Buca Eğitim Fakültesi Dergisi,
2025, sayı 66, ss. 3642-3674

Derleme Makale

The Journal of Buca Faculty of Education,
2025, issue 66, pp. 3642-3674

Review Article

Matematik Eğitiminde Yapay Zeka Destekli Değerlendirmeler: Çağdaş Araştırma Literatürünün Sistematik Biçimde İncelenmesi

AI-Powered Assessments in Mathematics Education: A Systematic Review of Contemporary Research Literature

Rusen Meylani [1]

[1]*Sorumlu Yazar, Dr. Öğretim Üyesi, Dicle Üniversitesi, rusen.meylani@dicle.edu.tr, (https://orcid.org/0000-0002-3121-6088)*

Geliş Tarihi: *04.08.2024* ***Kabul Tarihi:*** *28.10.2025*

ÖZ

Bu sistematik derleme çalışması, yapay zekâ (YZ) teknolojilerinin matematik değerlendirmelerinde kullanılmasını eleştirel bir bakış açısıyla incelemektedir. Geleneksel değerlendirme yöntemleri, özellikle matematik gibi bilişsel açıdan yoğun alanlarda, zamanında, kişiselleştirilmiş ve pedagojik açıdan anlamlı geri bildirim sunmada yetersiz kalmaktadır. PRISMA protokolü doğrultusunda yapılandırılan bu çalışma, güncel literatürü tematik analiz yoluyla sistematik bir biçimde sentezlemiş ve MAXQDA 24 yazılımı kullanılarak kodlama yapılmıştır. Bulgular, Akıllı Öğretim Sistemleri (ITS) ve uyarlanabilir hesap makineleri gibi YZ tabanlı araçların değerlendirmelerde nesnellik, doğruluk ve öğrenci katılımı açısından önemli katkılar sunduğunu ortaya koymaktadır. Ayrıca, bu teknolojilerin sistemik eşitsizlikleri azaltma ve bireysel ihtiyaçlara göre öğretimi uyarlama potansiyeline sahip olduğu görülmektedir. Ancak çalışmada, algoritmik şeffaflık eksikliği, veri güvenliği riskleri, altyapı eşitsizlikleri ve etik belirsizlikler gibi ciddi sorunlara da dikkat çekilmektedir. YZ'nin matematik değerlendirmelerine etkili entegrasyonu yalnızca teknolojik yeniliklerle değil, aynı zamanda etik temelli tasarım, bağlama duyarlı öğretmen eğitimi ve sürekli eleştirel sorgulama ile mümkün olabilir. Bu bağlamda çalışma, yapay zekâ destekli değerlendirme pratiklerinin eğitsel değerini, eşitlik boyutunu ve sürdürülebilirliğini sorgulayan özgün bir katkı sunmaktadır.

Anahtar Kelimeler: Yapay zekâ, matematik değerlendirme, eleştirel pedagoji, kişiselleştirilmiş öğrenme, eğitimde eşitlik, etik, akıllı öğretim sistemleri, sistematik derleme.

ABSTRACT

This systematic review critically examines the integration of artificial intelligence (AI) technologies in mathematics assessment, interrogating both the promises and pitfalls of this educational transformation. Traditional assessment practices often fall short in delivering timely, personalized, and pedagogically meaningful feedback, particularly in cognitively demanding domains such as mathematics. Drawing upon contemporary literature and guided by PRISMA standards, this study synthesizes findings using thematic analysis via MAXQDA 24. The review reveals that AI tools—including Intelligent Tutoring Systems and adaptive calculators—offer substantial benefits in terms of precision, objectivity, and learner engagement. Moreover, AI-driven systems demonstrate potential for advancing educational equity by addressing systemic biases and tailoring instruction to individual needs. However, the study also foregrounds serious challenges: algorithmic opacity, data privacy risks, infrastructural inequities, and ethical ambiguities that threaten the pedagogical integrity of AI-enhanced evaluations. The findings underscore that effective AI integration in mathematics assessment requires more than technological enthusiasm—it necessitates ethically grounded design, context-aware teacher training, and continuous critical inquiry. By confronting both the opportunities and limitations of AI, this paper contributes to a more nuanced understanding of how intelligent technologies reshape assessment in ways that are equitable, educative, and sustainable.

Keywords: Artificial ıntelligence, mathematics assessment, critical pedagogy, personalization, educational equity, ethics in aı, ıntelligent tutoring systems, systematic review.

INTRODUCTION

1.1. AI in Educational Evaluations

Artificial intelligence (AI) is increasingly recognized as a transformative force in education, particularly in the domain of assessment. AI technologies have reshaped traditional evaluation paradigms by automating processes, providing real-time and personalized feedback, and enabling adaptive learning experiences tailored to the needs of individual students. These capabilities support both educators and learners by streamlining instructional decision-making and fostering more effective pedagogical practices (Owan, 2023; Meylani, 2024b, 2024d, 2025b).

In education more broadly, AI has enabled dynamic and interactive assessment procedures that reflect the diversity of students' learning preferences and progress trajectories. Tools such as automated grading systems and feedback generators allow for more efficient evaluation while also addressing issues of timeliness, reliability, and scalability (Chisom, 2024; Meylani, 2024a, 2025a). These tools have been particularly beneficial in large-scale education systems where human grading capacity is limited and standardization is essential.

While these developments have positively influenced educational assessment practices across disciplines, a critical need remains to investigate how such technologies specifically function within subject-specific domains—particularly mathematics—where the complexity of reasoning, procedural fluency, and conceptual understanding requires more nuanced evaluation approaches. It is within this disciplinary context that the integration of AI promises both unique affordances and particular challenges that merit focused analysis (Meylani, 2025b; Meylani & Kutluca, 2025).

1.2. AI in Mathematics Assessments

Mathematics education places a premium on problem-solving, logical reasoning, and procedural accuracy. Assessments in this field not only measure correctness but also diagnose misconceptions, monitor cognitive development, and guide pedagogical interventions. However, traditional paper-based or summative assessments often fall short in capturing the depth of student understanding and in providing timely, targeted feedback (Meylani, 2025a, 2025d).

Integrating AI into mathematics assessments has introduced a range of possibilities for addressing these limitations. AI tools such as Intelligent Tutoring Systems (ITS), AI-driven calculators, automated feedback platforms, and diagnostic algorithms have been applied to enhance formative and summative assessments. These systems personalize feedback based on individual student input, adjust question difficulty in real-time, and provide analytics that inform instructional decisions (Xu & Ouyang, 2022; Lye, 2024; Meylani, 2024b, 2025b, 2025c). Furthermore, empirical studies confirm that AI-enhanced mathematics assessments improve not only the efficiency and objectivity of evaluation but also positively impact student performance and engagement (Agarwal, 2024; Bedizel, 2023; Meylani, 2024c; Kaplan & Meylani, 2025).

Nevertheless, the incorporation of AI into mathematics education is not without its complications. Unique disciplinary challenges—such as assessing mathematical notation, interpreting student reasoning steps, and maintaining alignment with curricular goals—require AI systems to be finely attuned to the epistemological and pedagogical nuances of mathematics. In addition, concerns regarding algorithmic fairness, data security, and teacher readiness remain persistent obstacles to broader adoption (Matzakos, 2023; Hermann, 2021; Meylani, 2025a; Meylani & Kaplan, 2025).

Given these opportunities and challenges, a focused investigation of how AI is being integrated into mathematics assessment is warranted. The current literature contains a growing number of studies on this topic, yet a comprehensive synthesis that critically examines the applications, benefits, and limitations of AI in mathematics-specific assessment contexts is still lacking (Meylani, 2025b; Meylani & Kutluca, 2025).

1.3. Study Aims and Research Questions

This study aims to systematically analyze how AI technologies are being implemented in mathematics assessments, with particular attention to their pedagogical roles, technological capacities, ethical implications, and impact on learning outcomes. By conducting a qualitative synthesis of contemporary literature using a systematic review methodology, the study provides a critical account of how AI is reshaping mathematics assessment practices in theory and in application.

The following research questions guide the inquiry:

1. How do AI technologies enhance the accuracy and objectivity of mathematics assessments compared to traditional methods?
2. In what ways do AI-powered assessments provide individualized feedback and learning experiences to improve student engagement and performance in mathematics?
3. What are the challenges and limitations associated with implementing AI technologies in mathematics assessments, and how can these be addressed?
4. How do educators perceive the integration of AI in mathematics assessments, and what factors influence their adoption of these technologies?
5. What are the implications of AI-powered assessments for promoting equity and inclusivity in mathematics education?

These questions aim to uncover not only the practical effects of AI adoption in mathematics assessments but also the theoretical and ethical dimensions that must be considered for responsible and effective implementation. The review contributes to the literature by identifying patterns, gaps, and future directions, and by offering evidence-based insights for practitioners and policymakers seeking to navigate this evolving landscape.

METHODOLOGY

This study employed a systematic review methodology grounded in the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) framework to critically examine contemporary research on artificial intelligence (AI) applications in mathematics assessments (Page et al., 2021). The review focused on synthesizing qualitative findings and employed thematic analysis using MAXQDA 24 software to identify conceptual trends, pedagogical implications, and implementation challenges across the selected studies.

2.1. Search Strategy

A comprehensive literature search was conducted across five major academic databases: ERIC, Scopus, Web of Science, IEEE Xplore, and Google Scholar. These databases were selected for their wide coverage of educational technology, assessment, and AI-related publications. The search was limited to works published between January 2010 and December 2024, ensuring a focus on recent developments.

The search strategy involved the use of Boolean operators and keyword combinations to capture relevant studies. Search terms included:

- "Artificial Intelligence" AND "Mathematics Education" AND "Assessment"
- "AI" AND "Mathematics Assessment"
- "Machine Learning" AND "Math Testing"
- "Adaptive Learning Systems" AND "Mathematics"

Search filters were applied to include only peer-reviewed journal articles, conference papers, and scholarly reports written in English. Duplicate records were manually removed before screening.

2.2. Inclusion and Exclusion Criteria

To ensure the conceptual and methodological relevance of the review, a set of explicit inclusion and exclusion criteria were applied:

Inclusion Criteria:

- The study focused specifically on AI applications in mathematics assessment contexts.
- It was empirical, reporting original qualitative or mixed-method findings.
- It was peer-reviewed or published in a credible academic outlet.
- It was written in English and published between 2010 and 2024.

Exclusion Criteria:

- Studies that addressed AI in education without specific reference to **assessment** practices.
- Works focusing solely on instruction or curriculum, rather than evaluation.
- Non-empirical sources (e.g., opinion articles, editorials, book reviews).
- Studies inaccessible in full text or written in languages other than English.

2.3. PRISMA Flow and Study Selection

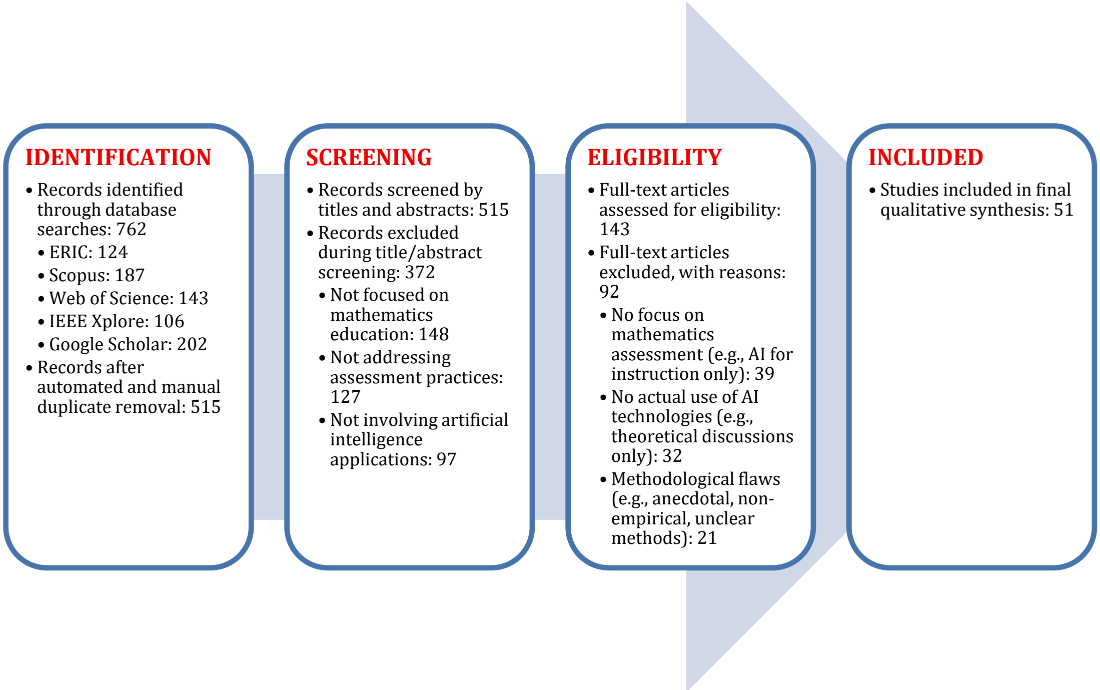
The PRISMA framework guided a four-stage selection process to ensure transparency and replicability:

1. Identification: A total of 762 records were retrieved from the database searches.
2. Screening: After the removal of 247 duplicates, 515 titles and abstracts were screened. Of these, 372 were excluded for not meeting the inclusion criteria.
3. Eligibility: 143 full-text articles were reviewed for eligibility. An additional 92 were excluded due to irrelevance to mathematics assessments, lack of AI integration, or methodological weaknesses.
4. Inclusion: A final set of 51 studies was selected for qualitative synthesis.

