



2025; 21(3):169-194

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Omdurman Islamic University Journal for Human
Sciences (OIUJ)-(HUSC)

مجلة جامعة أم درمان الإسلامية للعلوم الإنسانية والاجتماعية

<https://journal.oiu.edu.sd/index.php/oij>

<https://doi.org/10.52981/oijhs.v21i3.3424>



ISSN: 5361-1858

The Effectiveness of AI-Enhanced Generative Chatbots and Adaptive Feedback in Developing Mathematical Reasoning and Probability Problem-Solving Skills among Sudanese Secondary School Students: An Experimental Study (Grade 11- Port Sudan, Sudan)

Dr. Mohamed Ali Issa Musa¹

¹ Assistant Professor of Educational Technology / Faculty of Education Omdurman Islamic University-sudan

Email: mohamedali@oiu.edu.sd

To cite this article:

Dr. Mohamed Ali Issa Musa , The Effectiveness of AI-Enhanced Generative Chatbots and Adaptive Feedback in Developing Mathematical Reasoning and Probability Problem-Solving Skills among Sudanese Secondary School Students: An Experimental Study (Grade 11- Port Sudan, Sudan), Omdurman Islamic University Journal, Sudan

ISSN: 5361-1858

<https://doi.org/10.52981/oijhs.v21i3.3424>

Abstract:

This study aimed to evaluate the effectiveness of using generative chatbots enhanced with artificial intelligence and adaptive feedback in developing mathematical reasoning skills and solving probability problems among secondary school students in Port Sudan, Sudan. The researcher employed a quasi-experimental design, with a sample of 48 third-year secondary students randomly divided into experimental and control groups. A validated and reliable achievement test was developed to measure the targeted skills (recall, understanding, application), and data were collected during the 2024-2025

academic year. The data were analyzed using the Statistical Package for the Social Sciences (SPSS). The results revealed statistically significant differences in favor of the experimental group across all three skills, with the most notable improvement observed in application skills (mean = 72.3). Additionally, gender-based differences were identified, with female students outperforming their male counterparts in post-test performance following the use of generative chatbots. The study recommended the integration of generative AI tools in mathematics instruction, the design of training programs for teachers, and the expansion of future research to include other subjects and educational regions.

Keywords: Effectiveness – Generative Chatbots – Artificial Intelligence – Adaptive Feedback – Reasoning Skills – Probability – Secondary Students

1. General Framework

1.1. Introduction

Amidst the rapid advancements in artificial intelligence technologies, innovative educational tools have emerged that reshape the educational landscape. Among the most prominent are **Generative Chatbots** enhanced with **Large Language Models (LLMs)**, which are increasingly being deployed to support self-directed learning and provide immediate feedback to learners. These technologies serve as interactive tools capable of simulating a teacher through intelligent generative dialogue, opening new horizons for more personalized and adaptive education tailored to individual student needs.

Conversely, mathematics education – particularly in teaching mathematical reasoning and probability problem-solving – faces multifaceted challenges. These

include conceptual complexity and weak connections to real-life contexts, adversely affecting students' comprehension and academic achievement. Recent studies indicate that teaching strategies incorporating **adaptive feedback** and intelligent interaction may effectively enhance learners' mathematical thinking and logical problem-solving skills.

Within this context, this study aims to examine the efficacy of AI-enhanced generative chatbots and adaptive feedback in developing mathematical reasoning and probability problem-solving skills among third-grade secondary students in Port Sudan, Sudan. The study employs an experimental methodology, given its ability to precisely determine the impact of the independent variable on dependent variables.

This study derives its significance from its alignment with the educational needs of developing contexts like Sudan. It presents an applicable model for advancing mathematics education using cutting-edge AI technologies, consistent with global trends toward digital transformation in education.

2.1. Research Problem :

Mathematical reasoning and probability problem-solving skills represent higher-order competencies that form the foundation for understanding mathematical concepts and their real-life applications. These skills foster students' logical and analytical thinking while enhancing their ability to tackle novel and unfamiliar situations. Nevertheless, secondary school students' performance in these skills remains below expected levels, as reflected in assessment results and teacher

evaluations – particularly in educational environments facing resource and technological constraints.

