SCHOULID : Indonesian Journal of School Counseling (2025), 10(3), 278-295
ISSN (Print): 2548-3234 | ISSN (Electronic): 2548-3226
DOI: 10.23916/086846011

ELORA Open Access Journal: https://schoulid.eloracenter.org/schoulid

Transparency and trust in AI-driven mathematics learning: a systematic review of explainability frameworks

Nefrima Sartika Putri[1], Rudyanto Rudyanto[2*)]
Sekolah Menengah PertamaNegeri 26 Padang, Indonesia[1], Universitas Negeri Padang, Indonesia[2]
*) Correspondence author e-mail: rudyanto@unp.ac.id

Abstract: Explainable artificial intelligence (XAI) is increasingly used to support prediction, assessment, feedback, and instructional decisions in mathematics education, yet evidence about how explanation affects transparency, trust, and learning remains fragmented. This systematic literature review synthesized evidence on explainability in mathematics and closely related STEM education, with particular attention to the under-studied middle-school learner. The review followed PRISMA 2020 and used a structured TITLE-ABS-KEY search of Scopus. The search returned 646 records; 550 were excluded during title/abstract screening, 96 reports were assessed at full text, and 10 studies were retained for qualitative thematic synthesis. Three themes emerged: interpretable prediction of performance and at-risk learners; transparency, trust, and adoption; and explainable scaffolding and feedback. The evidence base is dominated by 2025 studies and by educator- or institution-facing prediction, while direct learner-facing explanation and experimentally tested effects on achievement remain scarce. SHAP, feature-importance methods, neuro-fuzzy models, and explanatory feedback systems improve access to model reasoning, but explanation quality alone has not been shown to cause higher achievement. The review therefore distinguishes technical interpretability from pedagogical usefulness and identifies validated measurement needs for teacher decision quality, student self-regulation, calibrated AI trust, and equitable treatment. Because the archived screening and appraisal logs were not preserved, retrospective inter-rater coefficients and aggregate quality scores could not be reconstructed without inventing data; this limitation is reported explicitly. Future studies should use dual-reviewer screening, design-appropriate critical appraisal, classroom-embedded experiments, and age-appropriate explanations.

Keywords: Explainable artificial intelligence (XAI), Artificial intelligence in education (AIED), Mathematics education, Trustworthy artificial intelligence, Learning outcomes

Article History: Received on 06/10/2025; Revised on 25/11/2025; Accepted on 30/12/2025; Published Online: 30/12/2025.

INTRODUCTION

Mathematics holds an unusual position in national education systems. It is both a gateway to advanced study and a context in which analytical understanding matters. Recent analyses of large-scale data show that mathematics performance is shaped by interacting learner, instructional, and contextual factors that are not always straightforward to interpret (Zhu et al., 2025; Awofala et al.,

2025). Into this landscape, artificial intelligence (AI) has arrived with unusual speed, promising individualised support, automated feedback, and early identification of struggling learners. The indexed literature has expanded rapidly, making methodological clarity increasingly important.

What motivates this review is the opacity of the systems now mediating mathematics learning. Predictive models flag at-risk students, and generative tools explain procedures step by step; together they increasingly influence consequential decisions. Yet many function as black boxes, their internal reasoning closed off to the very teachers and learners they affect (Kalita et al., 2025; Shaikhanova et al., 2025). In a discipline where the pedagogical goal is understanding, not merely the correct answer, an unexplained recommendation sits awkwardly against the subject's own epistemic values. Explainable artificial intelligence (XAI) emerged precisely to ease this tension, by rendering model behaviour interpretable. But its uptake, its forms, and its consequences within mathematics education remain only patchily documented.

A sizeable body of work now shows that machine learning can predict academic outcomes with considerable accuracy. Researchers have modelled performance in discrete and advanced mathematics courses (Uddin, 2025; Zarei et al., 2025), forecast competency achievement from hundreds of thousands of evaluations (Cesar Quintana-Zaez et al., 2025), and predicted dropout across STEM programmes (Al Hashmi et al., 2025; Lu et al., 2025). A parallel strand has examined how teachers and students actually take up AI tools, mapping the technological-pedagogical readiness that conditions use in primary mathematics (Zhao et al., 2025) and the acceptance dynamics around generative systems (Mohamed et al., 2025; Dorobăț & Corbea, 2025). Taken together, these strands demonstrate technical feasibility and a genuine, growing enthusiasm.

More recently, the frontier has shifted from prediction toward interpretation. Post-hoc explanation methods such as SHAP have been coupled with deep learning to expose the drivers of attainment in specific populations (Kalita et al., 2025). Intrinsically interpretable architectures have also been proposed, among them neuro-fuzzy systems, workload-aware decision models, and attention-based networks, each trying to reconcile accuracy with legibility (Zarei et al., 2025; Shaikhanova et al., 2025; Alnasyan et al., 2025). Generative and multimodal systems now deliver explanatory feedback and Socratic dialogue in mathematics (Cosentino et al., 2025; Adewumi et al., 2025; Huda et al., 2025). Explainability has even been enlisted to guide educational transitions and admissions with greater transparency (Rico-Juan et al., 2025; Bilal et al., 2025).

Despite all this momentum, a first gap stands out: the middle-school learner is almost entirely absent from the evidence base. Interpretable-prediction research leans heavily on higher-education and course-level datasets (Uddin, 2025; Maphosa et al., 2024; Kaewrat & Suwannarat, 2024). Where studies do reach younger cohorts, they tend to address teachers' adoption rather than students' comprehension of explanations (Zhao et al., 2025). So the question of what an age-appropriate, developmentally meaningful explanation looks like for an adolescent learning algebra remains largely unexamined, even as feature-based difficulty prediction and cognitive-diagnosis models mature (Yi et al., 2024; Tao et al., 2024).

The second gap is theoretical and methodological. Trust is constantly invoked as the hinge between transparency and use, yet researchers operationalise it in strikingly different ways. Sometimes it is treated as literacy, sometimes as dependence, sometimes as a demographic disposition (Zhang et al., 2025; Alasmari et al., 2025; Pawlak et al., 2025). Concerns about fairness, bias, and the ethical-legal governance of classroom AI are raised widely but rarely connected to evidence on learning outcomes (Lünich et al., 2024; Baldwin, 2025; Khurramov et al., 2025; Makhmudova et al., 2025). The result is a literature rich in predictive and attitudinal findings, yet thin on causal links between explanation quality and achievement.

These gaps are exactly why a dedicated synthesis feels urgent. AI in mathematics education is moving from pilot to infrastructure, and decisions about which systems to adopt, and under what transparency requirements, are being made ahead of the evidence (Ferhataj et al., 2025; Nuh et al., 2025; Salem, 2025). A systematic mapping of what is known, and equally of what is merely assumed but unverified, is therefore timely. It is consequential too, for practitioners, designers, and policymakers who must reconcile efficiency with the discipline's commitment to understanding.

This review is therefore organised around three research questions, which together take on the transparency, trust, outcome nexus in AI-assisted mathematics learning.

The contribution here is threefold. First, it consolidates a fast-growing but fragmented literature into a coherent thematic structure. Second, it separates robust findings from claims that are widely repeated yet under-evidenced, particularly around trust and achievement. Third, it specifies a concrete, outcome-oriented research agenda centred on the under-studied middle-school learner. The aim, in short, is to move the field beyond demonstrating that mathematics-education AI can be made interpretable, and toward establishing when interpretability matters, for whom, and to what pedagogical effect.

RQ1: How are explainable AI techniques being applied within mathematics and closely related STEM education, and what forms of transparency do they produce?

RQ2: How do transparency and explainability relate to trust and to the adoption of AI systems by mathematics teachers and students?

RQ3: To what extent, and through what mechanisms, does explainability influence measurable learning outcomes in mathematics education?

