Computers & Education: Artificial Intelligence

Exploring the Feasibility and Limits of Generative AI–Based Adaptive Instruction in K–12 Mathematics

--Manuscript Draft--

Manuscript Number:	
Full Title:	Exploring the Feasibility and Limits of Generative AI–Based Adaptive Instruction in K–12 Mathematics
Article Type:	Research Paper
Keywords:	Adaptive education.; K-12.; Artificial intelligence.; Knowledge graphs
Corresponding Author:	Elisa de Oliveira Flemer Inteli BRAZIL
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Inteli
Corresponding Author's Secondary Institution:	
First Author:	Elisa de Oliveira Flemer
First Author Secondary Information:	
Order of Authors:	Elisa de Oliveira Flemer
	Flavia Santoro
Order of Authors Secondary Information:	
Abstract:	The rapid adoption of generative artificial intelligence (GenAI) in education has prompted a renewed debate on the role of the teacher in increasingly automated learning environments. Although adaptive learning systems have long aimed to personalize instruction, their integration with generative models capable of producing explanations, exercises, and feedback remains underexplored—especially in K–12 contexts. This study presents the design and classroom evaluation of an adaptive GenAI-based learning platform that autonomously delivers mathematics instruction using a knowledge-graph architecture to structure conceptual relationships and guide real-time content generation. Following a design-science methodology, the system was developed, deployed, and tested with eighth-grade students in a private school. Quantitative results indicated short-term performance changes on a brief assessment following interaction with the system, along with qualitative evidence of increased engagement and confidence in problem-solving. However, students consistently emphasized the irreplaceable value of human interaction, citing emotional connection and conversational nuance as missing elements of AI-led instruction. Technical observations revealed current limitations in cost, latency, and generative reliability. These findings suggest that GenAI, while capable of adaptive and context-aware teaching, functions most effectively as a complementary rather than autonomous educational agent. The study concludes by outlining the design and pedagogical implications of hybrid human–AI learning models that preserve both personalization and the human dimension of teaching.
Opposed Reviewers:	
Additional Information:	
Question	**Response**

Exploring the Feasibility and Limits of Generative AI–Based Adaptive Instruction in K–12 Mathematics

Authors
Elisa de Oliveira Flemer (corresponding author) – elisa.flemer@sou.inteli.edu.br
Flavia Maria Santoro – flavia@inteli.edu.br

Institute
Inteli – Instituto de Tecnologia e Liderança,
Av. Prof. Almeida Prado, 520 - Butantã, São Paulo - SP, 05508-070, Brazil

Exploring the Feasibility and Limits of Generative AI–Based Adaptive Instruction in K–12 Mathematics

No Author Given

No Institute Given

Abstract. The rapid adoption of generative artificial intelligence (GenAI) in education has prompted a renewed debate on the role of the teacher in increasingly automated learning environments. Although adaptive learning systems have long aimed to personalize instruction, their integration with generative models capable of producing explanations, exercises, and feedback remains underexplored—especially in K–12 contexts. This study presents the design and classroom evaluation of an adaptive GenAI-based learning platform that autonomously delivers mathematics instruction using a knowledge-graph architecture to structure conceptual relationships and guide real-time content generation. Following a design-science methodology, the system was developed, deployed, and tested with eighth-grade students in a private school. Quantitative results indicated short-term performance changes on a brief assessment following interaction with the system, along with qualitative evidence of increased engagement and confidence in problem-solving. However, students consistently emphasized the irreplaceable value of human interaction, citing emotional connection and conversational nuance as missing elements of AI-led instruction. Technical observations revealed current limitations in cost, latency, and generative reliability. These findings suggest that GenAI, while capable of adaptive and context-aware teaching, functions most effectively as a complementary rather than autonomous educational agent. The study concludes by outlining the design and pedagogical implications of hybrid human–AI learning models that preserve both personalization and the human dimension of teaching.

Keywords: Adaptive education. · K-12. · Artificial intelligence. · Knowledge graphs.

1 Introduction

If a student can now ask an algorithm to explain, correct, and even encourage them, what happens to the teacher? The emergence of generative artificial intelligence (GenAI) has transformed this question from hypothetical to immediate in classrooms around the world.

Today, awareness and daily use of generative AI among students are approaching ubiquity. A 2025 Pew Research Center survey found that one in four

U.S. teenagers (26%) had used ChatGPT for schoolwork—twice the share reported in 2023 [39]. At the global level, a large-scale study covering 109 countries estimated that more than 70% of students have already used ChatGPT for academic purposes [1].

Beyond simple adoption, students are increasingly using GenAI as a learning companion. Indeed, qualitative evidence suggests that using such tools for brainstorming, clarifying complex topics and even refining writing often improve understanding and expression rather than replace original thinking [4]. Moreover, a recent meta-analysis of 51 experimental studies found that the use of ChatGPT leads to substantial gains in learning performance ($g = 0.867$) and moderate improvements in higher-order thinking and learning perception [46].

This marks a shift from novelty to evidence-driven transformation, suggesting that GenAI is not only reshaping how students access knowledge but also redefining the boundaries of instruction itself. Meanwhile, a growing body of research [8, 37, 31] highlights the potential of AI to operationalize long-standing educational ideals, like flipped classrooms, project-based learning and, especially, adaptive education.

Originally conceptualized by [41], adaptive education refers to instruction that adjusts contingently to the responses of each learner rather than following a fixed sequence. Contemporary definitions extend this principle through artificial intelligence and data analytics, framing adaptive learning as the use of AI and data analytics to tailor learning experiences to the needs, preferences, and progress of individual learners [21]. SRI Education further defines adaptive systems as those that adjust the path and pace of learning to optimize student outcomes [42]. Across these views, adaptivity encompasses both pedagogical principles and computational mechanisms for individualized self-paced progression.

Decades of research have demonstrated that adaptive learning can significantly improve educational outcomes. Personalized strategies are correlated with reduced dropout rates and increased engagement [13]; adaptive sequencing leads to measurable gain of performance in mathematics [24]; and modality adaptation produces substantial efficiency improvements [38]. Moreover, adaptive interfaces and affective modeling have been shown to strengthen motivation and agency, even when direct achievement gains are modest [34, 20].

Despite this promise, adaptive learning remains disproportionately studied in higher education, where students enjoy greater autonomy and technological access, while K–12 contexts remain underexplored [21, 42]. Moreover, most existing systems are validated only through simulations or dataset-based evaluations rather than real classrooms [49, 51, 34]. The resulting gap between theoretical potential and classroom reality persists, particularly in addressing curriculum alignment, teacher integration, and ethical transparency.

Against this backdrop, the present study advances the field by demonstrating and evaluating the feasibility of a generative-AI-based adaptive learning platform when deployed as the primary instructional interface in a real classroom. Rather than positioning AI as a supplementary tutor, the system autonomously deliv-

ered an entire mathematics micro-unit, generating explanations, exercises, and visual aids through large language and diffusion models. The platform merged generative automation with a knowledge-graph architecture that organized learning content by conceptual precedence, allowing AI to extract key concepts, represent them as interconnected nodes, and generate adaptive content aligned with the progression of each student.