Through his professional experience as a secondary mathematics teacher, the researcher observed during instruction that a significant number of students exhibit:

- Marked weakness in mathematical reasoning skills,
- Difficulty solving probability problems,
- Heavy reliance on rote memorization without deep conceptual understanding,
- Limited classroom interaction, and
- Low motivation levels.

These challenges are partially attributed to traditional teaching methods and insufficient opportunities for immediate, effective feedback.

Amid recent advancements in artificial intelligence, generative chatbots have emerged as promising educational tools capable of providing interactive, personalized learning environments – especially when integrated with adaptive feedback systems that accommodate individual differences and flexibly guide learners.

Hence, the study's central problem is defined by the following primary question:

What is the efficacy of using AI-enhanced generative chatbots and adaptive feedback in developing mathematical reasoning and probability problem-solving skills among third-grade secondary students in Port Sudan, Sudan?

3.1. Significance of the Study

This study derives its significance from the scarcity of research examining the efficacy of LLM-enhanced generative chatbots and adaptive feedback in teaching advanced mathematical topics (e.g., mathematical reasoning and probability) within the Sudanese educational context, particularly at the secondary level. Accordingly, the study's importance manifests in the following aspects:

1.3.1. To the researcher's knowledge, this study is likely among the first conducted in Sudan (specifically Port Sudan) investigating the impact of AI-enhanced generative chatbots in teaching complex mathematical topics to secondary students, endowing it with high research and practical relevance.

2.3.1. The study may provide a practical framework for secondary mathematics teachers to enhance instructional methods and deliver immediate student support through AI integration, creating interactive learning environments responsive to individual differences.

3.3.1. It contributes new scientific knowledge to Arab and African educational literature regarding the integration of generative AI technologies in mathematics education – specifically in mathematical reasoning and probability – thereby advancing smart education initiatives.

4.3.1. Should results demonstrate the efficacy of generative chatbots in improving mathematical achievement and thinking skills, this could pave the way for broader adoption in resource–constrained educational settings as an effective means to enhance education quality and outcomes.

4.1. Study Objectives

This study aims to achieve the following primary objective:

To examine the efficacy of generative chatbots (LLM–Chatbots) equipped with adaptive feedback in teaching probability concepts to secondary certificate students in Port Sudan, Red Sea State, Sudan.

To fulfill this objective, the following sub–objectives are derived:

1.4.1. Determine the efficacy of LLM–Chatbots in developing foundational *recall skills* for probability concepts (definitions, mathematical rules).

2.4.1. Measure the efficacy of LLM–Chatbots in enhancing students' understanding of *interconceptual relationships* in probability.

3.4.1. Evaluate the capability of LLM–Chatbots to support *practical application* skills.

4.4.1. Identify statistically significant gender–based differences in achievement post–intervention, analyzing contextual factors explaining these differences.

5.1. Study Hypotheses

1.5.1. Statistically significant differences exist between pre–test and post–test results in developing probability *recall skills* using LLM–Chatbots with adaptive feedback among secondary students in Port Sudan, Sudan.

2.5.1. Statistically significant differences exist between pre-test and post-test results in enhancing understanding of *interconceptual probability relationships* using LLM-Chatbots among secondary students in Port Sudan, Sudan.

3.5.1. Statistically significant differences exist between pre-test and post-test results in improving *practical application skills* for probability problems using LLM-Chatbots among secondary students in Port Sudan, Sudan.

4.5.1. Statistically significant gender-based differences exist in achievement after using LLM-Chatbots among secondary students in Port Sudan, attributable to gender.

6.1. Study Methodology:

To achieve the objectives and test the hypotheses of the study, the quasi-experimental method was selected as the most appropriate approach for the nature of the study topic.

7.1 Study Limitations

1.7.1. Conceptual Scope: Efficacy of AI-enhanced generative chatbots with adaptive feedback in developing mathematical reasoning and probability problem-solving skills among Sudanese secondary students.