This process is illustrated in the PRISMA flow diagram (Figure 1), which documents each stage of inclusion and exclusion.

Figure 1

The PRISMA Flow Diagram



2.4. Data Extraction

Data from the 40 eligible studies were extracted systematically using a structured coding template. The extracted data included:

- **Bibliographic Information**: Author(s), publication year, journal or source.
- **Study Design**: Methodological approach, setting, education level, and participant details (if applicable).
- **AI Technology Employed**: Tools and systems such as Intelligent Tutoring Systems, automated grading software, AI-driven calculators, or adaptive testing platforms.
- **Assessment Context**: Purpose of assessment (e.g., diagnostic, formative, summative), scope, and instructional alignment.
- **Reported Findings**: Benefits, challenges, and implications of AI integration in mathematics assessments.

This structured approach ensured consistency and accuracy across the data extraction process and prepared the studies for subsequent coding in MAXQDA.

2.5. Data Analysis Using MAXQDA 24

A thematic analysis was conducted using **MAXQDA 24**, which allowed for systematic identification and categorization of patterns within the data. The analysis followed the six-phase approach by Braun and Clarke (2006):

1. **Familiarization**: All studies were read thoroughly and initial impressions were noted.
2. **Initial Coding**: Meaningful segments of text were assigned descriptive codes related to AI technologies, assessment functions, instructional implications, and challenges.
3. **Theme Development**: Codes were clustered into conceptual themes and sub-themes based on their recurrence and interrelations.

4. **Theme Refinement**: Themes were refined for internal coherence and external distinctiveness, ensuring they were analytically robust.
5. **Naming and Defining Themes**: Each theme was labeled and defined clearly to reflect its conceptual boundaries.
6. **Final Analysis**: Thematic findings were synthesized in relation to the research questions and literature base.

To strengthen the credibility and confirmability of the analysis, **all 40 studies were independently coded by the researcher at two distinct times one month apart** using the finalized framework. Coding reliability was calculated using **Cohen's kappa**, yielding a value of **$\kappa = 0.82$**, which indicates **substantial agreement** (Landis & Koch, 1977).

2.6. Quality Assessment

The methodological quality of each study was appraised using a rubric adapted from the Critical Appraisal Skills Programme (CASP, 2018) and other established qualitative research guidelines. Four criteria were evaluated:

1. Clarity and specificity of research questions
2. Appropriateness of the research design
3. Depth of data analysis and interpretative rigor
4. Relevance and transferability of findings

Studies scoring poorly on more than one of these criteria were excluded during the eligibility stage. The 40 included studies demonstrated satisfactory to high methodological rigor, making them suitable for qualitative synthesis.

FINDINGS

This section presents the findings from a qualitative analysis of the current literature on the application of AI in mathematics assessments. The analysis aimed to identify key themes and sub-themes that capture AI's multifaceted role in enhancing educational evaluations.

Figure 2

The Themes and Sub-Themes that Emerged from the Qualitative Analysis.

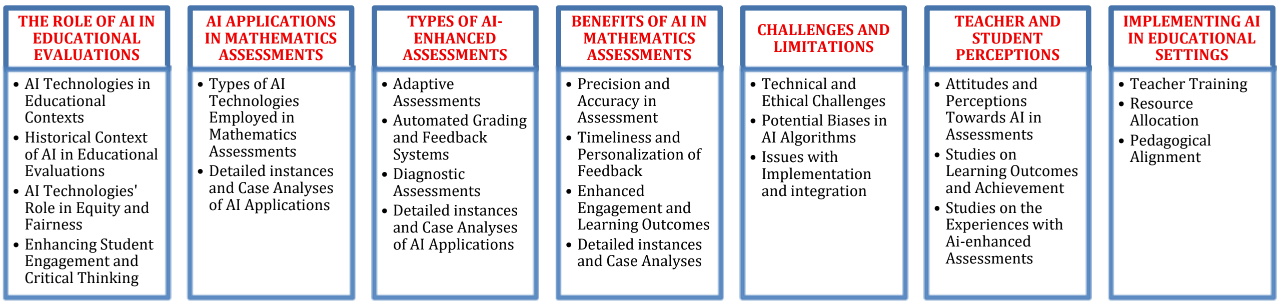


Figure 2 visually represents the themes and sub-themes that emerged from the qualitative analysis, while Table 1 provides a detailed breakdown of these themes and sub-themes, including references that support each. The themes highlight various aspects of AI's integration into educational evaluations, from the role of AI technologies in educational contexts to the specific applications and benefits of AI in mathematics assessments.

Table 1

The Summary Table that Depicts Themes and Sub-Themes Emerging from the Qualitative Review Along with the References Supporting Each Theme and Sub-Theme.

Main Theme	Sub-Theme	Citations	n	%
The Role of AI in Educational Evaluations	AI Technologies in Educational Contexts	Jiang & Pardos (2021); Grubaugh (2024)	2	3.92%
	Historical Context of AI in Educational Evaluations	Liang (2023); Bodemer (2023)	2	3.92%
	AI Technologies' Role in Equity and Fairness	Rieskamp (2023); Smyrnova-Trybulska et al. (2023); Odeyemi (2024)	3	5.88%
	Enhancing Student Engagement and Critical Thinking	Addy (2024); Khan (2023); Melkonian et al. (2022)	3	5.88%
The Role of AI in Educational Evaluations Total			10	19.61%
AI Applications in Mathematics Assessments	Types of AI Technologies	Hwang (2022); Stefanova (2024); Meylani (2024a); Pepin et al. (2025)	4	7.84%
	Case Analyses of AI Applications	Sunarto (2024); Dabingaya (2022)	2	3.92%
AI Applications in Mathematics Assessments Total			6	11.76%
Types of AI-Enhanced Assessments	Adaptive Assessments	Geary et al. (2019)	1	1.96%
	Automated Grading and Feedback Systems	Messer (2024)	1	1.96%
	Diagnostic Assessments	Molnár & Csapó (2019)	1	1.96%
	Case Analyses of AI Applications	Hiremath (2024); Lin (2024)	2	3.92%
Types of AI-Enhanced Assessments Total			5	9.80%
Benefits of AI in Mathematics Assessments	Precision and Accuracy	Owan (2023); Institute of Education Sciences (2023)	2	3.92%
	Timeliness and Personalization of Feedback	Lakomkin et al. (2018)	1	1.96%
	Engagement and Learning Outcomes	Sunarto (2024)	1	1.96%
	Case Analyses	Remoto (2023); Sarwari (2024)	2	3.92%
Benefits of AI in Mathematics Assessments Total			6	11.76%
Challenges and Limitations	Technical and Ethical Challenges	Tsopra et al. (2021); Esmaeilzadeh et al. (2021)	2	3.92%
	Potential Biases in AI Algorithms	Matzakos (2023); European Commission (2022)	2	3.92%
	Implementation and Integration	Meylani (2024c); Hermann (2021); College of Education at the University of Illinois (2024)	3	5.88%
Challenges and Limitations Total			7	13.73%
Teacher and Student Perceptions	Attitudes and Perceptions	Chiu et al. (2022); Dai et al. (2020); Walter (2024)	3	5.88%
	Learning Outcomes and Achievement	Seo et al. (2021); Hutson et al. (2022); Gerlich (2025)	3	5.88%
	Experiences with AI Assessments	Wu & Yu (2023); Muslimin (2024); Ibrahim (2024); Baez (2023)	4	7.84%
Teacher and Student Perceptions Total			10	19.61%
Implementing AI in Educational Settings	Teacher Training	Lee & Perret (2022); Meylani (2024b)	2	3.92%
	Resource Allocation	Gupta & Kitcharoen (2024); OECD (2024)	2	3.92%
	Pedagogical Alignment	Aghaziarati (2023); Mayo (2024); U.S. Department of Education (2023)	3	5.88%
Implementing AI in Educational Settings Total			7	13.73%
Grand Total			51	100.00%

3.1 Interpretation of the Summary Table

The Role of AI in Educational Evaluations is one of the two most heavily cited themes in the paper, accounting for 10 out of 51 total citations (19.61%). This indicates a strong foundational emphasis on the theoretical and structural transformation that artificial intelligence brings to educational assessment. The sub-themes are relatively balanced in citation count, with "AI Technologies' Role in Equity and Fairness" and "Enhancing Student Engagement and Critical Thinking" each receiving 3 citations (5.88%), while "AI Technologies in Educational Contexts" and "Historical Context" received 2 citations (3.92%) each. This suggests that the discussion prioritizes both the ethical-political implications of AI (e.g., fairness) and its pedagogical affordances (e.g., critical thinking), while maintaining a comprehensive conceptual grounding in the history and functions of AI in education.

AI Applications in Mathematics Assessments garnered a total of 6 citations (11.76%), with a skewed distribution favoring the sub-theme "Types of AI Technologies" which alone received 4 citations (7.84%). In contrast, the "Case Analyses" sub-theme accounted for only 2 citations (3.92%). This suggests that the paper placed more analytical and descriptive focus on the categorization and affordances of different AI tools (e.g., ITS, AI calculators, generative models) rather than detailing numerous classroom implementations. The emphasis on technological typologies indicates a desire to map out the landscape of AI tools before engaging deeper with their real-world pedagogical applications.

Types of AI-Enhanced Assessments was supported by 5 citations (9.80%), reflecting a somewhat lighter engagement with this theme. Three of the sub-themes—"Adaptive Assessments," "Automated Grading," and "Diagnostic Assessments"—each received only 1 citation (1.96%), while "Case Analyses" received 2 citations (3.92%). This minimal citation spread suggests that while these assessment types are acknowledged, they are not yet deeply problematized or richly discussed within the manuscript. The limited literature integration may point to either gaps in the literature or a need for greater elaboration in future revisions.

Benefits of AI in Mathematics Assessments also accounted for 6 citations (11.76%), indicating a balanced but moderate attention to the positive outcomes of AI integration. Sub-themes such as "Precision and Accuracy" and "Case Analyses" each received 2 citations (3.92%), while "Timeliness and Personalization of Feedback" and "Engagement and Learning Outcomes" received only 1 citation (1.96%) each. This reflects an acknowledgment of AI's instructional advantages but also suggests a cautious approach to overemphasizing its benefits without parallel discussion of critical limitations. The relatively low citation volume in this theme may reflect an effort to avoid technological determinism.

Challenges and Limitations attracted 7 citations (13.73%), making it the third most emphasized theme in the paper. The sub-theme "Implementation and Integration" led with 3 citations (5.88%), revealing that infrastructural and organizational challenges are considered particularly pressing. Both "Technical and Ethical Challenges" and "Potential Biases in AI Algorithms" were each supported by 2 citations (3.92%), indicating attention to both the ethical ramifications and operational constraints of AI systems. The balanced spread of citations across sub-themes suggests that this section attempts to provide a multi-dimensional critique—spanning policy, practice, and ethics—rather than focusing on a single risk domain.