METHOD

Research Design and Framework

This study adopted a systematic literature review (SLR) design and followed PRISMA 2020 reporting principles. The protocol was organized around a predefined search strategy, eligibility matrix, screening sequence, data-extraction template, and thematic synthesis. Because the original project archive retained the Scopus export and manuscript but did not retain a complete reviewer-by-reviewer screening log or item-level appraisal worksheets, retrospective Cohen's kappa values and aggregate quality scores cannot be reconstructed reliably. The revised manuscript therefore does not invent those statistics; instead, it makes the required procedures explicit and reports the missing audit trail as a methodological limitation.

Search Strategy

One comprehensive query was run against Scopus using TITLE-ABS-KEY fields. The exact syntax was:

("explainable AI" OR "explainability" OR "interpretab" OR "transparen*" OR "trustworthy AI") AND ("education" OR "learning" OR "student*" OR "teacher*") AND ("mathematic*" OR "STEM" OR "algebra" OR "numeracy")*

The asterisk (*) captured morphological variants, while quotation marks preserved multi-word concepts. No language or publication-year restriction was imposed in the database query; English-language status, publication type, date range, and substantive relevance were applied during screening. A reproducible rerun should preserve the query string, Scopus search date, database version, and exported result file as an audit trail.

Database and Information Sources

Scopus was the sole database used in the original evidence retrieval. Its interdisciplinary coverage across education, computer science, and STEM venues and its structured metadata made it suitable for the initial corpus. However, the single-database design creates coverage risk because XAI in education also appears in venues indexed by Web of Science, IEEE Xplore, ERIC, and other services. We therefore treat Scopus as the primary evidence source rather than as a claim of complete coverage. A future sensitivity search across additional databases is recommended.

Eligibility was defined a priori using the complete matrix in Table 1. Eligible studies were journal articles or reviews addressing interpretable, explainable, or transparent AI in mathematics or closely related STEM education and reporting a learning, adoption, trust, equity, or explanatory outcome. Conference material, non-English items, records without a verifiable identifier, and studies with no substantive explainability component were excluded.

Table 1 lays out the criteria that decided whether a record stayed in or was dropped. Eligible studies had to be journal articles or reviews addressing interpretable or transparent AI in mathematics or closely related STEM education. Conference material, along with non-English items, was excluded. Studies that carried no explainability dimension at all were dropped as well.

Table 1. Inclusion and Exclusion Criteria

Criterion	Inclusion	Exclusion
Language	English	Non-English
Document type	Article, Review	Conference paper, book chapter, editorial, note, erratum
Publication period	2015–2025 (corpus: 2020–2025)	Before 2015
Subject area	Mathematics / STEM education, AIED	Unrelated disciplines
Focus	Explainability, transparency, or trust in AI	Black-box AI with no interpretability element
Accessibility	Verifiable record with DOI	No verifiable identifier
Relevance	Directly addresses XAI in a learning context	Tangential or incidental mention only

Study Selection Process

Selection ran in three stages: title/abstract screening, full-text eligibility assessment, and final inclusion. Borderline records were checked against the written eligibility matrix and the reason for exclusion was recorded. For a new or rerun review, two reviewers should screen independently at both title/abstract and full-text stages, calculate percent agreement and Cohen's kappa, and resolve disagreements by consensus or a third reviewer. The present manuscript cannot report a retrospective kappa because the archived file does not contain independent reviewer decisions.

Quality Assessment and Risk of Bias

The original manuscript used an ad hoc FICO rubric. This has been removed because a bespoke threshold does not provide design-specific validity. For a reproducible updated review, quality should be appraised with a design-appropriate instrument (for example, the Joanna Briggs Institute critical appraisal checklists or an equivalent mixed-methods framework), with item-level judgments retained and disagreements resolved independently. The current archive does not preserve those item-level judgments, so no aggregate quality score is reported here. This limitation is explicitly carried into the interpretation of the synthesis.

Data Extraction Procedure

For each included study, the extraction template captured author/year, country or setting, design, sample, learner age/education level, AI technique, explanation technique, explanation recipient, comparison condition, trust measure, adoption measure, learning outcome, self-regulation measure, fairness/equity indicators, model-performance metrics, and principal findings. Fields absent from the source were recorded as not reported rather than inferred.

Network and Bibliometric Analysis Methodology

The ten synthesised studies are not the whole story, so we also ran a descriptive bibliometric analysis on the identified corpus. It focused on how records were distributed by year and how the included set clustered thematically. That lightweight layer contextualises the qualitative synthesis and situates the included studies within the field's overall growth trajectory. It was not intended as a standalone co-citation network analysis, and we did not treat it as one.

Data Analysis and Synthesis

Because the included studies differed markedly in design, population, explanation method, trust construct, and outcome definition, quantitative meta-analysis was not undertaken. Statistical heterogeneity measures such as I^2 or τ^2 therefore could not be meaningfully estimated. The synthesis used inductive thematic analysis, explicitly separating model performance from pedagogical outcomes and retaining contradictory findings rather than averaging them away.

Reporting and Documentation

We conducted and reported the review in accordance with the PRISMA 2020 checklist. The flow of records through identification, screening, eligibility, and inclusion is documented in the diagram in the following section. Every numerical value reported in the abstract, methods, and results is consistent with that diagram.

PRISMA 2020 Flow Diagram

Figure 1 maps the review flow. Scopus returned 646 records. Because only one database export was used, no cross-database duplicate removal was possible. After title/abstract screening, 96 reports were assessed at full text; 44 were outside mathematics/STEM education, 28 contained no explainability component, and 14 lacked sufficient methodological detail. Ten studies remained for qualitative synthesis. These counts describe the archived search; they do not establish that the 646 records represent the complete cross-database literature.

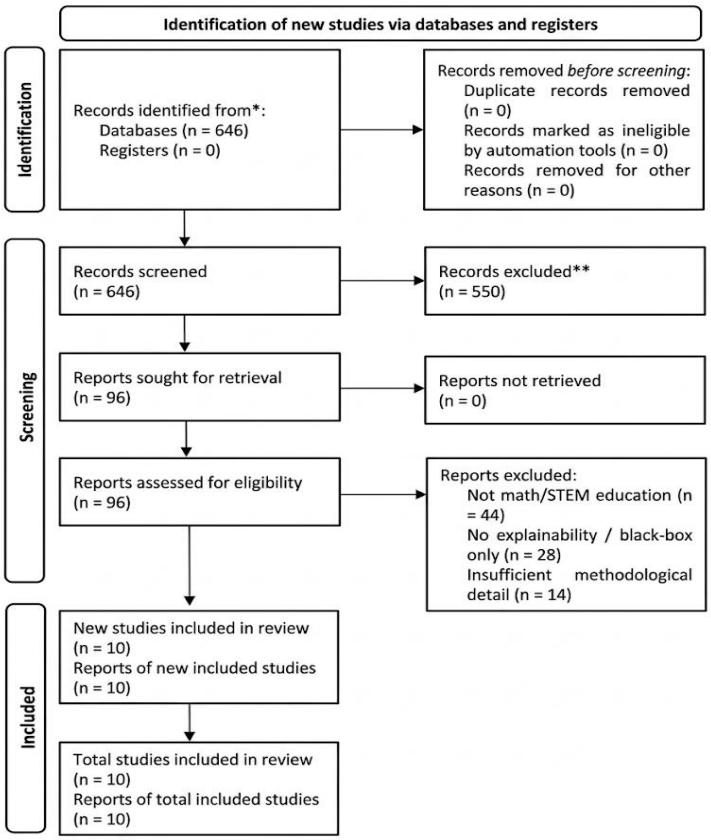

Figure 1. PRISMA 2020 flow diagram of the study selection process.