2.7.1. Temporal Scope: Findings are contextualized within the 2024–2025 academic year.

3.7.1. Geographical Scope: Red Sea State, Port Sudan, Sudan.

8.1. Operational Definitions

1.7.1. Efficacy: Operationalized as *measured improvement* in mathematical reasoning and probability problem-solving skills through pre-test/post-test comparison following LLM-Chatbot implementation.

2.8.1. Generative Chatbots (LLM-Chatbots): AI-powered interactive platforms

delivering:

- Adaptive explanations of probability concepts
- Dynamically generated exercises based on student proficiency

3.8.1. Artificial Intelligence: Machine learning algorithms capable of:

- Processing natural language inputs
- Generating contextual educational responses

4.8.1. Adaptive Feedback: *Systematic process* where quantitative/qualitative post-task data optimizes AI performance through machine learning.

5.8.1. Reasoning Skills: An AI system's ability to:

- Analyze data
- Apply logical rules
- Make evidence-based decisions

6.8.1. Probability: Curricular content per Sudan's Secondary Certificate syllabus.

7.8.1. Secondary Students (Port Sudan): Grade 12 students enrolled in public/private schools in Port Sudan during 2024–2025.

2. Theoretical Framework and Literature Review

1.2 Theoretical Framework

2.2. Introduction

In the era of the Fourth Industrial Revolution, where AI technologies reshape vital sectors, education emerges as a prime field for leveraging these innovations to overcome historical barriers to educational equity. Despite significant advances in generative AI applications (e.g., ChatGPT, Bard) in developed nations, their

efficacy in resource–constrained educational contexts—such as many African regions—remains questionable, particularly for subjects requiring abstract and deep thinking like mathematics.

In Sudan’s Red Sea State, where shortages of qualified advanced mathematics teachers and inadequate technological infrastructure pose structural challenges, innovative solutions for adaptive, supportive education are urgently needed. This study explores the potential of LLM–powered generative chatbots as compensatory tools to enhance mathematical reasoning and complex problem–solving skills in probability—a foundational subject for university disciplines like engineering and sciences.

Beyond its local relevance, this study contributes to the global discourse on localizing AI technologies in non–ideal educational systems. It offers a practical model for measuring adaptive feedback efficacy in contexts lacking traditional individualized learning support. By bridging modern theoretical frameworks with field application, this study illuminates pathways for researchers and policymakers toward more inclusive, technologically responsive education.

2.3.1 Generative AI in Education

Generative AI models (e.g., GPT–4, PaLM) have revolutionized education through their capacity to simulate human interaction and deliver dynamic, learner–adaptive content (Huang et al., 2023). Recent studies demonstrate their ability to:

- **Generate tiered mathematics exercises** classified by difficulty based on student performance analysis (Zhou et al., 2023), enhancing differentiated instruction.

- **Provide step-by-step scaffolding**, reducing teacher workload in overcrowded classrooms (Molnar, 2023).

2.3.2 Adaptive Feedback

Adaptive feedback is pivotal for optimizing learning in cumulative subjects like mathematics. According to Hattie and Timperley's (2023) model, feedback is effective when it:

- **Focuses on solution strategy refinement** rather than mere error correction.
- **Is delivered during cognitive processing** (not post-task).
Such feedback strengthens direct cognitive engagement between learner and content.

2.3.3 Probability Instruction

Probability remains among the most challenging mathematical domains due to its abstract nature (Batanero et al., 2023, "Teaching Probability: Challenges and Solutions"). Key challenges include:

- **Theory-practice disconnect**, e.g., applying Bayes' theorem to real-world data analysis (Garfield et al., 2023).

2.3.4 Education in Resource-Constrained Regions: Opportunities and Barriers

Educational environments in resource-limited settings (e.g., Red Sea State, Sudan) face structural barriers including digital divides and infrastructural gaps

(UNESCO, 2023, "Digital Learning in Marginalized Regions"). Despite this, smartphones have become primary conduits for digital education in Sub-Saharan Africa (Zhou & Churchill, 2023), opening avenues for **low-cost AI-powered educational solutions**.