Following a design-science methodology[16], the research proceeds through four stages: (1) identifying the problem of limited personalization in middle-school instruction; (2) designing and implementing an adaptive, GenAI-powered platform grounded in knowledge-graph reasoning; (3) demonstrating it through a quasi-experimental pilot with eighth-grade students; and (4) evaluating both learning outcomes and affective responses. By uniting structured knowledge representation with generative modeling, this work examines the technical, pedagogical, and emotional boundaries of AI-led instruction.

Specifically, the study addresses three research questions:

- **RQ1:** To what extent can a generative AI platform operationally function as a primary instructional interface in a K–12 classroom setting?
- **RQ2:** How do students perceive and emotionally respond to learning with GenAI as their main instructor?
- **RQ3:** What technical and pedagogical limitations emerge when GenAI is deployed in real classrooms?

Through this investigation, the paper contributes empirical evidence to a debate often dominated by speculation. We evaluate the scope and fragility of generative personalization, i.e., its latency, cost, and unpredictability, while aim to shed light in where AI enhances learning and where the irreplaceable presence of the teacher must remain.

The remainder of this paper is structured as follows. Section 2 reviews the theoretical foundations and related work on adaptive learning, student modeling, and curriculum-aligned knowledge representation. Section 3 presents the design-science methodology. Section 4 details the system architecture. Section 5 describes the classroom experiment. Section 6 reports quantitative and qualitative results. Section 7 discusses implications and limitations, and Section 8 concludes with directions for future research.

2 Theoretical Background and Related Work

This section consolidates the main theoretical foundations and recurring research themes identified in recent work on AI-driven adaptive learning for K–12 education. Although adaptive technologies have expanded rapidly in higher education, evidence suggests that K–12 settings remain underexplored and often lack empirical validation [10, 21, 40, 42]. This is a critical gap, since personalization has been repeatedly linked to improvements in engagement and retention [21], and is frequently proposed as a strategy to mitigate dropout patterns associated with low motivation [13].

2.1 Adaptive Learning as Algorithm-Driven Personalization in K–12

Adaptive learning systems aim to tailor instruction in response to learner progress, typically through continuous measurement of performance and dynamic selection of content, difficulty, or learning paths. The conceptual roots of automated instruction trace back to early teaching machines [41]. Contemporary systems extend this vision through machine learning models that update learner state estimates online and use them to personalize instructional decisions [21, 40].

Despite the intuitive appeal of adaptivity, K–12 imposes distinctive constraints: learners often require stronger scaffolding, more explicit motivational support, and closer alignment with formal curricula. In addition, core-subject coverage and curricular standards matter more directly in this population. As computational thinking becomes integrated into compulsory education frameworks worldwide, including Europe [5], the United States [33], and Brazil [6], adaptive systems increasingly need to provide transparent, curriculum-aligned instruction rather than only predicting performance.

2.2 Student Modeling Paradigms

Student modeling is the analytic foundation of adaptivity, enabling systems to infer what a learner knows, how that knowledge evolves, and what content should come next. The literature converges on three major paradigms—knowledge tracing, cognitive diagnosis, and latent embedding approaches—with a growing trend toward hybridization to combine predictive power with pedagogical interpretability [22, 36].

Knowledge Tracing Knowledge tracing (KT) models a learner's evolving mastery over time, typically using sequential interaction data. Traditional and Bayesian KT approaches remain relevant when interpretability is required. For example, knowledge proficiency tracking integrates learning and forgetting dynamics using explicit update rules grounded in learning curve and forgetting curve theories [18, 32, 11]. Bayesian Knowledge Tracing is also used to support explainable recommendations by estimating concept mastery with interpretable parameters (e.g., learning, slip, guess) [45].

Recent work increasingly favors deep KT models, especially Transformer-based architectures that leverage attention mechanisms for sequence modeling. SAINT applies an encoder–decoder structure to represent exercise sequences and student responses [30], while CL4KT introduces contrastive learning to improve robustness and reduce reliance on external concept tags [23]. MonaCoBERT adapts BERT-style modeling with monotonic and convolutional attention, incorporating difficulty signals inspired by Classical Test Theory [22, 7]. Some models explicitly target interpretability by incorporating psychometric constructs such as Item Response Theory [36, 12]. Memory-augmented neural KT remains relevant as well, with architectures such as DKVMN encoding concept states through key–value memory updates [44].

Overall, deep KT tends to yield strong predictive performance, but often raises explainability concerns in educational contexts that demand accountability and human-understandable reasoning [23, 30, 22].

Cognitive Diagnosis Models Cognitive Diagnosis Models (CDMs) aim to infer fine-grained mastery over skills or concepts based on response patterns, offering more interpretable diagnoses than many deep KT systems. Recent CDMs incorporate Bayesian uncertainty modeling (useful under sparse data) [3], graph-structured representations for concept-aware reasoning [25, 17], and misconception-level analysis [47]. Hybrid approaches can integrate behavior and preference signals into diagnosis; for instance, PreferenceCD augments DINA-style diagnosis with both a Q-matrix and an M-matrix, supported by topic modeling (TF–IDF + LDA), to connect assessment performance with reading behavior [19].

Beyond probabilistic and graph-based variants, fuzzy and multi-skill formulations address partial mastery and compensatory reasoning. FuzzyCDF leverages SI-GAM aggregation to handle polytomous responses and flexible skill combinations [15]. Lightweight regression-based approaches also remain useful when systems require simple, fast diagnosis under limited computational constraints [48].

Latent Embedding and Matrix Factorization Latent embedding methods (including matrix factorization) model learners and items in a shared latent space to support prediction and recommendation, often with less manual feature engineering than skill-tagged approaches. Wse-MF illustrates this strategy by learning latent student and exercise vectors while incorporating student- and item-specific weighting to reduce overconfident predictions on difficult items for low-ability learners [43]. Other work uses embeddings to infer learning style proxies and recommend resources accordingly [35]. Hybrid models also exist: RCES combines matrix factorization with sequential neural architectures (Bi-LSTM / Transformer) to model test preparation trajectories and adapt recommendations over time [29].

These approaches scale well and reduce reliance on expert labeling, but often provide limited pedagogical transparency compared to explicitly skill-based diagnosis.

2.3 Domain Knowledge Representation for Curriculum-Aligned Adaptation

Personalization requires not only a student model but also a representation of domain knowledge that supports inference, prerequisites, and sequencing. The literature reveals two recurring axes: whether knowledge is represented as a graph vs. non-graph structure, and whether the representation is expert-defined vs. data-driven.

Graph-Based Representations Graph-based systems encode dependencies between concepts and can support prerequisite-aware sequencing. Expert-defined graphs offer strong curricular alignment but require intensive manual work. Examples include concept graphs extracted from curriculum documents with teacher-defined importance weighting [50], multilevel knowledge graphs that compute concept importance via structural and semantic features [25], and curriculum-anchored platforms based on official national structures [26].

Data-driven graph construction improves scalability by inferring relationships from behavioral logs and semantic features. Presage constructs multilevel semantic graphs from open educational resources [2], while DMP_AI uses language modeling and heterogeneous graphs to connect resources and learners [49]. Graph-temporal fusion approaches combine structural and temporal signals to track evolving mastery [17]. Systems may also infer implicit graphs through attention weights and co-occurrence signals [30, 36].

Non-Graph-Based Representations Non-graph representations include Q-matrices, topic labels, and latent vectors. Q-matrix-driven systems remain common because they are interpretable and align with formal curricula [43, 18, 45]. Other systems rely on structured labels or interface-driven representations without modeling concept dependencies directly [20, 34, 48]. Generative approaches may use problem categories and equation extraction without explicit concept graphs [52].

