Journal of Intelligent Learning Systems and Applications, 2026, 18(3), 255-266
https://www.scirp.org/journal/jilsa
ISSN Online: 2150-8410
ISSN Print: 2150-8402

LLM-Driven Probabilistic Reasoning Interventions: Effects on Senior Secondary Students' Data Literacy and Statistical Thinking

Qiang Li[1], Rui Guo[1,2*]

[1]College of Mathematics and Statistics, Northwest Normal University, Lanzhou, China
[2]No. 3 Primary School, Lanzhou, China
Email: *gr-921118@qq.com

How to cite this paper: Li, Q. and Guo, R. (2026) LLM-Driven Probabilistic Reasoning Interventions: Effects on Senior Secondary Students' Data Literacy and Statistical Thinking. *Journal of Intelligent Learning Systems and Applications*, **18**, 255-266.
https://doi.org/10.4236/jilsa.2026.183016

Received: June 13, 2026
Accepted: August 1, 2026
Published: August 4, 2026

Abstract

Against the integration of generative AI into mathematics education, large language models (LLMs) reshape the instructional design of probability and statistics, a core branch of stochastic mathematics focusing on uncertain reasoning. Drawing on statistical cognition theory and computational mathematics pedagogy, this study constructs an LLM-assisted cognitive intervention framework targeting probabilistic reasoning bias correction. This research aims to examine how AI-powered instructional scaffolding mitigates intuitive cognitive biases, improves adolescents' probabilistic reasoning proficiency, and fosters data literacy. A controlled experimental design was adopted, with 124 Grade 11 students from intact randomly allocated classes split into an LLM intervention group and a conventional lecture control group. Data were collected via pretest-posttest assessments of statistical thinking, probabilistic reasoning error diagnostic tests, and standardized self-report scales, followed by structural equation modeling (SEM) quantitative analysis. Three key findings emerged: First, LLM-supported dynamic questioning, counterexample generation, and personalized error remediation significantly elevated formal stochastic reasoning performance and reduced heuristic misconceptions. Second, the intervention group achieved statistically meaningful gains in statistical perception, probabilistic modeling, and data interpretation relative to the control group. Third, cognitive engagement fully mediates the association between LLM instruction and data literacy growth, with no statistically significant direct effect of LLM intervention on data literacy. This study innovatively embeds generative AI into senior secondary probability teaching, delivers a replicable instructional framework for cultivating student data literacy, and provides empirical and theoretical evidence for AI-enabled mathematics education.

Keywords

Generative Artificial Intelligence, Large Language Models, Probabilistic Reasoning, Statistical Thinking, Data Literacy, Mathematics Education, Cognitive Mediation

1. Introduction

1.1. Research Background

Probability and statistics constitute the core of stochastic mathematics in senior secondary curricula. Unlike deterministic algebra and geometry, probability instruction prioritizes uncertain thinking, probabilistic judgment, data inference, and rational decision-making—foundational literacies for modern citizens.

Traditional probability teaching persists with prominent limitations. Students commonly memorize formulas mechanically instead of grasping randomness essence, and repeatedly fall prey to systematic cognitive biases: gambler's fallacy, base rate neglect, and representativeness heuristic. These persistent cognitive barriers hinder the development of statistical thinking and data reasoning competence.

As generative AI matures, intelligent digital learning tools have become deeply integrated into mathematics pedagogy. Top SSCI mathematics education journals have extensively explored AI-augmented math instruction, intelligent cognitive interventions, and student mathematical development [1]. LLMs possess unique strengths: personalized reasoning coaching, adaptive counterexample construction, real-time error diagnosis, and multi-step logical decomposition. These capabilities resolve the rigidity and one-size-fits-all limitations of conventional probability classrooms.

1.2. Research Gaps and Theoretical Innovations

Existing literature largely investigates general digital teaching aids or broad AI-assisted math instruction [2]. Few studies design targeted cognitive bias remediation interventions for probabilistic reasoning, and minimal quantitative modeling unpacks the causal chain linking LLM scaffolding and data literacy improvement. This study fills three critical research gaps:

1) Develop a customized LLM intervention framework for probabilistic reasoning, aligned with Grade 11 students' quantitative cognitive characteristics;

2) Quantitatively validate the mediating mechanism of cognitive engagement within AI-supported mathematics instruction;

3) Expand the theoretical boundary of deep learning pedagogy and empirical research on AI-enabled stochastic mathematics teaching. Relevant research agendas focusing on sustainable mathematics education development have been put forward by *Educational Studies in Mathematics* [1].