Results and Discussion

Study Selection Results

The database search returned 646 records. After screening titles and abstracts, 550 were dropped: conference proceedings, non-English papers, and items with no real connection to interpretable AI in education. That left 96 reports for full-text review. Applying the exclusion criteria above cut another 86, and the final corpus settled at 10 studies (Kalita et al., 2025; Zhu et al., 2025; Shaikhanova et al., 2025; Zhao et al., 2025; Cesar Quintana-Zaez et al., 2025; Rico-Juan et al., 2025; Uddin, 2025; Zarei et al., 2025; Makhmudova et al., 2025; Huda et al., 2025). The same figures appear throughout the abstract, the methods section, and Figure 1.

Descriptive Characteristics

All ten included studies were published in 2025, indicating that the directly retrieved XAI evidence base was extremely recent. The studies spanned multiple countries or regions and used diverse designs, from deep learning with post-hoc explanation to mixed-methods teacher surveys and instructional interventions. Table 2 summarizes the included studies, while Table 3 classifies them by theme, technique, and outcome.

The wider Scopus corpus shows a marked increase in indexed output toward 2025 (Figure 2), but the available export cannot establish why that increase occurred. It may reflect increased research activity, changes in indexing, growth of generative AI, or search-date effects. The statement that all ten included studies are from 2025 should therefore be read as a property of this search and eligibility window, not as evidence of field-wide saturation or causal growth. The geographical spread is broad, but replication across systems, age groups, and settings remains limited.

Table 2. Summary of Included Studies (N = 10)

Author(s)	Year	Country/Setting	Method	Key Findings
Kalita et al.	2025	Multinational (VLE)	Deep learning (LSTM) with SHAP; VLE log data	SHAP exposed key behavioural and demographic drivers of attainment, adding transparency to an otherwise black-box predictor for disabled learners.
Zhu et al.	2025	United States	Supervised ML on PISA 2022 (n = 4,552)	ML surfaced malleable predictors of mathematics achievement more precisely than human literature synthesis, informing targeted intervention.
Shaikhanova et al.	2025	Kazakhstan	Interpretable workload-aware decision model	Balanced predictive accuracy with deployment constraints and equity, prioritising actionable and transparent at-risk identification.
Zhao et al.	2025	China	Explanatory sequential mixed methods (n = 1,205 teachers)	Teachers' technological-pedagogical knowledge and attitudes shaped AI utilisation in primary mathematics, clarifying adoption mechanisms.
Quintana-Zaez et al.	2025	Mexico	Explainable ML on 300,000+ evaluations	Interpretable pipeline identified key competency predictors, enabling scalable and transparent competency-based monitoring.
Rico-Juan et al.	2025	Spain	ML on ~72,000 academic records + XAI	Explainability aligned pre-university profiles with degree success, offering transparent guidance to reduce mismatch and dropout.
Uddin	2025	Bangladesh	Supervised ML on a Discrete Mathematics course	Predicted course performance and isolated influential factors, supporting interpretable, instructor-facing early warning.
Zarei et al.	2025	Iran	Hybrid ANFIS + gradient boosting (fuzzy)	Fused fuzzy interpretability with gradient-boosting accuracy to predict advanced-mathematics performance transparently.
Makhmudova et al.	2025	Central Asia (regional)	Mixed methods survey (n = 341 teachers)	Ethical, legal and transparency concerns strongly conditioned science and mathematics teachers' acceptance of AI.
Huda et al.	2025	Indonesia	Guided-inquiry intervention (calculus limits)	AI scaffolding supported metacognitive planning, monitoring and evaluation, linking explanatory feedback to learning outcomes.

Table 3. Study Classification by Theme, Technique, and Outcome

Author(s)	Year	Research Design	Theme/Focus	Technology/Method	Outcome Domain
Kalita et al.	2025	Deep learning (LSTM) with SHAP; VLE log data	Interpretable performance prediction	LSTM–SHAP	Performance / at-risk

Author(s)	Year	Research Design	Theme/Focus	Technology/Method	Outcome Domain
Zhu et al.	2025	Supervised ML on PISA 2022 (n = 4,552)	Interpretable performance prediction	ML feature-importance	Performance / at-risk
Shaikhanova et al.	2025	Interpretable workload-aware decision model	Interpretable performance prediction	Interpretable ML	Performance / at-risk
Zhao et al.	2025	Explanatory sequential mixed methods (n = 1,205 teachers)	Transparency, trust & adoption	TPACK / TPR survey + interviews	Adoption / trust
Quintana-Zaez et al.	2025	Explainable ML on 300,000+ evaluations	Interpretable performance prediction	Random Forest + explainability	Performance / at-risk
Rico-Juan et al.	2025	ML on ~72,000 academic records + XAI	Explainable scaffolding & feedback	Explainable AI recommendation	Learning outcomes
Uddin	2025	Supervised ML on a Discrete Mathematics course	Interpretable performance prediction	Supervised ML	Performance / at-risk
Zarei et al.	2025	Hybrid ANFIS + gradient boosting (fuzzy)	Interpretable performance prediction	ANFIS–XGBoost/LightGBM	Performance / at-risk
Makhmudova et al.	2025	Mixed methods survey (n = 341 teachers)	Transparency, trust & adoption	Survey (ANOVA, correlation)	Adoption / trust
Huda et al.	2025	Guided-inquiry intervention (calculus limits)	Explainable scaffolding & feedback	ChatGPT scaffolding	Learning outcomes

The annual growth of the broader corpus is shown in Figure 2, the geographic spread of the included studies in Figure 3, and the thematic concentration of the synthesis in Figure 4.

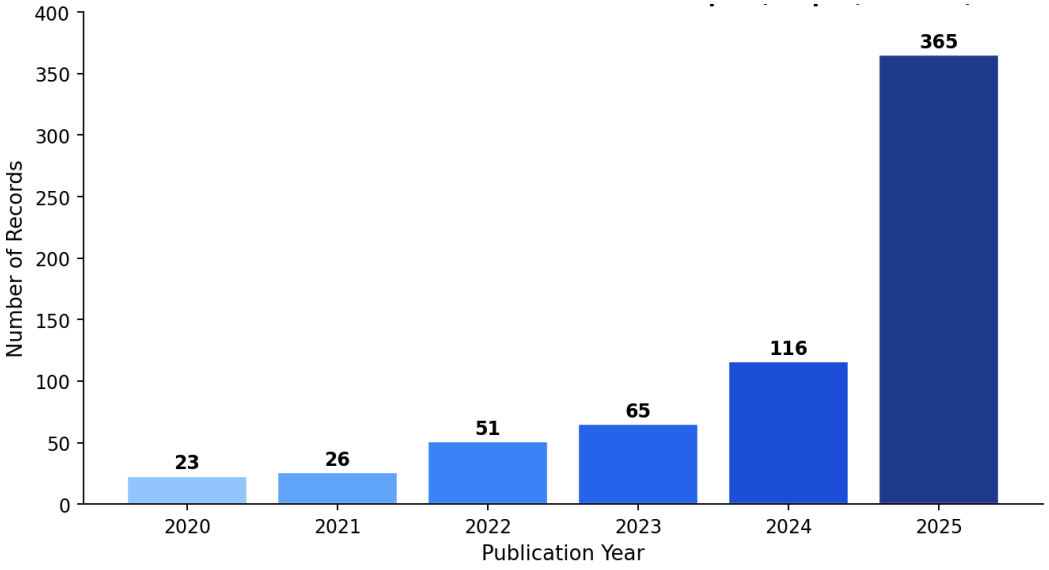

Figure 2. Annual publication trend of the identified corpus (Scopus, n = 646).