2.2. Previous Studies

1.2.2. Ali Muhammad Gharib Abdullah (2025)

Title: *The Impact of an AI Applications Training Program for Secondary Mathematics Teachers on Developing Students' Computational Thinking Skills*

Publisher: Faculty of Education, New Valley University, *Journal of Educational Research*

Methodology: Mixed-methods (qualitative/quantitative)

Sample: 30 secondary mathematics teachers

Key Findings:

- AI applications significantly enhanced students' computational thinking skills
- Recommended developing action plans for AI integration in math instruction and teacher training

2.2.2. Ali Muhammad Gharib Abdullah (2025)

(Note: Duplicate reference – identical to 1.2.2)

3.2.2. Rania Mahmoud & Ali Hassan (2023)

Title: *Efficacy of Adaptive Feedback in Enhancing Mathematical Reasoning Skills Among Secondary Students*

Publisher: *Journal of Educational Technology*, Egyptian Society for Educational

Technology

Methodology: Experimental

Sample: 180 Cairo students

Key Findings:

- 37% improvement in mathematical reasoning skills for experimental group
- Recommended institutional adoption of adaptive feedback systems

4.2.2. Abdullah Al-Khalifi (2022)

Title: *AI's Role in Developing Probability Problem-Solving Skills: A Secondary School Experiment*

Publisher: *King Saud University Journal – Education*

Methodology: Quasi-experimental

Sample: 120 students

Key Findings:

- Experimental group outperformed control by 29% in post-tests
- Recommended developing Arabic-language AI tools

5.2.2. Nora Al-Otaibi (2023)

Title: *AI-Enhanced Adaptive Learning and Its Impact on Mathematics Achievement*

Publisher: *International Conference on Smart Education*

Methodology: Descriptive-analytical

Sample: 200 Dubai public school students

Key Findings:

- 50% higher engagement with adaptive systems
- Recommended teacher training in technology integration

6.2.2. Abdullah Al-Khalifi (2022)

(Note: Duplicate reference – identical to 4.2.2)

7.2.2. Chen & Wang (2023)

Title: *Adaptive AI Systems' Role in Enhancing Probability Problem-Solving Skills*

Methodology: Descriptive

Sample: 500 Chinese students

Key Findings:

- Intermediate-achieving students benefited most from adaptive systems
- Recommended AI tools accommodating individual differences

8.2.2. Kim & Park (2023)

Title: *Traditional vs. AI-Enhanced Methods in Probability Instruction: A Meta-Analysis*

Methodology: Analytical (meta-analysis)

Sample: 20 previous studies

Key Findings:

- AI methods more effective for complex topics (e.g., probability)
- Used random-effects model for data analysis

2.3. Analysis of Previous Research

The current study aligns with Abdullah Al-Khalifi (2022) and Rania Mahmoud & Ali Hassan (2023) in evaluating the efficacy of AI chatbots for enhancing probability problem-solving skills, emphasizing interactive support and immediate feedback. It converges with Nora Al-Otaibi (2023) in prioritizing *AI-enhanced adaptive learning*. Methodologically, this study employs a **quasi-experimental design**, consistent with Al-Khalifi (2022) and Mahmoud & Hassan (2023), while adopting Ali Abdullah's (2025) mixed-methods approach to assess AI tool efficacy.

The sample population (secondary students) mirrors Al-Khalifi (2022) and Al-Otaibi (2023), and assessment via standardized tests follows Mahmoud & Hassan (2023). Expected outcomes parallel Al-Khalifi (2022) and Mahmoud & Hassan (2023), anticipating significant performance improvement post-chatbot implementation.

Key Advancements Beyond Prior Research:

1. **Contextualized Assessment:** Developed a *standardized achievement test* aligned with Sudan's curriculum while enabling cross-study comparability.
2. **Multi-Dimensional Skill Analysis:** Addressed prior limitations by evaluating three skill tiers (recall, comprehension, application) and measuring gender effects – a gap in existing literature.

3. **Resource–Constrained Innovation:** Pioneered an Arabic–language generative chatbot model for probability instruction in under–resourced settings.
4. **Methodological Integration:** Combined quasi–experimental design with gender–difference analysis, revealing novel insights about gender–based technology interaction (e.g., female students’ significant performance gains).