Teacher and Student Perceptions matched the first theme in total citations, with 10 citations (19.61%) distributed across three sub-themes. "Experiences with AI Assessments" stood out with 4 citations (7.84%), the highest single-sub-theme citation count in the entire paper. "Attitudes and Perceptions" and "Learning Outcomes and Achievement" were each supported by 3 citations (5.88%), indicating robust engagement with affective, cognitive, and behavioral dimensions of stakeholder responses to AI integration. The heavy citation volume here demonstrates a strong commitment to grounding the manuscript in the lived experiences and beliefs of both teachers and students, rather than in abstract system-level claims alone.

Implementing AI in Educational Settings accounted for 7 citations (13.73%), tying with "Challenges and Limitations" in overall citation weight. "Pedagogical Alignment" received 3 citations (5.88%), more than "Teacher Training" and "Resource Allocation," which received 2 citations (3.92%) each. This distribution suggests that the alignment of AI tools with curricular and instructional goals is perceived as slightly more critical than foundational readiness and infrastructural support. Nevertheless, the overall balance indicates an understanding that successful implementation is an ecosystemic challenge requiring attention to training, resources, and pedagogical integrity.

In total, the manuscript cites 51 unique instances, evenly distributed across 7 major themes and 23 sub-themes. The two most heavily cited themes, each representing 19.61% of the total citations, are **The Role of AI in Educational Evaluations** and **Teacher and Student Perceptions**. The sub-theme with the highest individual citation count is **Experiences with AI Assessments** at 7.84%, emphasizing its centrality to current discourse. Conversely, areas like **Automated Grading, Diagnostic Assessments, and Timeliness of Feedback** are underrepresented with only 1 citation (1.96%) each, revealing opportunities for expansion or deeper analysis. The citation landscape shows that the manuscript balances theoretical, technological, and humanistic concerns, though further elaboration in undercited domains would strengthen overall coverage.

RESULTS

4.1. Theme 1. The Role of AI in Educational Evaluations

Artificial Intelligence (AI) plays a transformative role in educational evaluations by restructuring how learning is assessed, feedback is delivered, and student performance is interpreted. Rather than merely supporting existing frameworks, AI-driven technologies redefine assessment paradigms by introducing adaptive, data-rich, and personalized mechanisms that challenge the limitations of traditional methods. The widespread adoption of AI across educational systems reflects not only technological advancement but also a paradigmatic shift in pedagogical values, emphasizing learner-centered, equity-driven, and formative assessment models. However, this transformation demands critical scrutiny, especially in relation to transparency, fairness, and educational validity.

4.1.1. AI Technologies in Educational Contexts

AI technologies have introduced multiple layers of automation, adaptability, and precision into educational assessments. Systems now support diagnostic evaluations, provide tailored feedback, generate individual learning pathways, and automate grading processes (Jiang & Pardos, 2021). These technologies collect and analyze vast datasets, allowing educators to access real-time insights into student learning patterns. Grubaugh (2024) underscores that such tools not only facilitate operational tasks but also offer actionable intelligence that aligns instruction with learner needs. The critical value lies not in automating human judgment but in enabling nuanced, data-informed decisions that respond to diverse educational contexts.

4.1.2. Historical Context of AI in Educational Evaluations

The integration of AI into educational assessments stems from broader historical developments in educational technology and learning analytics. Traditional assessments, often constrained by logistical and temporal limitations, gradually evolved with the introduction of automated scoring systems and digital testing environments. Over time, AI systems displaced many labor-intensive processes, such as manual grading and longitudinal data tracking, streamlining evaluation mechanisms while expanding their scope (Liang, 2023). Bodemer (2023) argues that AI systems introduced methodological shifts that recast assessment not merely as a summative act but as a continuous and formative inquiry into student development. This historical progression reflects a deeper educational realignment—moving from static measurement toward adaptive assessment ecosystems.

4.1.3. AI Technologies' Role in Equity and Fairness

AI has altered the discourse surrounding fairness and objectivity in assessment by offering models that mitigate human bias and standardize grading practices. AI-driven predictive models support teachers in curriculum refinement and intervention planning by identifying students in need of targeted support (Rieskamp, 2023). According to Smyrnova-Trybulska et al. (2023), these systems enhance fairness through consistent scoring algorithms and data-informed personalization strategies. However, the assumption that algorithmic assessment inherently promotes equity must be questioned. As Odeyemi (2024) highlights, while AI tools resolve some disparities, they risk introducing new biases if algorithmic training data or model design embeds existing systemic inequities. Equity in AI-enhanced assessment thus depends on deliberate design choices, transparency, and continual recalibration.

4.1.4. Enhancing Student Engagement and Critical Thinking

Beyond procedural efficiency, AI fosters deeper student engagement and cultivates critical thinking by offering interactive, tailored learning experiences. Through adaptive technologies, educators adjust instructional delivery to match diverse cognitive profiles and learning trajectories (Addy, 2024). Khan (2023) asserts that AI tools—when grounded in sound pedagogical design—shift students from passive recipients to active co-constructors of knowledge. Melkonian et al. (2022) emphasize that AI-powered assessments function not solely as evaluative instruments but as cognitive scaffolds that illuminate learning gaps and promote metacognitive awareness. However, engagement through AI must not be reduced to digital novelty; its pedagogical legitimacy depends on alignment with curricular goals and meaningful learning outcomes.

4.2. Theme 2. AI Applications in Mathematics Assessments

The integration of Artificial Intelligence (AI) into mathematics assessments reconfigures both the processes and philosophies of evaluating mathematical understanding. AI technologies do not simply enhance traditional practices; they introduce a paradigm grounded in adaptivity, interactivity, and personalization. By embedding intelligent systems into assessment frameworks, educators restructure the ways students engage with mathematical content and how evidence of learning is captured. However, these innovations warrant critical examination in light of concerns related to educational equity, cognitive development, and the pedagogical coherence of AI-infused tools. The mechanization of assessment must remain secondary to its educative purpose—supporting meaningful mathematical learning rather than merely optimizing procedural tasks.

4.2.1. Types of AI Technologies Employed in Mathematics Assessments

A diverse array of AI technologies has been deployed to transform mathematics assessments, each offering specific affordances to address learner variability and cognitive challenges. Intelligent Tutoring Systems (ITS), as highlighted by Hwang (2022), deliver targeted instruction and real-time feedback based on students' mathematical competencies. These systems do not function as static repositories of content but adaptively respond to individual errors and misconceptions, enabling differentiated learning experiences. By analyzing response patterns and knowledge gaps, ITS recalibrate problem difficulty and instructional support to suit the learner's current stage of development, aligning with schema-based learning models that emphasize structured knowledge retrieval (Meylani, 2024a).

In parallel, AI-enhanced calculators support mathematical cognition by offering step-by-step feedback and adaptive hints (Stefanova, 2024). These tools extend beyond conventional computational devices by employing AI algorithms that identify reasoning errors, guide problem-solving strategies, and reinforce conceptual understanding. Stefanova (2024) underscores that such calculators promote critical mathematical thinking by scaffolding the problem-solving process rather than supplying final answers.

Emerging generative AI tools, such as ChatGPT, have entered the mathematics education space with the potential to enhance discourse and interactivity. These models support mathematical exploration through interactive dialogue, clarification prompts, and personalized guidance. Although not developed exclusively for mathematics, their capacity to simulate conversation and model reasoning offers instructional potential when situated within well-designed pedagogical frameworks. Their growing use necessitates systematic scrutiny to assess alignment with curricular goals, accuracy of mathematical reasoning, and their influence on student epistemologies (Pepin et al., 2025).

4.2.2. Detailed Instances and Case Analyses of AI Applications

Case studies reveal the extent to which AI redefines not only how assessments are conducted but also how students engage with mathematics. Sunarto (2024) presents the development of an AI-based differentiated instruction model tailored for mathematics classrooms. The model supports autonomous learning and encourages students to develop critical thinking by solving contextualized mathematical problems with minimal teacher intervention. This approach shifts the instructional dynamic, allowing teachers to act as facilitators while the AI system manages instructional flow and content adaptation.

Dabingaya (2022) further documents the success of AI-powered adaptive learning platforms in mathematics education. These systems continually evaluate student understanding and adjust instructional trajectories to address conceptual weaknesses. In doing so, they operationalize formative assessment in real time, providing educators with actionable insights while empowering students to take ownership of their learning journey. The implication extends beyond mere efficiency; AI tools restructure the pedagogical ecology by redistributing agency among learners, systems, and educators.

These case analyses underscore that AI applications in mathematics assessment must be assessed not only on technical functionality but also on their capacity to align with educational values. While AI-driven platforms expand instructional possibilities, they must operate within frameworks that ensure mathematical rigor, conceptual depth, and ethical responsibility.

4.3. Theme 3. Types of AI-Enhanced Assessments

AI-enhanced assessments in mathematics education represent more than technological upgrades; they signify a redefinition of how mathematical knowledge is elicited, interpreted, and used for instructional decisions. These systems automate and personalize the assessment process, offering feedback and insights that shift assessment from a summative endpoint to a formative, ongoing process. However, this evolution demands a critical lens. The integration of AI into assessment must be evaluated for its capacity to preserve educational integrity, ensure equitable access, and reinforce deep mathematical reasoning. Without such critical interrogation, AI-enhanced assessments risk reducing complex cognitive activities to algorithmically simplified tasks.

4.3.1. Adaptive Assessments

Adaptive assessments rely on AI algorithms to dynamically adjust question difficulty based on a student's previous responses. This mechanism ensures a tailored evaluation experience, where each student encounters content aligned with their zone of proximal development (Geary et al., 2019). Such systems collect continuous data to locate a student's performance level with greater precision than static tests. Adaptive assessments restructure the assessment landscape by recognizing learning as a continuum rather than a fixed point. While this offers clear pedagogical benefits, it also introduces complexities regarding algorithmic transparency, test validity, and interpretability of outcomes—factors that remain under-theorized in current literature.

4.3.2. Automated Grading and Feedback Systems

Automated grading systems powered by AI process student responses, detect errors, and generate immediate feedback. Messer (2024) emphasizes their value in reducing teacher workload while maintaining feedback immediacy, which supports more timely instructional interventions. These systems promote formative assessment practices by enabling students to reflect on their performance and correct misconceptions without delay. Yet, the pedagogical utility of such systems depends on the quality of feedback and the accuracy of pattern recognition. Superficial or context-insensitive feedback risks trivializing complex mathematical reasoning. Therefore, automated grading systems must be continually evaluated for their instructional alignment and explanatory depth.

4.3.3. Diagnostic Assessments

AI-driven diagnostic assessments evaluate student competencies by identifying strengths, weaknesses, and misconceptions with a high level of granularity. Molnár and Csapó (2019) show that such systems facilitate targeted instruction by offering insight into learners' cognitive profiles. These diagnostics inform instructional design and enable differentiated teaching strategies. However, the assumption that diagnostic precision guarantees improved learning requires caution. Overreliance on algorithmically generated diagnostics may marginalize teacher judgment and contextual knowledge, particularly in culturally or linguistically diverse classrooms. Ensuring that diagnostic insights complement rather than replace pedagogical intuition remains essential.

4.3.4. Detailed Instances and Case Analyses of AI Applications

Case studies provide critical insights into how AI assessments function in authentic educational contexts. Hiremath (2024) introduces a multimodal AI approach to assessing handwritten responses, combining handwriting recognition, text detection, and language modeling. This technique streamlines evaluation and broadens accessibility for students who prefer written articulation. While the method enhances scalability and supports large-scale implementation, it raises questions about how well such systems capture the nuances of student reasoning expressed through non-standard formats or culturally influenced mathematical language.

Lin (2024) explores AI-powered grading for peer-assessed tasks in digital learning environments. These automated grading models offer consistent and scalable assessment processes, reducing reliance on subjective human scoring. However, the substitution of human evaluators with algorithmic systems introduces epistemic risks—particularly when evaluating creativity, strategy use, or mathematical modeling tasks. Lin's study underscores the importance of designing AI grading systems that account for qualitative dimensions of mathematical work, rather than optimizing for procedural correctness alone.