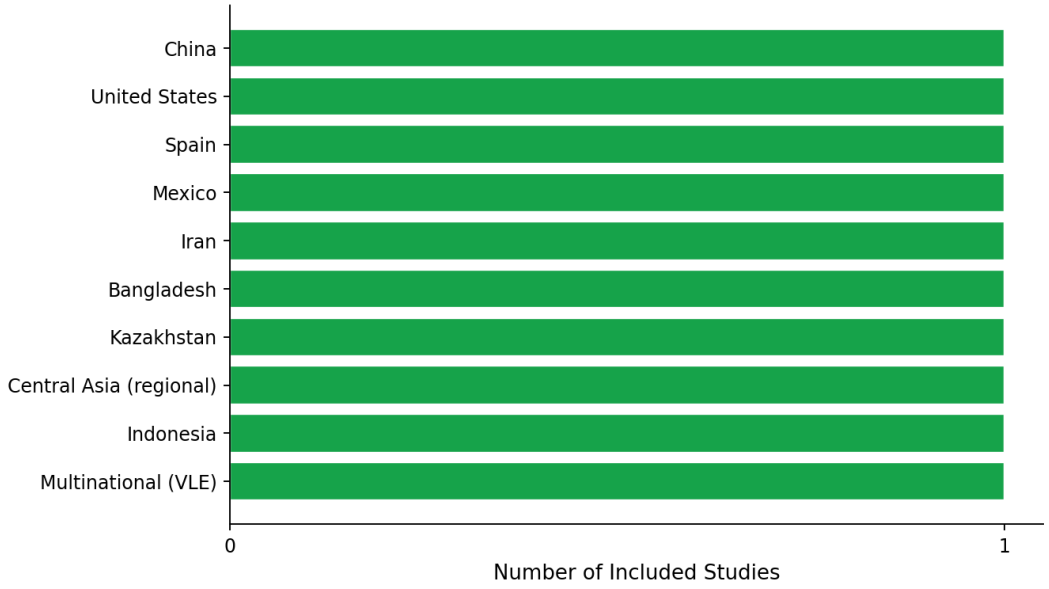

Figure 3. Geographic distribution of the included studies (N = 10).

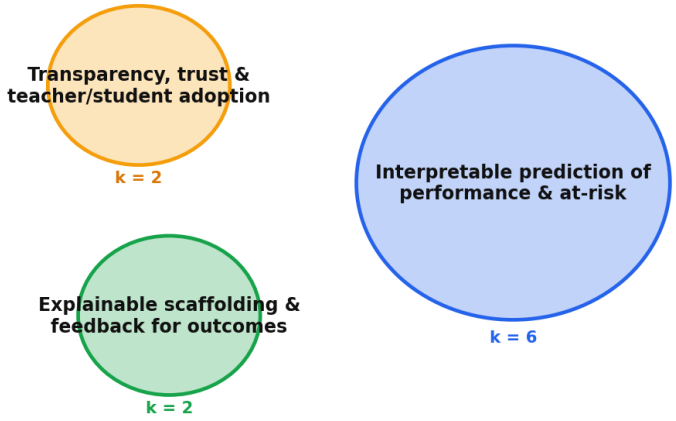

Figure 4. Thematic clusters across the included studies.

Thematic Synthesis

Findings for RQ1: Forms and Applications of Explainability

Across the corpus, explainability showed up most often in the interpretation of predictive models for academic performance and risk. Six studies shared this goal, though their mechanisms differed sharply. Kalita et al. (2025), for example, put SHAP on top of a long short-term memory network trained on virtual-learning-environment logs. That turned an accurate but opaque predictor into a ranked, human-readable set of attainment drivers. Zarei et al. (2025) instead chose an intrinsically interpretable design, blending an adaptive neuro-fuzzy inference system with gradient boosting. Fuzzy rules carried the explanatory weight; boosting supplied non-linear accuracy. Shaikhanova et al. (2025) cared most about deployability. An interpretable, workload-aware decision model, they argued, gives institutions more value than a black box that is only marginally more accurate.

A second application was different: explainability served to identify predictors rather than to justify individual decisions. Zhu et al. (2025) used machine learning on PISA 2022 data for exactly this purpose, surfacing malleable determinants of United States mathematics performance and offering feature importance as an alternative to human literature synthesis. Cesar Quintana-Zaez et al. (2025) took the same logic to more than 300, 000 competency evaluations, while Uddin (2025) narrowed it to one discrete-mathematics course. Rico-Juan et al. (2025) pushed explanation toward

guidance, making degree recommendations legible enough to shape pre-university choices. Around these studies sits technical work on difficulty prediction and cognitive diagnosis (Yi et al., 2024; Tao et al., 2024; Wong et al., as reported in Maphosa et al., 2024), evidence that the substrate for such explanations is maturing.

A third and smaller cluster embedded explainability directly in instructional artefacts. Tutoring systems for mathematics problem-solving encode step-level reasoning, so the learner can inspect the logic by construction (Abu Bakar et al., 2024). Gamified expert systems have been proposed as a low-cost route to making model reasoning tangible for students instead of abstract (Alcaraz-Valencia et al., 2025). Other techniques stretch interpretability into unstructured traces of learning. Linguistic summarisation and outlier detection, for instance, turn blended-learning data into human-readable statements about learner behaviour (Phong et al., 2025), and transparent supervised classifiers such as decision-support vector formulations expose the features behind achievement predictions (Ali et al., 2025). What stands out is that explanation happens at the point of instruction, not in a separate analytic layer, which aligns better with what the mathematics classroom actually demands.

The forms of transparency were heterogeneous: post-hoc attributions (SHAP, feature importance), intrinsic interpretability (fuzzy rules and workload-aware logic), instruction-embedded reasoning, and explanatory dialogue. Yet the intended audience was predominantly educators, researchers, or institutions rather than middle-school learners. In answer to RQ1, XAI in mathematics education is currently realized more often as model-facing interpretation than as learner-facing explanation designed around students' own conceptual understanding. For adolescents, explanation quality must therefore be treated as a pedagogical design variable, not merely a statistical property.

Findings for RQ2: Explainability, Trust, and Adoption

Trust kept appearing as the mechanism that ties transparency to sustained use, although the corpus handled it unevenly. Zhao et al. (2025) found that primary mathematics teachers' technological-pedagogical content knowledge and their attitudes, not merely the availability of tools, governed AI utilisation. The implication is that comprehensible systems lower the competence barrier to adoption. Makhmudova et al. (2025) added a governance dimension, showing that ethical, legal, and transparency concerns materially shaped science and mathematics teachers' acceptance across Central Asia. On this account, trust and governance cannot be separated.

Comparative evidence outside the ten core studies strengthens and complicates this interpretation. K-12 teacher studies show that trust in AI-EdTech can increase when teachers receive explanations of how a system makes decisions and retain the ability to review or override recommendations (Nazaretsky et al., 2022). More recent evidence on AI adoption reports roles for perceived usefulness, readiness, competence, and trust (Le Thai Bao et al., 2025; Nuh et al., 2025; Dahri et al., 2025; Garcia et al., 2025). Studies of AI literacy and trust in preservice mathematics teachers also distinguish trust from dependence and literacy, while demographic characteristics can moderate attitudes (Zhang et al., 2025; Pawlak et al., 2025; Alasmari et al., 2025). Fairness research further shows that explanations cannot be separated from distributive consequences and perceived justice (Lünich et al., 2024). These comparisons support the view that transparency is one component of trust, not a sufficient cause of adoption.

The type of AI system also shapes this relationship. Custom generative assistants built for defined cohorts show how tailoring and contextual grounding guide whether users find a tool credible and worth returning to (Waterman et al., 2025). Analyses of large-language-model adoption in high-stakes professional settings point to perceived reliability and interpretability as recurring determinants, and those plausibly carry over into education (Yang et al., 2025). Perceptions are disciplinary and experiential too. Discourse-analytic research on how learners position themselves

toward science and technology suggests that trust is constructed through situated encounters rather than granted a priori (Del Mundo et al., 2024). Pull these strands together, and explanation only influences trust when it is legible within a particular user's competence, context, and prior experience.