Divergences from Existing Studies:

Study	Distinction
Kim & Park (2023)	Focused on meta–analysis; current study conducts field experimentation
Ali Abdullah (2025)	Targeted teacher training; current research examines student outcomes
Al–Khalifi (2022)	Used Python–based chatbots; current study employs Arabic generative AI models

Scholarly Significance:

This research provides:

- The **first empirical application** of Arabic generative chatbots for probability education in Sudanese secondary schools.
- A **replicable framework** for AI integration in resource–limited contexts.

- **Evidence-based policy pathways** (e.g., gender-responsive AI tools), extending beyond Al-Otaibi's (2023) teacher-training recommendations.

1. Research Methodology

The researchers employed a **quasi-experimental design** to establish causal relationships governing the studied phenomenon. This approach involves designing an experimental scenario to modify reality and observe phenomena under controlled conditions, using experimentation as a tool to test hypotheses. The methodology was selected for its alignment with the study's objectives.

2. Study Population

The target population comprised all elements relevant to the research problem. Secondary students in Port Sudan, Red Sea State, Sudan, were selected (N = 48).

3. Study Sample

A random sample of **48 students** was selected from Sudanese universities. Participants were randomly divided into two equivalent groups:

- **Experimental group:** 24 students
- **Control group:** 24 students

This sampling strategy ensures representativeness and supports the study's objectives.

4. Research Instrument

An **achievement test** was administered twice:

- **Pre-test (Diagnostic):** Assessed baseline equivalence between groups in prior knowledge.
- **Post-test:** Measured progress in mathematical reasoning and probability problem-solving skills after AI chatbot intervention.

The test evaluated three Bloom's taxonomy levels: *Recall*, *Comprehension*, and *Application*.

a. Test Construction

A written test with four questions was designed to:

- Measure higher-order cognitive skills
- Assess knowledge retention
- Identify individual differences

Items were mapped to curricular objectives using a test specification table.

b. Face Validity

Initial test versions were reviewed by expert panelists to:

- Refine linguistic phrasing
- Ensure item-objective alignment
- Verify appropriate cognitive level targeting

c. Reliability

Cronbach's alpha coefficient was calculated as follows:

math

$$\alpha = \frac{N}{N-1} \left(1 - \frac{\sum \sigma_{item}^2}{\sigma_{total}^2}\right)$$

Where:

- N = Number of items
- $\sum \sigma_{item}^2$ = Sum of item variances
- σ_{total}^2 = Total test variance

Result: $\alpha = 0.94$ (indicating excellent internal consistency).

d. Item Discrimination & Difficulty

Discrimination Index:

math

$$D = \frac{C_H - C_L}{N_H}$$

(C_h = Correct responses in high-scoring group; C_l = Correct responses in low-scoring group; N_h = High-group size)

- All items showed strong discrimination ($D > 0.4$)
- Item 3 demonstrated exceptional discrimination ($D = 0.90$)

Difficulty Index:

math

$$P = \frac{\text{Number of correct responses}}{\text{Total students}}$$

- All items fell within optimal range ($0.3 \leq P \leq 0.7$)

e. Test Duration

Determined via pilot testing:

- Average time = $(30 \text{ min} + 50 \text{ min})/2 = 40 \text{ minutes}$

6. Statistical Methods

a. Arithmetic Mean:

math

$$\bar{X} = \frac{\sum X_i}{n}$$

(X_i = Individual scores; n = Sample size)

b. Standard Deviation:

math

$$\sigma = \sqrt{\frac{\sum (X_i - \bar{X})^2}{n}}$$

c. t-Test:

math

$$t = \frac{\bar{X} - \mu}{S / \sqrt{n}}$$

(Student's t-distribution used for hypothesis testing)

1. Introduction

This chapter presents pre-test/post-test results, analyzes findings, and interprets them in light of the study hypotheses.

2. Results Presentation

Hypothesis 1: *Statistically significant differences exist between pre-test/post-test*

results in developing probability recall skills using LLM-Chatbots with adaptive feedback among Port Sudan secondary students.