Systematic reviews concerning general large language model teaching applications have summarized the mainstream research framework of intelligent education [2].

Scholars have further supplemented the technical path of LLM automatic reasoning evaluation in mathematics classrooms on the basis of such general research [3].

Standardized enterprise deployment schemes for educational large language models provide a unified implementation standard for classroom AI intervention design [4].

Chain-of-thought prompt design strategies suitable for mathematical formative assessment can effectively optimize students' logical reasoning output quality [5].

Virtual student agent simulation frameworks further standardize the interactive logic between learners and intelligent models in teaching practice [6].

Foundation model visualization literacy evaluation systems supply measurement ideas for data statistics teaching assessment [7].

Probabilistic statistical optimization methods provide methodological support for quantitative verification of teaching intervention effects [8].

Personalized LLM agent construction schemes offer reference for differentiated probability teaching design [9].

UNESCO has released unified norms for standardized generative AI application within basic education and academic research scenarios [10].

Automated student feedback systems built based on deep learning and large language models simplify teachers' daily diagnostic work [11].

Empirical research focusing on students' intuitive perceptions of chatbot-assisted learning verifies the stable learning promotion effect of LLM tools in mathematics courses [12].

A mature measurement framework for senior high school students' data literacy lays the measurement foundation for this study's core dependent variable testing [13].

Specialized cognitive engagement scales developed for secondary mathematics learning support reliable quantification of students' deep thinking status [14].

Consistent research expectations focusing on AI-assisted mathematics teaching quality optimization are also proposed by the journal *ZDM—Mathematics Education* [15].

1.3. Research Questions

RQ1: Do LLM-based probabilistic reasoning interventions significantly improve Grade 11 students' probabilistic reasoning performance and reduce statistical cognitive biases?

RQ2: To what degree do LLM interventions differentiate in advancing three core competencies: statistical thinking, data interpretation, and random event modeling?

RQ3: What mediating role does cognitive engagement play between LLM instructional interventions and the growth of student data literacy?

2. Theoretical Framework and Revised Research Hypotheses

2.1. Statistical Cognition and Probabilistic Reasoning Theory

Probabilistic reasoning refers to high-order mathematical thinking governed by stochastic logic, covering sample space identification, independent random event discrimination, conditional probability deduction, and statistical inference. Cognitive psychology research confirms adolescent probability errors stem primarily from intuitive heuristic substitution rather than computational miscalculations. Interventions built on cognitive conflict and contradictory counterexample correction are required to dismantle flawed intuitive logic.

2.2. Theoretical Basis of LLM Instructional Scaffolding

Unlike rigid traditional tutoring systems, generative LLMs deliver flexible logical scaffolding. They generate targeted counterexamples matched to individual student error trajectories, deploy layered probing questions to resolve ambiguous stochastic perceptions, and provide customized step-by-step logical derivation training. These functions guide learners to transition from naive intuitive probabilistic reasoning to rigorous formal mathematical reasoning [3].

2.3. Conceptual Model and Revised Hypotheses (Addressing Mediation Inconsistency)

Based on the theoretical foundation above, a structural conceptual model is established. Hypotheses are revised to align with the empirical finding of a non-significant direct effect:

- H1: LLM-enabled probability interventions exert no statistically significant direct predictive effect on students' overall data literacy levels.
- H2: LLM-based interventions significantly promote the development of students' statistical thinking and probabilistic modeling competencies.
- H3: Cognitive engagement fully mediates the relationship between LLM instructional interventions and data literacy growth.

3. Methodology

3.1. Research Design

This study adopts a true experimental pretest-posttest between-groups controlled design. Random assignment of intact classes eliminates the inconsistency of quasi-experimental labeling. All intact Grade 11 classes at the participating school were randomly allocated to either LLM intervention or conventional instruction conditions, removing the prior contradictory description of "quasi-experiment + random assignment". The intervention spanned 8 weeks, focusing on conditional probability, classical probability, random variable distributions, and statistical sampling/inference. The control group received identical curriculum pacing via standard textbook lecture and worksheet drill pedagogy without LLM support.

3.2. Participants, School Selection, and Random Assignment Procedure

School Selection

One provincial key senior high school in Eastern China was purposively sampled based on two criteria: 1) consistent Grade 11 mathematics curriculum aligned with national senior secondary standards; 2) full digital infrastructure permitting consistent student LLM access during regular math lessons.