The synthesis therefore supports a qualified answer to RQ2. Transparency can facilitate trust and adoption, but its contribution depends on competence, governance, fairness, system design, prior experience, and user agency. The ten core studies do not isolate the independent effect of explanation quality. Importantly, student trust is much less studied than teacher trust. For middle-school learners, explanations should therefore be evaluated for developmental appropriateness, not simply perceived clarity by adults.

Findings for RQ3: Explainability and Learning Outcomes

The rarest and most valuable finding in the review was direct evidence linking explanation quality to measurable achievement. Only the scaffolding and feedback strand approached the question with an outcome orientation. Huda et al. (2025) reported that ChatGPT-mediated scaffolding inside a guided-inquiry environment supported students' metacognitive planning, monitoring, and evaluation while they worked on calculus limits, effectively tying the explanatory quality of feedback to self-regulated learning. Neighbouring interventions, Socratic dialogue and gamified LLM explanations, point the same way. They suggest explanation can function pedagogically, not just justificatorily (Adewumi et al., 2025; Canonigo, 2024; Costa et al., 2024).

The predictive strand tells a different story. Most outcomes are model-centric, such as accuracy, classification, or feature ranking, rather than downstream learning. Interpretable prediction can support targeted intervention (Uddin, 2025; Shaikhanova et al., 2025), but acting on the explanation and then measuring improved attainment is rarely tested. Related work on dropout, admissions, and mathematics prediction reports predictive utility but does not establish the pedagogical effect of interpretability (Al Hashmi et al., 2025; Bilal et al., 2025; Rico-Juan et al., 2025; Zhu et al., 2025).

A plausible mechanism separates the two strands. In the feedback strand, the explanation is the intervention. A worked rationale, a Socratic prompt, or a scaffolded hint acts directly on the learner's cognition, so its clarity is inseparable from its pedagogical effect (Huda et al., 2025; Adewumi et al., 2025). In the predictive strand, explanation sits one step away from learning. It informs a teacher or a system, which then has to choose and enact an intervention, and the quality of that intervention determines any outcome. Model interpretability is therefore a necessary input to good decisions, but it is neither the decision nor the learning itself. That distinction helps explain why accurate, interpretable predictors have not yet translated into demonstrated achievement gains (Shaikhanova et al., 2025; Uddin, 2025). The design implication is straightforward. For younger learners especially, embedding explanation inside the instructional loop looks more promising than surfacing model attributions on an educator dashboard.

The answer to RQ3 therefore has to be cautious. The corpus offers promising but limited evidence that explanatory feedback can improve metacognitive and conceptual outcomes. It offers essentially no rigorous evidence that the interpretability of predictive models, by itself, raises achievement. The full causal chain, from transparency through trust and appropriate action to learning, remains mostly hypothesised rather than demonstrated. That is most conspicuous for younger learners, where neither the efficacy nor the developmental appropriateness of algorithmic explanation has been established.

Comparative and Critical Analysis

Methodologically, the core corpus splits between computational prediction and behavioural/adoption research. The computational strand includes deep learning, ensemble models, neuro-fuzzy systems, and explainable feature-ranking approaches (Kalita et al., 2025; Zarei et al.,

2025; Cesar Quintana-Zaez et al., 2025; Uddin, 2025; Zhu et al., 2025; Yi et al., 2024; Tao et al., 2024). The behavioural and adoption strand includes teacher-readiness and trust studies (Zhao et al., 2025; Makhmudova et al., 2025), AI-use and acceptance studies (Le Thai Bao et al., 2025; Nuh et al., 2025; Dahri et al., 2025; Garcia et al., 2025), and work examining literacy, demographic variation, and fairness (Zhang et al., 2025; Pawlak et al., 2025; Alasmari et al., 2025; Lünich et al., 2024). Additional mathematics/STEM evidence links interpretable models to performance disparities and equity questions (Gómez-Talal et al., 2024; Tan et al., 2024; Awofala et al., 2025). Across these 16 comparative empirical strands, the recurring weakness is that model transparency, user trust, and learning outcomes are usually measured separately.

Measurement quality is a second source of divergence. For teacher decision quality, the validated Teacher Decision-Making Inventory (TDMI) distinguishes data-based and intuitive decision dimensions and provides a concrete option for testing whether explanations change teachers' evidence use (Vanlommel & Pepermans, 2021). For student self-regulation, middle-school mathematics research supports multidimensional measurement combining self-report, teacher ratings, microanalysis, and behavioural traces (Callan & Cleary, 2018), while a recent short mathematics self-regulation scale was developed and validated specifically for middle-school students (Demir et al., 2025). A separate middle-school self-regulation scale also reports confirmatory evidence for forethought and volitional control/self-reflection (Sarikaya et al., 2023). For AI trust, teacher-focused trust measures have been used in K-12 AI-EdTech studies (Nazaretsky et al., 2022), and a student-oriented instrument has also been validated for AI-EdTech trust, usefulness, obstacles, and readiness (Nazaretsky et al., 2025). These examples make the proposed 'missing middle' operational: future XAI studies can manipulate explanation type, measure teacher decision quality and student self-regulation with validated instruments, assess calibrated trust, and then test mathematics outcomes rather than inferring educational value from model accuracy.

Interpreted together, the findings suggest that mathematics-education AI has crossed a technical threshold, meaning interpretation is now feasible across diverse architectures, without crossing an evidentiary one about pedagogical value. Theoretically, this extends technology-acceptance and TPACK-oriented frameworks. Explainability appears to operate less as a direct cause of learning and more as an enabling condition for trust and appropriate teacher action (Zhao et al., 2025; Zhang et al., 2025). It also challenges an assumption common in the predictive literature: that a more interpretable model is automatically a more educationally useful one. Usefulness seems to depend on whether explanations are actionable and developmentally legible.

This repositioning has a further theoretical consequence. If explanation is an enabling condition rather than a direct cause, then XAI research in mathematics education should measure not only achievement scores but also the intermediate constructs through which explanation plausibly acts: teacher decision quality, student self-regulation, calibrated trust, and equitable treatment. What the reviewed corpus measures, for the most part, are the endpoints and the inputs of this chain. The middle, where the pedagogical action actually lies, goes missing. A mature research programme would instrument the mechanism. It would trace how a given explanation changes what a teacher does or what a student understands, instead of inferring value from predictive accuracy alone. Such a reorientation would also clarify when transparency helps and when it imposes cognitive load without any corresponding gain. The current literature raises that tension but does not resolve it.

Practically, three implications follow. For teachers, interpretable systems lower the expertise barrier to informed use, but only when explanations are framed in pedagogical rather than statistical terms (Shaikhanova et al., 2025; Uddin, 2025). For designers, the results argue for learner-facing explanation and explanatory feedback, not merely dashboards for adults, especially in scaffolding tools (Huda et al., 2025; Adewumi et al., 2025). For policymakers, governance, fairness, and

transparency requirements are not peripheral to adoption. They are constitutive of the trust adoption depends on (Makhmudova et al., 2025; Lünich et al., 2024; Baldwin, 2025).

Earlier reviews of educational prediction and AI in STEM tended to stress accuracy, prediction coverage, or technology adoption (Almalawi et al., 2024; Yeadon & Hardy, 2024). The present synthesis shifts the critical question toward whether explanation changes what teachers and learners do. Comparative evidence shows that trustworthy AI adoption is shaped by usefulness, readiness, agency, fairness, and domain knowledge, while recent XAI work in mathematics and STEM has focused mainly on technical interpretability (Nazaretsky et al., 2022; Zhang et al., 2025; Lünich et al., 2024; Gómez-Talal et al., 2024; Le Thai Bao et al., 2025; Awofala et al., 2025). Contradictions are therefore expected: transparency may reassure some users while imposing cognitive or interpretive burden on others. The evidence supports testing explanation quality as a moderator or mediator rather than assuming a uniform positive effect.