Table 1: Statistical Differences and t-values

Test	N	Mean	SD	t	df	p-value	The percentage of improvement
Pre-test	24	25.8	22.3	8.2	46	<0.001	204.65%
Post-test	24	78.6	18.9	15.7	46	<0.001	

Results show statistically significant differences (* $p < 0.001$, $df=46$) favoring the post-test. These align with Al-Khalifi (2022) (+29% improvement) and Mahmoud & Hassan (2023) but contradict Abdullah (2025), which focused on teacher training.

The percentage of improvement in memory level is approximately **204.65%**. *This very high percentage indicates a **massive improvement** in test scores after implementing the experiment (using chatbots with adaptive feedback) compared to the pre-test. The students' average performance in the memory test after the experiment **more than doubled** compared to their performance before it (an increase of over 200%).*

Hypothesis 2: *Statistically significant differences exist in enhancing understanding of probability concept relationships using LLM-Chatbots.*

Table 2: Statistical Analysis

Test	N	Mean	SD	t	df	p-value	percentage of improvement
Pre-test	24	18.2	15.4	5.6	46	<0.001	259.3%
Post-test	24	65.4	21.7	9.3	46	<0.001	

Significant differences ($p < 0.001$) confirm post-test superiority. This aligns with Kim & Park (2023) on AI's efficacy for complex concepts.

Percentage of improvement: 259.3%

*This tremendous improvement indicates that students' average performance in the memory test ***increased 2.6-fold*** after using chatbots with adaptive feedback compared to their pre-intervention performance.*

*The high statistical significance ($p < 0.001$) confirms that this improvement is **statistically meaningful** (not random).*

Hypothesis3: *Statistically significant differences exist in improving practical application skills.*

Table 3: Pre–Post Comparison

Test	n	Mean	SD	t	df	p-value	percentage of improvement
Pre-test	24	20.5	19.1	7.8	46	<0.001	252.68%
Post-test	24	72.3	16.8	12.4	46	<0.001	

Post-test shows significant gains ($p < 0.001$), consistent with Al-Khalifi (2022) and Kim & Park (2023) but diverges from Abdullah (2025), which measured computational thinking indirectly.

percentage of improvement: 252.68%

- *This substantial improvement indicates that students' average performance ***increased approximately 2.5-fold*** following the intervention.*
- **The statistical values ($t=12.4$, $p < 0.001$) confirm the improvement is* **highly statistically significant.***
- *The smaller standard deviation (SD) in the post-test (16.8 vs. pre-test 19.1) suggests **greater homogeneity** in student performance after the intervention.*

Hypothesis 4: *Statistically significant gender-based achievement differences exist post-intervention.*

Table 4: Gender Performance Comparison

Criteria	Males (n=28)	Females (n=20)
Mean	38.9	58.2
SD	24.1	39.5
Median	52	85
Minimum	1	0
Maximum	75	100

Females significantly outperformed males (Mean: 58.2 vs. 38.9). This contradicts Mahmoud & Hassan (2023), which reported no gender differences.

Conclusion

I. Study Findings

The pre-test/post-test results revealed statistically significant improvements across all measured domains:

1. Recall Skills Development ($\alpha \leq 0.05$)

- Mean increased from **25.8** (SD = 22.3) to **78.6** (SD = 18.9)
- $t = 15.7 \rightarrow$ **204.6% improvement**

2. Conceptual Understanding Enhancement

- Mean rose from **18.2** (SD = 15.4) to **65.4** (SD = 21.7)
- $t = 9.3 \rightarrow$ **259.3% gain in theoretical comprehension**

3. Practical Application Proficiency

- Mean improved from **20.5** (SD = 19.1) to **72.3** (SD = 16.8)
- ***t* = 12.4 → 252.7% advancement in applied skills**

4. Gender-Based Achievement Gap

- Females significantly outperformed males (**58.2** vs. 38.9)
- Greater score dispersion among females (SD = 39.5 vs. 24.1)

II. Recommendations

Based on the empirical evidence, the researchers propose:

1. Systemic Integration of LLM-Chatbots

Institutionalize adaptive feedback-enabled generative chatbots as core pedagogical tools for probability instruction, given demonstrated efficacy across cognitive domains (recall, comprehension, application).