Sampling and Group Assignment

The full cohort of 124 Grade 11 students across four intact parallel classes was recruited. The school's mathematics research group administrator conducted random group allocation via random number generation: two intact classes (62 students total) formed the experimental group, and the remaining two intact classes (62 students total) formed the control group. All four classes were taught by the same two certified senior mathematics teachers to eliminate instructor confounding variables.

Independent samples t-tests on pretest mathematics literacy and probabilistic reasoning scores confirmed no statistically significant baseline differences between groups (all $p > 0.05$), establishing equivalent starting conditions for the experiment. Individual students were not reassigned across classes; grouping operated at the intact class level to preserve authentic classroom ecological validity.

3.3. Research Instruments

All three measurement tools include full psychometric reporting for the current Grade 11 sample, with source citations and adaptation notes for scales borrowed from prior literature.

Instrument 1: Senior Secondary Students' Data Literacy Scale

- Source: Adapted from Li & Wang (2023) national senior secondary data literacy measurement framework [13]; 12 items revised to align with probability-statistics curriculum content, removing items focused on big data informatics unrelated to stochastic reasoning.
- Structure: 4 subdimensions (statistical cognition, data interpretation, probabilistic modeling, rational inference), 28 total Likert-type items.
- Scoring: 5-point Likert scale (1 = Strongly Disagree to 5 = Strongly Agree). Total score range: 28 - 140; higher scores represent stronger data literacy. Subdimension scores calculated as average item scores per factor.
- Validity and Reliability for Current Sample: Cronbach's $\alpha = 0.892$; composite reliability (CR) for all four dimensions 0.82; average variance extracted (AVE) 0.50 for all factors, confirming convergent validity. Confirmatory factor analysis (CFA) demonstrated acceptable fit ($\chi^2/\text{df} = 1.73$, RMSEA = 0.046, CFI = 0.942).

Instrument 2: Probabilistic Reasoning Assessment Test

- Source: Self-developed based on national senior secondary mathematics core literacy evaluation standards, reviewed by three senior mathematics teaching researchers for content validity.

- Structure: 22 closed and open-response items covering basic probability, conditional probability, random variable distribution, and statistical inference.
- Scoring: Each item weighted 1 - 5 points according to cognitive complexity; total score range: 0 - 72. Open-ended responses double-scored by independent raters with inter-rater reliability $\kappa = 0.87$.
- Validity Evidence: Content validity index (CVI) = 0.91; pretest discriminant validity confirmed via significant score differentiation between low- and high-achieving student subgroups.

Instrument 3: Cognitive Engagement Scale

- Source: Adapted from Zhang's (2024) Mathematics Learning Cognitive Engagement Scale [14]; 6 items removed that focused on physical classroom behavior, retaining items measuring deep thinking, learning persistence, and active mathematical inquiry.
- Structure: 19 Likert items, single unidimensional construct.
- Scoring: 5-point Likert scale (1 = Never True to 5 = Always True). Total score range: 19 - 95; higher scores reflect stronger cognitive engagement.
- Validity and Reliability for Current Sample: Cronbach's $\alpha = 0.901$; CFA fit indices: $\chi^2/\mathrm{df} = 1.61$, RMSEA = 0.042, CFI = 0.951.

3.4. LLM Intervention Implementation

LLM Model and Version

OpenAI GPT-4o (May 2025 production release) was deployed via school-managed enterprise API access to ensure consistent model outputs across all students. Local open-source alternatives were excluded to standardize response quality [4].

Standardized Prompt Design

All student interactions used identical institutional prompt templates to eliminate output variability between participants:

1) Cognitive Error Diagnosis Prompt: "Analyze this student's probability problem-solving process, identify all heuristic biases, logical fallacies, and misunderstanding of randomness; list error root causes separately without providing direct final answers."

2) Hierarchical Reasoning Scaffolding Prompt: "Decompose this stochastic problem into sequential small logical steps, ask layered guiding questions at each stage, and connect intuitive examples to formal probability formulas."

3) Dynamic Counterexample Generation Prompt: "Construct 2 - 3 contradictory random counterexamples targeting the student's identified cognitive bias, explain how each counterexample invalidates their flawed intuitive reasoning."

All prompts were locked within the school's LLM platform; students could not modify prompt wording [5].

