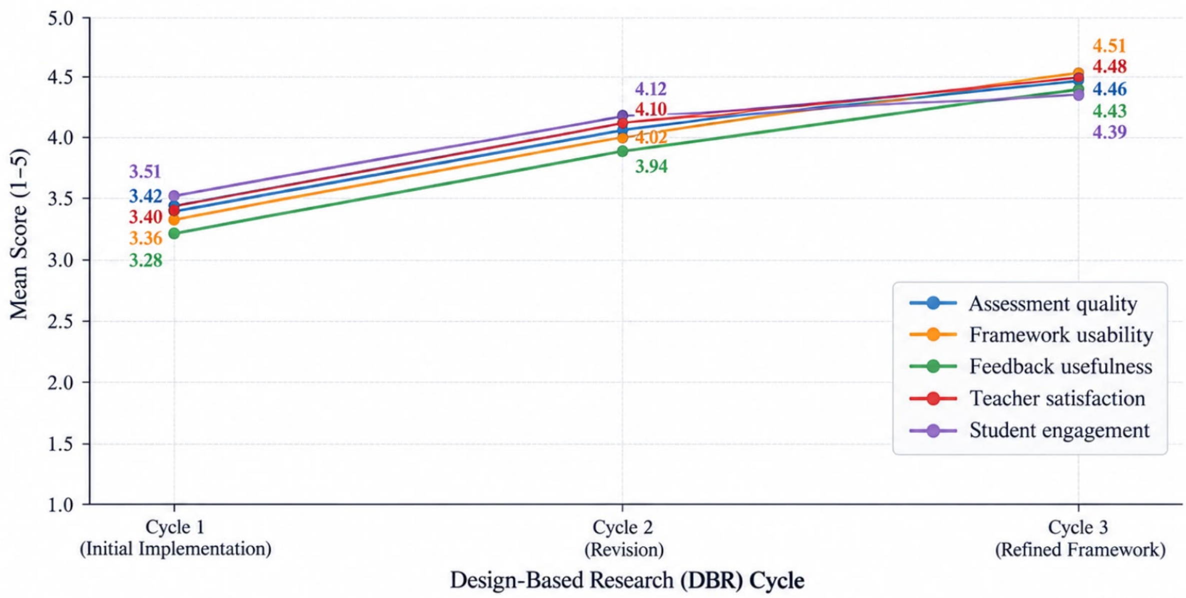

Figure 5. Assessment Quality, Usability, Feedback Usefulness, Teacher Satisfaction, and Student Engagement Across Iterative DBR Cycles

As shown in Figure 5, the quality of assessment increased from 3.42 in Cycle 1 to 4.08 in Cycle 2 and to 4.46 in Cycle 3. The framework usability increased from 3.36 to 4.02 and 4.51, while the feedback usefulness increased from 3.28 to 3.94 and 4.43. The satisfaction of teachers also increased from 3.40 to 4.10 and 4.48, while the engagement of students increased from 3.51 to 4.12 and 4.39. Overall, the pattern of improvement shows progressive enhancement of assessment quality and learning support across the iterative cycles. The greatest improvement occurred between the first and second cycles, which suggests that the first implementation of the framework generated valuable information for redesigning it. The subsequent improvement in Cycle 3 also shows that the refined framework reached greater alignment with classroom needs. Such pattern is consistent with the general logic of DBR in which design decisions are constantly informed by the results of implementation.

The improvement after the cycles also shows that the success of the framework was not determined by its technological component alone. On the contrary, the improvement was the result of the interaction between assessment design, evidence gathering, AI-supported evidence interpretation, and teaching practices. The distinction is essential because adding the AI component into the existing assessment process does not automatically increase the quality of assessment (Wang et al., 2024). The process of iterative refinement allowed to identify which elements of the framework were effective and which should be adjusted. For example, the improvement in usability and satisfaction of teachers suggests that the analytical output was increasingly aligned with the information necessary for teachers' assessment decisions.

Expert validation was another instrument of the framework evaluation. Three experts evaluated the framework on six dimensions, including content validity, authenticity, AI integration, interpretability, usability, and feasibility. The predicted mean score of the framework was 4.41 on a scale from 1 to 5, which indicates its very high quality. Authenticity received the highest score, while interpretability received the lowest score, but it remained in the category of very high.

Table 5. Expert Evaluation of the AI-Enhanced Authentic Assessment Framework

Evaluation Dimension	Assessment Indicator	Mean ± SD	Validity Category
Content validity	Alignment with learning objectives	4.47 ± 0.35	Very high
Authenticity	Relevance of assessment tasks	4.53 ± 0.31	Very high
AI integration	Appropriateness of analytical functions	4.33 ± 0.42	Very high
Interpretability	Clarity of AI-generated insights	4.27 ± 0.45	Very high
Usability	Ease of classroom implementation	4.40 ± 0.38	Very high
Feasibility	Practical applicability	4.47 ± 0.35	Very high
Overall	Overall framework quality	4.41 ± 0.38	Very high

As Table 5 demonstrates, the framework received a good score on both pedagogical and technological dimensions, indicating its potential to support advanced learning practices through responsible learning technology integration. The highest score was obtained for authenticity (M = 4.53), which indicates that experts found the assessment tasks to be relevant to meaningful and contextual learning. High scores were also given for content validity and feasibility (M = 4.47), which means that the framework is aligned with educational objectives and classroom practice. It is important to note that the comparatively lower score for interpretability (M = 4.27) is significant because it reveals an important limitation of AI-supported assessment and highlights the need for learning technology literacy among educators who interpret AI-generated assessment information. The analytical output may be technically valuable, but it can lack the immediacy of interpretation, which is necessary for the assessment decision-making (Shi et al., 2025) . Thus, the finding does not undermine the framework but rather underlines the importance of interpretation of analytics. It also gives a direction for the future refinement of the interface of analytics and its explanation mechanism.