Across the literature, manual labeling persists as a bottleneck even as models become more sophisticated, motivating research that automates curriculum-aligned knowledge extraction while preserving pedagogical meaning [23, 49, 19].

2.4 Evaluation Practices and Remaining Gaps

Most studies emphasize offline evaluation on benchmark datasets, reporting predictive metrics such as AUC or RMSE [23, 22, 3, 30, 43]. In-situ deployments exist but are often exploratory and may lack control groups or longitudinal follow-up [49, 26, 38, 34]. Only a small portion of the literature reports controlled experiments with clear causal attribution, including large-scale A/B tests and randomized controlled trials [29, 24].

Across the field, several gaps persist: reliance on manual knowledge engineering, limited explainability in high-performing deep models, scarce classroom-grade validation, and inconsistent measurement of motivational and behavioral outcomes. Addressing these gaps motivates research that combines curriculum-aligned structure, explainable learner modeling, and robust evaluation designs suitable for real K–12 environments [40, 21, 42].

3 Methodology

This study follows a *design-science research* (DSR) methodology, in which a purposeful artifact is constructed to address a relevant problem and evaluated in its

intended context [16]. The methodological objective is to investigate the feasibility, learner experience and limitations of *autonomous, generative instruction* in K–12 mathematics through the design and deployment of a novel adaptive learning artifact.

3.1 Problem context and research objective

The research is motivated by a persistent limitation in middle-school instruction: while curricula demand mastery of heterogeneous concepts, teachers operate under time and attention constraints that limit individualized explanation, pacing, and feedback. Recent advances in generative AI suggest the possibility of automating portions of instructional delivery, yet the viability of such systems as primary teaching resources in K–12 classrooms remains largely unexplored.

Hence, the study addresses the following objective: to examine whether a generative-AI-based platform can autonomously deliver curriculum-aligned mathematics instruction and to identify the pedagogical, technical, and experiential boundaries that emerge when such a system is used in a real classroom.

3.2 Artifact as the unit of design

In accordance with DSR principles, the central contribution of this research is the artifact itself. The artifact is an adaptive learning platform that integrates a curriculum-anchored knowledge graph and a generative-model agent system capable of producing explanations, exercises, feedback, and visual aids at runtime.

The artifact operationalizes the notion of *autonomous teaching* by removing real-time teacher mediation during instruction. Adaptivity is achieved through a closed feedback loop: student interactions update the learner model, which conditions content generation and sequencing constrained by the knowledge graph. The artifact was intentionally scoped to a middle-school mathematics micro-unit to allow controlled observation of system behavior and learner response.

3.3 Design process and simulation-based DSR cycles

Following Hevner et al. [16], iterative design cycles were conducted *prior* to classroom deployment, exclusively through simulation-based testing. These cycles focused on refining the internal logic of the artifact rather than learner outcomes.

During this phase, the system was exercised using synthetic student profiles and scripted interaction traces to simulate diverse learning behaviors (e.g., rapid mastery, persistent misconceptions, off-topic questions). Iterative refinements targeted:

- Prompt structure and constraints for explanation, exercise, and feedback generation;
- Coordination between planning and execution agents to preserve pedagogical coherence across lesson blocks;

– Failure modes such as hallucinated prerequisites, overlong explanations, or inconsistent visual outputs;

These simulation cycles followed a build–simulate–inspect–adjust pattern, enabling controlled iteration without exposing real students to unstable behaviors. Importantly, no classroom-based iterations were conducted; the simulation phase served as the sole iterative design cycle prior to deployment.

3.4 Demonstration through classroom deployment

After stabilization through simulation, the artifact was demonstrated once in its target environment: an eighth-grade mathematics classroom in a private school. In DSR terms, this stage constitutes a *demonstration* rather than an iterative field deployment [16].

The classroom application was designed as a short pilot in which students interacted with the system as the primary instructional resource for a curriculum-aligned micro-unit. Teacher involvement was limited to logistical support and participant selection, ensuring that instruction was delivered autonomously by the artifact. The purpose of this deployment was to expose the system to authentic classroom conditions and observe its performance under real constraints such as time pressure, connectivity variability, and heterogeneous learner behavior.

3.5 Evaluation strategy

Evaluation in this study focuses on assessing the artifact's utility, feasibility, and experiential impact, rather than on establishing generalizable learning effects. Three complementary evidence types were collected:

1. **Learning evidence:** Pre- and post-instruction assessments provided an initial indication of short-term learning change under autonomous instruction.
2. **Experiential evidence:** Student feedback and researcher observations captured perceptions of clarity, engagement, frustration, trust in explanations, and perceived absence or presence of human teaching qualities.
3. **Operational evidence:** System logs documented latency, request volume, error cases, and inference cost, offering insight into the practical viability of generative instruction in middle-school settings.

Detailed descriptions of participants, instruments, procedures, and analysis methods are provided in a dedicated section to maintain a clear separation between methodological framing and experimental protocol.

3.6 Scope and validity

Given the single classroom deployment and exploratory intent, the study does not aim to produce population-level causal claims. Instead, it contributes design knowledge by clarifying which aspects of autonomous generative instruction are currently feasible, fragile, or pedagogically insufficient in K–12 contexts. The findings are intended to inform future iterations of hybrid human–AI instructional models and to guide subsequent, larger-scale evaluations.

4 System Architecture and Design

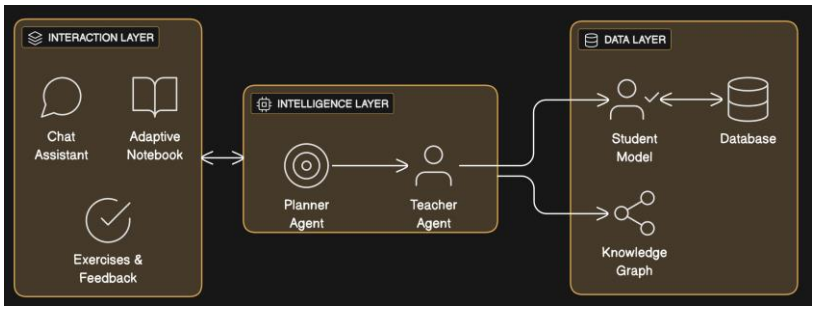

Fig. 1: Three-layer architecture of the adaptive learning system, showing the interaction, intelligence, and data layers connected through generative feedback loops.

The system was designed as an AI-first adaptive learning environment, merging generative modeling and graph-based reasoning to deliver personalized instruction on scale. Its central premise is that both knowledge representation and instructional generation can be automated through large language models (LLMs) while remaining grounded in explicit curricular structure. By combining symbolic organization (the knowledge graph) with generativity (text and image synthesis), the platform creates a dynamic feedback loop between what a student knows and how the system teaches.

4.1 Data Layer

The data layer provides the structural foundation for adaptivity by maintaining both a domain-level representation of knowledge and a longitudinal record of learner states. It serves two main purposes: (i) organizing curricular content into a machine-interpretable knowledge graph, and (ii) storing dynamic learner data that capture progress, mastery, and evolving strengths and weaknesses.