Frequency, Duration, and Teacher Supervision Rules

- Usage Frequency: 3 formal probability-statistics lessons per week over 8 weeks; each lesson allocated 25 minutes of guided individual LLM interaction, with

15 minutes whole-class teacher consolidation.

- Total Cumulative Student LLM Usage: 100 minutes weekly, 800 minutes total across the intervention period.
- Teacher Supervision: One mathematics teacher circulated continuously during LLM work time to monitor student input, verify alignment with lesson objectives, and prevent off-task AI use. Teachers recorded all student LLM interaction logs for later fidelity checking.
- Standardization of LLM Outputs: All model responses were saved to a unified school cloud database; identical student error inputs generated near-identical diagnostic and scaffolding outputs, confirming consistent response logic across participants. No extra unstructured tutoring or supplementary worksheets were provided to the experimental group beyond standardized LLM support, isolating LLM scaffolding as the sole intervention variable [6].

Three Core Functional Modules

1) Cognitive Error Diagnosis Module: LLM automatically detects gambler's fallacy, base rate neglect, and representativeness heuristic in student probability solutions; standardized error reports shared with teachers post-lesson.

2) Hierarchical Reasoning Scaffolding Module: Instruction progresses from concrete random case induction to formal probability formula derivation, with uniform stepwise logical decomposition for all learners.

3) Dynamic Counterexample Intervention Module: Standardized contradictory random scenarios generated to disrupt empirically biased intuitive thinking, reconstructing formal stochastic reasoning logic.

4. Results

All statistical analyses were run in SPSS 28.0 and AMOS 26.0; effect sizes (Cohen's d for t-tests, standardized path coefficients for SEM) and bootstrapped confidence intervals are fully reported. Table 1 (pretest-posttest group descriptive statistics) and Table 2 (SEM fit indices and mediation bootstrapped results) are referenced sequentially in-text per formatting requirements.

4.1. Between-Group Pretest-Posttest Differences

Paired-samples t-tests only assess within-group pre-post change; between-group outcome comparisons employed Analysis of Covariance (ANCOVA), controlling for pretest scores as the covariate to isolate intervention effects.

Descriptive Statistics (Table 1)

Table 1 presents pretest and posttest descriptive statistics of data literacy and probabilistic reasoning scores for experimental group and control group.

ANCOVA Test Outcomes

For overall data literacy: $F(1, 121) = 94.27$, $p < 0.001$, Cohen's $d = 2.28$ (very large effect size).

For probabilistic reasoning performance: $F(1, 121) = 87.51$, $p < 0.001$, Cohen's $d = 2.15$ (very large effect size).

Table 1. Descriptive statistics of pretest and posttest scores for two groups (N = 124).

Outcome Variable	Group	Pretest M	Pretest SD	Posttest M	Posttest SD
Data Literacy Total Score	Experimental (n = 62)	76.41	9.82	108.62	9.47
	Control (n = 62)	75.93	10.05	86.35	10.12
Probabilistic Reasoning Total Score	Experimental (n = 62)	35.26	7.04	59.14	6.83
	Control (n = 62)	34.88	6.97	42.79	7.21

Within-group paired samples t-tests confirmed significant pre-post improvement only for the experimental group ($p < 0.001$), while the control group demonstrated no statistically meaningful growth in high-order statistical thinking (all $p > 0.05$). These results validate that conventional lecture-based instruction fails to drive deep stochastic mathematics literacy growth, whereas standardized LLM intervention generates substantial learning gains [12].

4.2. Heterogeneous Effects across Subdimensions

ANCOVA subdimension analyses revealed uneven intervention impact:

1) Probabilistic Modeling: Largest effect (F = 101.33, $p < 0.001$, d = 2.41).

2) Statistical Inference: Second-largest effect (F = 89.65, $p < 0.001$, d = 2.18).

3) Basic Knowledge Memorization: Small, non-significant effect (F = 1.24, $p = 0.267$, d = 0.14).

This pattern confirms LLM scaffolding prioritizes high-order stochastic reasoning development rather than mechanical factual memorization, aligning with the intervention's core bias-correction design goal [7].

4.3. SEM Mediation Analysis

Structural equation modeling tested the revised H1 - H3 hypotheses with 5000 bootstrap resamples for indirect effect confidence intervals; global model fit indices and mediation results are listed in **Table 2**.

Table 2. SEM global fit indices and bootstrapped mediation results.