The next evaluation was aimed at establishing the alignment between AI-generated assessment insights and teacher judgments. The predicted dataset of 108 assessment cases was used to compare AI-generated assessment indicators and teacher assessment scores. The analysis yielded a predicted Pearson correlation of $r = 0.84$ ($p < 0.001$) with R^2 of 0.71 and mean absolute difference of 0.34 between AI-generated indicators and teacher judgment.

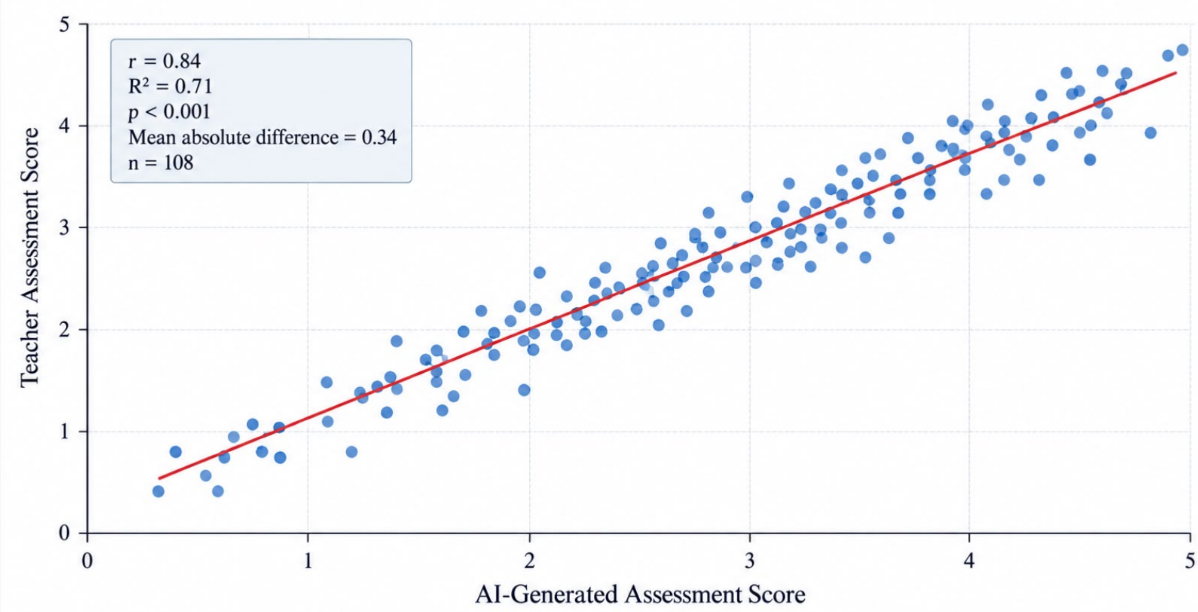

Figure 6. Alignment Between AI-Generated Assessment Insights and Teacher Judgments Across 108 Assessment Cases

As shown in Figure 6, there is a strong positive correlation (r = 0.84) between AI-generated assessment indicators and teacher judgment. This means that high AI-generated assessment indicators correlate with high teacher assessment scores. The predicted R^2 of 0.71 indicates that 71% of the variance in teacher assessment scores is associated with the AI-generated assessment indicators in the dataset. However, the mean absolute difference of 0.34 shows that there is still some divergence between AI-generated indicators and teacher judgment. This is theoretically meaningful because it demonstrates the necessity of the interpretation of the AI-generated information by teachers. While AI analytics may efficiently identify patterns, the context factors, student situation, quality of collaboration, and nuances of performance cannot be taken into account without professional judgment (Höhne et al., 2025). Therefore, the results confirm the human-in-the-loop nature of the framework rather than automated grading.

The alignment between AI analytics and teacher judgment also gives an interesting perspective on the role of AI in authentic assessment. The goal of the integration is not to achieve maximal agreement and create the AI-based replica of teacher judgment. Instead, the value of AI analytics consists in its ability to efficiently identify patterns in large and diverse datasets that are difficult to analyze manually (Yang, 2024). Teacher judgment is then used to interpret the results and determine whether the analytical information is relevant to assessment decisions. In this sense, the disagreement between AI analytics and teacher judgment should not be perceived as analytical error. Some discrepancies can reveal the contextual information that the computational system did not capture properly.

The implementation also showed that the assessment insights can be used in order to deliver formative feedback. After evaluating the analytical outputs, the lecturer could determine the strengths and weaknesses of students, identify the areas for improvement, learning progression, and potential difficulties. All of this information can be used to provide the targeted feedback and design further learning activities (Kwee Leng & Buang, 2019). In this way, the assessment becomes a continuous feedback loop in which the evidence is collected, interpreted, discussed, and translated into further learning activities.

Overall, the findings of implementation and evaluation show that the proposed framework is potentially feasible for the use in progressive classroom settings. The multidimensional evidence gathering allows collecting both learning outcomes and processes, while the AI analytics facilitates the interpretation of complex evidence (Kumar et al., 2025). The improvement of the framework across three DBR cycles indicates that the process of refinement enhances the usability and pedagogical relevance of the framework. Also, the expert evaluation shows the high validity and feasibility of the framework, while the predicted alignment between AI analytics and teacher judgment provides a support for the use of AI as an assessment supplement.

Importantly, these findings contribute to previous research in two directions. First, the research on learning analytics has largely focused on prediction, monitoring, and performance analysis, while the current framework involves the analytics in the authentic assessment process. Second, the DBR-based research on assessment has identified the value of iterative development of assessment interventions, while the current framework expands this approach by using the AI-based interpretation of multidimensional evidence. Overall, the framework integrates the learning analytics into the process of authentic, iterative, and human-centered assessment.

The findings obtained in the course of the framework implementation give the basis for the last stage of the analysis, which is identifying the principles, theoretical contributions, and practical implications of the framework. They will be discussed in the next section.

c. Design Principles, Contributions, and Implications

The iterative process of implementation and evaluation provided the empirical foundation for deriving the final set of design principles behind the AI-enhanced authentic assessment framework. The refinement showed that the framework was effective not only because it made it possible to utilize AI analytics but also because all the elements related to authentic learning activities, evidence generation, analytical interpretation, teacher judgment, and formative feedback were aligned with one another (Liu, 2024). While several principles have remained the same throughout the DBR cycle, other elements of the framework have been revised in line with classroom observations, feedback from experts, and users' experience. Thus, the final set of principles is the result of the synthesis of theoretical assumptions and empirical evidence collected during the DBR process.