4.4. Theme 4. Benefits of AI in Mathematics Assessments

AI technologies in mathematics assessments yield a variety of pedagogical and operational benefits that directly affect both teaching efficacy and student learning outcomes. These systems enhance the precision of evaluations, reduce feedback latency, personalize instruction, and foster student engagement. However, these gains should not be interpreted uncritically. While AI tools offer efficiency and individualization, their educational impact depends on the quality of implementation, the contextual alignment with curriculum goals, and the equity of access. A meaningful examination of benefits must interrogate not only technological capacity but also educational value and ethical alignment.

4.4.1. Precision and Accuracy in Assessment

AI systems increase the accuracy and objectivity of mathematics assessments by eliminating inconsistencies often introduced by human judgment. As noted by Owan (2023), AI-powered tools evaluate student responses based on pre-defined criteria, offering consistent and reproducible scoring that strengthens reliability. Such systems allow for fine-grained analysis of problem-solving patterns and identify subtle misconceptions that traditional grading may overlook. However, the reliance on automated precision must not obscure the complexity of mathematical thinking, especially in open-ended or exploratory tasks. Overstandardization risks prioritizing what is easily measurable rather than what is educationally meaningful. As highlighted by the Institute of Education Sciences (2023), while recent advances in AI have made autoscoring of student open-ended math responses possible, challenges remain in accurately capturing the nuances of students' mathematical reasoning, particularly when they mix symbols, equations, and words in their answers.

4.4.2. Timeliness and Personalization of Feedback

AI technologies deliver immediate, tailored feedback that supports continuous learning and self-regulation. Lakomkin et al. (2018) emphasize the pedagogical value of timely feedback, showing that students who receive real-time insights into their errors are more likely to revise their thinking and engage in productive struggle. Personalized feedback fosters metacognitive awareness and strengthens conceptual understanding. However, the quality and pedagogical appropriateness of AI-generated feedback remain uneven across systems. Feedback that lacks contextual depth or is overly prescriptive may limit critical reasoning. The personalization afforded by AI must be rooted in dialogic pedagogy, not in one-size-fits-all algorithms.

4.4.3. Enhanced Engagement and Learning Outcomes

AI-enhanced assessments in mathematics contribute to deeper engagement and improved academic performance. According to Sunarto (2024), students interacting with adaptive AI platforms report higher motivation, increased persistence, and more frequent engagement with mathematical content. These environments foster active learning by offering interactive tasks that respond to student input and encourage exploration. The personalization of content and the sense of autonomy promote a growth-oriented learning culture. Yet, enhanced engagement should not be conflated with mere screen time or gamification. The educational value of engagement lies in whether the task supports conceptual rigor, promotes inquiry, and aligns with long-term learning objectives.

4.4.4. Detailed Instances and Case Analyses

Case analyses illustrate how AI systems operationalize these benefits in real-world mathematics education settings. Remoto (2023) presents evidence that AI-assisted learning tools elevate students’ numerical competence by scaffolding complex procedures and offering intuitive visual representations. These tools guide students through multi-step problems, supporting both procedural fluency and conceptual depth. Similarly, Sarwari (2024) explores the broader communicative and collaborative affordances of AI in education, including improved intercultural dialogue and shared problem-solving. While these gains indicate the expansive potential of AI, they also highlight the need for critical examination of which student populations benefit, under what conditions, and with what pedagogical oversight.

4.5. Theme 5. Challenges and Limitations

Despite the promise of AI-enhanced mathematics assessments, significant challenges and limitations persist. These include technical constraints, ethical concerns, algorithmic biases, and implementation issues that threaten to undermine the educational benefits of AI systems. The discourse surrounding AI in education often emphasizes innovation and personalization, yet fails to sufficiently engage with the structural, sociocultural, and epistemological implications of automated assessment. Addressing these challenges requires not only technical solutions but also sustained pedagogical reflection and institutional accountability.

4.5.1. Technical and Ethical Challenges

AI integration in mathematics assessment depends on robust digital infrastructure, reliable algorithmic performance, and comprehensive cybersecurity protocols. Tsopra et al. (2021) identify critical technical barriers, including interoperability with legacy systems, limited scalability in low-resource environments, and the need for continuous maintenance and calibration. Without these conditions, assessment outcomes risk becoming inaccurate or inconsistent, thereby eroding trust in the system.

Ethical concerns further complicate implementation. Esmaeilzadeh et al. (2021) highlight the importance of data privacy, informed consent, and algorithmic transparency in educational technologies. When AI systems collect, store, and analyze sensitive student data, they create new responsibilities for educators and policymakers to protect learners' rights. Ethical frameworks must govern not only the design and use of these systems but also the interpretation and consequences of their outputs. Without enforceable standards, AI risks replicating and amplifying existing educational inequalities.

4.5.2. Potential Biases in AI Algorithms

Biases embedded in AI algorithms compromise fairness in mathematics assessments. Matzakos (2023) argues that biased training datasets, flawed algorithmic design, and exclusionary assumptions result in systems that reproduce rather than mitigate social disparities. These biases manifest in misaligned scoring models, inaccurate performance predictions, and inequitable feedback distribution. In educational settings, such distortions disproportionately affect marginalized students, distorting their academic profiles and reinforcing systemic disadvantages.

The ethical imperative to detect and correct algorithmic bias must inform all stages of AI development and deployment. Transparent model design, diverse training data, and inclusive stakeholder engagement are essential to ensure fair and equitable assessments. Educators must critically examine not only how AI scores students but also whose knowledge and experiences are validated or excluded in the process. As highlighted by the European Commission's ethical guidelines on AI in education, involving educators and learners in the design and implementation of AI systems is crucial to uphold fairness and inclusivity in educational settings (European Commission, 2022).

4.5.3. Issues with Implementation and Integration

Successful integration of AI into mathematics assessments demands alignment with pedagogical goals, teacher readiness, and institutional capacity. Adopting new AI-driven assessment tools, integrating them into existing pedagogical practices, and ensuring their consistent and effective use often depends on contextual factors such as modality, teacher capacity, and system support structures (Meylani, 2024c). Also, Hermann (2021) documents how gaps in teacher training, limited technological fluency, and resistance to innovation undermine implementation efforts. When educators lack the tools or confidence to effectively use AI systems, technological potential translates into classroom dysfunction. Misalignment between AI systems and curricular standards further compounds this issue, leading to fragmented instruction and reduced instructional coherence.

Resource limitations pose additional barriers. Schools with insufficient infrastructure or support struggle to adopt and maintain AI systems, exacerbating educational inequities between high- and low-resourced institutions. These disparities challenge the often-unexamined narrative that AI universally enhances education. Strategic planning, policy support, and long-term professional development are essential to mitigate these risks and support sustainable implementation. As highlighted by the College of Education at the University of Illinois (2024), the cost of implementing AI technologies can be prohibitive for many schools, particularly those in underserved communities, thereby widening the educational gap.

4.6. Theme 6. Teacher and Student Perceptions

Teacher and student perceptions profoundly influence the implementation, acceptance, and educational effectiveness of AI technologies in assessment. While institutional discourse often centers on innovation and efficiency, the lived experiences of users expose both the opportunities and limitations of AI integration. Positive perceptions correlate with enhanced engagement and uptake, whereas skepticism—particularly regarding ethical concerns and pedagogical disruption—acts as a barrier to effective assimilation. Understanding these perceptions requires critical attention to emotional, cultural, and pedagogical dimensions, rather than reducing users to passive recipients of technological change.

4.6.1. Attitudes and Perceptions Towards AI in Assessments

Teacher and student attitudes shape the trajectory of AI adoption in educational assessments. Chiu et al. (2022) report that students participating in co-created AI curricula developed greater confidence in learning with AI and expressed more positive attitudes toward its educational relevance. This suggests that meaningful involvement in curriculum design strengthens both AI literacy and learner empowerment. However, such outcomes rely on the depth of engagement and the quality of the AI tools—not merely their presence.

Dai et al. (2020) show that students' confidence in using AI tools and their perceived relevance to real-life contexts strongly influence their sense of preparedness for the AI era. These perceptions link directly to motivation and learning efficacy. Yet, enthusiasm alone does not guarantee educational quality. Without critical pedagogy to contextualize AI use, students may internalize instrumentalist views of learning that prioritize efficiency over understanding. Positive attitudes must be cultivated through informed practice, sustained dialogue, and critical reflection on technology's role in education. As highlighted by Walter (2024), AI in education demands a focus on creativity and technology fluency to foster innovation and critical thought, necessitating a paradigm shift towards more dynamic, interactive, and student-centered learning environments.

4.6.2. Studies on Learning Outcomes and Achievement

The relationship between AI integration and student learning outcomes remains complex and context-dependent. Seo et al. (2021) demonstrate that AI systems enhance student satisfaction and achievement by improving learner-instructor interaction in online environments. These systems provide timely feedback and simulate presence, contributing to improved engagement and learning quality. However, the effectiveness of such tools depends on how well they support dialogic interaction, conceptual rigor, and emotional engagement—core elements of meaningful learning.

Hutson et al. (2022) further highlight AI's potential in improving learning outcomes, reducing costs, and increasing completion rates in higher education. While these metrics suggest efficiency gains, they do not necessarily capture deeper learning, critical thinking, or the development of disciplinary habits of mind (Gerlich, 2025). Educational institutions must avoid conflating performance indicators with pedagogical success. AI must serve learning, not redefine it according to technocratic metrics alone.

4.6.3. Studies on the Experiences with AI-Enhanced Assessments

Empirical studies reveal a wide range of experiences with AI-enhanced assessments. Wu and Yu (2023) show that AI chatbots increase academic achievement by offering real-time guidance, facilitating interaction, and fostering engagement. These findings reinforce the importance of responsive systems that provide learners with cognitive and emotional support. Yet, overdependence on AI-generated prompts risks undermining students' autonomy and initiative, particularly when critical thinking and creativity are required.

Muslimin (2024) documents how AI-powered tools influence attitudes and writing performance in EFL contexts. The study indicates improved outcomes when AI systems support personalized learning and offer formative feedback. However, the pedagogical design of the AI tool—its ability to adapt, diagnose, and guide meaningfully—determines its success. Ibrahim (2024) further examines Nigerian university lecturers' use of AI in assessments, highlighting concerns about academic integrity and the personalization of evaluation processes. These concerns reflect broader uncertainties about algorithmic judgment and human oversight.

Baez (2023) identifies teacher and student self-efficacy as critical determinants of AI perception. Where mutual trust and motivational alignment exist, AI is perceived as a supportive extension of the classroom. Without such conditions, AI technologies remain peripheral or provoke resistance. Ultimately, perceptions are shaped not just by tool efficacy, but by the broader pedagogical culture and relational dynamics in which AI is embedded.

4.7. Theme 7. Implementing AI in Educational Settings

The effective implementation of AI technologies in educational contexts requires more than access to tools. It demands systemic change—encompassing teacher training, infrastructure, curricular alignment, and institutional vision. While AI offers significant pedagogical opportunities, its impact depends on thoughtful integration into existing educational ecosystems. Uncritical deployment risks reinforcing digital divides, overwhelming teachers, and disconnecting technology from learning goals. Implementation must be situated within a broader educational framework that prioritizes pedagogical integrity, equity, and sustainability.

4.7.1. Teacher Training

Teacher preparation remains central to any meaningful AI integration. Lee and Perret (2022) demonstrate that professional development programs enhance high school teachers' capacity to embed AI tools within STEM curricula. Effective training programs address not only technical competencies but also ethical considerations, instructional design, and content-specific applications. Without pedagogical grounding, AI use remains superficial—limited to efficiency gains rather than conceptual enrichment.

Training must also foster critical engagement. Teachers require not only operational proficiency but also the ability to evaluate the implications of AI use for student learning, assessment validity, and equity. Developing teacher agency ensures that AI remains a pedagogical tool—not a driver of instructional decisions detached from context (Meylani, 2024b).

4.7.2. Resource Allocation

Equitable and sustainable AI implementation depends on strategic resource allocation. Gupta and Kitcharoen (2024) underscore the need for ongoing professional support, technical infrastructure, and institutional investment. AI tools require maintenance, calibration, and integration into broader learning management systems. Without consistent support structures, AI initiatives remain fragmented and often collapse after initial enthusiasm fades.