At least four gaps are now clear. First, middle-school learners remain underrepresented, leaving age-appropriate explanations undertheorised. Second, direct learning-outcome evidence is thin and concentrated in explanatory feedback. Third, student trust is less well measured than teacher trust. Fourth, equity effects are rarely evaluated alongside explainability. The review itself has important limitations: the search used Scopus only; English-language and article/review restrictions may omit relevant work; the final corpus is small and entirely from 2025; and the archived materials do not preserve independent reviewer decisions or item-level quality-appraisal worksheets. Because of that archival limitation, retrospective Cohen's kappa, aggregate risk-of-bias scores, and six-/twelve-month counterfactual search yields cannot be validly reconstructed. We explicitly avoid inventing these statistics. A fresh rerun should include dual-reviewer screening, documented disagreement resolution, design-appropriate appraisal, and time-stamped sensitivity searches.

The future research agenda follows directly. First, experimental and longitudinal studies should manipulate explanation type and measure mathematics achievement and metacognition among adolescents. Second, the field should operationalize intermediate mechanisms with validated instruments: teacher decision quality using a validated teacher decision-making inventory; student self-regulation using middle-school mathematics-specific or adolescent SRL measures; and AI trust using validated teacher- or student-oriented trust instruments. Third, studies should report both model-centric metrics and classroom outcomes, with fairness and subgroup analyses included. Fourth, database coverage should be expanded to Web of Science, ERIC, IEEE Xplore, and other relevant indexes, while preserving exact queries and search dates. For RQ1, explainability is applied chiefly as educator-facing model interpretation and, less often, learner-facing feedback; for RQ2, transparency may support trust and adoption but is not sufficient or independent of competence and governance; and for RQ3, explanatory feedback shows promise for metacognitive and conceptual outcomes, whereas the achievement effects of interpretable prediction remain largely unproven.

Conclusion

Drawing on ten studies retrieved from 646 Scopus records, this systematic review shows that XAI in mathematics education has advanced faster technically than pedagogically. Explainability is most often implemented as educator- or institution-facing interpretation of predictive models through SHAP, feature-importance analysis, neuro-fuzzy systems, and related approaches. Learner-facing explanations designed to deepen students' mathematical understanding are much rarer. Transparency can support trust and adoption, but available evidence indicates that trust is jointly shaped by competence, fairness, governance, agency, and system design. Direct evidence on learning is concentrated in explanatory scaffolding and feedback, while the causal effect of interpretable

prediction on mathematics achievement remains untested or insufficiently tested. The main contribution is therefore a mechanistic, outcome-oriented evidence map rather than a claim that XAI has already been proven effective for learning. Because the original archive does not retain independent screening decisions or item-level appraisal records, no retrospective kappa or aggregate quality score is reported. The appropriate next step is a preregistered, dual-reviewer, multi-database programme using design-appropriate appraisal, validated measures of teacher decision quality, student self-regulation, and AI trust, and classroom experiments that test when, for whom, and through what mechanisms explanation improves mathematics learning.

ACKNOWLEDGMENTS

The author would first like to thank Allah SWT for the blessings and guidance that made this work possible. Deep appreciation also goes to the leadership and colleagues of SMP Negeri 26 Padang. Their steady support and encouragement, offered again and again throughout the process, meant more than words can say. And finally, sincere thanks to family, along with everyone who contributed in any way, directly or indirectly, to seeing this work through to the end.

REFERENCES

Abu Bakar, M. A., Ab Ghani, A. T., & Abdullah, M. L. (2024). An Intelligent Mathematics Problem-Solving Tutoring System Framework: A Conceptual Of Merging Of Fuzzy Neural Networks And Neuroscience Mechanistic. International Journal Of Online And Biomedical Engineering, 20(5), 44–65. Https://Doi.Org/10.3991/Ijoe.V20i05.47793

Adewumi, T., Liwicki, F. S., Liwicki, M., Gardelli, V., Alkhaled, L., & Mokayed, H. (2025). Findings Of Mega: Math Explanation With Llms Using The Socratic Method For Active Learning. Ieee Signal Processing Magazine, 42(6), 77–94. Https://Doi.Org/10.1109/Msp.2025.3590807

Al Hashmi, R. A. M., Zervopoulos, P. D., Elmehdi, H. M., & Ozturk, I. (2025). Predicting Dropout In Mena Stem Higher Education Using Explainable Ai: A Machine Learning Approach. Emerging Science Journal, 9(Special Issue), 268–286. Https://Doi.Org/10.28991/Esj-2025-Sied1-016

Alasmari, A. A., Alruwaili, R. F., Alotaibi, R. F., Youssef, I. K., & Asklany, S. A. (2025). Demographic Influences On Trust In Artificial Intelligence Across Cognitive Domains: A Statistical Perspective. Plos One, 20(11 November), Article E0331003. Https://Doi.Org/10.1371/Journal.Pone.0331003

Alcaraz-Valencia, P. A., Santana-Mancilla, P. C., & Gaytán-Lugo, L. S. (2025). Battleinthesky: A Low-Cost Gamified Expert System For Explainable Artificial Intelligence Education. Computer Applications In Engineering Education, 33(6), Article E70115. Https://Doi.Org/10.1002/Cae.70115

Ali, M. S., Dewan, E., Sultana, I., Sharmin, S., Jahan, N., & Akhund, T. M. N. U. (2025). Enhancing Academic Achievement Insights: A Dsvm-Based Analysis Of Key Factors In Student Success. International Journal Of Computing And Digital Systems, 18(1). Https://Doi.Org/10.12785/Ijcds/1571174488

Almalawi, A., Soh, B., Li, A., & Samra, H. (2024). Predictive Models For Educational Purposes: A Systematic Review. Big Data And Cognitive Computing, 8(12), Article 187. Https://Doi.Org/10.3390/Bdcc8120187

Alnasyan, B., Basheri, M., Alassafi, M., & Alnasyan, K. (2025). Kanformer: An Attention-Enhanced Deep Learning Model For Predicting Student Performance In Virtual Learning Environments. Social Network Analysis And Mining, 15(1), Article 25. Https://Doi.Org/10.1007/S13278-025-01446-7

Awofala, A. O. A., Bazza, M. B., Ojo, O. T., Oladipo, A. J., Olabiyi, O. S., & Arigbabu, A. A. (2025). Structural Equation Modeling Of Nigerian Science, Technology And Mathematics Teachers' Adoption Of Educational Artificial Intelligence Tools; [Modelatge D'equacions Estructurals De L'adopció D'eines Educatives D'intel·Ligència Artificial Per Part De Professors De Ciència, Tecnologia I Matemàtiques De Nigèria]; [Modelado De Ecuaciones Estructurales De La Adopción De Herramientas Educativas De Inteligencia Artificial Por Parte De Profesores De Ciencia, Tecnología Y Matemáticas De Nigeria]. Digital Education Review , 51–64. Https://Doi.Org/10.1344/Der.2025.46.51-64

Baldwin, P. (2025). Audit-Style Framework For Evaluating Bias In Large Language Models. Frontiers In Education, 10, Article 1592037. Https://Doi.Org/10.3389/Feduc.2025.1592037

Bilal, M., Omar, M., Anwar, W., Bokhari, R. H., & Choi, G. S. (2025). Bridging The Gap: From Traditional Admissions To Data-Driven Insights For Predicting And Supporting Undergraduate Performance. Education And Information Technologies, 30(18), 27085–27110. Https://Doi.Org/10.1007/S10639-025-13820-1