Knowledge Graph Construction Traditional adaptive systems are heavily dependent on expert-defined taxonomies and manually labeled relationships between concepts and exercises. Such manual encoding, while precise, is prohibitively time-consuming and limits scalability across curricula. To address this bottleneck, the present system leverages a large language generative model (Google Gemini 2.5 Flash) to automatically extract both *conceptual* and *assessment* elements directly from instructional materials.

The model operates through schema-constrained prompts that request structured outputs defining:

- **Concept nodes** — each containing a canonical name, a concise summary, and short teaching instructions;
- **Prerequisite relations** — directed links representing the conceptual dependencies between topics (*e.g.*, "Fractions" → "Decimals" → "Percentages");
- **Exercise associations** — question nodes attached to relevant concepts via HAS_EXERCISE edges, containing item type, expected response, and explanatory feedback.

Graph Functionality Within instruction, the knowledge graph functions as the central reasoning substrate. It supports:

1. **Navigation** — determining which concepts are available for study based on completed prerequisites;
2. **Prerequisite enforcement** — ensuring learners only advance after demonstrating mastery in antecedent nodes;
3. **Semantic retrieval** — enabling the system to locate conceptually similar nodes or exercises using text embeddings (Gemini Embedding 001).

Student Model Parallel to the knowledge graph, the system maintains a student model within a relational database that tracks cognitive and behavioral states over time. Each student profile includes identifiers for the concepts covered, mastery scores, and two evolving vectors: *strengths* (areas of demonstrated proficiency) and *weaknesses* (recurring misconceptions or difficulties).

After each learning interaction—such as completing an exercise, marking a block as understood, or participating in a tutoring chat—the model updates these attributes automatically. The update logic is also mediated by the LLM, which analyzes conversation transcripts and activity logs to refine the learner profile. The result is a continuously adapting representation of both the domain and the individual learner, establishing the foundation upon which all higher-level adaptive behaviors operate.

4.2 Intelligence Layer

The intelligence layer constitutes the adaptive core of the platform, where large language models (LLMs) generate, interpret, and continuously refine the learning process. It combines generative planning, multimodal content creation, and analytical adaptation into a unified reasoning pipeline that transforms static curricular materials into interactive, learner-specific instruction.

Generative Planning, Multi-Level Sequencing, and Exercise Selection A major design challenge in large-scale generative instruction is the limited context window of current LLMs. To mitigate information loss across extended curricular sequences, the system adopts a *multi-level planning* strategy that separates high-level pedagogical design from fine-grained block generation.

At the upper level, a **Planner Agent** produces a structured lesson blueprint for each concept retrieved from the knowledge graph. Each plan is a sequence of typed pedagogical steps—*context*, *highlight*, *misconception*, *explanation*, and *question*—corresponding to specific instructional purposes such as activation of prior knowledge, definition emphasis, correction of misconceptions, elaboration, and formative assessment.

Exercises are integral to this process. Each concept node in the knowledge graph is linked to multiple exercise nodes annotated with difficulty, type, and learning objective. During lesson planning, the Planner Agent queries the graph to retrieve the available exercises and selects those most relevant to the current student profile. Selection is based on the learner's profile, ensuring that the questions are not redundant or disproportionately difficult. The agent also determines the order in which exercises will appear within the lesson sequence to balance reinforcement and challenge.

At the lower level, a **Teacher Agent** executes each planned step independently, generating the actual textual and visual materials presented to students, including the contextual framing and explanatory scaffolds that accompany exercises. Through this dual-layer planning, the system integrates generative instruction with structured assessment, producing lessons that continuously respond to individual learner trajectories.

Multimodal Content Generation The system employs distinct models for specialized generation tasks, forming a coordinated generative ensemble:

- **Text Generation:** The main LLM (Google Gemini 2.5 Flash) is responsible for producing natural-language explanations, examples, and formative feedback. It follows structured prompts that include the pedagogical type of the block and the current learner profile to ensure contextual relevance and age-appropriate phrasing.
- **Image Generation:** A lightweight diffusion-based model (OpenAI GPT-4.1 Mini) transforms automatically generated image prompts into didactic illustrations such as geometric diagrams or visual analogies. These images are rendered asynchronously and integrated into the corresponding instructional blocks.

4.3 Interaction Layer

The interaction layer transforms the system's generative and analytical processes into tangible learning experiences. Adaptation unfolds directly within the student-facing interface—the adaptive notebook—which acts as both the medium of instruction and the mechanism for continuous feedback.

The notebook presents instructional content as sequential learning blocks derived from the intelligence layer's lesson plan. Each block—*context*, *highlight*, *explanation*, *misconception*, or *question*—is generated independently and revealed progressively. This progressive disclosure strategy manages the cognitive load

and encourages reflection, as students must mark each block as understood before accessing the next. Every interaction—time spent, exercise attempts or clarification requests—feeds back into the learner model, allowing the system to dynamically adjust pacing and emphasis.

Embedded in this interface is a contextual chat assistant that enables students to ask questions about any block in real time. The system sends the block content and conversation history to the language model, which generates concise, supportive explanations that guide reasoning rather than simply providing answers. These dialogs emulate personalized tutoring while simultaneously generating valuable data on the learner's understanding. The model periodically analyzes these exchanges, updating the student's strengths and weaknesses profile.

When students complete the exercises, the system performs semantic answer validation, generating immediate explanatory feedback. Correctness judgments and interaction patterns contribute to mastery estimates for each concept node in the knowledge graph. Once mastery surpasses a defined threshold, new concepts are unlocked, creating a personalized learning trajectory.

Through these mechanisms, adaptation becomes intrinsic to interaction. The notebook continuously transforms student behavior into data that inform how the next block, question, or explanation is generated. This closed feedback loop ensures that instruction evolves with each learner's progress, allowing generative and analytical intelligence to manifest as a fluid, personalized learning experience.

5 Experiment

This section describes the classroom experiment conducted to demonstrate and evaluate the artifact in its intended educational context. Consistent with the design-science methodology adopted in this study, the experiment was conceived as a pilot deployment rather than a large-scale controlled trial.

The pilot was carried out in a partner school with a group of six eighth-grade students (three girls and three boys) during a regular math period. To capture a range of learner profiles, participants were selected by their teacher and categorized as two low-performing, two mid-performing, and two high-performing students based on their prior term grades. The activities addressed two topics from the regular curriculum—*Perimeter* and *Unit and Measurement Conversions*—chosen for their concreteness and suitability for visual and computational explanation.

5.1 Procedure

The session lasted approximately 90 minutes and consisted of three sequential phases:

1. **Pre-test:** A three-question assessment evaluated students' prior understanding of the target topics.

2. **Intervention:** Students interacted with the adaptive generative-AI platform for roughly one hour. Students worked individually or in pairs and were encouraged to navigate freely, while researchers observed engagement and recorded behavioral reactions.
3. **Post-test:** A second three-question instrument, structurally equivalent to the pre-test, measured short-term learning gains.

Quantitative analysis compared the results of the pre- and post-test to estimate improvement. Qualitative observations and students' spontaneous comments were examined to identify trends in engagement, curiosity, and frustration, without formal coding procedures.

6 Results

This section presents the results of the exploratory classroom deployment of the platform. Given the small sample size and pilot nature of the study, the analyses are descriptive and interpretive rather than inferential. Quantitative results focus on changes in performance between pre- and post-tests, while qualitative results examine behavioral patterns, student perceptions, and technical observations recorded during the session.