Resource allocation also shapes equity. Institutions with limited financial and technical resources face structural disadvantages in implementing AI-enhanced assessments. This gap reinforces existing educational inequalities and undermines AI's transformative potential. Targeted investment and policy support are necessary to create conditions for equitable access and scalable impact. As highlighted by the OECD (2024), the integration of AI in education can exacerbate existing disparities if not accompanied by deliberate strategies to ensure equitable access and inclusion.

4.7.3. Pedagogical Alignment

Successful AI implementation hinges on alignment with pedagogical goals, curricular standards, and learning outcomes. Aghaziarati (2023) and Mayo (2024) emphasize the importance of fostering a culture of experimentation and integrating AI-focused modules into teacher education programs. AI use must support—not replace—pedagogical purpose. When technology drives practice, rather than serving instructional objectives, learning risks becoming fragmented and instrumental.

Alignment also ensures coherence between assessment design and instructional strategies. Without such coherence, AI systems produce outputs that teachers struggle to interpret or act upon. Embedding AI within reflective pedagogical practice ensures that technological innovation translates into educational value. Ultimately, pedagogical alignment transforms AI from an external intervention into an integrated element of effective teaching and assessment. As highlighted by the U.S. Department of Education, Office of Educational Technology (2023), aligning AI models with a shared vision for education is crucial to ensure that these technologies support, rather than hinder, pedagogical goals.

DISCUSSION

5.1. Discussion of Key Findings

This discussion critically interprets the findings of the systematic review, relating them explicitly to the research questions and broader theoretical and practical issues in mathematics education and AI-enhanced assessment. Rather than merely summarizing the results, the discussion evaluates the implications, tensions, and limitations identified in the existing literature and proposes areas for future innovation and research (Meylani, 2024b, 2025b).

5.1.1. Enhancing Accuracy and Objectivity in Mathematics Assessments

The findings reveal that AI technologies considerably enhance the precision and consistency of mathematics assessments, addressing longstanding concerns about human error, grading biases, and inconsistency (Owan, 2023; Messer, 2024; Meylani, 2025b; Meylani & Kutluca, 2025). AI systems, particularly Intelligent Tutoring Systems (ITS) and automated grading tools, provide impartial and replicable evaluation processes that are difficult to achieve through traditional human grading alone.

However, this perceived objectivity must be interpreted cautiously. Several studies, including those by Matzakos (2023) and Tsopra et al. (2021), underscore that AI systems themselves are products of human decisions and may embed latent biases in their algorithms. Although AI reduces certain types of subjective bias, it simultaneously risks introducing algorithmic bias, particularly if training datasets are not sufficiently representative. Therefore, while the promise of greater objectivity is significant, it remains contingent upon rigorous algorithm auditing, diverse data inclusion, and continual monitoring — aspects that current studies address only partially (Meylani, 2025a).

5.1.2. Personalization and Learning Pathways: Opportunities and Limits

AI-powered assessments show strong potential for providing individualized feedback and supporting differentiated instruction, which are vital for fostering equitable mathematics learning environments (Sunarto, 2024; Remoto, 2023; Meylani, 2025b, 2025c). Adaptive testing and real-time feedback mechanisms allow instruction to be more closely aligned with students' current levels of understanding and cognitive needs (Meylani, 2024b).

Nevertheless, there is an overreliance on narrow definitions of "personalization" in many reviewed studies. Personalization often remains limited to item difficulty adjustment or provision of hints (Geary et al., 2019; Lakomkin et al., 2018), with limited attention to deeper aspects of personalized pedagogy such as culturally responsive feedback, learner agency, or metacognitive strategy development. A critical perspective suggests that while AI is well-positioned to tailor difficulty, truly transformative personalization must also involve fostering learners' self-regulation and reflection — dimensions largely overlooked in current AI-assisted mathematics assessment systems (Meylani, 2024c, 2025a).

5.1.3. Addressing Challenges and Ethical Risks

Although AI integration enhances assessment quality, significant technical, pedagogical, and ethical challenges persist. Issues such as algorithmic opacity, data privacy, technological infrastructure gaps, and pedagogical misalignment frequently constrain AI adoption in real-world educational settings (Esmaeilzadeh et al., 2021; Hermann, 2021; Meylani, 2024a, 2025a).

Surprisingly, the reviewed studies exhibit a relative lack of deep ethical engagement. While most acknowledge ethical risks, few propose concrete frameworks or mechanisms for safeguarding fairness, transparency, or student autonomy. The literature tends to treat ethical concerns as peripheral rather than central to system design and implementation. Future research must reposition ethics as foundational, not supplementary, to the development and deployment of AI-based assessments in mathematics education (Meylani, 2025a; Meylani & Kutluca, 2025).

Moreover, the studies seldom problematize the sociotechnical realities of educational inequality: for instance, how AI-based assessments might inadvertently disadvantage students in under-resourced schools that lack access to advanced technological infrastructure. This dimension demands urgent attention, as unchecked AI adoption risks amplifying, rather than mitigating, existing educational disparities (Odeyemi, 2024; Smyrnova-Trybulska et al., 2023; Meylani, 2025c).

5.1.4. Teacher and Student Perceptions: Complexities and Contradictions

The literature documents generally positive attitudes among educators and students toward AI integration (Chiu et al., 2022; Dai et al., 2020; Meylani, 2025b, 2025d). Perceptions of increased engagement, improved feedback quality, and enhanced learning motivation are commonly reported outcomes. However, these findings require a more critical lens.

First, enthusiasm for AI is not uniformly distributed: variability exists based on factors such as digital literacy, pedagogical philosophy, trust in technology, and prior experience with AI tools

(Ibrahim, 2024; Baez, 2023; Meylani, 2025b; Kaplan & Meylani, 2025). Some teachers express concerns about losing professional agency or reducing nuanced pedagogical judgment to algorithmic outputs. Similarly, while many students appreciate instant feedback, others view AI assessments as impersonal or mechanistic, potentially undermining deeper conceptual engagement (Meylani, 2024c).

These contradictions indicate that the successful integration of AI-enhanced assessments hinges on technological functionality and robust teacher preparation, participatory design processes, and alignment with human-centered educational values (Meylani, 2024a, 2025a).

5.1.5. Contributions to Equity and Inclusivity: Aspirations versus Realities

The review highlights that AI holds potential to promote equity in mathematics assessment by minimizing subjective bias and offering personalized support (Rieskamp, 2023; Smyrnova-Trybulska et al., 2023; Meylani, 2025b; Meylani & Kutluca, 2025). However, the realization of this potential is neither automatic nor guaranteed.

Equity gains are contingent upon careful system design, continuous monitoring for biases, and proactive inclusion of diverse learners' needs. Alarmingly, few studies actively address how AI might inadvertently exacerbate inequities — for instance, through biased training data, differential access to AI-enhanced tools, or cultural mismatch in adaptive algorithms (Meylani, 2024b, 2024d, 2025a, 2025c).

Future research must move beyond optimistic assumptions and empirically investigate under what conditions, for which students, and with what unintended consequences AI-powered assessments either support or hinder educational equity (Meylani, 2025b; Meylani & Kaplan, 2025).

5.2. Research Gaps and Future Research Directions

While this systematic review highlights significant advancements in the use of artificial intelligence (AI) for mathematics assessments, it also reveals critical gaps that must be addressed to ensure ethical, effective, and equitable integration. Future research must be strategically oriented toward addressing these deficiencies, rather than merely expanding the volume of publications.

5.2.1. Research Gaps Identified

- **Shallow Conceptualization of Personalization:** Although many studies claim to promote individualized learning through AI, personalization is often operationalized narrowly in terms of item difficulty adjustment or speed of response (Lye, 2024; Xu & Ouyang, 2022). Richer dimensions of personalization—such as adaptive scaffolding, culturally responsive support, or fostering metacognitive skills—remain largely unexplored (Meylani, 2024b).
  Gap: Future studies must theorize and empirically validate broader, more meaningful models of personalization in mathematics assessment that extend beyond surface-level adaptivity.
- **Limited Ethical and Sociopolitical Analysis:** While most studies acknowledge ethical risks, few engage critically with the deeper sociopolitical implications of AI assessments, such as how systems may reinforce existing educational inequities (Matzakos, 2023; Hermann, 2021).
  Gap: There is an urgent need for interdisciplinary studies that integrate educational technology research with critical race theory, feminist pedagogy, disability studies, and data justice frameworks to interrogate the broader ethical and social impacts of AI in education.

- **Overreliance on Technological Solutionism:** Many reviewed studies implicitly frame AI as a technical fix for longstanding educational problems without critically examining the pedagogical contexts into which AI is introduced (Geary et al., 2019; Odeyemi, 2024).
  Gap: Future research must critically evaluate the pedagogical compatibility and consequences of AI tools, rather than assuming that technological sophistication automatically translates into educational improvement.
- **Underexplored Teacher Agency and Professional Judgment:** Current studies often treat teachers as passive users of AI outputs rather than active interpreters, critics, and adaptors of AI-generated data (Chiu et al., 2022).
  Gap: Research must focus on understanding how teachers interpret, resist, modify, or co-design AI assessment systems, emphasizing teacher agency as central to meaningful technology integration.
- **Lack of Longitudinal and Contextually Diverse Studies:** Most existing research relies on short-term pilot studies conducted in technologically privileged contexts. Longitudinal data examining the sustained impacts of AI assessments across different socio-economic, linguistic, and cultural environments are rare.
  Gap: There is a pressing need for long-term, multi-contextual studies that examine how AI assessment tools perform and evolve over time, and under varying conditions of infrastructure, culture, and policy support.

5.2.2. Future Research Directions

- **Designing Pedagogically Aligned AI Assessment Systems:** Researchers should collaborate with educators to develop AI tools that not only adjust to learner performance but are also aligned with broader pedagogical goals, such as promoting mathematical reasoning, critical thinking, creativity, and collaboration.
  Suggested Approach: Mixed-methods studies combining AI development with participatory action research involving teachers and students at every stage.
- **Centering Equity and Inclusivity in AI Development:** Future research must systematically incorporate equity audits during AI system design, deployment, and evaluation. Rather than treating equity as an afterthought, it must be an explicit design criterion from inception.
  Suggested Approach: Develop and validate equity-focused evaluation frameworks that assess AI assessment tools for bias across race, gender, language, disability status, and socio-economic background.
- **Operationalizing Ethical AI Practices in K-12 Settings:** While ethical guidelines exist at the theoretical level, future research should create and empirically test operational models for ethical AI use in real-world school environments.
  Suggested Approach: Field trials that implement specific ethical governance structures—such as student consent protocols, data minimization practices, and algorithmic transparency reports—paired with qualitative evaluation of their feasibility and impact.
- **Empowering Teacher-Centric Innovations:** Future investigations should focus on co-design models that empower teachers to adapt and customize AI assessment tools to fit their pedagogical intentions and student needs.
  Suggested Approach: Experimental studies where teachers are given tools and frameworks to modify AI feedback parameters, followed by evaluations of pedagogical effectiveness and teacher satisfaction.
- **Developing Culturally Responsive AI Assessment Models:** AI systems must be able to recognize and value diverse cultural expressions of mathematical reasoning. Future work should explore how culturally responsive pedagogical frameworks can be embedded within AI algorithms.

Suggested Approach: Collaborative design studies that engage educators and communities from diverse cultural backgrounds in developing and testing AI assessment tools.

5.2.3. Final Reflection on Research Priorities

Future research must resist technological determinism and adopt a critically reflexive stance that interrogates not only what AI technologies can do but also what they ought to do within mathematics education. The goal is not merely to innovate technologically but also ethically, pedagogically, and equitably. Only such a critical, multi-dimensional research agenda will ensure that AI integration in mathematics assessment genuinely advances educational quality, justice, and sustainability for all learners.

5.3. Suggestions for Policy and Practice

Building on the findings of this review, this section offers detailed recommendations aimed at guiding policymakers, educational leaders, curriculum developers, and classroom practitioners in the ethical and effective integration of artificial intelligence (AI) technologies in mathematics assessments. Given the opportunities and challenges identified, deliberate strategies must be implemented to ensure that AI tools contribute to educational quality, equity, and sustainability rather than exacerbating existing disparities or introducing new risks.

5.3.1. Policy Recommendations

- **Establish Ethical Governance Frameworks for AI in Education:** National and institutional policies should mandate the development and adoption of ethical guidelines that govern the use of AI in educational assessments. These frameworks must address data privacy, algorithmic transparency, bias mitigation, and student consent. Clear regulatory structures are essential to ensure that AI systems adhere to fundamental educational and societal values rather than operating solely on technological imperatives.

Action Points:

  - Require all AI-based assessment platforms to publicly disclose algorithmic decision-making processes and data sources.
  - Implement regular audits for AI tools to assess fairness, accuracy, and unintended biases.
  - Create mechanisms for student and teacher appeals regarding AI-generated assessment results.

- **Encourage Professional Development on AI Literacy for Educators:** Policy frameworks should **embed** AI literacy and critical data practices into teacher certification and professional development programs. Educators must possess not only technical familiarity with AI tools but also critical understanding of their pedagogical implications, ethical risks, and limitations.

Action Points:

  - Fund national-level professional development initiatives focusing on AI in education.
  - Integrate AI ethics, algorithmic bias, and critical data use into teacher education curricula.
  - Require demonstrated AI literacy competencies as part of professional teacher standards.

- **Incentivize Equity-Driven AI Innovation:** Government agencies and funding bodies should incentivize the development of AI assessment systems that explicitly prioritize educational equity, inclusivity, and **accessibility** for marginalized student populations.

Action Points:

- o Offer grants or tax incentives for developers who design AI tools aligned with Universal Design for Learning (UDL) principles.
- o Establish diversity and inclusion benchmarks that AI vendors must meet to qualify for adoption in public educational systems.
- o Support public-private partnerships to ensure that AI innovations address the needs of underserved schools and communities.

- **Require Transparency and Accountability in Procurement:** Educational institutions must implement transparent procurement processes when adopting AI assessment technologies. Decision-makers should **critically** evaluate tools based on evidence of effectiveness, equity impact, ethical safeguards, and pedagogical fit.

Action Points:

- o Standardize evaluation rubrics that prioritize ethical, pedagogical, and equity considerations alongside technical performance.
- o Include teachers and students in the decision-making process for selecting AI-based assessment platforms.
- o Require vendors to provide independent validation studies demonstrating the efficacy and fairness of their products.

5.3.2. Practice Recommendations

- **Embed AI Assessment Tools Within Human-Centered Pedagogical Frameworks:** Teachers should use AI-powered assessments as supportive tools, not substitutes for professional judgment. AI outputs must be interpreted critically and integrated into broader instructional designs that emphasize critical thinking, problem-solving, and conceptual understanding.

Action Points:

- o Combine AI-generated feedback with human-facilitated discussions to promote deeper reflection and metacognition among students.
- o Use AI tools to support formative assessment cycles, not just summative evaluations.
- o Regularly cross-check AI assessment outputs with traditional methods (e.g., student interviews, open-ended responses) to verify validity.

- **Foster Critical AI Literacy Among Students:** Classroom practices should promote student awareness of how AI systems function, their potential biases, and their limitations. Students must develop critical digital literacy skills to navigate AI-mediated learning environments responsibly.

Action Points:

- o Incorporate short lessons on "Understanding AI" as part of mathematics instruction.
- o Discuss openly with students the ways AI assessments generate feedback and potential areas for error or bias.
- o Encourage students to critically question and cross-validate AI-provided feedback when appropriate.

- **Promote Culturally Responsive AI Use:** Educators must ensure that AI assessment systems respect and reflect the cultural, linguistic, and cognitive diversity of their students. AI should be leveraged to amplify, not suppress, students’ diverse ways of knowing and expressing mathematical understanding.

Action Points:

- o Customize AI feedback settings where possible to account for diverse language backgrounds and learning styles.

- o Use AI assessments in conjunction with culturally responsive pedagogy to ensure assessments recognize multiple ways of mathematical reasoning and problem-solving.
- o Advocate for the development of AI tools that incorporate multilingual support and diverse mathematical representations.

- **Implement Ongoing Monitoring and Reflective Practice:** Teachers and administrators should engage in continual monitoring of AI assessment outcomes, looking for patterns of differential performance or unintended biases. Reflection must be an embedded part of AI tool use, not an afterthought.

Action Points:

- o Maintain records of student performance trends across AI assessments to identify possible disparities early.
- o Conduct regular team discussions among educators to share observations and refine the use of AI tools.
- o Invite student feedback about their experiences with AI assessments and incorporate this feedback into practice revisions.

5.3.3. Concluding Note on Policy and Practice Integration

Critically integrating AI into mathematics assessments requires a dual commitment: on the one hand, macro-level structural reforms in policy to ensure ethical, equitable system design and implementation, and on the other, micro-level innovations in classroom practice to preserve human agency, critical thinking, and pedagogical richness. Only through sustained, critical engagement across these levels can AI technologies fulfill their promise of transforming mathematics education for all learners, rather than entrenching new forms of inequality.

CONCLUSION

6.1. Summary of Key Insights

This systematic review critically synthesized contemporary research on artificial intelligence (AI) integration in mathematics assessments, highlighting its potential to enhance assessment precision, support individualized learning experiences, and address equity in education. The analysis revealed that AI-powered tools, such as Intelligent Tutoring Systems and adaptive feedback mechanisms, can improve both the objectivity and responsiveness of mathematics evaluations. Furthermore, these technologies foster more dynamic learning environments, supporting deeper engagement and critical thinking.

Despite these benefits, the review identified critical challenges that temper AI's transformative promise. Issues related to algorithmic bias, ethical governance, infrastructural readiness, and pedagogical misalignment persist across many implementations. AI technologies, while innovative, are not neutral instruments—they are embedded in broader sociopolitical and cultural contexts that may perpetuate or intensify existing educational inequities. Accordingly, realizing the full potential of AI in mathematics assessments demands critical reflection, contextual adaptation, and sustained monitoring to ensure these tools serve inclusive, pedagogically sound educational purposes.

6.2. Merits and Contributions of the Study

This study contributes to the literature by offering a structured and critical synthesis of AI's role in mathematics assessment, advancing understanding in four key areas. First, it maps the current landscape of AI applications in mathematics assessments, clarifying how various technologies are used to support instructional goals. Second, it exposes tensions between the ideals of personalized learning and the often superficial adaptations offered by current systems.

Third, the review spotlights ethical concerns, particularly those related to data privacy, algorithmic fairness, and educator agency—areas underexplored in much of the empirical literature. Lastly, it frames AI not as a panacea, but as a complex educational intervention that must be governed by ethical principles and informed pedagogical goals. By moving beyond techno-optimistic narratives, the study provides a grounded foundation for future inquiry and responsible practice.

6.3. Limitations of the Study

Several limitations must be acknowledged in interpreting the findings. First, the review was restricted to English-language sources, which may have excluded relevant studies from non-English-speaking contexts. Second, its focus on qualitative research may underrepresent quantitative trends and outcomes in AI assessment. Third, the breadth of AI technologies and diversity of educational settings covered in the studies introduce variability that complicates generalization. Additionally, the potential biases inherent in AI algorithms were not deeply evaluated in most primary sources, limiting the ability to draw firm conclusions about fairness. Finally, variations in teacher and student digital competencies and institutional support structures were not always fully explored in the reviewed studies, despite their centrality to effective AI adoption. These limitations indicate the need for continued empirical research across diverse contexts to capture the full complexity of AI integration in mathematics education.

DECLARATIONS

Author Contributions: The author conceptualized the study, conducted the systematic literature review, performed the data coding and thematic analysis, interpreted the findings, and wrote and revised the manuscript in its entirety.

Funding: The author received no specific grant from any funding agency, commercial, or not-for-profit sectors for the completion of this research.

Conflict of Interest: The author declares no potential conflicts of interest with respect to the research, authorship, or publication of this article.

Ethical Approval: This study involved the analysis of previously published literature and did not include human participants or animals. Ethical approval was therefore not required. This study has not been derived from any conference paper, project or thesis.

Data Availability Statement: All data analyzed in this study were derived from publicly accessible scholarly sources cited in the reference list. No new datasets were generated or analyzed.

Acknowledgments: The author extends gratitude to peer reviewers and editorial staff for their valuable feedback during the manuscript preparation and review process.

REFERENCES

Addy, W. (2024). AI in credit scoring: A comprehensive review of models and predictive analytics. *Global Journal of Engineering and Technology Advances, 18*(2), 118–129. https://doi.org/10.30574/gjeta.2024.18.2.0029

Agarwal, P. (2024). Assessing the challenges and opportunities of artificial intelligence in Indian education. *International Journal for Global Academic & Scientific Research, 3*(1), 36–44. https://doi.org/10.55938/ijgasr.v3i1.71

Aghaziarati, A. (2023). Artificial intelligence in education: Investigating teacher attitudes. *Aitechbesosci, 1*(1), 35–42. https://doi.org/10.61838/kman.aitech.1.1.6

Baez, C. (2023). Exploring the perception of AI in learning: Unveiling the role of student and teacher motivation and self-efficacy. https://doi.org/10.31219/osf.io/wqvrh

Bedizel, N. (2023). Evolving landscape of artificial intelligence (AI) and assessment in education: A bibliometric analysis. *International Journal of Assessment Tools in Education, 10*(Special Issue), 208–223. https://doi.org/10.21449/ijate.1369290

Bodemer, O. (2023). Artificial intelligence in governance: A comprehensive analysis of AI integration and policy development in the German government. https://doi.org/10.36227/techrxiv.24639588.v1

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Chisom, O. (2024). Review of AI in education: Transforming learning environments in Africa. *International Journal of Applied Research in Social Sciences, 5*(10), 637–654. https://doi.org/10.51594/ijarss.v5i10.725

Chiu, T., Meng, H., Chai, C., King, I., Wong, S., & Yam, Y. (2022). Creation and evaluation of a pretertiary artificial intelligence (AI) curriculum. *IEEE Transactions on Education, 65*(1), 30–39. https://doi.org/10.1109/TE.2021.3085878

College of Education at the University of Illinois. (2024, October 24). *AI in schools: Pros and cons.* https://education.illinois.edu/about/news-events/news/article/2024/10/24/ai-in-schools--pros-and-cons

Critical Appraisal Skills Programme (CASP). (2018). *CASP qualitative checklist.* https://casp-uk.net/wp-content/uploads/2018/03/CASP-Qualitative-Checklist-2018_fillable_form.pdf

Dabingaya, M. (2022). Analyzing the effectiveness of AI-powered adaptive learning platforms in mathematics education. *Interdisciplinary Journal Papier Human Review, 3*(1), 1–7. https://doi.org/10.47667/ijphr.v3i1.226

Dai, Y., Chai, C., Lin, P., Jong, M., Guo, Y., & Jian-jun, Q. (2020). Promoting students' well-being by developing their readiness for the artificial intelligence age. *Sustainability, 12*(16), 6597. https://doi.org/10.3390/su12166597

Djajasoepena, R. (2024). Utilization of artificial intelligence to support the development of teaching and project modules. *JCSSE, 4*(1), 7–11. https://doi.org/10.35806/jcsse.v4i1.440

Esmaeilzadeh, P., Mirzaei, T., & Dharanikota, S. (2021). Patients' perceptions toward human–artificial intelligence interaction in health care: Experimental study. *Journal of Medical Internet Research, 23*(11), e25856. https://doi.org/10.2196/25856

European Commission. (2022, October 28). *An ethical use of artificial intelligence (AI) in education. AI Watch.* https://ai-watch.ec.europa.eu/news/ethical-use-artificial-intelligence-ai-education-2022-10-28_en

Geary, D., Hoard, M., Nugent, L., Chu, F., Scofield, J., & Hibbard, D. (2019). Sex differences in mathematics anxiety and attitudes: Concurrent and longitudinal relations to mathematical competence. *Journal of Educational Psychology, 111*(8), 1447–1461. https://doi.org/10.1037/edu0000355

Gerlich, M. (2025). AI tools in society: Impacts on cognitive offloading and the future of critical thinking. *Societies, 15*(1), 6. https://doi.org/10.3390/soc15010006

Grubaugh, S. (2024). The future of elementary social studies: Harnessing AI's potential through evidence-based practices. *Technium Social Sciences Journal, 58,* 87–93. https://doi.org/10.47577/tssj.v58i1.10991

Gupta, D. (2024). Navigating the future of education: The impact of artificial intelligence on teacher–student dynamics. *EATP,* 6006–6013. https://doi.org/10.53555/kuey.v30i4.2332

Hermann, I. (2021). Artificial intelligence in fiction: Between narratives and metaphors. *AI & Society, 38*(1), 319–329. https://doi.org/10.1007/s00146-021-01299-6

Hiremath, A. (2024). Transforming handwritten answer assessment: A multi-modal approach combining text detection, handwriting recognition, and language models. https://doi.org/10.21203/rs.3.rs-4301899/v1

Hutson, J., Jeevanjee, T., Graaf, V., Lively, J., Weber, J., Weir, G., & Edele, S. (2022). Artificial intelligence and the disruption of higher education: Strategies for integrations across disciplines. *Creative Education, 13*(12), 3953–3980. https://doi.org/10.4236/ce.2022.1312253

Hwang, S. (2022). Examining the effects of artificial intelligence on elementary students' mathematics achievement: A meta-analysis. *Sustainability, 14*(20), 13185. https://doi.org/10.3390/su142013185

Ibrahim, A. (2024). Artificial intelligence (AI): Perception and utilization of AI technologies in educational assessment in Nigerian universities. *Edukasiana Jurnal Inovasi Pendidikan, 3*(3), 367–380. https://doi.org/10.56916/ejip.v3i3.763

Institute of Education Sciences. (2023, October 16). *Math autoscoring is finally here—Let's tap its potential for students and teachers.* U.S. Department of Education. https://ies.ed.gov/director/remarks/10-16-2023.asp

Jiang, W., & Pardos, Z. (2021). Towards equity and algorithmic fairness in student grade prediction. https://doi.org/10.1145/3461702.3462623

Kaplan, R., & Meylani, R. (2025). Dimensions of artificial intelligence literacy: A qualitative synthesis of contemporary research literature. *Journal of Computer and Education Research, 13*(26), 790–825. https://doi.org/10.18009/jcer.1648380

Khan, H. (2023). Re: Artificial intelligence technology in surgery. *Australian and New Zealand Journal of Surgery, 94*(3), 489–489. https://doi.org/10.1111/ans.18809

Kitcharoen, P. (2024). Enhancing teachers' AI competencies through artificial intelligence of things professional development training. *International Journal of Interactive Mobile Technologies (iJIM), 18*(2), 4–15. https://doi.org/10.3991/ijim.v18i02.46613

Lakomkin, N., Dhamoon, M., Carroll, K., Singh, I., Tuhrim, S., Lee, J., … & Mocco, J. (2018). Prevalence of large vessel occlusion in patients presenting with acute ischemic stroke: A 10-year systematic review of the literature. *Journal of NeuroInterventional Surgery, 11*(3), 241–245. https://doi.org/10.1136/neurintsurg-2018-014239

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159–174. https://doi.org/10.2307/2529310

Lee, I., & Perret, B. (2022). Preparing high school teachers to integrate AI methods into STEM classrooms. *Proceedings of the AAAI Conference on Artificial Intelligence, 36*(11), 12783–12791. https://doi.org/10.1609/aaai.v36i11.21557

Lee, J. (2024). Development of a content framework of artificial intelligence integrated education considering ethical factors. *International Journal on Advanced Science, Engineering and Information Technology, 14*(1), 205–213. https://doi.org/10.18517/ijaseit.14.1.19558

Liang, Y. (2023). Balancing: The effects of AI tools in educational context. *Frontiers in Humanities and Social Sciences, 3*(8), 7–10. https://doi.org/10.54691/fhss.v3i8.5531

Lin, Z. (2024). Exploring an effective automated grading model with reliability detection for large-scale online peer assessment. *Journal of Computer Assisted Learning, 40*(4), 1535–1551. https://doi.org/10.1111/jcal.12970

Lye, C. (2024). Generative artificial intelligence in tertiary education: Assessment redesign principles and considerations. *Education Sciences, 14*(6), 569. https://doi.org/10.3390/educsci14060569

Mahligawati, F. (2023). Artificial intelligence in physics education: A comprehensive literature review. *Journal of Physics: Conference Series, 2596*(1), 012080. https://doi.org/10.1088/1742-6596/2596/1/012080

Matzakos, N. (2023). Learning mathematics with large language models. *International Journal of Emerging Technologies in Learning (iJET), 18*(20), 51–71. https://doi.org/10.3991/ijet.v18i20.42979

Mayo, S. (2024). Co-creating with AI in art education: On the precipice of the next terrain. *Education Journal, 13*(3), 124–132. https://doi.org/10.11648/j.edu.20241303.15

Melkonian, S., Crowder, J., Adam, E., White, M., & Peipins, L. (2022). Social determinants of cancer risk among American Indian and Alaska Native populations: An evidence review and map. *Health Equity, 6*(1), 717–728. https://doi.org/10.1089/heq.2022.0097

Messer, M. (2024). Automated grading and feedback tools for programming education: A systematic review. *ACM Transactions on Computing Education, 24*(1), 1–43. https://doi.org/10.1145/3636515

Meylani, R. (2024a). Innovations with schema theory: Modern implications for learning, memory, and academic achievement. *International Journal for Multidisciplinary Research, 6*(1). https://doi.org/10.36948/ijfmr.2024.v06i01.13785

Meylani, R. (2024b). A critical glance at adaptive learning systems using artificial intelligence: A systematic review and qualitative synthesis of contemporary research literature. *Batı Anadolu Eğitim Bilimleri Dergisi, 15*(3), 3519–3547. https://doi.org/10.51460/baebd.1525452

Meylani, R. (2024c). Artificial intelligence in the education of teachers: A qualitative synthesis of the cutting-edge research literature. *Journal of Computer and Education Research, 12*(24), 600–637. https://doi.org/10.18009/jcer.1477709

Meylani, R. (2024d). An in-depth literature review of remote, online, and hybrid learning with case studies of successful and failed attempts. *International Journal of Studies in Education and Science, 5*(1), 29–41. https://doi.org/10.46328/ijses.87

Meylani, R. (2024e). Blueprint for the 21st-century online learning environment in STEM education through a systematic review and qualitative synthesis. *Edelweiss Applied Science and Technology, 8*(6), 8196–8226. https://doi.org/10.55284/25768484.xii86.8763

Meylani, R. (2025a). A critical glance at technology's role in mathematics education for a sustainable future: Advancing SDG 4 – Quality education through a systematic review and qualitative synthesis. *Journal of Lifestyle and SDGs Review, 5*(2), e04566. https://doi.org/10.47172/2965-730X.SDGsReview.v5.n02.pe04566

Meylani, R. (2025b). Artificial intelligence in mathematics teacher education: A systematic review and qualitative synthesis of contemporary research literature. *International Journal of Technology in Education and Science, 1,* 63–91. https://doi.org/10.46328/ijoftes.258

Meylani, R. (2025c). Gamification and game-based learning in mathematics education for advancing SDG 4: A systematic review and qualitative synthesis of contemporary research literature. *Journal of Lifestyle and SDGs Review, 5*(2), e04567. https://doi.org/10.47172/2965-730X.SDGsReview.v5.n02.pe04567

Meylani, R. (2025d). Integration of TI-84 and TI-89 model graphing calculators in mathematics education: Precalculus instruction using the TPACK framework. *Journal of Computer and Education Research, 13*(25), 254–282. https://doi.org/10.18009/jcer.1589181

Meylani, R., & Kaplan, R. (2025). Novel approaches for teaching mathematical modeling to preservice teachers: A systematic review. *EMTEKA: Jurnal Pendidikan Matematika, 6*(1), 505–519. https://doi.org/10.24127/emteka.v6i1.8362

Meylani, R., & Kutluca, T. (2025). AI-powered discourse in mathematics education in support of SDG 4: A systematic review of contemporary research literature. *Discourse and Communication for Sustainable Education, 16*(2), 25–43. https://doi.org/10.2478/dcse-2025-0014

Molnár, G., & Csapó, B. (2019). Making the psychological dimension of learning visible: Using technology-based assessment to monitor students' cognitive development. *Frontiers in Psychology, 10.* https://doi.org/10.3389/fpsyg.2019.01368

Muslimin, A. (2024). Evaluating Cami AI across SAMR stages: Students' achievement and perceptions in EFL writing instruction. *Online Learning, 28*(2). https://doi.org/10.24059/olj.v28i2.4246

Nasir, M. (2024). Utilizing artificial intelligence in education to enhance teaching effectiveness. *Proceedings of ICE, 2*(1), 280–285. https://doi.org/10.32672/pice.v2i1.1367

Odeyemi, O. (2024). Reviewing the role of AI in fraud detection and prevention in financial services. *International Journal of Science and Research Archive, 11*(1), 2101–2110. https://doi.org/10.30574/ijsra.2024.11.1.0279

OECD. (2024). *The potential impact of artificial intelligence on equity and inclusion in education.* OECD Publishing. https://www.oecd.org/en/publications/the-potential-impact-of-artificial-intelligence-on-equity-and-inclusion-in-education_15df715b-en.html

Owan, V. (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., ... & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372,* n71. https://doi.org/10.1136/bmj.n71

Pepin, B., Buchholtz, N., & Salinas-Hernández, U. (2025). Mathematics education in the era of ChatGPT: Investigating its meaning and use for school and university education. *Digital Experiences in Mathematics Education, 11*(1), 1–8. https://doi.org/10.1007/s40751-025-00173-0

Remoto, J. (2023). ChatGPT and other AIs: Personal relief and limitations among mathematics-oriented learners. *Environment and Social Psychology, 9*(1). https://doi.org/10.54517/esp.v9i1.1911

Rieskamp, J. (2023). Approaches to improve fairness when deploying AI-based algorithms in hiring: Using a systematic literature review to guide future research. https://doi.org/10.24251/hicss.2023.026

Sarwari, A. (2024). Assessment of the impacts of artificial intelligence (AI) on intercultural communication among postgraduate students in a multicultural university environment. *Scientific Reports, 14*(1). https://doi.org/10.1038/s41598-024-63276-5

Seo, K., Tang, J., Roll, I., Fels, S., & Yoon, D. (2021). The impact of artificial intelligence on learner–instructor interaction in online learning. *International Journal of Educational Technology in Higher Education, 18*(1). https://doi.org/10.1186/s41239-021-00292-9

Smyrnova-Trybulska, E., Przybyła-Kasperek, M., & Kommers, P. (2023). Educational, social, and ethical aspects of AI in the opinion of IT and pedagogy students: A case study. In *International Conferences on E-Society 2023 and Mobile Learning 2023* (p. 171). https://doi.org/10.33965/es_ml2023_202302l021

Stefanova, T. (2024). Possibilities for using AI in mathematics education. *MEM, 53,* 117–125. https://doi.org/10.55630/mem.2024.53.117-125

Sunarto, M. (2024). Developing a differentiated learning model based on artificial intelligence: Implementation in the mathematics classroom. *Jurnal Kependidikan: Jurnal Hasil Penelitian dan Kajian Kepustakaan di Bidang Pendidikan, Pengajaran, dan Pembelajaran, 10*(2), 416. https://doi.org/10.33394/jk.v10i2.11135

Tsopra, R., Luchinat, C., Alberghina, L., Lehrach, H., Vanoni, M., Dreher, F., … & Burgun, A. (2021). A framework for validating AI in precision medicine: Considerations from the European ITFOC consortium. *BMC Medical Informatics and Decision Making, 21*(1). https://doi.org/10.1186/s12911-021-01634-3

U.S. Department of Education, Office of Educational Technology. (2023). *Artificial intelligence and the future of teaching and learning: Insights and recommendations.* https://www.ed.gov/sites/ed/files/documents/ai-report/ai-report.pdf

Walter, Y. (2024). Embracing the future of artificial intelligence in the classroom: The relevance of AI literacy, prompt engineering, and critical thinking in modern education. *International Journal of Educational Technology in Higher Education, 21*(15). https://doi.org/10.1186/s41239-024-00448-3

Wu, R., & Yu, Z. (2023). Do AI chatbots improve students' learning outcomes? Evidence from a meta-analysis. *British Journal of Educational Technology, 55*(1), 10–33. https://doi.org/10.1111/bjet.13334

Xie, X. (2023). Influence of AI-driven inquiry teaching on learning outcomes. *International Journal of Emerging Technologies in Learning (iJET), 18*(23), 59–70. https://doi.org/10.3991/ijet.v18i23.45473

Xu, W., & Ouyang, F. (2022). The application of AI technologies in STEM education: A systematic review from 2011 to 2021. *International Journal of STEM Education, 9*(1). https://doi.org/10.1186/s40594-022-00377-5

Yang, Y. (2023). Enhancing students' metacognition via AI-driven educational support systems. *International Journal of Emerging Technologies in Learning (iJET), 18*(24), 133–148. https://doi.org/10.3991/ijet.v18i24.45647

GENİŞLETİLMİŞ ÖZET

Giriş

Matematik eğitimi, öğrencilerin eleştirel düşünme, problem çözme ve analitik akıl yürütme becerilerini geliştirmede temel bir alan olarak kabul edilmektedir. Ancak bu becerilerin güvenilir ve bütüncül bir biçimde ölçülmesi, günümüz eğitim sistemlerinde hâlâ önemli bir sorun alanıdır. Geleneksel değerlendirme yöntemleri çoğunlukla doğruluk odaklıdır ve öğrencinin kavramsal anlayışını, akıl yürütme sürecini ya da öğrenme ilerlemesini derinlemesine yansıtmakta yetersiz kalmaktadır. Bu durum, öğretmenlerin zamanında, kişiselleştirilmiş ve pedagojik olarak anlamlı geri bildirim üretmesini güçleştirmekte; değerlendirme sistemini öğrenmeyi destekleyen bir araç olmaktan uzaklaştırmaktadır.

Bu noktada Yapay Zekâ (YZ) teknolojileri, matematik değerlendirmelerinde yenilikçi ve dönüştürücü bir potansiyel sunmaktadır. Akıllı Öğretim Sistemleri (ITS), otomatik puanlama yazılımları, uyarlanabilir ölçme araçları, tanılayıcı algoritmalar ve YZ destekli hesap makineleri, öğrencinin bireysel performansına göre anında geri bildirim sağlayabilmekte, hataları analiz edebilmekte ve öğretim sürecine veri temelli katkılar sunabilmektedir. Ancak YZ'nin eğitimsel kullanımı yalnızca teknik bir yenilik değildir; etik, pedagojik ve toplumsal boyutlarıyla da değerlendirilmesi gereken karmaşık bir dönüşüm alanıdır. Bu nedenle çalışma, YZ destekli matematik değerlendirmelerini, eleştirel pedagoji perspektifiyle inceleyerek bu teknolojilerin eğitimde fırsat eşitliği, etik şeffaflık ve öğretmen özerkliği üzerindeki etkilerini sorgulamaktadır.

Yöntem

Bu araştırma, PRISMA 2020 protokolü çerçevesinde yürütülmüş sistematik bir derleme çalışmasıdır. Çalışmada 2010–2024 yılları arasında yayımlanmış çağdaş literatür incelenmiştir. Veri tabanı aramaları Scopus, Web of Science, ERIC, IEEE Xplore ve Google Scholar üzerinden gerçekleştirilmiş; "Artificial Intelligence", "Mathematics Education", "Assessment", "Adaptive Learning Systems" ve "Machine Learning" anahtar sözcükleri kullanılmıştır.

Toplam 762 kayıt incelenmiş, yinelenenler çıkarıldıktan sonra 51 çalışma nitel sentez için uygun bulunmuştur. Veriler, MAXQDA 24 yazılımı aracılığıyla tematik analiz yöntemiyle çözümlenmiştir. Kodlama sürecinde YZ tabanlı değerlendirme uygulamalarının pedagojik işlevi, etik boyutları, öğretmen–öğrenci algıları, teknolojik altyapı farklılıkları ve algoritmik önyargılar gibi temalar öne çıkmıştır. Kodlamalar iki aşamada bağımsız olarak yapılmış, Cohen's $\kappa = 0.82$ değeriyle yüksek güvenirlik elde edilmiştir.

Nitel araştırmalarda geçerliği artırmak amacıyla her çalışmanın yöntemsel kalitesi CASP (2018) ölçütlerine göre değerlendirilmiştir. Yetersiz metodolojik derinliğe sahip yayınlar kapsam dışı bırakılmıştır. Böylece analiz, yüksek nitelikli ve karşılaştırılabilir araştırmalardan oluşan bir veri kümesi üzerinde yürütülmüştür.

Bulgular

Analiz sonucunda yedi ana tema ve yirmi üç alt tema belirlenmiştir. Bulgular, YZ'nin matematik değerlendirmelerinde sağladığı olanakların yanında, uygulamada karşılaşılan sınırlılıkları da ortaya koymaktadır.

1. **YZ'nin Eğitimsel Değerlendirmelerdeki Rolü:** YZ sistemleri, öğrenme sürecine ait verileri toplayıp analiz ederek öğretmene anlık geri bildirim sağlamaktadır. Bu teknolojiler yalnızca ölçme işlemini otomatikleştirmekle kalmayıp, öğrencinin öğrenme yolculuğunu sürekli izleyebilme olanağı tanımaktadır. Bununla birlikte algoritmik şeffaflık eksikliği ve veri güvenliğine dair belirsizlikler, değerlendirme süreçlerinin güvenilirliğini tehdit eden temel sorunlar olarak öne çıkmıştır.

2. **Matematik Değerlendirmelerinde YZ Uygulamaları:** Matematik eğitiminde kullanılan ITS, uyarlanabilir hesap makineleri, otomatik değerlendirme yazılımları ve ChatGPT benzeri üretken modeller, öğrencinin bilişsel düzeyine göre dinamik bir geri bildirim döngüsü oluşturabilmektedir. Ancak yapılan durum çalışmaları, bu araçların pedagojik bütünlüğünün öğretmen rehberliğiyle desteklenmediğinde sınırlı etki yarattığını göstermiştir.
3. **YZ Destekli Ölçme Türleri:** Uyarlanabilir testler, tanılayıcı ölçmeler ve otomatik geri bildirim sistemleri, öğrencinin gelişim düzeyine uygun içerikler sunmaktadır. Bununla birlikte, test geçerliği ve algoritmaların yorumlanabilirliği gibi konuların literatürde yeterince tartışılmadığı belirlenmiştir.
4. **YZ'nin Sağladığı Faydalar**: YZ tabanlı değerlendirmeler, doğruluk ve nesnellik düzeyini artırmakta, anlık ve kişiselleştirilmiş geri bildirimler sunarak öğrencinin öğrenmeye katılımını güçlendirmektedir. Çeşitli araştırmalar, bu tür sistemlerin öğrencilerin motivasyonunu ve başarı düzeyini artırdığını göstermektedir. Bununla birlikte, geribildirimlerin yüzeysel veya bağlamdan kopuk olması durumunda eleştirel düşünmeyi sınırladığı saptanmıştır.
5. **Zorluklar ve Sınırlılıklar:** YZ'nin eğitimdeki en büyük zorlukları teknik altyapı, etik ilkeler, algoritmik önyargı ve öğretmen yeterlikleriyle ilgilidir. Özellikle düşük donanımlı okullarda sistemlerin sürdürülebilirliği zayıftır. Ayrıca, veri gizliliği ve öğrenci rızası gibi etik konulara yönelik standartların eksikliği, uygulamaların güvenilirliğini azaltmaktadır.
6. **Öğretmen ve Öğrenci Algıları:** Çoğu öğretmen ve öğrenci YZ'yi faydalı bir yenilik olarak görmekte, özellikle geri bildirim hızından memnuniyet duymaktadır. Ancak bazı öğretmenler pedagojik özerkliğin azaldığı, değerlendirme süreçlerinin "mekanikleştiği" endişesini dile getirmiştir. Öğrencilerde ise yapay zekâya aşırı güvenin eleştirel düşünmeyi zayıflatabileceği gözlemlenmiştir.
7. **Eğitim Ortamlarında YZ'nin Uygulanması:** Etkili entegrasyonun ön koşulları arasında öğretmen eğitimi, kaynak dağılımı ve pedagojik uyum öne çıkmaktadır. Öğretmenlerin yalnızca teknik bilgi değil, aynı zamanda eleştirel yapay zekâ okuryazarlığı geliştirmeleri gerekmektedir. OECD (2024) raporlarına göre, bu tür profesyonel gelişim eksikliği eşitsizliği derinleştirebilmektedir.

Tartışma

Çalışma bulguları, YZ'nin matematik değerlendirmelerinde doğruluk ve nesnellik düzeyini artırmakla birlikte, etik ve pedagojik riskler barındırdığını göstermektedir. Algoritmik önyargılar, özellikle veri kümelerinde temsil edilmeyen öğrenci gruplarını dezavantajlı hâle getirebilir. Bu nedenle, "YZ nesneldir" önermesi eleştirel biçimde sorgulanmalıdır.

YZ'nin sunduğu kişiselleştirme olanakları da çoğu zaman yüzeysel düzeyde kalmaktadır. Soruların zorluk düzeyini ayarlamak ya da otomatik ipucu sunmak, gerçek anlamda bireyselleştirilmiş öğrenme değildir. Gerçek kişiselleştirme, öğrencinin kültürel bağlamını, öz düzenleme becerilerini ve öğrenme stratejilerini dikkate alan bütüncül modeller gerektirir.

Ayrıca etik kaygıların literatürde sıklıkla "ek" bir unsur olarak ele alındığı görülmüştür. Oysa eğitimde YZ'nin adil ve sürdürülebilir biçimde kullanılabilmesi için etik ilkelerin tasarım aşamasına dâhil edilmesi şarttır. Veri güvenliği, algoritmik şeffaflık, öğrenci onayı ve öğretmen gözetimi bu bağlamda temel ilkeler arasında yer almalıdır.

Sonuç ve Öneriler

Bu çalışma, yapay zekânın matematik eğitiminde değerlendirme süreçlerini yeniden biçimlendirme gücüne sahip olduğunu, ancak bu dönüşümün yalnızca teknik yenilikle değil, etik temelli tasarım, öğretmen yetkinliği ve pedagojik uyum ile mümkün olabileceğini göstermektedir.

Sonuçlar, üç düzeyde öneriler sunmaktadır:

1. **Politika Düzeyinde:** Eğitim kurumları ve bakanlıklar, YZ tabanlı değerlendirmelere ilişkin açık etik yönergeler oluşturmalı; algoritmik şeffaflık, veri gizliliği ve eşit erişim ilkelerini yasal çerçeveye dâhil etmelidir.
2. **Öğretmen Eğitimi Düzeyinde:** Hizmet öncesi ve hizmet içi programlarda öğretmenlere yalnızca araç kullanımı değil, YZ'nin felsefi ve etik boyutlarını da kapsayan eleştirel eğitimler verilmelidir.
3. **Uygulama Düzeyinde:** YZ sistemleri, öğretmen yargısının yerini almamalı; insan-merkezli pedagojik kararların destekçisi olarak tasarlanmalıdır.

Genel olarak, yapay zekâ destekli değerlendirme sistemleri eğitimde fırsat eşitliğini güçlendirme potansiyeline sahip olsa da, dikkatli bir tasarım ve sürekli denetim olmaksızın mevcut eşitsizlikleri yeniden üretebilir. Bu nedenle çalışma, teknolojik iyimserliğe dayalı yaklaşımlar yerine, eleştirel, etik ve kapsayıcı bir vizyonun benimsenmesi gerektiğini vurgulamaktadır.

Sonuç olarak, yapay zekâ yalnızca matematik değerlendirmesinin biçimini değil, aynı zamanda anlamını da dönüştürmektedir. Bu dönüşümün eğitimsel değeri, teknolojinin "ne kadar akıllı" olduğuna değil, "ne kadar adil, kapsayıcı ve pedagojik" olduğuna bağlıdır.