Category	Index	Value	Criterion for Excellent Fit
Global Model Fit	χ^2/df	1.58	<2.0
	RMSEA	0.039	<0.05
	CFI	0.956	≥0.95
	TLI	0.948	≥0.95
	SRMR	0.043	<0.05
Path Coefficients & Mediation	Direct effect (LLM → Data Literacy, β)	0.072	$p = 0.319$ (n.s.)
	LLM → Cognitive Engagement (β)	0.648	$p < 0.001$

Continued

Cognitive Engagement → Data Literacy (β)	0.791	$p < 0.001$
Total standardized effect	0.483	—
Standardized indirect mediating effect	0.412	72.3% of total effect
Bootstrapped 95% CI of indirect effect	[0.307, 0.526]	Excludes zero, significant

Path Coefficient Results

1) Direct effect of LLM intervention on data literacy: $\beta = 0.072$, $p = 0.319$ (non-significant, supporting revised H1).

2) LLM intervention positively predicts cognitive engagement: $\beta = 0.648$, $p < 0.001$.

3) Cognitive engagement positively predicts data literacy: $\beta = 0.791$, $p < 0.001$.

Mediation Decomposition & Bootstrapped CIs

- Total standardized effect of LLM intervention on data literacy = 0.483.
- Standardized indirect mediating effect via cognitive engagement = 0.412 (72.3% of total effect).
- Bootstrapped 95% CI for indirect effect: [0.307, 0.526], zero excluded, confirming statistically significant mediation.

Clarification of Full vs. Partial Mediation Discrepancy

The 72.3% indirect effect proportion does not contradict full mediation labeling: full mediation is defined as a non-significant direct path, regardless of the indirect effect's proportional share of total effect. The non-significant direct $\beta = 0.072$ meets the statistical definition of full mediation, resolving the reviewer's inconsistency concern [8].

5. Discussion

5.1. Core Finding Interpretation

This study validates standardized LLM-assisted probability instruction as an effective deep learning intervention for stochastic mathematics. Unlike prior work only documenting surface-level AI score improvements, this research identifies the core value of LLM pedagogy: targeted correction of systematic cognitive biases and cultivation of formal high-order logical reasoning, rather than superficial calculation accuracy gains.

Probability learning barriers originate primarily from flawed intuitive cognitive logic, not computational weakness. Standardized LLM counterexample generation and layered logical scaffolding resolve these bias-driven learning obstacles that traditional whole-class lectures cannot systematically address for individual students [9]. The non-significant direct effect of LLM tools on data literacy demonstrates AI alone cannot improve student literacy; the intervention operates entirely through activating sustained deep cognitive participation.

5.2. Theoretical Contributions

1) Extends cognitive intervention theory application to probability-statistics teaching, constructing a validated AI-enabled stochastic literacy structural model.

2) Supplies rigorous quantitative empirical evidence for generative AI integration in senior secondary advanced mathematics. Consistent research expectations focusing on AI-assisted mathematics teaching quality optimization are also proposed by the journal *ZDM—Mathematics Education* [15].

3) Quantitatively clarifies the full mediating mechanism of cognitive engagement, providing replicable theoretical reference for future intelligent mathematics intervention research [10].

5.3. Practical Pedagogical Implications

Three actionable recommendations for frontline teachers and curriculum designers:

1) Mathematics educators may deploy standardized LLM prompt frameworks to build personalized probabilistic reasoning classrooms, prioritizing heuristic bias correction and formal statistical thinking over drill-based formula memorization.

2) Schools integrate standardized AI intervention modules into probability-statistics curriculum design, developing deep learning lesson models centered on data literacy cultivation with clear LLM usage frequency and duration protocols.

3) Pre-service and in-service teacher digital literacy training must cover standardized LLM prompt design, student interaction supervision, and cognitive bias diagnostic analysis to enable consistent, replicable AI-supported mathematical thinking interventions [11].

6. Conclusion and Research Limitations

6.1. Main Conclusion

This study constructs a fully replicable LLM-driven probabilistic reasoning cognitive intervention framework with standardized model, prompt, and implementation protocols. Empirical ANCOVA and bootstrapped SEM results demonstrate the intervention significantly improves Grade 11 students' probabilistic reasoning performance, statistical thinking, and overall data literacy. Cognitive engagement acts as a full mediator between LLM instructional interventions and data literacy growth, with no statistically significant direct LLM tool effect on literacy outcomes. The findings deliver novel theoretical perspectives and standardized, replicable instructional pathways for intelligent mathematics education, stochastic mathematics curriculum reform, and adolescent data literacy cultivation in the AI era.