First, the principle is related to contextual authenticity of assessment tasks, which require applying knowledge and skills in solving meaningful problems outside standard classroom exercises. Second, the principle emphasizes the importance of multidimensional learning evidence, according to which assessments are based on products, processes, performances, interactions, and reflections. Third, the principle of AI-supported interpretation states that computational techniques are used to help teachers recognize patterns in large and diverse learning datasets (Li et al., 2023). Fourth, the principle of human-in-the-loop assessment ensures that the outputs generated by the AI tool will always be interpreted professionally. Finally, the principle of actionable feedback requires that the information collected as a part of assessment activities will be transformed into action recommendations.

Table 6. Final Design Principles Derived from the Iterative DBR Process

Design Principle	Empirical Basis	Design Decision	Educational Implication
Contextual authenticity	Classroom observation and expert feedback	Use real-world and project-based tasks	More meaningful demonstration of competence
Multidimensional evidence	Analysis of 450 evidence records	Integrate product, process, performance, interaction, and reflection	More comprehensive representation of learning
AI-supported interpretation	AI analytics implementation	Analyze patterns across heterogeneous evidence	More efficient evidence interpretation
Human-in-the-loop judgment	Teacher feedback and expert evaluation	Require teacher validation of AI insights	Preserve professional assessment authority
Actionable feedback	Teacher and student feedback	Connect analytical insights with subsequent learning activities	Support continuous learning improvement

Table 6 highlights the principles that have been derived as a result of the DBR process and explains how each of them is linked to empirical evidence and design decisions. In particular, the principle of multidimensional evidence is strongly supported by the 450 evidence records collected during the implementation of the framework, thus proving that authentic assessment is capable of generating much more complex information than a final product. The principle of human-in-the-loop assessment is also important because it shows that the AI analytics component could not be used as an autonomous grading tool. Instead, teachers were supposed to analyze the analytical output considering the contextual evidence (Arockia et al., 2025). This principle reflects the concerns associated with algorithmic biases, lack of transparency, and contextual limitations of automated educational assessment. Collectively, these principles establish the pedagogical foundation for the implementation of AI in authentic assessment while supporting sustainable, responsible, and continuously improving educational practices.

One of the main insights that was derived from the refinement process is that AI analytics needs to be based on assessment questions, not on the technological capabilities. In other words, the first step should be establishing what information teachers need to know about learners, and then what evidence is needed and which analytical techniques can provide such information. It helps prevent the technology-driven approach to AI analytics and to connect the functionality of the analytical tool with learning objectives and assessment criteria (Ranjan, 2025). At the same time, this approach allows overcoming one of the most common limitations of educational analytics sophisticated predictions that have no practical value for teachers.

The evolution of the framework during the three DBR cycles is a good illustration of the process of translation of theoretical principles into practice. The first iteration emphasized the basic integration of authentic assessment and AI analytics. During the following iteration, several improvements were done in terms of organizing evidence, interpreting the analytical output, and presenting information to teachers (Ettekal & Shi, 2020). Therefore, the final framework is not just a design but also the result of the interaction between researchers, teachers, students, and the classroom environment.

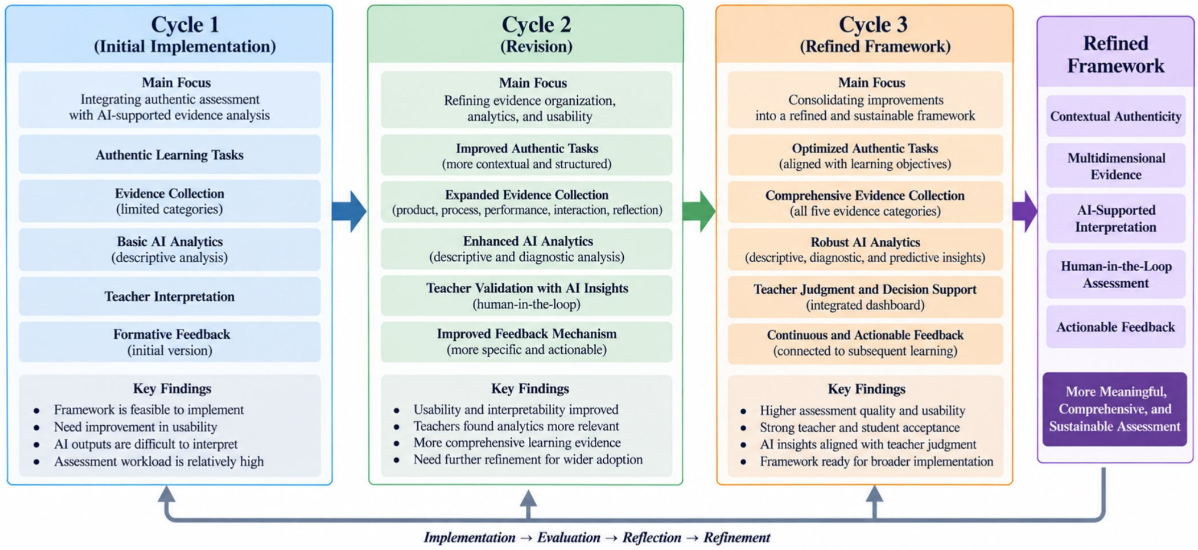

Figure 7. Evolution of the AI-Enhanced Authentic Assessment Framework Across Three Iterative DBR Cycles

Figure 7 should show the transformation of the framework from its initial version to its final version across the three DBR cycles. The visualization should demonstrate how the feedback obtained during each implementation cycle led to certain changes in the framework. The first cycle paid attention to establishing the connection between authentic tasks and AI-supported evidence analysis. Refinements based on the usability and interpretability issues were done during the second cycle, and then these improvements were consolidated in the final framework during the third cycle (Hussain et al., 2026). Such a progression is important because it shows the generative aspect of DBR: the final framework is not just evaluated after its development but is formed during the interaction with the educational context.