6.1 Quantitative Analysis

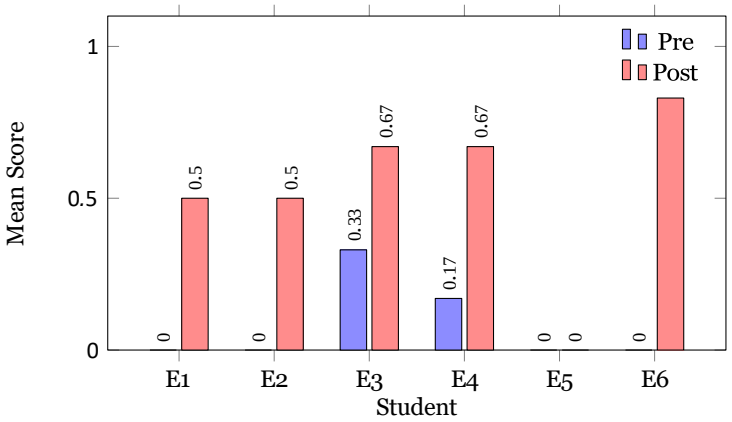

Fig. 2: Pre- and post-test mean scores per student.

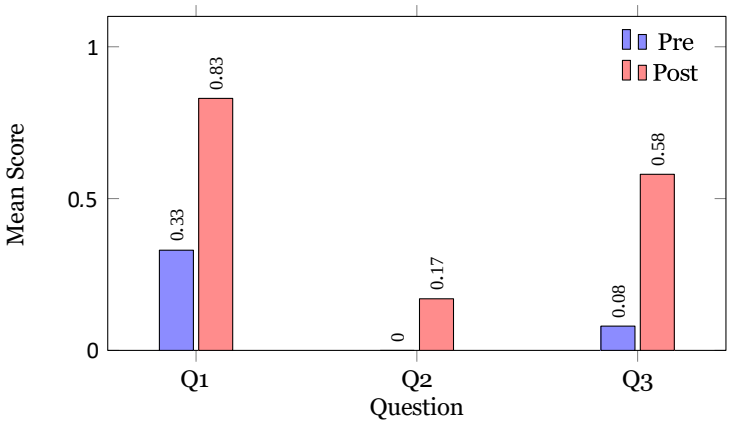

Fig. 3: Mean pre- and post-test scores per question.

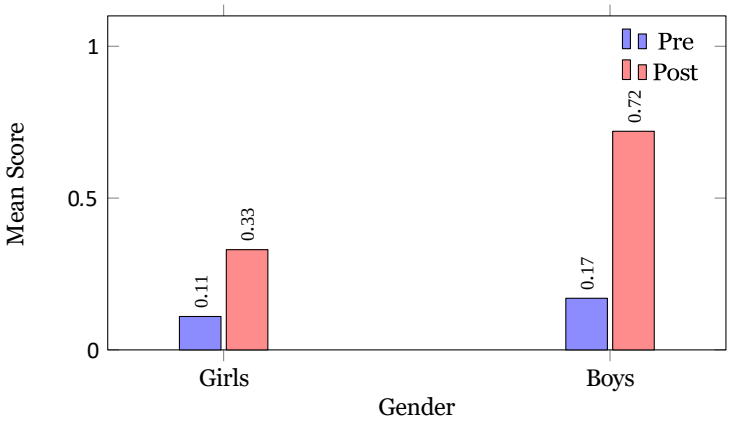

Fig. 4: Mean pre- and post-test scores by gender.

Figure 2 presents the pre- and post-test mean scores for each of the six students in the experimental group. All participants completed both assessments. Overall performance increased from a mean score of 0.08 in the pre-test to 0.48 in the post-test, indicating a substantial improvement after interacting with the system.

At the individual level, heterogeneous learning trajectories can be observed. Students E1, E2, and E6 entered the intervention with zero scores in the pre-test and showed notable gains in the post-test, particularly E6, who achieved the highest post-test mean (0.83). Students E3 and E4, who already demonstrated partial prior knowledge, also improved, although their relative gains were smaller in magnitude. Student E5 maintained a zero score in both phases, suggesting that

the intervention was insufficient to overcome initial difficulties for this learner within the short duration of the session.

These observations suggest that the platform may lower participation barriers for students with low initial performance, encouraging task attempts within a short intervention, although its impact is uneven and likely mediated by individual differences in engagement, prior knowledge, and interaction patterns.

Figure 3 analyzes performance aggregated by question. The largest improvement occurred in Question 1, whose mean score increased from 0.33 to 0.83. This question targeted the direct application of a known formula, which is consistent with the type of procedural guidance emphasized by the platform's generated explanations and visual supports. Question 3, which a step of formula application with unit conversion, also showed a substantial increase, from 0.08 to 0.58, indicating partial conceptual consolidation after the intervention.

In contrast, Question 2 drew on content introduced in prior weeks of the curriculum, specifically involving circumference calculation, and remained challenging, with a post-test mean of only 0.17. This outcome indicates that some conceptual difficulties persisted after the intervention, which may reflect higher abstraction demands, limited instructional scaffolding, or constraints in the generated explanations. The uneven performance across questions highlights the need for content-specific analysis in adaptive learning research, as instructional strategies may differentially support procedural and conceptual topics.

Figure 4 presents the results disaggregated by gender. Pre-test scores were similarly low for both groups (girls 0.11, boys 0.17), indicating comparable starting points. In the post-test, however, boys achieved a higher average score (0.72) than girls (0.33). While this difference should not be overinterpreted given the very small sample size, it suggests that interaction styles or engagement patterns may have differed during the session.

Taken together, the quantitative results indicate that the platform supported measurable learning progress for most participants, particularly in terms of task completion and confidence in attempting answers. However, performance gains were uneven across students and questions, highlighting both the potential and the limitations of fully generative instruction in short, real-world deployments.

6.2 Qualitative Analysis

Qualitative observations and post-session interviews provide additional context for interpreting the quantitative findings. During the pre-test, four out of six students either left questions unanswered or erased their work after initial attempts, indicating low confidence and reluctance to commit to answers. In the post-test, this behavior was observed only once, suggesting increased willingness to engage with problems, even when uncertainty remained.

All participants reported regular use of generative AI tools such as ChatGPT for studying, primarily as tutoring aids or for reviewing content rather than as substitutes for classroom instruction. Students emphasized that they value face-to-face explanations and the ability to ask spontaneous questions to teachers, particularly when dealing with confusion or frustration. Several noted

that teacher-led classes can sometimes feel slow or repetitive, but still considered the teacher essential for emotional support and clarification.

Students also reported frequent use of online educational resources, including recorded lessons provided by their textbook platform, which they described as useful for reinforcement and self-paced review. In this context, the generative-AI platform was perceived as more similar to a supplementary learning tool than a replacement for formal instruction.

Feedback on the system itself was generally positive, particularly regarding the clarity of step-by-step explanations and the availability of immediate feedback. However, students consistently pointed out usability issues. Long loading times between activities disrupted the learning flow, excessive explanatory text led to fatigue, and some generated images were described as confusing or imprecise. These limitations occasionally reduced engagement and required intervention from the researchers.

Despite these issues, most participants stated that they would use the platform again, especially for review or exam preparation. Notably, none of the students could suggest modifications that would make them prefer the system as their sole learning resource, reinforcing the perception that human instruction remains central to their learning experience.