2. Teacher Capacity Building

Develop specialized training programs focusing on:

- Curriculum-aligned chatbot implementation
- Strategic deployment of adaptive feedback mechanisms
- Personalized scaffolding techniques

3. Future Research Agenda

- Large-scale replications across mathematical domains (e.g., statistics, algebra)
 - Investigation of gender-performance dynamics through sociocultural lenses
 - Exploration of gender-learning method interaction effects
4. **Abdullah, A. M. G.** (2025). The impact of an AI applications training program for secondary mathematics teachers on developing students' computational thinking skills. *Journal of Educational Research*, 12 (3), 45-62. Faculty of Education, New Valley University.
 5. **Al-Hassan, S., & Ahmed, M.** (2023). Gender differences in AI-assisted mathematics learning: Evidence from Middle Eastern contexts. *Journal of Educational Technology & Society*, 26 (4), 205–220. <https://doi.org/10.2307/jeductechsoci.26.4.205>
 6. **Al-Khalifi, A.** (2022). AI's role in developing probability problem-solving skills: A secondary school experiment. *King Saud University Journal - Education* 15 (1), 33–50.
 7. **Al-Otaibi, N.** (2023). AI-enhanced adaptive learning and its impact on mathematics achievement. *Proceedings of the International Conference on Smart Education* (pp. 112–130). UAE: Springer. https://doi.org/10.1007/978-3-031-24601-8_9
 8. **Batanero, K., Chernov, I. J., Engel, J., Lee, H. S., & Sánchez, E.** (2023). Teaching probability: Challenges and solutions. *International Journal of Mathematical Education in Science and Technology*, 54 (2), 157–176. <https://doi.org/10.1080/0020739X.2022.2153921>
 9. **Chen, L., Wang, Y., & Xie, H.** (2023). The impact of AI-powered chatbots on secondary students' mathematics learning: A randomized controlled trial. *Journal of Educational Psychology*, 115 (4), 789–802. **García, M., & Rodríguez, P.** (2023). Machine learning models for personalized probability instruction: A quasi-experimental study. *Journal of STEM Education*, 24 (3), 77–95.
 10. **Hassan, A., & Mahmoud, R.** (2023). Efficacy of adaptive feedback in enhancing mathematical reasoning skills among secondary students. *Journal of Educational Technology*, 30 (2), 88–105. Egyptian Society for Educational Technology.
 11. **Hattie, J., & Timperley, H.** (2023). The power of enhanced feedback: A framework for adaptive education. *Review of Educational Research*, 93 (1), 48–72. <https://doi.org/10.3102/003465432211055>
 12. **Huang, J. J., Song, H. Y., & Zhang, H. Y.** (2023). Generative AI models in education: Applications and implications. *Journal of Educational Technology Research and Development*, 71 (2), 115–134. <https://doi.org/10.1007/s11423-023-10201-8>

13. **Kim, Y., & Park, J.** (2023). Traditional LMS vs. AI chatbots: A comparative analysis in mathematics education. *Educational Technology Research and Development*, 71 (6), 2345–2360. <https://doi.org/10.1007/s11423-023-10245-w>
14. **Kumar, S., & Lee, H.** (2024). Enhancing mathematical reasoning skills using generative AI: A case study in high schools. *International Journal of Artificial Intelligence in Education*, 34 (1), 112–130. <https://doi.org/10.1007/s40593-023-00369-8>
15. **Molnar, E.** (2023). The role of explainable AI in classroom education: Opportunities and limitations. *British Journal of Educational Technology*, 54 (1), 14–30. <https://doi.org/10.1111/bjet.13345>
16. **Smith, J., & Johnson, R.** (2024). Adaptive feedback in probability education: A comparative study of AI vs. human tutors. *Computers & Education*, 198, 104–120.
17. **UNESCO.** (2023). *Digital learning in marginalized regions: Global report 2023*. United Nations Educational, Scientific and Cultural Organization. <https://unesdoc.unesco.org/ark:/48223/pf0000382997>