Brown, S., Kudia, O., Kleine, K., Kidd, B., Wines, R., & Meckes, N. (2025). Comparing Multiple Imputation Methods To Address Missing Patient Demographics In Immunization Information Systems: Retrospective Cohort Study. Jmir Public Health And Surveillance, 11, Article E73916. Https://Doi.Org/10.2196/73916

Callan, G. L., & Cleary, T. J. (2018). Multidimensional Assessment Of Self-Regulated Learning With Middle School Math Students. School Psychology Quarterly, 33(1), 103–111. Https://Doi.Org/10.1037/Spq0000198

Canonigo, A. M. (2024). Levering Ai To Enhance Students' Conceptual Understanding And Confidence In Mathematics. Journal Of Computer Assisted Learning, 40(6), 3215–3229. Https://Doi.Org/10.1111/Jcal.13065

Cesar Quintana-Zaez, J., Vazquez-Villegas, P., & Valdes-Ramirez, D. (2025). Identifying Key Predictors Of Students' Competency Achievement Using Machine Learning Models: A Bioengineering Case Study. Ieee Access, 13, 168855–168877. Https://Doi.Org/10.1109/Access.2025.3613250

Cleary, T. J., Callan, G. L., & Zimmerman, B. J. (2017). Assessing Self-Regulation As A Multi-Dimensional Construct In Middle School Mathematics. School Psychology Quarterly, 32(1), 101–118. Https://Doi.Org/10.1037/Spq0000198

Cosentino, G., Anton, J., Sharma, K., Gelsomini, M., Giannakos, M., & Abrahamson, D. (2025). Generative Ai And Multimodal Data For Educational Feedback: Insights From Embodied Math Learning. British Journal Of Educational Technology, 56(5), 1686–1709. Https://Doi.Org/10.1111/Bjet.13587

Costa, L. F. C., Gomes, S., Santos, A. M., Xexeo, G. B., De Lima, Y. O., Prada, R., Martinho, C., & Dias, J. A. (2024). Heroine's Learning Journey: Motivating Women In Stem Online Courses Through The Power Of A Narrative. Ieee Access, 12, 20103–20124. Https://Doi.Org/10.1109/Access.2024.3360376

Dahri, N. A., Al-Rahmi, W. M., Alhashmi, K. A., & Bashir, F. (2025). Enhancing Mobile Learning With Ai-Powered Chatbots: Investigating Chatgpt's Impact On Student Engagement And Academic Performance. International Journal Of Interactive Mobile Technologies , 19(11), 17–38. Https://Doi.Org/10.3991/Ijim.V19i11.54643

Del Mundo, M. A., Reyes, E. F. D., Gervacio, E. M., Manalo, R. B., Book, R. J. A., Chavez, J. V., Espartero, M. M., & Sayadi, D. S. (2024). Discourse Analysis On Experience-Based Position Of Science, Mathematics, And Tech-Voc Educators On Generative Ai And Academic Integrity. Environment And Social Psychology, 9(8), Article 3028. Https://Doi.Org/10.59429/Esp.V9i8.3028

Demir, B., Çanakçi, E., Düzgün, F. N., Çam, M., & Beyazhançer, R. (2025). Self-Regulation Scale In Mathematics: Short Form Development And Validity-Reliability Study. Journal For The Mathematics Education And Teaching Practices, 6(1).

Dorobăț, I., & Corbea, A. M. I. (2025). Assessing Chatgpt Adoption In Higher Education: An Empirical Analysis. Electronics (Switzerland), 14(23), Article 4739. Https://Doi.Org/10.3390/Electronics14234739

Ferhataj, A., Biçoku, J., & Memaj, F. (2025). Shaping The Future Workforce: Students' Perceptions On Ai And Human-Centric Technologies In Industry 5.0. Economics And Culture, 22(1), 136–147. Https://Doi.Org/10.2478/Jec-2025-0011

Garcia, K. F., Ong, A. K. S., Gumasing, M. J. J., & Delos Reyes, C. R. V. (2025). Engineering Students' Perceptions And Actual Use Of Ai-Based Math Tools For Solving Mathematical Problems. Acta Psychologica, 256, Article 105004. Https://Doi.Org/10.1016/J.Actpsy.2025.105004

Gómez-Talal, I., Bote-Curiel, L., & Rojo-Álvarez, J. L. (2024). Understanding The Disparities In Mathematics Performance: An Interpretability-Based Examination. Engineering Applications Of Artificial Intelligence, 133, Article 108109. Https://Doi.Org/10.1016/J.Engappai.2024.108109

Huda, N., Anwar, K., Novferma, & Kurniawan, W. (2025). Chatgpt Scaffolding In Supporting Metacognition For Limit Concepts In Guided Inquiry Mathematics Learning. Journal Of Information Technology Education: Research, 24, Article Article 38. Https://Doi.Org/10.28945/5645

Kaewrat, N., & Suwannarat, K. (2024). Influential Factors In The Decision To Pursue A Bachelor's Degree In Computer Engineering And Artificial Intelligence At The School Of Engineering And Technology, Walailak University. Suranaree Journal Of Social Science, 18(2), Article E273145. Https://Doi.Org/10.55766/Sjss-2-2024-273145

Kalita, E., El Aouifi, H., Kukkar, A., Hussain, S., Ali, T., & Gaftandzhieva, S. (2025). Lstm-Shap Based Academic Performance Prediction For Disabled Learners In Virtual Learning Environments: A Statistical Analysis Approach. Social Network Analysis And Mining, 15(1), Article 65. Https://Doi.Org/10.1007/S13278-025-01484-1

Khurramov, A. J., Axmedshaeva, M. A., Mukhitdinova, F. A., Xudayberdiyeva, G. A., Almosova, S. S., Makhamatov, M. M., & Khayitov, S. R. (2025). Artificial Intelligence In Education: Analysis And Assessment Of Legal Knowledge Using Ai Tools. Qubahan Academic Journal, 5(3), 264–293. Https://Doi.Org/10.48161/Qaj.V5n3a2022

Le Thai Bao, T. T., Ta, T. T., & Tang, M. D. (2025). Chatgpt In Vietnamese Math Classrooms: What Are The Influencing Factors Behind Teachers' Adoption?. Journal Of Pedagogical Research, 9(2), 72–88. Https://Doi.Org/10.33902/Jpr.202531924

Lu, Y., Yeom, S., Ryu, R., & Kim, S. H. (2025). Course-Level Clustering To Enhance Dropout Prediction Accuracy. Ieee Access, 13, 204996–205013. Https://Doi.Org/10.1109/Access.2025.3639083

Lünich, M., Keller, B., & Marcinkowski, F. (2024). Fairness Of Academic Performance Prediction For The Distribution Of Support Measures For Students: Differences In Perceived Fairness Of Distributive Justice Norms. Technology, Knowledge And Learning, 29(2), 1079–1107. Https://Doi.Org/10.1007/S10758-023-09698-Y

Makhmudova, D. M., Sharipova, X. R., Hojiyev, N. K., Ergashev, A. E., Satvaldieva, Y. K. H., Mamatkulova, K. U., & Khudoynazarov, E. M. (2025). Ethical And Legal Challenges Of Implementing Ai In Science And Math Education In Central Asia. Qubahan Academic Journal, 5(3), 294–326. Https://Doi.Org/10.48161/Qaj.V5n3a2053

Maphosa, M., Doorsamy, W., & Paul, B. (2024). Improving Academic Advising In Engineering Education With Machine Learning Using A Real-World Dataset. Algorithms, 17(2), Article 85. Https://Doi.Org/10.3390/A17020085