From a technical perspective, system logs revealed that the one-hour session generated 144 API requests for six students, resulting in a cost of USD 14.23 associated primarily with image generation. Internet instability and inference latency were also observed, leading researchers to pair students at shared devices to reduce concurrent requests. While this mitigated technical failures, it limited individual adaptivity and illustrates how infrastructural constraints can directly shape pedagogical outcomes.

Overall, the qualitative findings complement the quantitative results by highlighting increased engagement and confidence, while also exposing practical, emotional, and technical barriers that constrain the viability of fully autonomous AI-led instruction in K–12 settings.

7 Discussion

This study set out to explore the feasibility of generative AI as an autonomous instructional agent in a K–12 mathematics classroom. The findings contribute to ongoing debates in the literature on adaptive learning, generative instruction, and the role of human teachers by providing empirical evidence from a real classroom deployment rather than simulations or higher-education contexts.

Engagement, confidence, and learning behavior

One of the most salient outcomes of the intervention was the increase in students' willingness to engage with mathematical problems. During the pre-test, most students either left questions blank or erased their work, a behavior that almost disappeared in the post-test. This shift suggests increased confidence and reduced

fear of making mistakes—conditions that are widely recognized as essential for learning, particularly in mathematics education.

Prior work on adaptive learning systems reports similar behavioral effects, where personalization and immediate feedback increase student engagement even when learning gains are modest [13, 34, 20]. The present results align with these findings, indicating that generative explanations and on-demand feedback can lower barriers to participation and encourage persistence. However, as observed in other classroom-based studies [40, 49], such engagement gains do not uniformly translate into strong conceptual mastery for all learners, particularly within short interventions.

These results address RQ1 by showing that a generative AI platform can support sustained learning activity and task completion under autonomous instructional conditions, but not consistently or comprehensively across students and content types.

Student perception and the role of the teacher

Despite recognizing the usefulness of the system, students consistently framed generative AI as a supplementary tool rather than a replacement for human instruction. This perception echoes a growing body of research showing that students value AI for clarification, review, and efficiency, but continue to associate meaningful learning with human interaction [8, 9, 14].

In line with studies in both higher education and K–12 settings, participants emphasized the importance of conversational nuance, emotional support, and real-time responsiveness—qualities that current AI systems struggle to reproduce [37, 40]. Although generative explanations were often perceived as clear, students reported that they lacked the adaptive empathy and situational awareness that characterize effective teaching.

These findings provide a direct response to **RQ2**: while students are comfortable interacting with generative AI and perceive it as helpful, they do not emotionally or pedagogically accept it as a primary instructor. This reinforces the argument that AI's educational value lies in augmentation rather than substitution.

Pedagogical and technical limitations of autonomous instruction

The experiment also surfaced several pedagogical and technical constraints that limit the viability of fully autonomous generative instruction. Uneven performance across questions suggests that certain concepts—particularly those requiring abstraction or cumulative reasoning—are not adequately supported by generic generative explanations. Similar limitations have been reported in prior work on AI-generated exercises and explanations, where surface-level fluency masks deeper conceptual gaps [52, 27, 28].

From a usability perspective, excessive text, latency, and visually inconsistent image generation disrupted the learning flow. These issues are consistent

with findings from recent classroom deployments of AI-supported systems, which highlight that even small delays or unclear visuals can significantly affect engagement in K–12 settings [34, 40]. The need to pair students to mitigate latency further constrained individual adaptivity, underscoring how infrastructural limitations can directly shape pedagogical outcomes.

When extrapolated to a typical instructional schedule (6 hours per day, 20 days per month, over 8 months), the per-student API cost of approximately USD 2,276.80 per year represents a non-trivial share of overall education spending. By comparison, government expenditure per student in primary education in Brazil is roughly USD 3,745 per year. Thus, API costs alone could approach 61% of Brazil's per-student annual spending—before accounting for staffing, facilities, and other operational expenditures.

These observations collectively address **RQ3**, revealing that current generative AI systems face intertwined pedagogical, technical, and economic barriers that prevent their deployment as self-contained instructional solutions in K–12 education.

Implications for adaptive learning research

Taken together, the findings challenge narratives that frame generative AI as a near-term replacement for teachers. Instead, they support a growing consensus in the literature that the future of AI in education lies in hybrid human–AI models [21, 42, 40]. In such models, generative systems can enhance personalization, provide alternative explanations, and support formative feedback, while teachers retain responsibility for motivation, emotional support, and pedagogical judgment.

From a design-science perspective, the study contributes design knowledge by clarifying which aspects of autonomous instruction are currently feasible (engagement, immediate feedback, adaptive sequencing) and which remain fragile (deep conceptual scaffolding, affective interaction, and cost efficiency). These insights can inform future work that integrates generative AI into teacher-centered workflows rather than attempting full instructional automation.

In summary, the results suggest that generative AI is not yet ready to function as a standalone teacher in K–12 classrooms. Its current strength lies in complementing human instruction—supporting review, experimentation, and personalization—while preserving the relational core of teaching that students continue to value.

8 Conclusion

This study examined whether a generative-AI-based platform could function as a primary instructional resource in a middle-school mathematics classroom. The results indicate that, while the system supported increased student engagement and confidence in attempting exercises, it did not replicate the relational, contextual, and affective dimensions of human teaching. Students valued the adaptive

explanations and immediate feedback, but consistently expressed a preference for the presence of a teacher to clarify doubts, provide encouragement, and respond to nuanced questions.

These findings must be interpreted in light of the exploratory scope of the study. The classroom deployment involved a small group of six students from a single school and focused on a short instructional sequence covering only two mathematics topics. In addition, the absence of a control group limits direct comparison with traditional instruction, and the short duration of the intervention restricts conclusions about long-term learning effects. Qualitative insights were derived from informal observations and student feedback rather than from structured coding procedures. As such, the results are not intended to support broad generalizations, but rather to provide situated evidence about feasibility and learner experience.

Technical and infrastructural constraints further shaped both the outcomes and their interpretation. High inference costs, network latency, and limited server capacity reduced usability and constrained individual adaptivity, occasionally requiring students to share devices. Moreover, occasional inaccuracies in generated explanations and visual aids highlight the need for stronger pedagogical control and validation mechanisms when deploying generative systems in educational contexts. These limitations underscore the gap between current generative AI capabilities and the demands of reliable, large-scale classroom use—particularly in middle-school settings.

Taken together, the findings suggest that generative AI is not yet ready to operate as a fully autonomous instructional agent in K–12 education. Its current strengths lie in complementing human instruction by supporting review, offering alternative explanations, and enabling adaptive feedback, rather than replacing the teacher's role. Future research should therefore prioritize hybrid human–AI instructional models, larger and controlled classroom studies, longitudinal evaluation of learning outcomes, and technical advances aimed at reducing latency, cost, and generative inconsistency. Sustainable progress in this domain will depend on aligning technical efficiency with pedagogical integrity while preserving the human connection at the center of learning.

Acknowledgments

During the preparation of this work, the authors used ChatGPT-4o to assist with English translation and language readability. The authors subsequently reviewed and edited the content as necessary and assume full responsibility for the final publication.