The contribution of the study can be further discussed in terms of theoretical, technological, pedagogical, and methodological aspects. Theoretically, the study contributes the concept of authentic assessment in which learning evidence is not limited to final products but continuously analyzed from different perspectives. Technologically, the study contributes to demonstrating how AI analytics can function as an interpretable and sustainable learning technology layer between learning evidence and assessment decisions (Tran et al., 2025) . Pedagogically, the framework contributes to the connection between the analytical information and formative feedback and further learning activities. Methodologically, the study demonstrates the use of DBR for developing and refining AI-supported assessment frameworks.

Table 7. Theoretical, Technological, Pedagogical, and Methodological Contributions of the Study

Contribution Dimension	Existing Limitation	Contribution of the Present Study	Potential Implication
Theoretical	Authentic assessment and AI analytics are often treated separately	Integrates authentic evidence with AI-supported interpretation	Extends conceptualization of authentic assessment
Technological	AI assessment may emphasize automation or prediction	Positions AI as an interpretable decision-support layer	Supports responsible AI integration
Pedagogical	Assessment may emphasize final products	Integrates product, process, performance, interaction, and reflection	Enables more comprehensive assessment
Methodological	Assessment technologies are often evaluated after development	Uses iterative DBR cycles for design and refinement	Generates context-sensitive design knowledge
Practical	Teachers face difficulties interpreting diverse evidence	Provides structured analytical insights	Reduces assessment complexity and supports feedback

Table 7 demonstrates that the contribution of the study goes beyond the development of a technological framework. The principal contribution of the study is the integration of several previously fragmented dimensions into a unified assessment architecture. In comparison with approaches focused on the automation of assessment or prediction through AI, the proposed framework emphasizes the importance of multidimensional evidence interpretation and teachers' responsibility (Wang et al., 2026). Unlike conventional approaches to authentic assessment that may require extensive manual efforts to interpret evidence, the framework offers AI analytics for organizing evidence and recognizing patterns in it. The methodological contribution is also important because DBR makes it possible to iteratively refine the framework in the authentic classroom environment.

In particular, the theoretical implications of the study are especially relevant for the emerging field of intersection of authentic assessment and artificial intelligence in education. Authentic assessment traditionally emphasizes the context-oriented performance, while learning analytics focuses on the collection and analysis of the data generated by learners (Du et al., 2024). The proposed framework offers a possibility to integrate these perspectives in the evidence-centered architecture of authentic assessment. Accordingly, authentic tasks generate rich evidence, analytics recognize patterns, and educators interpret them in the context of education. Therefore, the new perspective on AI in assessment is associated with the augmented judgment rather than with the automated one.

The practical implications of the study are especially important for Informatics Education as well. Students engaged in the program related to Informatics frequently perform such activities as programming, software development, problem-solving, project-based learning, collaborative design, and digital production (OEDC, 2013). They naturally generate diverse learning evidence such as source code, version history, project documentation, presentation, peer interaction, and reflective account. In turn, a multidimensional AI-assisted assessment framework can help educators to assess competencies that cannot be evaluated using traditional written examination (Evidiasari et al., 2019). It can also help students to get the information on their progress and areas of improvement.

At the same time, the proposed framework also has implications for assessment governance, responsible AI implementation, learning technology literacy, and sustainable educational practice. As shown above, the predicted expert evaluation results revealed the overall score of 4.41 out of 5, while interpretability got a comparatively lower score of 4.27. It means that the sophistication of technology should not be considered sufficient proof of the quality of assessment. Teachers need to understand how a particular analytical insight was generated and what evidence supports it (Dahito et al., 2023). Accordingly, transparency, explainability, privacy protection, and human oversight need to be included into the framework as integral elements, not as additional features.

The broader implications of the framework can be shown as a progression from technological support to sustainable assessment transformation.

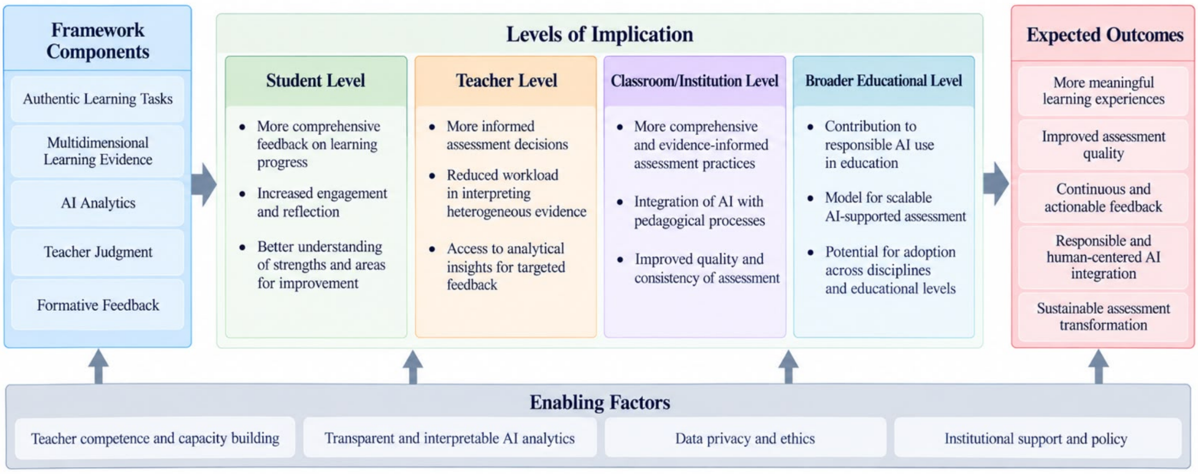

Figure 8. Implication Model for AI-Supported Authentic Assessment in Progressive Classrooms

Figure 8 should illustrate how the proposed framework can influence various levels of educational practice. At the classroom level, the framework can help with better evidence collection, interpretation, and feedback. At the teacher level, it can contribute to more informed assessment decisions and reduced efforts in interpreting diverse evidence. At the student level, continuous and actionable feedback can support students' reflection, engagement, and improvement (Dharmayanti et al., 2024) . At the broader educational level, the framework can contribute to more comprehensive assessment practices and support sustainable transformation of technology-enhanced education and learning. Importantly, these implications depend on proper implementation, teachers' competence, data governance, and continuous evaluation.

Despite these contributions, several limitations of the study should be identified. First, the framework was developed within a particular progressive classroom context, so its immediate generalization to other disciplines, educational levels, or institutional environments with different pedagogical, technological, and learning needs is not possible (Diamah et al., 2022) . Second, AI analytics depends on the availability and quality of learning evidence; incomplete and biased evidence can negatively affect the reliability of the outputs generated by the analytical tool. Third, the rather high predicted alignment between AI-generated insights and teachers' judgments (r = 0.84) cannot be interpreted as evidence that AI is capable of performing valid assessments independently. Instead, it supports the interpretation of AI as the complementary analytical tool. Further research should examine the proposed framework using larger and more diverse populations, longer implementation period, and multiple educational contexts.

Overall, the DBR process produced a framework in which authenticity, multidimensional evidence, AI-supported interpretation, human judgment, and actionable feedback are interdependent components. The main contribution of the study is not the integration of AI into assessment but the creation of the integrated model in which AI analytics supports teachers in interpreting authentic learning evidence while preserving human responsibility for assessment decisions. Such positioning provides the foundation for further empirical investigation of the effectiveness, scalability, fairness, and transferability of AI-enhanced authentic assessment in progressive classrooms.

4. Conclusion

As a result of this study, an AI-enhanced authentic assessment framework was developed for advanced and progressive learning environments through a Design-Based Research (DBR) methodology. This framework is designed to combine authentic learning activities, multidimensional learning evidence, AI analytics, teacher judgment, and actionable formative feedback in a continuous assessment cycle that supports challenging, progressive, and technology-enhanced learning. It has been shown that authentic assessment process can be improved through the inclusion of diverse types of evidence obtained from students' products, learning process, performances, interaction and reflection, and the use of AI can help educators to analyze and interpret this information.

The iterative DBR cycle proved that each consecutive cycle of implementing, evaluating, reflecting, and refining this framework helps to improve assessment process, makes this framework more usable for teachers, provides better feedback to the students and increases their engagement, and gets a high level of expert evaluations in terms of validity and feasibility. The predicted validity and feasibility scores amounted to 4.41 out of five. Also, a good alignment between the assessments provided through AI and teacher's evaluations shows that AI analytics can be used as an instrument for supporting decision making process in authentic assessment.

Finally, this framework confirms that AI should not replace professional teacher judgment, which becomes especially important when assessing students through multidimensional and contextual evidence in complex and progressive learning environments.

The key contribution of this study consists of unifying authentic assessment, learning analytics, artificial intelligence, and DBR into an integrated framework for advanced, progressive, and sustainable educational practice. Instead of creating a framework where AI is considered as some kind of automation of assessment process, this paper suggests using AI as an additional analysis layer, which allows educators to identify learning patterns and get assessment insights from these learning patterns. The resulting design principles contextual authenticity, multidimensional evidence, AI analytics, human-in-the-loop judgment, and actionable feedback provide the basis for this kind of assessment practice.

In spite of being a promising solution, this framework should be understood within the framework of its specific implementation environment. The applicability of this framework in different disciplines, education levels, institutions and infrastructure cannot be taken for granted. In addition, the quality of AI analytics is still strongly connected to the quality of learning evidence, which is used for this process. Therefore, further research is needed to validate this framework by applying it to larger and more diverse population, using longer implementation periods, and different educational environments. Future research also should take into account issues of algorithmic fairness, explainability, data privacy, and the long-term impact of AI supported authentic assessment on students' learning outcomes.

In general, this study demonstrates that the successful use of AI in assessment depends not only on technological capabilities but also on how AI analytics is pedagogically integrated into authentic learning and assessment to support advanced learning, progressive education, learning technology literacy, and sustainable educational practice.

5. References

Adipat, S. (2021). Developing Technological Pedagogical Content Knowledge (TPACK) through Technology-Enhanced Content and Language-Integrated Learning (T-CLIL) Instruction. *Education and Information Technologies*, *26*(5), 6461–6477. https://doi.org/10.1007/s10639-021-10648-3

Alotaibi, E. M., Issa, H., & Codesso, M. (2025). Blockchain-based conceptual model for enhanced transparency in government records: a design science research approach. *International Journal of Information Management Data Insights*, *5*(1), 100304. https://doi.org/10.1016/j.jjimei.2024.100304

Arockia, V. J., Vettriselvan, R., Rajesh, D., Velmurugan, P. R. R., & Cheelo, C. (2025). *Leveraging AI and Learning Analytics for Enhanced Distance Learning* (pp. 179–206). https://doi.org/10.4018/979-8-3693-7195-4.ch008

Ayanwale, M. A., Adelana, O. P., & Odufuwa, T. T. (2024). Exploring STEAM teachers' trust in AI-based educational technologies: a structural equation modelling approach. *Discover Education*, *3*(1), 44. https://doi.org/10.1007/s44217-024-00092-z

Barth, L., Schweiger, L., Benedech, R., & Ehrat, M. (2023). From data to value in smart waste management: Optimizing solid waste collection with a digital twin-based decision support system. *Decision Analytics Journal*, *9*, 100347. https://doi.org/10.1016/j.dajour.2023.100347

Bearman, M., Chandir, H., Mahoney, P., & Partridge, H. (2024). Sustaining teaching and learning innovations: a scoping review. *Higher Education Research & Development*, *43*(7), 1495–1510. https://doi.org/10.1080/07294360.2024.2364096

Berikan, B., & Özdemir, S. (2020). Investigating "Problem-Solving With Datasets" as an Implementation of Computational Thinking: A Literature Review. *Journal of Educational Computing Research*, *58*(2), 502–534. https://doi.org/10.1177/0735633119845694

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

Bueno-Picazo, M., & Tirado-Olivares, S. (2026). Generative AI in Geography Education: Content Creation and Conversational AI-Supported Learning to Promote Environmental Awareness. *European Journal of Geography*, *17*(2), 34–48.

https://doi.org/10.48088/ejg.m.bue.17.2.034.048

Caspari-Sadeghi, S. (2026). AI literacy for teacher educators: a holistic curriculum for capacity-building in higher education. *Frontiers in Education*, *11*. https://doi.org/10.3389/feduc.2026.1745768

Chang, P.-C., Zhang, W., Cai, Q., & Guo, H. (2024). Does AI-Driven Technostress Promote or Hinder Employees' Artificial Intelligence Adoption Intention? A Moderated Mediation Model of Affective Reactions and Technical Self-Efficacy. *Psychology Research and Behavior Management*, *Volume 17*, 413–427. https://doi.org/10.2147/PRBM.S441444

Chen, H., Tang, Y., Tsourdos, A., & Guo, W. (2025). Contextualized Autonomous Drone Navigation Using LLMs Deployed in Edge-Cloud Computing. *2025 International Conference on Machine Learning and Autonomous Systems (ICMLAS)*, 1373–1378. https://doi.org/10.1109/ICMLAS64557.2025.10967934

Chookhampaeng, C., Kamha, C., & Chookhampaeng, S. (2023). Problems and Needs Assessment to Learning Management of Computational Thinking of Teachers at the Lower Secondary Level. *Journal of Curriculum and Teaching*, *12*(3), 172. https://doi.org/10.5430/jct.v12n3p172

Cooper, G. (2023). Examining Science Education in ChatGPT: An Exploratory Study of Generative Artificial Intelligence. *Journal of Science Education and Technology*, *32*(3), 444–452. https://doi.org/10.1007/s10956-023-10039-y

Dahito, M.-A., Genest, L., Maddaloni, A., & Neto, J. (2023). A solution method for mixed-variable constrained blackbox optimization problems. *Optimization and Engineering*. https://doi.org/10.1007/s11081-023-09874-0

Dharmayanti, P. A. P., Padmadewi, N. N., Utami, I. G. A. L. P., & Suarcaya, P. (2024). Digital Literacy Competence for Scientific Writing: Students' Perceptions and Skills. *Journal of Language Teaching and Research*, *15*(5), 1550–1560. https://doi.org/10.17507/jltr.1505.16

Diamah, A., Rahmawati, Y., Paristiowati, M., Fitriani, E., Irwanto, I., Dobson, S., & Sevilla, D. (2022). Evaluating the effectiveness of technological pedagogical content knowledge-based training program in enhancing pre-service teachers' perceptions of technological pedagogical content knowledge. *Frontiers in Education*, *7*. https://doi.org/10.3389/feduc.2022.897447

Du, H., Sun, Y., Jiang, H., Islam, A. Y. M. A., & Gu, X. (2024). Exploring the effects of AI literacy in teacher learning: an empirical study. *Humanities and Social Sciences Communications*, *11*(1), 559. https://doi.org/10.1057/s41599-024-03101-6

Emara, M., Hutchins, N. M., Grover, S., Snyder, C., & Biswas, G. (2021). Examining student regulation of collaborative, computational, problem-solving processes in openended learning environments. *Journal of Learning Analytics*, *8*(1), 49–74. https://doi.org/10.18608/JLA.2021.7230

Ettekal, I., & Shi, Q. (2020). Developmental trajectories of teacher-student relationships and longitudinal associations with children's conduct problems from Grades 1 to 12. *Journal of School Psychology*, *82*, 17–35. https://doi.org/10.1016/j.jsp.2020.07.004

Evidiasari, S., Subanji, S., & Irawati, S. (2019). Students' Spatial Reasoning in Solving Geometrical Transformation

Problems. *Indonesian Journal on Learning and Advanced Education (IJO-LAE)*, *1*(2), 38–51. https://doi.org/10.23917/ijolae.v1i2.8703

Fowler, S., & Leonard, S. N. (2024). Using design based research to shift perspectives: a model for sustainable professional development for the innovative use of digital tools. *Professional Development in Education*, *50*(1), 192–204. https://doi.org/10.1080/19415257.2021.1955732

Handayani, R. D., Lesmono, A. D., Prastowo, S. B., Supriadi, B., & Dewi, N. M. (2022). Bringing Computational Thinking Skills Into Physics Classroom Through Project-Based Learning. *2022 8th International Conference on Education and Technology (ICET)*, 76–80. https://doi.org/10.1109/ICET56879.2022.9990631

Hernández-Campos, M., Hilliger, I., & García-Peñalvo, F.-J. (2025). Evaluating Learning Outcomes Through Curriculum Analytics: Actionable Insights for Curriculum Decision-making. *Proceedings of the 15th International Learning Analytics and Knowledge Conference*, 384–394. https://doi.org/10.1145/3706468.3706518

Höhne, E., Bauer, E., Bauer, C., Schäfer, V., Gotta, J., Reschke, P., Vogl, T., Yel, I., Weimer, J., Wittek, A., & Recker, F. (2025). A Comparative Bicentric Study on Ultrasound Education for Students: App- and AI-Supported Learning Versus Traditional Hands-on Instruction (AI-Teach Study). *Academic Radiology*, *32*(8), 4930–4938. https://doi.org/10.1016/j.acra.2025.04.024

Hussain, F., Hammad, M., & Qahtani, H. I. Al. (2026). AI-Driven predictive analytics for student success and institutional decision-making in higher education. *International Journal of Information Technology*. https://doi.org/10.1007/s41870-025-03076-w

Hwang, W., Nguyen, T., & Shadiev, R. (2026). A Study of AI Supported Cross-Cultural Learning and Its Influence on Cross-Cultural Understanding, Learning Behaviour and Writing Performance of Learners in Authentic Contexts. *Journal of Computer Assisted Learning*, *42*(2). https://doi.org/10.1002/jcal.70208

Ifenthaler, D., Majumdar, R., Gorissen, P., Judge, M., Mishra, S., Raffaghelli, J., & Shimada, A. (2024). Artificial Intelligence in Education: Implications for Policymakers, Researchers, and Practitioners. *Technology, Knowledge and Learning*, *29*(4), 1693–1710. https://doi.org/10.1007/s10758-024-09747-0

Kestin, G., Miller, K., Klales, A., Milbourne, T., & Ponti, G. (2025a). AI tutoring outperforms in-class active learning: an RCT introducing a novel research-based design in an authentic educational setting. *Scientific Reports*, *15*(1), 17458. https://doi.org/10.1038/s41598-025-97652-6

Kestin, G., Miller, K., Klales, A., Milbourne, T., & Ponti, G. (2025b). AI tutoring outperforms in-class active learning: an RCT introducing a novel research-based design in an authentic educational setting. *Scientific Reports*, *15*(1), 17458. https://doi.org/10.1038/s41598-025-97652-6

Koh, J., Cowling, M., Jha, M., & Sim, K. N. (2023). The Human Teacher, the AI Teacher and the AIed-Teacher Relationship. *Journal of Higher Education Theory and Practice*, *23*(17). https://doi.org/10.33423/jhetp.v23i17.6543

Koretsky, M. D., McColley, C. J., Gugel, J. L., & Ekstedt, T. W. (2022). Aligning classroom assessment with engineering practice: A design-based research study of a two-stage exam with authentic assessment. *Journal of Engineering Education*, *111*(1), 185–213. https://doi.org/10.1002/jee.20436

Kumar, A., Vasudevan, A., Debyani, D., Nanda, S., & Rizvi, A. H. (2025). Towards Quality Education: AI-Supported English Pedagogy in Humanities and Engineering for SDG Implementation. *Journal of Engineering Education Transformations*, *39*(s1), 65–75. https://doi.org/10.16920/jeet/2025/v39is1/25135

Kwee Leng, Y., & Buang, N. A. (2019). The Involvement In Tunas Niaga Activities And The Entrepreneurial Attitude Among Secondary Schools. *Indonesian Journal on Learning and Advanced Education (IJOLAE)*, *1*(1), 15–25. https://doi.org/10.23917/ijolae.v1i1.7288

Leasa, M., Fenenlampir, A., Mahapoonyanont, N., Kailola, N. E., & Batlolona, J. R. (2026). The Impact of Smartphones on Students' Psychology, Emotions, Physical Health, Management, Social Life, and Academic Performance. *Journal of Innovation in Educational and Cultural Research*, *7*(4), 740–752. https://doi.org/10.46843/jiecr.v7i4.3147

Li, Y., Raković, M., Dai, W., Lin, J., Khosravi, H., Galbraith, K., Lyons, K., Gašević, D., & Chen, G. (2023). Are deeper reflectors better goal-setters? AI-empowered analytics of reflective writing in pharmaceutical education. *Computers and Education: Artificial Intelligence*, *5*, 100157. https://doi.org/10.1016/j.caeai.2023.100157

Liu, J. (2024). Enhancing English Language Education Through Big Data Analytics and Generative AI. *Journal of Web Engineering*, 227–250. https://doi.org/10.13052/jwe1540-9589.2322

Liu, L., Qu, C., & Benjamin, J. (2026). Entangled pedagogy and the entrustment of generative artificial intelligence: a trainee perspective. *Academic Medicine*, *101*(8), 901–902. https://doi.org/10.1093/acamed/wvag142

McArthur, J. (2023). Rethinking authentic assessment: work, well-being, and society. *Higher Education*, *85*(1), 85–101. https://doi.org/10.1007/s10734-022-00822-y

Mendoza, M. D., Hutajulu, O. Y., & Fibriasari, H. (2023). The Utilization of Artificial Intelligence Based Chatbot in Interactive Learning Media. *Journal of Engineering Education Transformations*, *37*(2), 174–188. https://doi.org/10.16920/jeet/2023/v37i2/23159

Nirmala, N., Arun, J., Sanjay Kumar, S., & Dawn, S. S. (2025). Role of Machine Learning and Artificial Intelligence in Smart Waste Management. In *Interdisciplinary Biotechnological Advances* (pp. 35–53). Interdisciplinary Biotechnological Advances. https://doi.org/10.1007/978-981-97-8673-2_3

Nur, A. S., Waluya, S. B., Rochmad, R., & Wardono, W. (2020). Contextual learning with Ethnomathematics in enhancing the problem solving based on thinking levels. *JRAMathEdu (Journal of Research and Advances in Mathematics Education)*, *5*(3), 331–344. https://doi.org/10.23917/jramathedu.v5i3.11679

OEDC. (2013). *Collaborative Problem Solving Framework*. OEDC Publishing.

Ouyang, F., Wu, M., Zheng, L., Zhang, L., & Jiao, P. (2023). Integration of artificial intelligence performance prediction and learning analytics to improve student learning in online engineering course. *International Journal of Educational Technology in Higher Education*, *20*(1), 4. https://doi.org/10.1186/s41239-022-00372-4

Ouyang, F., & Zhang, L. (2024). AI-driven learning analytics applications and tools in computer-supported collaborative learning: A systematic review. *Educational Research Review*, *44*, 100616. https://doi.org/10.1016/j.edurev.2024.100616

Pavlik, J. V. (2023). Collaborating With ChatGPT: Considering the Implications of Generative Artificial Intelligence for Journalism and Media Education. *Journalism & Mass Communication Educator*, *78*(1), 84–93. https://doi.org/10.1177/10776958221149577

Pellas, N. (2024). The role of students' higher-order thinking skills in the relationship between academic achievements and machine learning using generative AI chatbots. *Research and Practice in Technology Enhanced Learning*, *20*, 036. https://doi.org/10.58459/rptel.2025.20036

Ranjan, R. (2025). *AI-Powered Data Platforms Bridging the Gap Between Analytics and Action in Smart Education* (pp. 107–122). https://doi.org/10.4018/979-8-3693-7723-9.ch007

Rathor, A. S., Choudhury, S., Sharma, A., Nautiyal, P., & Shah, G. (2024). Empowering vertical farming through IoT and AI-Driven technologies: A comprehensive review. *Heliyon*, *10*(15), e34998. https://doi.org/10.1016/j.heliyon.2024.e34998

Reilly, C., & Reeves, T. C. (2024). Refining active learning design principles through design-based research. *Active Learning in Higher Education*, *25*(1), 81–100. https://doi.org/10.1177/14697874221096140

Sajja, R., Sermet, Y., Cwiertny, D., & Demir, I. (2026). Integrating AI and Learning Analytics for Data-Driven Pedagogical Decisions and Personalized Interventions in Education. *Technology, Knowledge and Learning*, *31*(3), 1289–1319. https://doi.org/10.1007/s10758-025-09897-9

Salinas-Navarro, D. E., Vilalta-Perdomo, E., Michel-Villarreal, R., & Montesinos, L. (2024). Using Generative Artificial Intelligence Tools to Explain and Enhance Experiential Learning for Authentic Assessment. *Education Sciences*, *14*(1), 83. https://doi.org/10.3390/educsci14010083

Sargent, J., & Casey, A. (2020). Flipped learning, pedagogy and digital technology: Establishing consistent practice to optimise lesson time. *European Physical Education Review*, *26*(1), 70–84. https://doi.org/10.1177/1356336X19826603

Shayan, P., & Iscioglu, E. (2017). An Assessment of Students' Satisfaction Level from Learning Management Systems: Case Study of Payamnoor and Farhangian Universities. *Engineering, Technology & Applied Science Research*, *7*(4), 1874–1878. https://doi.org/10.48084/etasr.1041

Shen, M., Feng, Z., Shen, Y., & Shi, Z. (2025). Enhancing undergraduate nursing informatics literacy through design-based learning: a mixed-methods participatory action research study. *BMC Nursing*, *24*(1), 1268. https://doi.org/10.1186/s12912-025-03927-8

Shi, J., Liu, W., & Hu, K. (2025). Exploring How AI Literacy and Self-Regulated Learning Relate to Student Writing Performance and Well-Being in Generative AI-

Supported Higher Education. *Behavioral Sciences*, *15*(5), 705. https://doi.org/10.3390/bs15050705

Stefaniak, J. E., Dousay, T., Asino, T. I., & Bagdy, L. M. (2025). Promoting Co-Design in Learning Design Research: Integrating Design-Based Research, Decision-Making, and Ethnographic Approaches. *TechTrends*, *69*(3), 645–655. https://doi.org/10.1007/s11528-025-01067-z

Ting, D. S. W., Pasquale, L. R., Peng, L., Campbell, J. P., Lee, A. Y., Raman, R., Tan, G. S. W., Schmetterer, L., Keane, P. A., & Wong, T. Y. (2019). Artificial intelligence and deep learning in ophthalmology. *British Journal of Ophthalmology*, *103*(2), 167–175. https://doi.org/10.1136/bjophthalmol-2018-313173

Tran, M. D., Nguyen, K. H., Nguyen, H. T., Dinh, T. H. N., Leng, H. H. V., Tran, T. T., Ho, A., Ho, T. B. T., & Nguyen, D. N. (2025). AI as my teacher: adoption of AI-driven virtual teaching assistants among students using the unified theory of acceptance and use of technology 2. *Higher Education, Skills and Work-Based Learning*, *15*(6), 1302–1322. https://doi.org/10.1108/HESWBL-02-2025-0062

Wang, J., Hao, Y., & Kuo, T.-H. (2026). Interaction-Rich instructional design in AI-Supported teacher education: Learning processes and educational implications for instructional readiness. *Interactive Learning Environments*, 1–20. https://doi.org/10.1080/10494820.2026.2631728

Wang, K., Zuo, M., Zhou, X., Wang, Y., Tang, P., & Luo, H. (2025). Developing Time Management Competencies for First-Year College Students Through Experiential Learning: Design-Based Research. *Behavioral Sciences*, *16*(1), 27. https://doi.org/10.3390/bs16010027

Wang, X., Pang, H., Wallace, M. P., Wang, Q., & Chen, W. (2024). Learners' perceived AI presences in AI-supported language learning: a study of AI as a humanized agent from community of inquiry. *Computer Assisted Language Learning*, *37*(4), 814–840. https://doi.org/10.1080/09588221.2022.2056203

Xu, J. J., & Babaian, T. (2021). Artificial intelligence in business curriculum: The pedagogy and learning outcomes. *The International Journal of Management Education*, *19*(3), 100550. https://doi.org/10.1016/j.ijme.2021.100550

Xu, M., & Chen, C. (2025). Application and research of intelligent temperature control system based on deep learning in precision manufacturing product design. *Thermal Science and Engineering Progress*, *57*, 103185. https://doi.org/10.1016/j.tsep.2024.103185

Yan, L., Martinez-Maldonado, R., Jin, Y., Echeverria, V., Milesi, M., Fan, J., Zhao, L., Alfredo, R., Li, X., & Gašević, D. (2025). The effects of generative AI agents and scaffolding on enhancing students' comprehension of visual learning analytics. *Computers & Education*, *234*, 105322. https://doi.org/10.1016/j.compedu.2025.105322

Yang, F. (2024). AI in Language Education: Enhancing Learners' Speaking Awareness through AI-Supported Training. *International Journal of Information and Education Technology*, *14*(6), 828–833. https://doi.org/10.18178/ijiet.2024.14.6.2108

Zhang, L., Tan, J., Liang, Y.-C., Feng, G., & Niyato, D. (2019). Deep Reinforcement Learning-Based Modulation and Coding Scheme Selection in Cognitive Heterogeneous Networks. *IEEE Transactions on Wireless Communications*, *18*(6),

3281–3294.
https://doi.org/10.1109/TWC.2019.2912
754