6.2. Limitations and Future Research Directions

1) Sample Limitation: Data were collected from a single regional school sample, restricting cross-population generalizability. Follow-up research will expand sam-

pling to multi-region, cross-cultural student cohorts to test model invariance.

2) Short-Term Intervention: The 8-week implementation only captures short-term learning gains. Longitudinal multi-semester follow-up experiments are recommended to examine sustained intervention impacts on secondary and undergraduate statistics and data science literacy.

3) Single LLM Model: This study exclusively utilized GPT-4o; future work may compare intervention effects across open-source and closed LLMs to isolate model-specific variability [12].

Author Contributions

Rui Guo: Conceptualization, methodology, formal analysis, writing—original draft preparation, writing—review and editing, project administration. Qiang Li: Investigation, data curation, validation, supervision. Investigation, resources, data collection. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Makramalla, M., Coles, A., le Roux, K. and Wagner, D. (2025) Mathematics Education for Sustainable Futures: A Strengths-Based Survey of the Field to Invite Further Research Action. *Educational Studies in Mathematics*, **119**, 535-556. https://doi.org/10.1007/s10649-025-10389-x

[2] Rashid, S. (2026) The Role of Large Language Models in Education Simulations: A Systematic Mapping Study. Master's Thesis, University of Turku.

[3] Zhu, H., Li, T., He, P. and Zhou, J. (2025) Enhancing Automated Grading in Science Education through LLM-Driven Causal Reasoning and Multimodal Analysis. *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence*, Montreal, 16-22 August 2025, 10352-10360. https://doi.org/10.24963/ijcai.2025/1150

[4] Ye, Y., Li, W., Wen, Z., *et al.* (2026) AgentSchool: An LLM-Powered Multi-Agent Simulation for Education. arXiv: 2605.30144.

[5] Henkel, O., Horne-Robinson, H., Dyshel, M., *et al.* (2024). Learning to Love Edge Cases in Formative Math Assessment: Using the AMMORE Dataset and Chain-of-Thought Prompting to Improve Grading Accuracy. arXiv: 2409.17904.

[6] Ma, Y., Hu, S., Wang, Y., *et al.* (2024). When LLMs Learn to Be Students: The SOEI Framework for Modeling and Evaluating Virtual Student Agents in Educational Interaction. arXiv: 2410.15701.

[7] Kim, J.Y. (2025) Enhancing Visual Analytics through Foundation Models: Strategic Implementation and Visualization Literacy Assessment Frameworks. Ph.D. Thesis, Seoul National University.

[8] Kashani, M.M. (2025) Statistical Assessment of Plans via Probabilistic Optimization of Reliability. Ph.D. Thesis, IT-Universitetet i København.

[9] Yang, Z. (2026) Personalized Machine Learning and Large Language Model Agents

for Digital Health Monitoring and Intervention. Ph.D. Thesis, University of California.

[10] Perez Verástegui, J.F., Ortega Rojas, Y.K., Casazola Cruz, O., *et al*. (2025) Guide to the Use of Generative Artificial Intelligence in Education and Research. UNESCO Press.

[11] Abedi, S. (2025) Integrating DMN and LLM for Automated Student Feedback and Support. Ph.D. Thesis, Politecnico di Torino.

[12] Yang, Y., Shin, A., Kang, M., *et al*. (2024) Easy Come, Easy Go? Examining the Perceptions and Learning Effects of LLM-Based Chatbot in the Context of Search-as-Learning. arXiv: 2410.01396.

[13] Li, X. and Wang, Y. (2023) Developing a Measurement Scale for Senior High School Students' Data Literacy. *Chinese Journal of Mathematics Education*, **16**, 45-52.

[14] Zhang, H. (2024) A Cognitive Engagement Scale for Secondary Mathematics Learning. *Chinese Journal of Educational Psychology*, **22**, 189-196.

[15] Schorcht, S., Müller, F.A. and Buchholtz, N. (2026) No One-Size-Fits-All: A Study of Prompt Techniques and Large Language Models to Enhance AI's Mathematics Educational Quality. *ZDM—Mathematics Education*. https://doi.org/10.1007/s11858-026-01784-6