References

1. Amoah, A., Asiama, R.K., Kwablah, E.: Chatgpt early usage among students: A global evidence of determinants. Development and Sustainability in Economics and Finance 7, 100065 (Sep 2025). https://doi.org/10.1016/j.dsef.2025.100065, http://dx.doi.org/10.1016/j.dsef.2025.100065

2. Bazouzi, A., Le Capitaine, H., Miklos, Z., Foursov, M.: Precedability prediction between open educational resources. In: Proceedings of the 2024 International Conference on Information Technology for Social Good. p. 386–393. GoodIT '24, ACM (Sep 2024). https://doi.org/10.1145/3677525.3678686
3. Bi, H., Chen, E., He, W., Wu, H., Zhao, W., Wang, S., Wu, J.: Beta-cd: A bayesian meta-learned cognitive diagnosis framework for personalized learning. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 37, pp. 5018–5026 (2023). https://doi.org/10.1609/aaai.v37i4.25629
4. Black, R.W., Tomlinson, B.: University students describe how they adopt AI for writing and research in a general education course. Sci. Rep. **15**(1), 8799 (Mar 2025)
5. Bocconi, S., Inamorato dos Santos, A., Chioccariello, A., Cachia, R., Kampylis, P., Giannoutsou, N., Dagiene˙, V., Punie, Y., Wastiau, P., Engelhardt, K., Earp, J., Horvath, M., Jasute˙, E., Malagoli, C., Masiulionyte˙-Dagiene˙, V., Stupuriene˙, G.: Reviewing computational thinking in compulsory education: State of play and practices from computing education. Research report, European Commission: Joint Research Centre, Luxembourg (2022). https://doi.org/10.2760/126955
6. Brasil, Ministério da Educação: Base nacional comum curricular (2018), http://basenacionalcomum.mec.gov.br
7. Brookhart, S.M., Nitko, A.J.: Educational Assessment of Students. Pearson, Boston, MA, 8th edn. (2018)
8. Chan, C.K.Y., Hu, W.: Students' voices on generative AI: perceptions, benefits, and challenges in higher education. Int. J. Educ. Technol. High. Educ. **20**(1) (Jul 2023)
9. Chan, C.K.Y., Tsi, L.H.Y.: Will generative AI replace teachers in higher education? a study of teacher and student perceptions. Stud. Educ. Eval. **83**(101395), 101395 (Dec 2024)
10. Cukurova, M., Miao, X., Brooker, R.: Adoption of artificial intelligence in schools: Unveiling factors influencing teachers' engagement. In: Artificial Intelligence in Education. AIED 2023. Lecture Notes in Computer Science (2023)
11. Ebbinghaus, H.: Memory: A contribution to experimental psychology. Teachers College Press (1913). https://doi.org/10.1037/10011-000
12. Embretson, S.E., Reise, S.P.: Item Response Theory. Psychology Press (Sep 2013). https://doi.org/10.4324/9781410605269, http://dx.doi.org/10.4324/9781410605269
13. Ferreira, S.G., Ribeiro, G., Tafner, P.: Abandono e evasão escolar no brasil. Nota Técnica NT 2022/1, Instituto Mobilidade e Desenvolvimento Social (Imds), Rio de Janeiro (june 2022), https://imdsbrasil.org/publicacao/abandono-e-evasao-escolar-no-brasil/
14. Freeman, J.: Provide or punish? students' views on generative ai in higher education. Higher Education Policy Institute (2024)
15. Fumanal-Idocin, J., Takáč, Z., Horanská, , da Cruz Asmus, T., Dimuro, G., Vidaurre, C., Fernandez, J., Bustince, H.: A generalization of the sugeno integral to aggregate interval-valued data: An application to brain computer interface and social network analysis. Fuzzy Sets and Systems **451**, 320–341 (2022). https://doi.org/10.1016/j.fss.2022.10.003
16. Hevner, A.R., March, S.T., Park, J., Ram, S.: Design science in information systems research. MIS quarterly pp. 75–105 (2004)
17. Huang, M., Wei, T.: Gtfn: Knowledge tracing model based on graph temporal fusion networks. Int. J. Data Warehous. Min. **20**(1), 1–17 (2024). https://doi.org/10.4018/IJDWM.345406

18. Huang, Z., Liu, Q., Chen, Y., Wu, L., Xiao, K., Chen, E., Ma, H., Hu, G.: Learning or forgetting? a dynamic approach for tracking the knowledge proficiency of students. ACM Trans. Inf. Syst. **38**(2) (2020). https://doi.org/10.1145/3379507
19. Jiang, P., Wang, X.: Preference cognitive diagnosis for student performance prediction. IEEE Access **8**, 219775–219787 (2020). https://doi.org/10.1109/ACCESS.2020.3042775
20. Jiao, X., Yu, X., Peng, H., Zhang, X.: A smart learning assistant to promote learning outcomes in a programming course. Int. J. Softw. Sci. Comput. Intell. **14**(1), 1–23 (Nov 2022). https://doi.org/10.4018/IJSSCI.312557
21. Kabudi, T., Pappas, I., Olsen, D.H.: Ai-enabled adaptive learning systems: A systematic mapping of the literature. Computers and Education: Artificial Intelligence **2**, 100017 (2021). https://doi.org/10.1016/j.caeai.2021.100017
22. Lee, U., Park, Y., Kim, Y., Choi, S., Kim, H.: Monacobert: Monotonic attention based convbert for knowledge tracing. In: Generative Intelligence and Intelligent Tutoring Systems: 20th International Conference, ITS 2024, Thessaloniki, Greece, June 10–13, 2024, Proceedings, Part II. pp. 107–123. Springer-Verlag, Berlin, Heidelberg (2024). https://doi.org/10.1007/978-3-031-63031-6_10
23. Lee, W., Chun, J., Lee, Y., Park, K., Park, S.: Contrastive learning for knowledge tracing. In: Proceedings of the ACM Web Conference 2022. p. 2330–2338. WWW '22, Association for Computing Machinery, New York, NY, USA (2022). https://doi.org/10.1145/3485447.3512105
24. Leite, W.L., Kuang, H., Shen, Z., Chakraborty, N., Michailidis, G., D'Mello, S., Xing, W.: Heterogeneity of treatment effects of a video recommendation system for algebra. In: Proceedings of the Ninth ACM Conference on Learning @ Scale. p. 12–23. L@S '22, Association for Computing Machinery, New York, NY, USA (2022). https://doi.org/10.1145/3491140.3528275
25. Li, L., Wang, Z.: Knowledge graph-enhanced intelligent tutoring system based on exercise representativeness and informativeness. International Journal of Intelligent Systems **2023**(1) (2023). https://doi.org/10.1155/2023/2578286
26. Liu, T.C.: A case study of the adaptive learning platform in a taiwanese elementary school: Precision education from teachers' perspectives. Education and Information Technologies **27**(5), 6295–6316 (2022). https://doi.org/10.1007/s10639-021-10851-2
27. Logacheva, E., Hellas, A., Prather, J., Sarsa, S., Leinonen, J.: Evaluating contextually personalized programming exercises created with generative ai. In: Proceedings of the 2024 ACM Conference on International Computing Education Research - Volume 1. p. 95–113. ICER '24, Association for Computing Machinery (2024). https://doi.org/10.1145/3632620.3671103
28. Logacheva, E., Hellas, A., Prather, J., Sarsa, S., Leinonen, J.: Evaluating contextually personalized programming exercises created with generative ai. In: Proceedings of the 2024 ACM Conference on International Computing Education Research-Volume 1. pp. 95–113 (2024)
29. Loh, H., Shin, D., Lee, S., Baek, J., Hwang, C., Lee, Y., Cha, Y., Kwon, S., Park, J., Choi, Y.: Recommendation for effective standardized exam preparation. In: LAK21: 11th International Learning Analytics and Knowledge Conference. p. 397–404. LAK21, Association for Computing Machinery, New York, NY, USA (2021). https://doi.org/10.1145/3448139.3448177
30. Lu, G., Niu, K., Peng, X., Zhou, Y., Zhang, K., Tai, W.: Self-kt: Self-attentive knowledge tracing with feature fusion pre-training in online education. In: 2024 International Joint Conference on Neural Networks (IJCNN). pp. 1–8 (2024). https://doi.org/10.1109/IJCNN60899.2024.10651418