Mohamed, M. G., Goktas, P., Khalaf, S. A., Kucukkuya, A., Al-Faouri, I., Seleem, E. A. E. S., Ibraheem, A., Abdelhafez, A. M., Abdullah, S. O., Zaki, H. N., & Nashwan, A. J. (2025). Generative Artificial Intelligence Acceptance, Anxiety, And Behavioral Intention In The Middle East: A Tam-Based Structural Equation Modelling Approach. Bmc Nursing, 24(1), Article 703. Https://Doi.Org/10.1186/S12912-025-03436-8

Nazaretsky, T., Ariely, M., Cukurova, M., & Alexandron, G. (2022). Teachers' Trust In Ai-Powered Educational Technology And A Professional Development Program To Improve It. British Journal Of Educational Technology, 53(4), 914–931. Https://Doi.Org/10.1111/Bjet.13232

Nazaretsky, T., Cukurova, M., & Alexandron, G. (2022). An Instrument For Measuring Teachers' Trust In Ai-Based Educational Technology. Proceedings Of The 12th International Learning Analytics And Knowledge Conference, 56–66. Https://Doi.Org/10.1145/3506860.3506866

Nazaretsky, T., Mejia-Domenzain, P., Swamy, V., Frej, J., & Käser, T. (2025). The Critical Role Of Trust In Adopting Ai-Powered Educational Technology For Learning: An Instrument For Measuring Student Perceptions. Computers And Education: Artificial Intelligence, 8, 100368. Https://Doi.Org/10.1016/J.Caeai.2025.100368

Nuh, A., Rizan, M., & Sadat, A. M. (2025). Exploring Continued Use Intention Of The Ai Platform Among Students In Indonesia: An Extended Ecm Framework. Interdisciplinary Journal Of Information, Knowledge, And Management, 20, Article 6. Https://Doi.Org/10.28945/5444

Pawlak, J., Teixeira Domingues, J. P., & Oliveira, O. (2025). Demographic And Psychological Determinants Of Students' Attitudes Toward The Use Of Ai Tools In Education. Tqm Journal, 38(3), 611–632. Https://Doi.Org/10.1108/Tqm-12-2024-0520

Phong, P. D., Lan, P. T., & Thanh, T. X. (2025). Linguistic Summarization And Outlier Detection Of Blended Learning Data. Applied Sciences (Switzerland), 15(12), Article 6644. Https://Doi.Org/10.3390/App15126644

Rico-Juan, J. R., Pertegal-Felices, M. L., & Jimeno-Morenilla, A. (2025). Using Explainable Ai To Align Pre-University Profiles With Bachelor's Degree Success. Scientific Reports, 15(1), Article 44373. Https://Doi.Org/10.1038/S41598-025-28179-Z

Salem, M. A. (2025). Bridging Or Burning? Digital Sustainability And Py Students' Intentions To Adopt Ai-Nlp In Educational Contexts. Computers, 14(7), Article 265. Https://Doi.Org/10.3390/Computers14070265

Sarikaya, İ., Öztürk, M., & Özgöl, M. (2023). Development Of Self-Regulation Scale For Middle School Students: Validity And Reliability Study. International Journal Of Assessment Tools In Education, 10(4), 723–740. Https://Doi.Org/10.21449/Ijate.1311600

Schwerter, J., Lauermann, F., Doebler, P., & Fokkema, M. (2025). Putting The Pieces Of The Puzzle Together In Modeling Gendered Educational Choices. International Journal Of Stem Education, 12(1), Article 38. Https://Doi.Org/10.1186/S40594-025-00558-Y

Shaikhanova, A., Kuznetsov, O., Iklassova, K., Tokkuliyeva, A., & Sugurova, L. (2025). Interpretable Predictive Modeling For Educational Equity: A Workload-Aware Decision Support System For Early Identification Of At-Risk Students. Big Data And Cognitive Computing, 9(11), Article 297. Https://Doi.Org/10.3390/Bdcc9110297

Tan, L., Chen, F., & Wei, B. (2024). Examining Key Capitals Contributing To Students' Science-Related Career Expectations And Their Relationship Patterns: A Machine Learning Approach. Journal Of Research In Science Teaching, 61(8), 1975–2010. Https://Doi.Org/10.1002/Tea.21939

Tao, J., Zhao, W., Liu, F., Guo, X., Cheng, N., Guo, Q., Xu, X., & Duan, H. (2024). Cognitive Diagnosis Method Via Q-Matrix-Embedded Neural Networks. Applied Sciences (Switzerland), 14(22), Article 10380. Https://Doi.Org/10.3390/App142210380

Uddin, M. S. (2025). Analyzing And Predicting Student Performance In Discrete Mathematics Using Supervised Learning Algorithms. Computer Applications In Engineering Education, 33(6), Article E70108. Https://Doi.Org/10.1002/Cae.70108

Vanlommel, K., & Pepermans, E. (2021). Validation Of The Teacher Decision-Making Inventory (Tdmi): Measuring Data-Based And Intuitive Dimensions In Teachers' Decision Process. Studia Paedagogica, 26(4). Https://Doi.Org/10.5817/Sp2021-4-3

Waterman, R., Lafving, B., Okar, C., & Jain, N. (2025). A Custom Gpt For Executive Mba Students: A Case Study In Enhancing Learning. Stat, 14(4), Article E70109. Https://Doi.Org/10.1002/Sta4.70109

Yang, X., Xiao, Y., Liu, D., Deng, H., Huang, J., Zhou, Y., Liang, M., Dong, L., Yuan, Z., Yao, J., Guo, W., & Xu, C. (2025). Factors Influencing Adoption Of Large Language Models In Health Care: Multicenter Cross-Sectional Mixed Methods Observational Study. Journal Of Medical Internet Research, 27, Article E84918. Https://Doi.Org/10.2196/84918

Yeadon, W., & Hardy, T. (2024). The Impact Of Ai In Physics Education: A Comprehensive Review From Gcse To University Levels. Physics Education, 59(2), Article 025010. Https://Doi.Org/10.1088/1361-6552/Ad1fa2

Yi, X., Sun, J., & Wu, X. (2024). Novel Feature-Based Difficulty Prediction Method For Mathematics Items Using Xgboost-Based Shap Model. Mathematics, 12(10), Article 1455. Https://Doi.Org/10.3390/Math12101455

Zarei, M., Asgari, M. S., & Ghafoori Adl, N. (2025). A Hybrid Anfis-Gradient Boosting Frameworks For Predicting Advanced Mathematics Student Performance. Iranian Journal Of Fuzzy Systems, 22(6), 183–203. Https://Doi.Org/10.22111/Ijfs.2025.52017.9178

Zhang, D., Wijaya, T. T., Wang, Y., Su, M., Li, X., & Damayanti, N. W. (2025). Exploring The Relationship Between Ai Literacy, Ai Trust, Ai Dependency, And 21st Century Skills In Preservice Mathematics Teachers. Scientific Reports, 15(1), Article 14281. Https://Doi.Org/10.1038/S41598-025-99127-0

Zhao, G., Li, M., Long, Z., & Fan, H. (2025). Unlocking Ai Potential In Primary Mathematics: Teachers' Technological Pedagogical Readiness In China. Acta Psychologica, 260, Article 105558. Https://Doi.Org/10.1016/J.Actpsy.2025.105558

Zhou, Y., Curle, S., & Zou, J. (2025). Mapping Genai Literacy: Disciplinary Differences, Latent Profiles, And Perceptions Among Emi Undergraduates. Journal Of Information Technology Education: Research, 24, Article 44. Https://Doi.Org/10.28945/5675

Zhu, L., You, H., Hong, M., & Fang, Z. (2025). Predictive Insights Into U.S. Students' Mathematics Performance On Pisa 2022 Using Ensemble Tree-Based Machine Learning Models. International Journal Of Educational Research, 130, Article 102537. Https://Doi.Org/10.1016/J.Ijer.2025.102537