31. Mittal, U., Sai, S., Chamola, V., Sangwan, D.: A comprehensive review on generative ai for education. Ieee Access **12**, 142733–142759 (2024)
32. Newell, A., Rosenbloom, P.S.: Mechanisms of skill acquisition and the law of practice. In: Cognitive skills and their acquisition, pp. 1–55. Psychology Press (2013)
33. Next Generation Science Standards (NGSS): Next generation science standards search, https://www.nextgenscience.org/search-standards
34. Ooge, J., Vanneste, A., Szymanski, M., Verbert, K.: Designing visual explanations and learner controls to engage adolescents in ai-supported exercise selection. In: Proceedings of the 15th International Learning Analytics and Knowledge Conference. p. 1–12. LAK '25, Association for Computing Machinery, New York, NY, USA (2025). https://doi.org/10.1145/3706468.3706470
35. Pardamean, B., Suparyanto, T., Cenggoro, T.W., Sudigyo, D., Anugrahana, A.: Ai-based learning style prediction in online learning for primary education. IEEE Access **10**, 35725–35735 (2022). https://doi.org/10.1109/ACCESS.2022.3160177
36. Park, S., Lee, D., Park, H.: Enhancing knowledge tracing with concept map and response disentanglement. Know.-Based Syst. **302**(C) (2024). https://doi.org/10.1016/j.knosys.2024.112346
37. Prather, J., Denny, P., Leinonen, J., Becker, B.A., Albluwi, I., Craig, M., Keuning, H., Kiesler, N., Kohn, T., Luxton-Reilly, A., et al.: The robots are here: Navigating the generative ai revolution in computing education. In: Proceedings of the 2023 working group reports on innovation and technology in computer science education, pp. 108–159. Association for Computing Machinery (2023)
38. Sayed, W.S., Noeman, A.M., Abdellatif, A., Abdelrazek, M., Badawy, M.G., Hamed, A., El-Tantawy, S.: Ai-based adaptive personalized content presentation and exercises navigation for an effective and engaging e-learning platform. Multimedia Tools and Applications **82**(3), 3303–3333 (2022). https://doi.org/10.1007/s11042-022-13076-8, http://dx.doi.org/10.1007/s11042-022-13076-8
39. Sidoti, O., Park, E., Gottfried, J.: About a quarter of u.s. teens have used chatgpt for schoolwork – double the share in 2023 (January 15 2025), https://www.pewresearch.org/short-reads/2025/01/15/about-a-quarter-of-us-teens-have-used-chatgpt-for-schoolwork-double-the-share-in-2023/, accessed: December 19, 2025
40. Simon, P.D., Zeng, L.M.: Behind the scenes of adaptive learning: A scoping review of teachers' perspectives on the use of adaptive learning technologies. Education Sciences **14**(12), 1413 (2024). https://doi.org/10.3390/educsci14121413
41. Skinner, B.F.: Teaching machines. Science **128**(3330), 969–977 (1958). https://doi.org/10.1126/science.128.3330.969
42. SRI Education: Using technology to personalize learning in k–12 schools. Tech. rep., SRI International, Menlo Park, CA (2018)
43. Sun, X., Li, B., Sutcliffe, R., Gao, Z., Kang, W., Feng, J.: Wse-mf: A weighting-based student exercise matrix factorization model. Pattern Recognition **138**, 109285 (Jun 2023). https://doi.org/10.1016/j.patcog.2022.109285
44. Sun, X., Zhao, X., Li, B., Ma, Y., Sutcliffe, R., Feng, J.: Dynamic key-value memory networks with rich features for knowledge tracing. IEEE Transactions on Cybernetics **52**(8), 8239–8245 (2022). https://doi.org/10.1109/TCYB.2021.3051028
45. Takami, K., Flanagan, B., Dai, Y., Ogata, H.: Evaluating the effectiveness of bayesian knowledge tracing model-based explainable recommender. Int. J. Distance Educ. Technol. **22**(1), 1–23 (2024). https://doi.org/10.4018/IJDET.337600

46. Wang, J., Fan, W.: The effect of chatgpt on students' learning performance, learning perception, and higher-order thinking: insights from a meta-analysis. Humanities and Social Sciences Communications **12**(1) (May 2025). https://doi.org/10.1057/s41599-025-04787-y, http://dx.doi.org/10.1057/s41599-025-04787-y
47. Wang, J., Liang, K.: A cognitive diagnosis method in adaptive learning system based on preconceptions. Scientific Programming **2022**, 1–10 (2022). https://doi.org/10.1155/2022/5011804
48. Yang, H., Shankar, A., S., V.: Artificial intelligence-enabled interactive system modeling for teaching and learning based on cognitive web services. International Journal of e-Collaboration **19**(2), 1–18 (2023). https://doi.org/10.4018/ijec.316655
49. Yang, Z., et al.: Dmp_ai: An ai-aided k-12 system for teaching and learning in diverse schools. In: Ma, W., Li, C., Fan, C., U, L., Lu, A. (eds.) Blended Learning. Intelligent Computing in Education. ICBL 2024. Lecture Notes in Computer Science, vol. 14797, pp. 123–135. Springer, Singapore (2024). https://doi.org/10.1007/978-981-97-4442-8_9
50. Zhang, J., Xia, R., Wang, Q.: Design of data-driven learning path based on knowledge graph and tracing model. In: 2023 IEEE International Conference on High Performance Computing & Communications, Data Science & Systems, Smart City & Dependability in Sensor, Cloud & Big Data Systems & Application (HPCC/DSS/SmartCity/DependSys). pp. 813–820 (2023). https://doi.org/10.1109/HPCC-DSS-SmartCity-DependSys60770.2023.00118
51. Zhu, J., Liao, Y., Lu, D.: Design and optimization of personalized education system based on intelligent algorithms. Procedia Computer Science **243**, 514–522 (2024). https://doi.org/10.1016/j.procs.2024.09.063
52. Zong, M., Krishnamachari, B.: Solving math word problems concerning systems of equations with gpt-3. In: Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence and Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence and Thirteenth Symposium on Educational Advances in Artificial Intelligence. AAAI'23/IAAI'23/EAAI'23, AAAI Press (2023). https://doi.org/10.1609/aaai.v37i13.26896

Declaration of Interest Statement

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.
The author is an Editorial Board Member/Editor-in-Chief/Associate Editor/Guest Editor for this journal and was not involved in the editorial review or the decision to publish this article.
The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: