Journal of Educational Research and Learning Analytics
Published by Yayasan Cerdas Pedia Indonesia
Volume 2 Issue 2, 2026 ISSN (Online) ***3110-3987***
Journal Homepage: **https://jurnal.cerdaspedia.com/index.php/jerlra**

Artificial Intelligence Strengthens Secondary Students' Financial Literacy in Balkh, Afghanistan: Quasi-Experimental Evidence

Kecerdasan Buatan Memperkuat Literasi Keuangan di Sekolah Menengah Balkh, Afghanistan: Bukti Kuasi-Eksperimen

Sameullah Samim[1], Abdul Kabir Azizi[2]
[1-2]Faculty of Economics, Department of Banking and Finance, Balkh University, Afghanistan

Article Info

Corresponding Author:
Sameullah Samim
✉samiullahsamim2017@gmail.com

History:
Submitted: 03-05-2026
Revised: 10-11-2026
Accepted: 20-11-2026

Keyword:
Artificial Intelligence; Financial Literacy; Financial Education; Secondary School Students; Balkh Province.

Kata Kunci:
Kecerdasan Buatan; Literasi Keuangan; Pendidikan Keuangan; Siswa Sekolah Menengah; Provinsi Balkh.

Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY SA).

https://doi.org/10.65101/jerlra.v2i2.343

Abstract

Adolescent financial literacy remains weak in resource-constrained schools, where lecture-based instruction lacks diagnostic tracking, iterative practice, and timely feedback. This study evaluated whether artificial intelligence (AI) strengthens secondary students' financial literacy in Balkh Province, Afghanistan. Using a quasi-experimental pretest-posttest control group design across 14 public urban general and rural vocational schools in seven districts, 350 male students in grades 10 to 12 received AI-supported instruction across four tools (167 students) or conventional teaching (183 students). Scores on a 20-point four-domain test were analyzed using t-tests, analysis of covariance, and multivariable regression. Baseline-equivalent cohorts diverged markedly: AI-supported learners gained 3.65 points versus 1.58 conventionally, yielding a covariate-adjusted advantage of 2.07 points ($p < .001$; Hedges' $g = 1.44$), largest in budgeting, saving, and digital financial safety, and consistent across tools and settings. By disaggregating tools, domains, and urban-rural equity, this study informs pedagogy-led AI policy requiring teacher training and device access.

Abstrak

Literasi keuangan remaja masih lemah di sekolah terbatas sumber daya karena pembelajaran ceramah minim pemantauan diagnostik, latihan berulang, dan umpan balik cepat. Penelitian ini mengevaluasi apakah kecerdasan buatan (AI) memperkuat literasi keuangan siswa sekolah menengah di Provinsi Balkh, Afghanistan. Melalui desain kuasi-eksperimen pretest-posttest berkelompok kontrol pada 14 sekolah negeri umum perkotaan dan kejuruan pedesaan di tujuh distrik, 350 siswa laki-laki kelas 10 sampai 12 mengikuti pembelajaran empat alat AI (167 siswa) atau pengajaran konvensional (183 siswa). Skor tes 20 poin empat domain dianalisis melalui uji t, analisis kovarians, dan regresi multivariabel. Kelompok setara awal berbeda tajam: siswa berbantuan AI meningkat 3,65 poin berbanding 1,58 secara konvensional, menghasilkan keunggulan terkoreksi 2,07 poin ($p < ,001$; g Hedges = 1,44), terbesar pada penganggaran, menabung, dan keamanan keuangan digital, serta konsisten lintas alat dan lokasi. Pemetaan alat, domain, dan pemerataan kota-desa ini mendasari kebijakan AI berbasis pedagogi melalui pelatihan guru dan akses perangkat.

A. INTRODUCTION

1. Background

Large-scale international educational assessments increasingly treat adolescent financial literacy as a measurable schooling outcome rather than an informal household skill, yet empirical evidence consistently exposes severe disparities in students' capacity to manage income, savings, credit, and digital financial risks across socioeconomic and institutional strata.[1] Across both developed and developing economies, inadequate financial capability correlates with weak financial planning, insufficient precautionary savings, and costly borrowing errors.[2] Although secondary schools represent the most equitable institutional mechanism for cultivating this foundational competency at scale, classroom delivery frequently fails to bridge conceptual instruction with applied decision-making. Conventional personal finance instruction relies heavily on lecture-based teaching that lacks diagnostic tracking, iterative practice, and timely formative feedback—creating a persistent disconnect between curriculum exposure and measurable learning gains.[3] Simultaneously, while artificial intelligence (AI) and learning analytics platforms have expanded rapidly to provide automated assessment, adaptive scaffolding, and data-informed personalization, rigorous classroom-level evidence demonstrating whether these technologies improve applied life competencies in digitally uneven, resource-constrained school systems remains critically scarce.[4]

Prior scholarship relevant to this intersection spans three distinct research streams that seldom converge. First, causal evaluations of school-based financial education confirm that structured instruction improves financial knowledge and downstream behaviors, as established in meta-analytic syntheses of randomized experiments by Kaiser and colleagues and earlier reviews by Kaiser and Menkhoff, alongside policy

[1] OECD, *PISA 2018 Results (Volume IV): Are Students Smart About Money?* (Paris: OECD Publishing, 2020), https://doi.org/10.1787/48ebd1ba-en.
[2] Annamaria Lusardi and Olivia S. Mitchell, "The Economic Importance of Financial Literacy: Theory and Evidence," *Journal of Economic Literature* 52, no. 1 (2014): 5–44, https://doi.org/10.1257/jel.52.1.5; Annamaria Lusardi, "Financial Literacy and the Need for Financial Education: Evidence and Implications," *Swiss Journal of Economics and Statistics* 155 (2019): 1, https://doi.org/10.1186/s41937-019-0027-5.
[3] Aisa Amagir et al., "A Review of Financial-Literacy Education Programs for Children and Adolescents," *Citizenship, Social and Economics Education* 17, no. 1 (2018): 56–80, https://doi.org/10.1177/2047173417719555.
[4] Olaf Zawacki-Richter et al., "Systematic Review of Research on Artificial Intelligence Applications in Higher Education—Where Are the Educators?" *International Journal of Educational Technology in Higher Education* 16 (2019): 39, https://doi.org/10.1186/s41239-019-0171-0; Gwo-Jen Hwang et al., "Vision, Challenges, Roles and Research Issues of Artificial Intelligence in Education," *Computers and Education: Artificial Intelligence* 1 (2020): 100001, https://doi.org/10.1016/j.caeai.2020.100001.

evaluations by Urban and colleagues linking high school personal finance mandates to stronger adult credit outcomes.[5] Nevertheless, systematic reviews focusing specifically on children and adolescents—most notably by Amagir and colleagues—reveal substantial heterogeneity in program effectiveness, attributing modest transfer to passive pedagogical delivery and insufficient opportunities for applied practice and feedback.[6] Second, educational technology and learning analytics research documents consistent learning advantages for adaptive digital environments over conventional instruction. Foundational meta-analyses of intelligent tutoring systems by Ma and colleagues and by Kulik and Fletcher report moderate-to-large learning gains,[7] while comprehensive reviews by Chen, Chen, and Lin, Hwang and colleagues, and Zhai and colleagues map the progression of educational AI toward adaptive assessment, learner profiling, and data-driven instructional support.[8] Experimental evidence by Huang, Lu, and Yang further demonstrates that AI-enabled personalized recommendations measurably strengthen student engagement, motivation, and academic achievement.[9] Third, emerging scholarship on generative AI and large language models—including studies by Kasneci and colleagues and by Chan and Hu—highlights the capacity of conversational agents to generate realistic scenarios and on-demand explanations, while raising vital ethical and pedagogical concerns regarding output accuracy, algorithmic bias, student data privacy, and learner over-reliance.[10]

---

[5] Tim Kaiser et al., "Financial Education Affects Financial Knowledge and Downstream Behaviors," *Journal of Financial Economics* 145, no. 2 (2022): 255–72, https://doi.org/10.1016/j.jfineco.2021.09.022; Tim Kaiser and Lukas Menkhoff, "Does Financial Education Impact Financial Literacy and Financial Behavior, and If So, When?" *World Bank Economic Review* 31, no. 3 (2017): 611–30, https://doi.org/10.1093/wber/lhx018; Carly Urban et al., "The Effects of High School Personal Financial Education Policies on Financial Behavior," *Economics of Education Review* 78 (2020): 101786, https://doi.org/10.1016/j.econedurev.2018.03.006.

[6] Amagir et al., "Review of Financial-Literacy Education Programs," 56–80.

[7] Wenting Ma et al., "Intelligent Tutoring Systems and Learning Outcomes: A Meta-Analysis," *Journal of Educational Psychology* 106, no. 4 (2014): 901–18, https://doi.org/10.1037/a0037123; James A. Kulik and J. D. Fletcher, "Effectiveness of Intelligent Tutoring Systems: A Meta-Analytic Review," *Review of Educational Research* 86, no. 1 (2016): 42–78, https://doi.org/10.3102/0034654315581420.

[8] Lijia Chen, Pingping Chen, and Zhijian Lin, "Artificial Intelligence in Education: A Review," *IEEE Access* 8 (2020): 75264–78, https://doi.org/10.1109/ACCESS.2020.2988510; Hwang et al., "Vision, Challenges, Roles and Research Issues"; Xuesong Zhai et al., "A Review of Artificial Intelligence (AI) in Education from 2010 to 2020," *Complexity* 2021 (2021): 8812542, https://doi.org/10.1155/2021/8812542.

[9] Anna Y. Q. Huang, Owen H. T. Lu, and Stephen J. H. Yang, "Effects of Artificial Intelligence–Enabled Personalized Recommendations on Learners' Learning Engagement, Motivation, and Outcomes in a Flipped Classroom," *Computers & Education* 194 (2023): 104684, https://doi.org/10.1016/j.compedu.2022.104684.

[10] Enkelejda Kasneci et al., "ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education," *Learning and Individual Differences* 103 (2023): 102274, https://doi.org/10.1016/j.lindif.2023.102274; Cecilia Ka Yuk Chan and Wenjie Hu, "Students' Voices on

Synthesizing these three bodies of work exposes critical methodological, pedagogical, and contextual blind spots. As Zawacki-Richter and colleagues and Hwang and colleagues observe, AI-in-education research remains heavily concentrated in higher education and well-resourced school systems, frequently emphasizing technological novelty over pedagogical design and equity-oriented classroom evidence.[11] Controlled evaluations testing AI-supported financial literacy instruction among secondary school students are virtually absent, particularly in developing and fragile educational systems such as Afghanistan, where pronounced urban–rural divides in device availability, internet connectivity, multilingual student demographics, and teacher preparation directly shape instructional feasibility. Furthermore, prior empirical studies typically evaluate a single AI tool against a control condition using aggregate achievement scores, leaving unresolved whether different AI instructional modalities produce comparable learning gains, which specific cognitive domains (procedural calculation versus value-based conceptualization) respond most strongly to adaptive feedback, and how institutional enabling conditions mediate student outcomes.

In contrast to previous studies that predominantly evaluated isolated AI applications in well-resourced higher education environments or assessed financial education solely through aggregate knowledge scores, this study advances educational research and learning analytics by conducting a multi-tool, domain-disaggregated, and equity-oriented quasi-experimental evaluation of AI-supported financial literacy instruction across 14 urban general and rural vocational secondary schools in Balkh Province, Afghanistan. By modeling pretest-adjusted learning gains across four distinct AI modalities—an educational chatbot, an interactive financial learning application, AI-generated case studies, and an adaptive quiz platform—alongside institutional covariates including device access, internet connectivity, teacher AI training, and instructional dosage, this research isolates the pedagogical mechanisms through which structured AI feedback strengthens adolescent financial competence under uneven infrastructural conditions.

This study defends the analytical thesis that structured, pedagogy-aligned AI support substantially outperforms conventional teacher-led instruction in cultivating

---

Generative AI: Perceptions, Benefits, and Challenges in Higher Education," *International Journal of Educational Technology in Higher Education* 20 (2023): 43, https://doi.org/10.1186/s41239-023-00411-8.
[11] Zawacki-Richter et al., "Systematic Review of Research"; Hwang et al., "Vision, Challenges, Roles and Research Issues."

secondary students' applied financial literacy, and that these learning gains stem primarily from shared instructional functions—immediate formative feedback, repeated practice, and realistic financial scenarios—rather than the selection of a single proprietary tool. For the international educational research and learning analytics community, these findings deliver crucial cross-contextual evidence on the instructional transferability and equity potential of AI-supported learning within low-resource secondary and vocational systems. Empirically, across 350 baseline-equivalent students ($p = .753$), AI-supported instruction generated a covariate-adjusted advantage of 2.07 points on a 20-point assessment ($p < .001$; Hedges' $g = 1.44$; ANCOVA $R^2 = 0.729$; fully adjusted regression $b = 2.13$, $R^2 = 0.751$), with the largest gains concentrated in applied domains—budgeting ($g = 0.48$), saving ($g = 0.28$), and digital financial safety ($g = 0.25$)—consistent effectiveness across all four AI tools ($g = 1.20$–1.53) and across both urban ($g = 1.24$) and rural vocational ($g = 1.74$) schools, while device access ($b = 0.64$) and teacher AI training ($b = 0.48$) emerged as pivotal enabling conditions for evidence-based educational policy.

2. Research Questions

To examine the instructional efficacy, domain-level patterns, modality parity, and enabling conditions of AI-supported financial education, this study investigates four research questions linking digital instructional exposure, behavioral metrics, and institutional covariates to measurable learning outcomes:

a. To what extent does AI-supported financial literacy instruction yield greater posttest learning gains than conventional teacher-led instruction among secondary school students after controlling for baseline prior knowledge?
b. Across which financial literacy domains—budgeting, saving, needs versus wants, and digital financial safety—does AI-supported instruction generate the largest differential learning gains relative to conventional instruction?
c. Do the four AI instructional tools (educational chatbot, interactive financial learning application, AI-generated case studies, and adaptive quiz platform) differ significantly in learning effectiveness or student satisfaction, and is the AI advantage consistent across urban general and rural vocational schools?
d. How do digital exposure metrics and enabling conditions—specifically device access, internet connectivity, teacher AI training, prior AI learning experience,

number of AI sessions, interest in financial literacy, and AI learning satisfaction—relate to and moderate students' financial literacy gains?

3. Research Methods

Research Design and Approach. This study adopted a Quantitative Educational Research approach using a quasi-experimental pretest–posttest nonequivalent control group design.[12] This analytical design was selected because intact public school classrooms in Balkh Province did not permit individual-level random assignment due to administrative and scheduling constraints, whereas a controlled pretest–posttest framework enabled rigorous causal approximation through explicit baseline equivalence testing, pretest covariate adjustment via Analysis of Covariance (ANCOVA), multivariable contextual controls, and within-school condition balancing. All participating students completed a standardized financial literacy pretest before instruction, participated in either AI-supported instruction or conventional teacher-led instruction covering identical domains, and completed an equivalent posttest immediately afterward.

Data Sources and Participants. Data were collected in 2026 across 14 public secondary schools located in seven districts of Balkh Province, northern Afghanistan: eight urban general public lycees in Mazar-e-Sharif and six rural agricultural or multi-disciplinary vocational lycees in Dawlatabad, Balkh District, Charbolak, Chimtal, Sholgara, and Charkent. The study corpus comprised 350 male students enrolled in grades 10 to 12, aged 15 to 19 years ($M = 16.33$, $SD = 1.01$), representing four home-language groups: Dari (67.4%), Uzbeki, Pashto, and Turkmen. Through stratified sampling, exactly 25 eligible students were drawn from each school to ensure balanced institutional representation. Both instructional conditions were represented within every participating school to limit school-level confounding: 167 students (47.7%; 92 urban, 75 rural) received AI-supported instruction, and 183 students (52.3%; 108 urban, 75 rural) received conventional teacher-led instruction. Table 1 summarizes the institutional sampling frame and student allocation.

[12] William R. Shadish, Thomas D. Cook, and Donald T. Campbell, *Experimental and Quasi-Experimental Designs for Generalized Causal Inference* (Boston: Houghton Mifflin, 2002), 135–70.

Table 1. Sampling frame of participating schools and allocation of students to instructional conditions

No.	School	District	School Type	Setting	N	AI	Conv.
1	Molavi Mohammad Hanif Balkhi	Mazar-e-Sharif	General public	Urban	25	15	10
2	Haji Qari Abdul Rauf	Mazar-e-Sharif	General public	Urban	25	9	16
3	Balkh Hawaei	Mazar-e-Sharif	General public	Urban	25	9	16
4	Mohammad Yusuf Khairkhwah	Mazar-e-Sharif	General public	Urban	25	11	14
5	Ali Sina Balkhi	Mazar-e-Sharif	General public	Urban	25	13	12
6	Wazir Mohammad Akbar Khan	Mazar-e-Sharif	General public	Urban	25	11	14
7	Sultan Ghiyasuddin	Mazar-e-Sharif	General public	Urban	25	13	12
8	Haji Walid	Mazar-e-Sharif	General public	Urban	25	11	14
9	Dawlatabad Agricultural	Dawlatabad	Agricultural vocational	Rural	25	12	13
10	Balkh Agricultural	Balkh District	Agricultural vocational	Rural	25	11	14
11	Charbolak Multi-Disciplinary	Charbolak	Multi-disciplinary vocational	Rural	25	15	10
12	Chimtal Multi-Disciplinary	Chimtal	Multi-disciplinary vocational	Rural	25	11	14
13	Sholgara Multi-Disciplinary	Sholgara	Multi-disciplinary vocational	Rural	25	14	11
14	Charkent Multi-Disciplinary	Charkent	Multi-disciplinary vocational	Rural	25	12	13
	Total	**7 districts**			**350**	**167**	**183**

Source: Authors' own data analysis.
Note. AI = AI-supported instruction; Conv. = conventional instruction. School names are shortened.

Students assigned to the AI-supported condition learned financial literacy through one of four tools: an AI educational chatbot ($n = 42$), an interactive financial learning application ($n = 48$), AI-generated case studies ($n = 50$), or an adaptive quiz platform ($n = 27$), completing between 5 and 10 AI-supported sessions ($M = 7.44$, $SD = 1.72$). Students in the comparison group studied the exact same four financial literacy domains exclusively through conventional teacher-led classroom instruction without AI support.

Data Collection and Preprocessing Techniques. Primary learner data were collected through a researcher-developed 20-point financial literacy achievement test paired with structured student background and attitude instruments. Aligned with established financial literacy frameworks and instructional objectives, the test assessed four core domains—budgeting, saving, needs versus wants, and digital financial safety—

using 0–5 subscales that summed to a 0–20 total score. Content validity was verified by domain specialists in financial education, educational technology, and educational measurement, who reviewed item clarity, relevance, appropriateness, and curricular alignment before final revision. The revised instrument was pilot-tested with secondary students outside the main study sample to evaluate item clarity and suitability. Internal consistency reliability was subsequently evaluated using Cronbach's alpha for both the composite financial literacy test and each of the four domain subscales to determine the measurement consistency of the instrument prior to main data collection. Supplementary attitudinal measures captured student interest in financial literacy (1–5 rating, all students) and satisfaction with AI-supported learning (1–5 rating, AI group only), alongside categorical enabling indicators (setting, grade, age, home language, device access, internet access, prior AI learning, and teacher AI training). Table 2 details all study variables and their analytical roles.

Table 2. Study variables, scoring, and use in the analysis

Variable	Measurement	Range	Role in Analysis
Financial literacy (pretest, posttest)	Total test score	0–20	Primary outcome; covariate (pretest)
Budgeting	Subscale score	0–5	Domain outcome
Saving	Subscale score	0–5	Domain outcome
Needs versus wants	Subscale score	0–5	Domain outcome
Digital financial safety	Subscale score	0–5	Domain outcome
Learning gain	Posttest minus pretest	−20 to 20	Gain-score comparisons
Instruction group	AI-supported vs. conventional	2 groups	Independent variable
AI tool	Chatbot, app, case studies, quiz platform	4 tools	Tool comparison (AI group only)
Number of AI sessions	Count	5–10	Dose–response (AI group only)
Interest in financial literacy	Rating	1–5	Attitude outcome (all students)
AI learning satisfaction	Rating	1–5	Attitude outcome (AI group only)
Context variables	Setting, grade, age, language, device, internet, prior AI, teacher AI training	Categorical	Baseline equivalence; moderators; covariates

Source: Authors' own data analysis.

Data Analysis Techniques. Statistical analysis followed a sequential inferential pipeline aligned with the four research questions, conducted on the complete data set of 350 students with no missing values on the outcome variables. First, baseline equivalence across pretest scores and contextual characteristics was verified using Welch two-sample *t*-tests and Pearson chi-square (χ^2) tests. Second, within-group pretest-to-posttest learning gains were evaluated using paired-samples *t*-tests with standardized mean

change (d_z), whereas between-group gain differences were tested via Welch t-tests with Hedges' g as the bias-corrected effect size interpreted against Cohen's conventional benchmarks.[13] Third, the primary treatment effect was estimated using Analysis of Covariance (ANCOVA), specified as a linear regression of posttest score on instructional condition controlling for pretest score after confirming homogeneity of regression slopes (group × pretest interaction $p = .430$).

Fourth, domain-level effects were evaluated across the four 5-point subscales using parallel paired and Welch t-tests; effectiveness and satisfaction differences across the four AI tools and session-dosage brackets were analyzed using one-way Analysis of Variance (ANOVA) F-tests and Pearson correlation (r); and equity moderation across urban–rural settings and enabling conditions was tested through group-by-condition interaction regressions and a fully adjusted multiple regression controlling simultaneously for pretest score, rural setting, device access, internet connectivity, previous AI learning, teacher AI training, grade, and age. Finally, three robustness checks were performed to verify stability: a school-level paired difference analysis across all 14 schools, a distribution-free Mann–Whitney test, and an intraclass correlation (ICC) check to rule out school-level clustering inflation. All statistical tests were two-tailed with a significance threshold of $\alpha = .05$.

B. RESULTS

1. Sample Characteristics and Baseline Equivalence

Establishing pretreatment comparability is essential for causal inference in a quasi-experimental design without random assignment. Across all demographic, linguistic, infrastructural, and prior-knowledge indicators, the AI-supported ($n = 167$) and conventional ($n = 183$) groups exhibited statistical equivalence at baseline (all $p > .05$; Table 3). More than half of the participating students attended urban general schools (57.1%), spoke Dari as their primary home language (67.4%), and reported personal or household access to a digital device (67.1%), whereas fewer than half had internet connectivity (48.9%) or prior experience learning with AI tools (44.3%). Crucially,

[13] Larry V. Hedges, "Distribution Theory for Glass's Estimator of Effect Size and Related Estimators," *Journal of Educational Statistics* 6, no. 2 (1981): 107–28, https://doi.org/10.3102/10769986006002107; Jacob Cohen, *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. (Hillsdale, NJ: Lawrence Erlbaum Associates, 1988), 24–27; Daniël Lakens, "Calculating and Reporting Effect Sizes to Facilitate Cumulative Science: A Practical Primer for t-Tests and ANOVAs," *Frontiers in Psychology* 4 (2013): 863, https://doi.org/10.3389/fpsyg.2013.00863.

baseline financial literacy scores were virtually identical between conditions, with the AI-supported group averaging 10.08 ($SD = 1.92$) and the conventional group averaging 10.02 ($SD = 2.07$) out of 20 points, $t(347.9) = 0.32$, $p = .753$, Hedges' $g = 0.03$. Because both groups entered the instructional period at approximately 50% baseline mastery with balanced background profiles, posttest divergences cannot be attributed to preexisting disparities in student readiness, demographic composition, or school-level resource exposure.

Table 3. Profile of students by instructional condition and baseline equivalence tests

Characteristic	AI-supported (n = 167)	Conventional (n = 183)	Test of difference
Setting			$\chi^2(1) = 0.40$, $p = .527$
Urban	92 (55.1)	108 (59.0)	
Rural	75 (44.9)	75 (41.0)	
Grade			$\chi^2(2) = 1.10$, $p = .576$
Grade 10	53 (31.7)	57 (31.1)	
Grade 11	63 (37.7)	78 (42.6)	
Grade 12	51 (30.5)	48 (26.2)	
Age (years), *M* (*SD*)	16.35 (0.98)	16.31 (1.04)	$t(348) = 0.33$, $p = .740$
Home language			$\chi^2(3) = 2.08$, $p = .556$
Dari	107 (64.1)	129 (70.5)	
Uzbeki	39 (23.4)	38 (20.8)	
Pashto	14 (8.4)	10 (5.5)	
Turkmen	7 (4.2)	6 (3.3)	
Device access (yes)	107 (64.1)	128 (69.9)	$\chi^2(1) = 1.11$, $p = .292$
Internet access (yes)	82 (49.1)	89 (48.6)	$\chi^2(1) = 0.00$, $p = 1.000$
Previous AI learning (yes)	71 (42.5)	84 (45.9)	$\chi^2(1) = 0.28$, $p = .597$
Teacher received AI training (yes)	84 (50.3)	101 (55.2)	$\chi^2(1) = 0.65$, $p = .419$

Source: Authors' own data analysis.
Note. Cells show n (%) unless otherwise stated.

2. Overall and Covariate-Adjusted Learning Gains (RQ1)

Addressing the first research question, both instructional conditions generated statistically significant pretest-to-posttest improvement, yet students receiving AI-supported instruction achieved more than twice the mean learning gain of peers taught conventionally (Table 4; Figure 1). Mean scores in the AI-supported group advanced from 10.08 to 13.74 points (mean gain = 3.65, $SD = 1.45$), compared with an advance from 10.02 to 11.60 points in the conventional group (mean gain = 1.58, $SD = 1.43$). Distributional thresholds reinforce the breadth of this instructional shift: 97.6% of AI-supported students improved by at least 1 point compared with 78.7% of conventional

students, while 81.4% of AI-supported students achieved substantial gains of at least 3 points (15% of the full assessment scale) versus only 25.1% under conventional teaching. Conversely, score regression was nearly eliminated under AI support, occurring in just 1 student (0.6%) compared with 13 conventional students (7.1%).

Table 4. Financial literacy scores (0-20) before and after instruction

Group	n	Pretest *M* (*SD*)	Posttest *M* (*SD*)	Gain *M* (*SD*)	Gain ≥ 1 (%)	Gain ≥ 3 (%)
AI-supported	167	10.08 (1.92)	13.74 (2.53)	3.65 (1.45)	97.6	81.4
Conventional	183	10.02 (2.07)	11.60 (2.55)	1.58 (1.43)	78.7	25.1
All students	350	10.05 (2.00)	12.62 (2.75)	2.57 (1.77)	87.7	52.0

Source: Authors' own data analysis.
Note. Gain = posttest minus pretest.

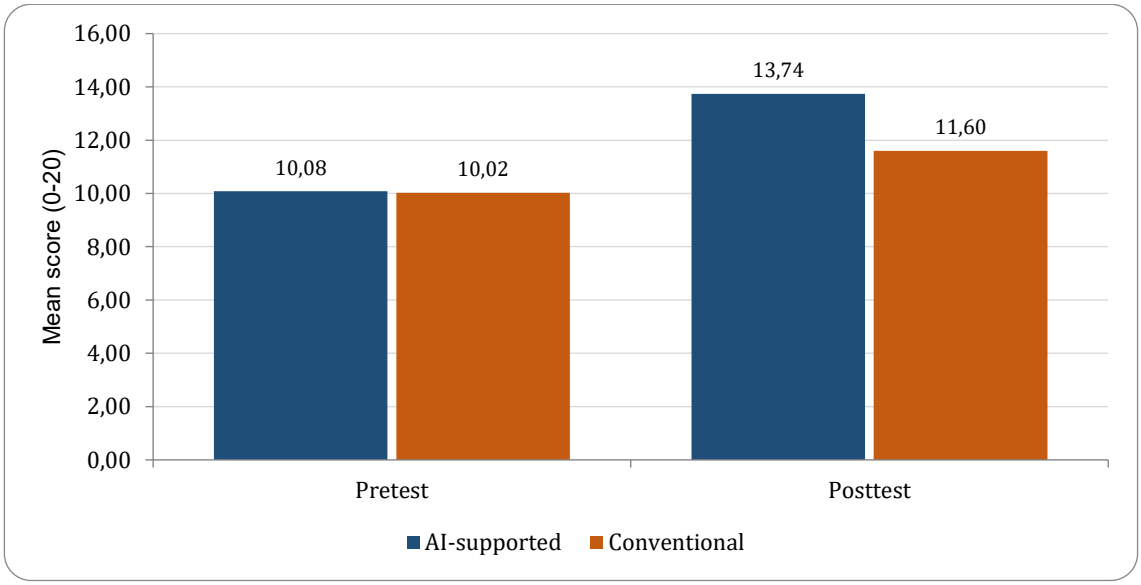


Figure 1. Mean financial literacy scores before and after instruction, by group

Source: Authors' own data analysis

As illustrated in Figure 1, the pretest baselines of the two cohorts overlapped almost completely before diverging sharply at posttest, visually confirming that the instructional divergence emerged strictly during the intervention window. Inferential comparisons in Table 5 quantify this divergence: within-group improvement was exceptionally large in the AI-supported condition ($t(166) = 32.51$, $p < .001$, $d_z = 2.52$) and large in the conventional condition ($t(182) = 14.97$, $p < .001$, $d_z = 1.11$). Crucially, the unadjusted between-group gain comparison favored AI-supported instruction by 2.07 points, 95% CI [1.77, 2.38], $t(343.9) = 13.46, p < .001$, representing a very large standardized effect size (Hedges' $g = 1.44$).

Table 5. Tests of change within groups and of differences between groups

Comparison	Mean difference [95% CI]	*t* (*df*)	*p*	Effect size
Pretest, AI vs. conventional	0.07 [-0.35, 0.49]	0.32 (347.9)	.753	g = 0.03
Posttest minus pretest, AI	3.65 [3.43, 3.87]	32.51 (166)	< .001	dz = 2.52
Posttest minus pretest, conventional	1.58 [1.37, 1.79]	14.97 (182)	< .001	dz = 1.11
Gain, AI vs. conventional	**2.07 [1.77, 2.38]**	**13.46 (343.9)**	**< .001**	**g = 1.44**

Source: Authors' own data analysis.
Note. dz = standardized mean change for paired data; g = Hedges' g; CI = confidence interval.

To control directly for individual prior knowledge, an Analysis of Covariance (ANCOVA) regressed posttest scores on instructional group and pretest scores after verifying homogeneity of regression slopes (group × pretest interaction, $p = .430$). In the covariate-adjusted model (Table 6), pretest scores strongly predicted posttest achievement ($b = 1.05$, $p < .001$), and AI-supported instruction retained an exact 2.07-point advantage over conventional instruction, 95% CI [1.77, 2.37], $t(347) = 13.45$, $p < .001$, partial $\eta^2 = 0.34$. Covariate-adjusted posttest means reached 13.70 for the AI-supported group versus 11.63 for the conventional group, with the model explaining 72.9% of total posttest variance ($R^2 = 0.729$).

Table 6. ANCOVA model predicting posttest financial literacy

Predictor	*b*	*SE*	*t*	*p*	95% CI
Intercept	1.13	0.40	2.82	.005	[0.34, 1.91]
Pretest	1.05	0.04	27.17	< .001	[0.97, 1.12]
AI group	**2.07**	**0.15**	**13.45**	**< .001**	**[1.77, 2.37]**

Source: Authors' own data analysis.
Note. $R^2 = 0.729$; *df = 347. Reference group for AI group = conventional instruction.*

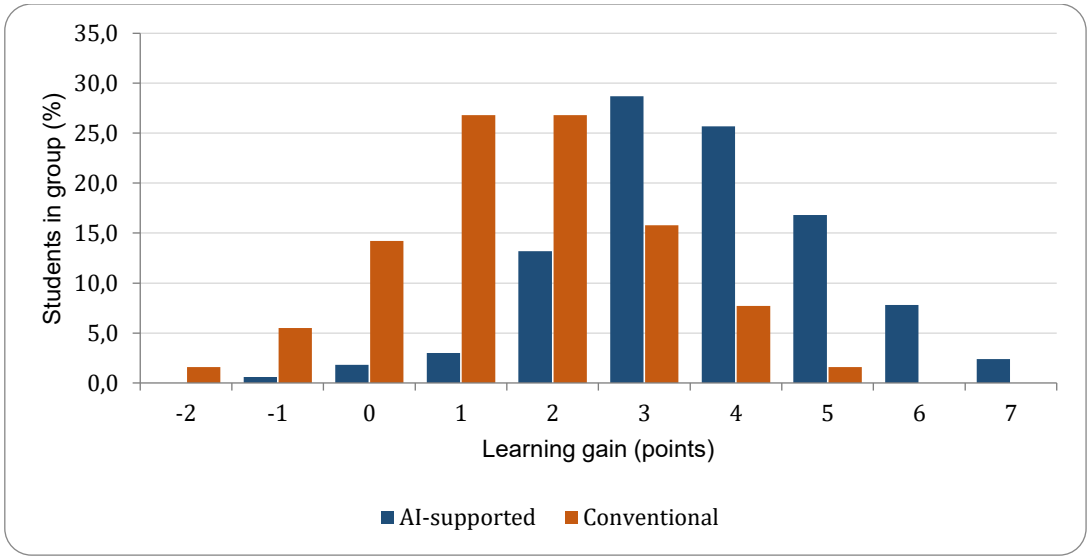


Figure 2. Distribution of learning gains (posttest minus pretest) by group

Source: Authors' own data analysis

Interpreting the gain distributions in Figure 2 confirms that the AI advantage reflects a systematic rightward shift across the entire learner distribution rather than an artifact driven by a small cluster of high-performing outliers.

3. Domain-Specific Financial Literacy Gains (RQ2)

Disaggregating total achievement across the four 5-point subscales addresses the second research question and reveals a pronounced split between applied procedural domains and conceptual domains (Tables 7 and 8; Figure 3). Students in the AI-supported condition achieved highly significant within-group improvements across all four domains (all $p < .001$), with the largest absolute progress in budgeting ($+1.04$), saving ($+0.98$), and digital financial safety ($+0.91$), followed by needs versus wants ($+0.73$). By contrast, conventional instruction produced modest gains in needs versus wants ($+0.58, p < .001$), digital financial safety ($+0.46, p < .001$), and saving ($+0.44, p = .004$), while failing to generate statistically meaningful improvement in budgeting ($+0.10, t(182) = 0.67, p = .503$).

Table 7. Domain scores (0-5) before and after instruction, by group

Domain	Group	Pretest *M* (*SD*)	Posttest *M* (*SD*)	Gain *M* (*SD*)	Paired *t* (*df* = 166 / 182)	*p*
Budgeting	AI	2.54 (1.41)	3.58 (1.35)	1.04 (1.93)	6.95	< .001
	Conv.	2.70 (1.42)	2.80 (1.45)	0.10 (1.98)	0.67	.503
Saving	AI	2.43 (1.35)	3.40 (1.44)	0.98 (1.86)	6.77	< .001
	Conv.	2.33 (1.38)	2.77 (1.47)	0.44 (2.02)	2.93	.004
Needs versus wants	AI	2.62 (1.37)	3.35 (1.40)	0.73 (1.79)	5.27	< .001
	Conv.	2.47 (1.41)	3.05 (1.33)	0.58 (1.90)	4.12	< .001
Digital financial safety	AI	2.50 (1.40)	3.41 (1.34)	0.91 (1.82)	6.45	< .001
	Conv.	2.52 (1.52)	2.98 (1.36)	0.46 (1.81)	3.48	< .001

Source: Authors' own data analysis.
Note. Paired *t* has 166 degrees of freedom for the AI group and 182 for the conventional group.

Table 8. Between-group differences in domain gains (AI-supported minus conventional)

Domain	Difference	95% CI	*t* (*df*)	*p*	*g*
Budgeting	**0.94**	**[0.53, 1.35]**	**4.48 (347)**	**< .001**	**0.48**
Saving	0.54	[0.13, 0.95]	2.60 (348)	.010	0.28
Needs versus wants	0.15	[-0.24, 0.54]	0.77 (348)	.444	0.08
Digital financial safety	0.45	[0.06, 0.83]	2.29 (344)	.022	0.25

Source: Authors' own data analysis.

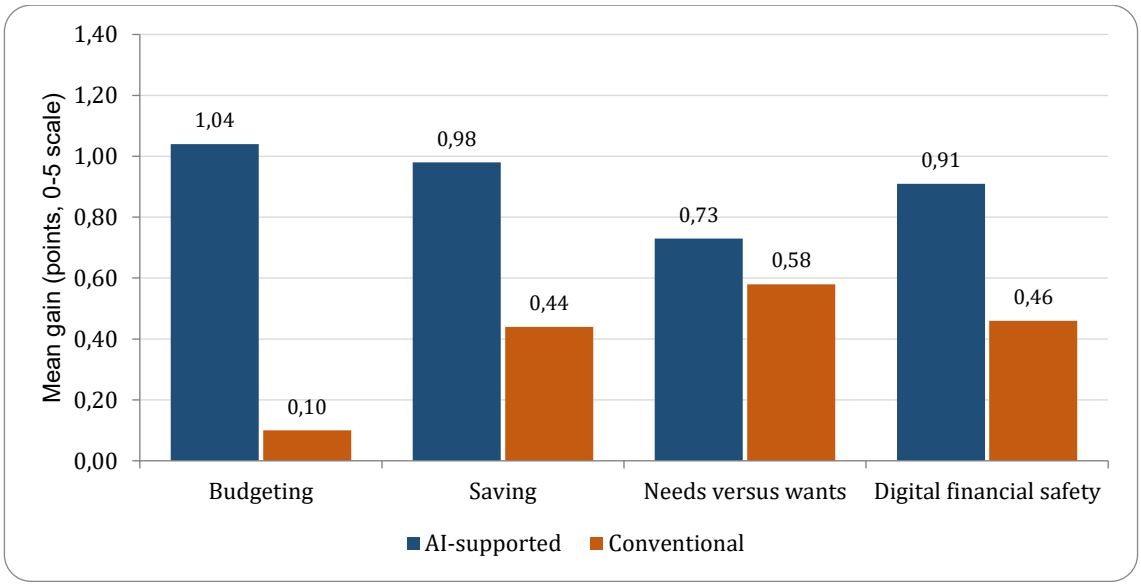


Figure 3. Mean gain in each financial literacy domain, by group

Source: Authors' own data analysis

Between-group domain contrasts (Table 8; Figure 3) establish that the overall AI advantage was concentrated in domains requiring quantitative allocation, trade-off evaluation, and applied risk recognition: budgeting yielded the largest differential advantage (+0.94 points, $t(347) = 4.48$, $p < .001$, $g = 0.48$), followed by saving (+0.54 points, $p = .010$, $g = 0.28$) and digital financial safety (+0.45 points, $p = .022$, $g = 0.25$). In contrast, the between-group difference in needs versus wants was small and statistically non-significant (+0.15 points, $t(348) = 0.77$, $p = .444$, $g = 0.08$), indicating that conventional classroom instruction already achieved near-parity on this conceptual classification skill.

4. **Effectiveness Across AI Tools, School Settings, and Districts (RQ3)**

To address the third research question, learning outcomes were disaggregated first by AI instructional modality and subsequently by urban–rural school setting and administrative district. Every one of the four AI tools outperformed conventional instruction by a very large margin (all $p < .001$), with effect sizes ranging from $g = 1.20$ to $g = 1.53$ (Table 9; Figure 4). Mean gains clustered tightly across the AI educational chatbot ($M = 3.74$, $g = 1.46$), interactive financial learning application ($M = 3.73$, $g = 1.53$), AI-generated case studies ($M = 3.68$, $g = 1.50$), and adaptive quiz platform ($M = 3.33$, $g = 1.20$). Crucially, one-way ANOVA tests revealed no statistically significant differences among the four tools in either learning gain, $F(3,163) = 0.53$, $p = .662$, or student satisfaction, $F(3,163) = 0.26$, $p = .854$. Although the adaptive quiz platform

posted the lowest descriptive mean gain and satisfaction rating, it also had the smallest subgroup sample ($n = 27$) and widest estimation variance, making that numerical dip tentative rather than evidence of inferiority.

Table 9. Learning outcomes by AI tool compared with conventional instruction

Instruction	n	Pretest *M* (*SD*)	Posttest *M* (*SD*)	Gain *M* (*SD*)	*g* vs. conv.	Satisfaction *M*
AI educational chatbot	42	9.88 (2.35)	13.62 (3.19)	3.74 (1.68)	1.46	3.83
Interactive financial learning app	48	9.90 (1.63)	13.62 (2.13)	3.73 (1.32)	1.53	3.85
AI-generated case studies	50	10.26 (1.88)	13.94 (2.20)	3.68 (1.28)	1.50	3.76
Adaptive quiz platform	27	10.41 (1.78)	13.74 (2.71)	3.33 (1.62)	1.20	3.67
Conventional instruction	183	10.02 (2.07)	11.60 (2.55)	1.58 (1.43)	-	-

Source: Authors' own data analysis.
Note. g = Hedges' g for the tool's gain compared with the conventional group's gain. Satisfaction on a 1-5 scale.

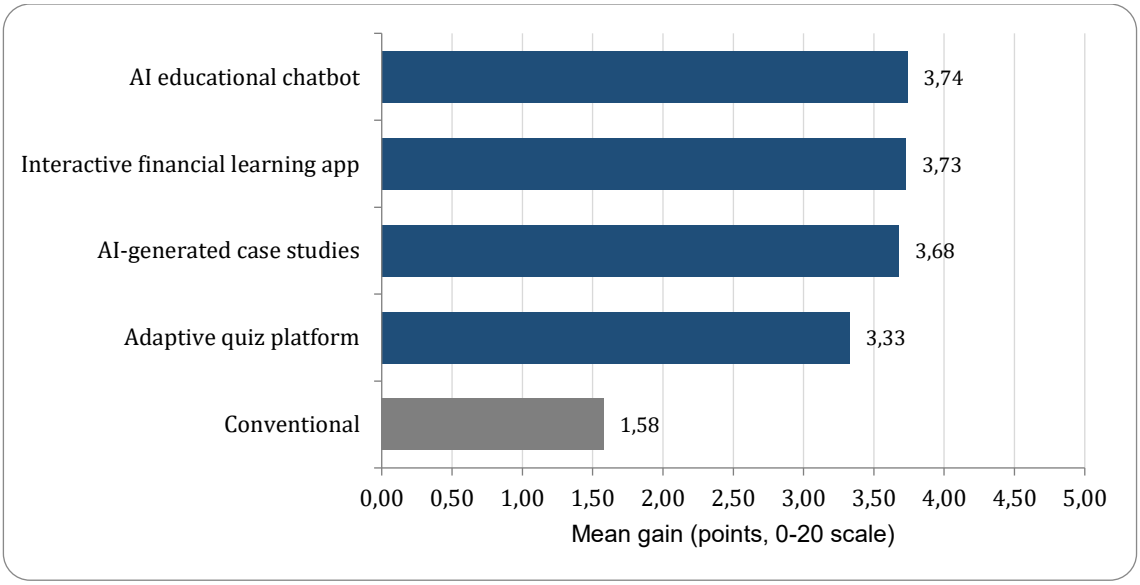


Figure 4. Mean learning gain by AI tool and for conventional instruction

Source: Authors' own data analysis

Across geographical and institutional contexts (Table 10; Figure 5), AI-supported instruction demonstrated robust consistency across both urban general lycees and rural vocational lycees. Urban students learned substantially more under AI support ($M = 3.63, SD = 1.51$) than under conventional instruction ($M = 1.79, SD = 1.45$), $t(190.1) = 8.77$, $p < .001$, $g = 1.24$. Notably, rural AI-supported students achieved practically identical gains ($M = 3.68, SD = 1.39$), whereas rural conventional students recorded the lowest learning gains in the study ($M = 1.28$, $SD = 1.35$), producing an even larger standardized AI advantage in rural vocational schools, $t(147.9) = 10.73$, $p < .001$, $g =$

1.74. While the formal group-by-setting interaction fell just short of statistical significance ($p = .073$), Figure 5 highlights a compelling equity pattern: AI-supported instruction fully closed the urban–rural achievement gap that otherwise widened under conventional teaching.

Table 10. Learning outcomes by school setting and instructional condition

Setting	Group	n	Pretest *M* (*SD*)	Posttest *M* (*SD*)	Gain *M* (*SD*)
Urban	AI-supported	92	10.04 (2.00)	13.67 (2.53)	3.63 (1.51)
	Conventional	108	10.02 (1.82)	11.81 (2.35)	1.79 (1.45)
Rural	AI-supported	75	10.13 (1.84)	13.81 (2.55)	3.68 (1.39)
	Conventional	75	10.01 (2.40)	11.29 (2.80)	1.28 (1.35)

Source: Authors' own data analysis.
Note. Group difference in gain: urban $t(190.1) = 8.77$, $p < .001$; rural $t(147.9) = 10.73$, $p < .001$.

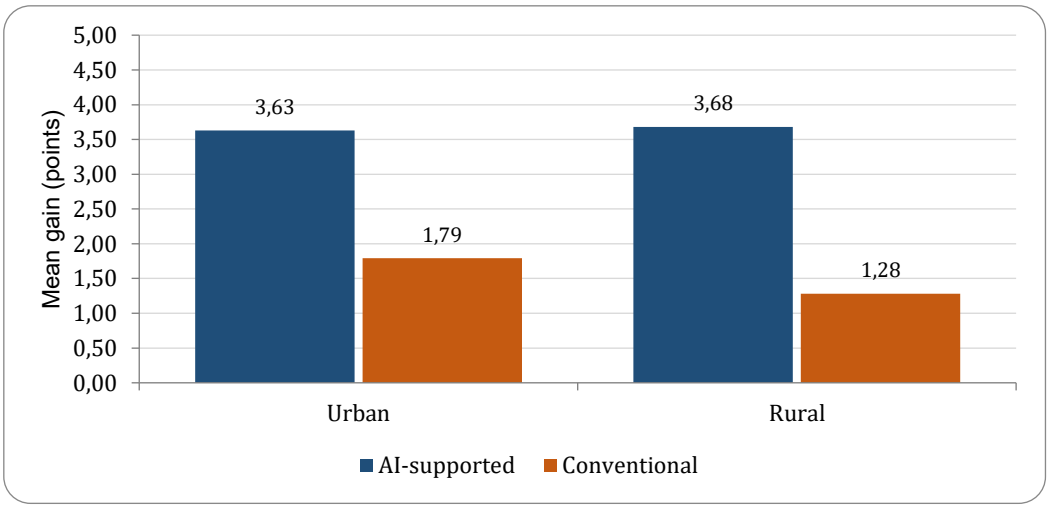


Figure 5. Mean learning gain by school setting and group

Source: Authors' own data analysis

District-level disaggregation confirms that this pattern was geographically pervasive rather than confined to a single district (Table 11; Figure 6). Across all seven districts of Balkh Province, AI-supported students outperformed conventional peers, with mean gain differentials spanning from +1.35 points in Charkent and +1.54 points in Chimtal to +3.02 points in Balkh District and +3.26 points in Dawlatabad.

Table 11. Learning outcomes by district

District	n	AI	Conv.	Pretest *M*	Posttest *M*	AI gain	Conv. gain	Difference
Mazar-e-Sharif	200	92	108	10.03	12.66	3.63	1.79	1.84
Dawlatabad	25	12	13	10.00	12.64	4.33	1.08	3.26
Balkh District	25	11	14	10.24	12.64	4.09	1.07	3.02
Charbolak	25	15	10	10.16	13.20	3.93	1.70	2.23
Chimtal	25	11	14	9.96	12.28	3.18	1.64	1.54
Sholgara	25	14	11	10.72	13.40	3.93	1.09	2.84

District	n	AI	Conv.	Pretest *M*	Posttest *M*	AI gain	Conv. gain	Difference
Charkent	25	12	13	9.36	11.16	2.50	1.15	1.35

Source: Authors' own data analysis.
Note. All values are means; AI and Conv. give the number of students in each condition.

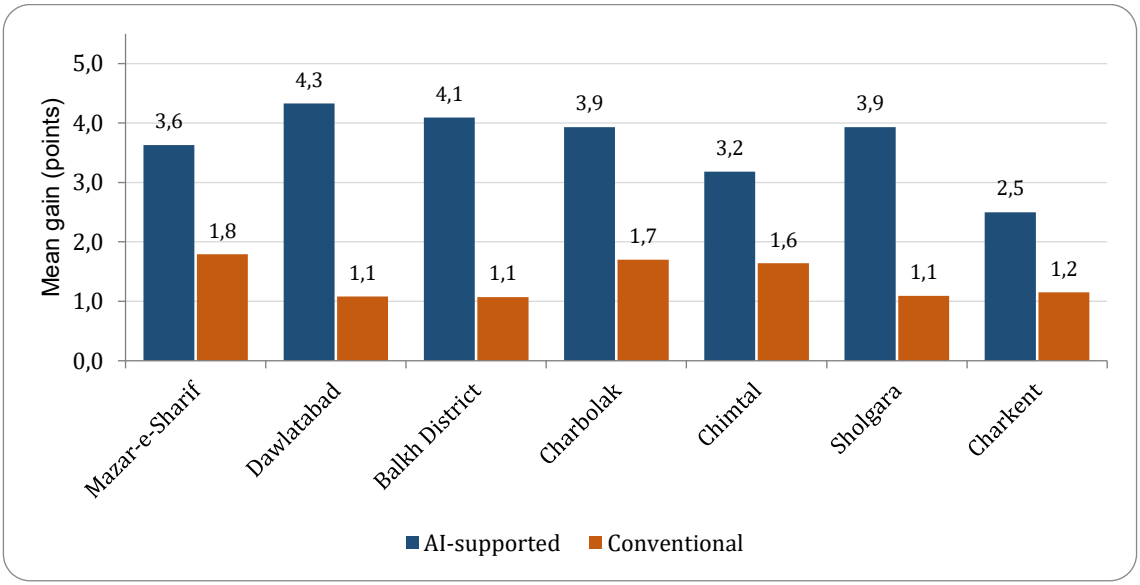


Figure 6. Mean learning gain by district and group

Source: Authors' own data analysis

5. Enabling Conditions, Multivariable Covariates, Session Dosage, and Student Attitudes (RQ4)

Addressing the fourth research question, subgroup moderation tests (Table 12) and multivariable regression modeling (Table 13) isolated how infrastructural, instructional, behavioral, and attitudinal metrics shaped student learning gains. Neither personal internet connectivity (interaction $p = .166$) nor previous AI learning experience (interaction $p = .262$) significantly influenced gains in either instructional condition. In contrast, student device access exhibited an asymmetric pattern: under conventional instruction, students with device access gained nearly twice as much as peers without devices (1.86 vs. 0.93 points, $t = 4.83$, $p < .001$), whereas within the AI-supported condition that access gap narrowed to a non-significant 0.34 points (3.78 vs. 3.43, $p = .167$; group × device interaction $p = .066$). Teacher AI training similarly elevated student learning, yielding statistically significant gains under conventional instruction (1.80 vs. 1.30, $p = .019$) and a parallel positive directional increment in the AI-supported group (3.83 vs. 3.47, $p = .106$; interaction $p = .663$).

Table 12. Mean learning gain by enabling condition and instructional condition

Condition	Level	AI n	AI gain *M* (*SD*)	Conv. n	Conv. gain *M* (*SD*)	Group × condition *p*
Device access	Yes	107	3.78 (1.34)	128	1.86 (1.48)	.066
	No	60	3.43 (1.62)	55	0.93 (1.05)	
Internet access	Yes	82	3.72 (1.39)	89	1.43 (1.43)	.166
	No	85	3.59 (1.51)	94	1.72 (1.42)	
Teacher received AI training	Yes	84	3.83 (1.50)	101	1.80 (1.41)	.663
	No	83	3.47 (1.38)	82	1.30 (1.41)	
Previous AI learning	Yes	71	3.51 (1.59)	84	1.63 (1.55)	.262
	No	96	3.76 (1.34)	99	1.54 (1.32)	
School setting	Rural	75	3.68 (1.39)	75	1.28 (1.35)	.073
	Urban	92	3.63 (1.51)	108	1.79 (1.45)	

Source: Authors' own data analysis.
Note. p values in the last column test whether the AI-supported advantage differs between levels of the condition (interaction in a regression of gain on group, condition, and their product).

When all contextual predictors were estimated simultaneously alongside baseline pretest scores (Table 13), the multivariable regression explained 75.1% of posttest variance ($R^2 = 0.751$, $df = 340$). AI-supported instruction remained the strongest instructional predictor of posttest financial literacy ($b = 2.13$, $\beta = 0.39$, $t = 14.25$, $p <$ .001, 95% CI [1.84, 2.43]), slightly exceeding its unadjusted and ANCOVA estimates. Beyond baseline knowledge ($b = 1.04$, $\beta = 0.76$, $p < .001$), only two institutional enabling factors exerted statistically significant independent effects: student device access ($b = 0.64$, $\beta = 0.11$, $t = 3.91$, $p < .001$) and teacher AI training ($b = 0.48$, $\beta =$ 0.09, $t = 3.02$, $p = .003$). Rural school location ($p = .552$), internet access ($p = .183$), prior AI experience ($p = .218$), grade level ($p = .437$), and age ($p = .157$) showed no independent predictive effect after covariate adjustment.

Table 13. Multiple regression predicting posttest financial literacy

Predictor	*b*	*SE*	β	*t*	*p*	95% CI
Intercept	1.96	1.29	-	1.52	.128	[-0.57, 4.49]
Pretest	1.04	0.04	0.76	27.51	< .001	[0.97, 1.11]
AI-supported instruction	**2.13**	**0.15**	**0.39**	**14.25**	**< .001**	**[1.84, 2.43]**
Rural school	-0.11	0.18	-0.02	-0.59	.552	[-0.46, 0.25]
Device access	0.64	0.16	0.11	3.91	< .001	[0.32, 0.97]
Internet access	-0.22	0.16	-0.04	-1.33	.183	[-0.53, 0.10]
Previous AI learning	-0.20	0.16	-0.04	-1.23	.218	[-0.51, 0.12]
Teacher AI training	0.48	0.16	0.09	3.02	.003	[0.17, 0.79]
Grade	0.11	0.14	0.03	0.78	.437	[-0.16, 0.37]
Age	-0.15	0.10	-0.05	-1.42	.157	[-0.35, 0.06]

Source: Authors' own data analysis.

Note. R^2 = 0.751; *df* = 340. Binary predictors are coded 1 = yes (or rural, or AI-supported).

Within the AI-supported cohort, instructional dosage exhibited a weak, statistically non-significant positive correlation with learning gains ($r = 0.13$, $p = .091$). Students completing 5–6 sessions gained 3.32 points, those completing 7–8 sessions gained 3.90 points, and those completing 9–10 sessions gained 3.75 points; differences across dosage brackets did not reach statistical significance, $F(2,164) = 2.52$, $p = .084$ (Figure 7). Crucially, Figure 7 demonstrates that even the minimum observed dosage of 5–6 AI-supported sessions generated learning gains more than double the conventional benchmark (1.58 points), with performance plateauing beyond 7–8 sessions.

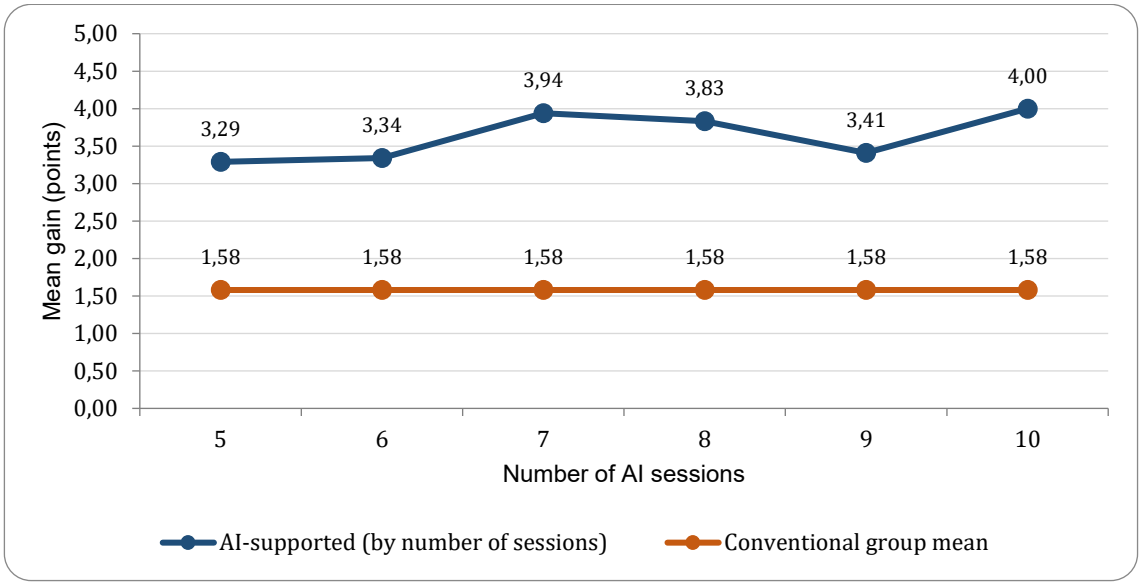


Figure 7. Mean learning gain of AI-supported students by number of sessions, with the conventional group mean for reference

Source: Authors' own data analysis

Finally, attitudinal data revealed a clear empirical decoupling between learner perceptions and objective test performance (Table 14; Figure 8). Both groups expressed moderate, statistically indistinguishable interest in financial literacy (AI: $M = 3.57$, $SD = 0.89$; conventional: $M = 3.60$, $SD = 1.05$; $t(346) = -0.31$, $p = .757$). Among AI-supported students, overall satisfaction with AI learning was relatively high ($M = 3.79$, $SD = 0.96$), with 68.3% assigning ratings of 4 or 5 on the 5-point scale. However, neither AI learning satisfaction ($r = 0.04$, $p = .586$) nor financial literacy interest (AI group: $r = -0.01$, $p = .905$; conventional group: $r = -0.09$, $p = .206$) correlated with learning gain, confirming that positive subjective reception did not drive—and could not serve as a

proxy for—cognitive achievement.

Table 14. Interest in financial literacy and satisfaction with AI-supported learning (1-5 scales)

Measure	n	*M* (*SD*)	Rating 4-5 (%)
Interest in financial literacy, AI	167	3.57 (0.89)	52.1
Interest in financial literacy, Conv.	183	3.60 (1.05)	59.0
Satisfaction with AI learning (AI only)	167	3.79 (0.96)	68.3

Source: Authors' own data analysis.
Note. Satisfaction was collected only from students in the AI-supported condition.

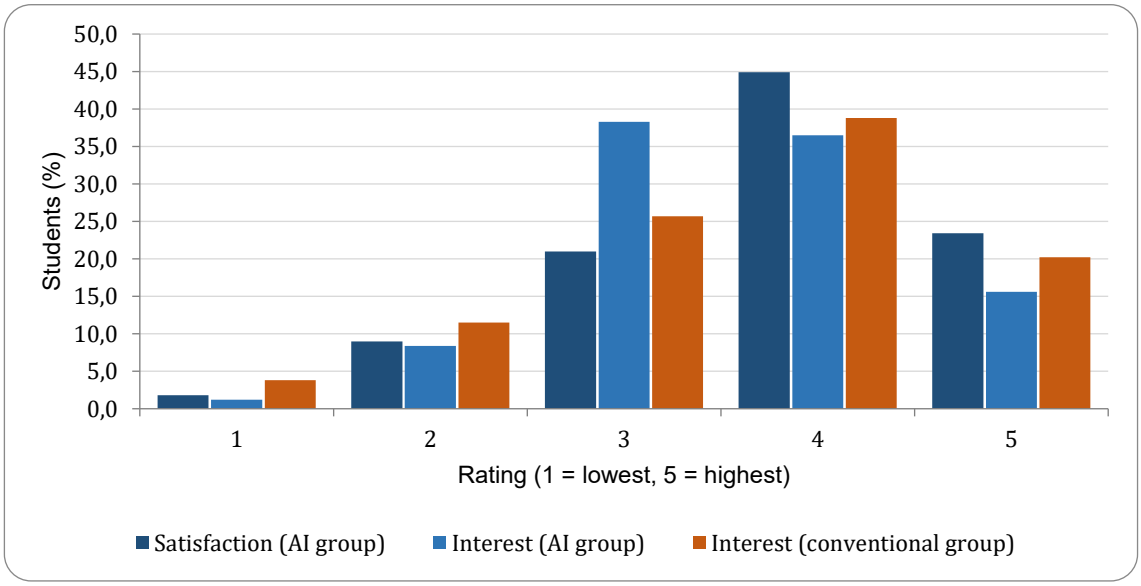


Figure 8. Distribution of interest and satisfaction ratings

6. Robustness and Sensitivity Verification

Three complementary sensitivity checks verified that the primary treatment effect was neither an artifact of distributional assumptions nor inflated by school-level clustering. First, exploiting the within-school allocation of both instructional conditions, school-level mean gain differences were positive in all 14 participating schools, ranging from 0.95 to 3.26 points with a school-level mean difference of 2.08 points, $t(13) = 11.70$, $p < .001$. Second, a nonparametric Mann–Whitney test confirmed the statistical superiority of AI-supported gains without assuming normality ($p < .001$). Third, the intraclass correlation (ICC) of learning gains across schools was approximately zero, demonstrating that variance clustering within individual schools did not bias standard errors or inflate inferential significance.

C. DISCUSSION

1. Pedagogical Mechanisms Behind Overall and Domain-Specific Learning Gains

The empirical evidence demonstrates that structured AI-supported instruction produces substantially larger gains in adolescent financial literacy than conventional teacher-led teaching. Starting from statistically equivalent baselines near half of the 20-point assessment ($M = 10.08$ vs. 10.02), AI-supported students achieved approximately 2.3 times the mean improvement of their conventionally taught peers (3.65 vs. 1.58 points), maintaining an exact covariate-adjusted advantage of 2.07 points ($g = 1.44$; partial $\eta^2 = 0.34$) that held across multivariable regression ($b = 2.13$, $R^2 = 0.751$), school-level aggregation across all 14 schools, and nonparametric sensitivity testing. In scholarly dialogue with prior research, these findings directly confirm causal meta-analyses showing that structured financial education elevates financial knowledge[14] and corroborate meta-analytic syntheses indicating that intelligent tutoring and adaptive digital systems outperform conventional classroom instruction.[15] Crucially, however, the standardized effect size observed here ($g = 1.44$) exceeds the moderate-to-large benchmarks typically reported in those meta-analyses. Rather than signaling an unqualified technological panacea, this larger magnitude likely reflects three contextual and methodological factors: low initial baseline mastery leaving substantial room for upward movement, tight instructional alignment between the AI practice tasks and the four tested domains, and immediate posttest measurement within a non-randomized design where short-term knowledge acquisition is naturally more pronounced than long-term behavioral retention.

Interpreting these gains through instructional and cognitive learning mechanisms clarifies why AI support outperformed lecture-based teaching. As systematic reviews of youth financial education emphasize, conventional classroom programs frequently underperform because passive exposition deprives learners of iterative practice, realistic decision contexts, and timely error correction.[16] Each of the four AI modalities examined in this study systematically repaired those instructional deficits by delivering immediate formative feedback, calibrated difficulty, and repeated exposure to realistic financial

[14] Kaiser et al., "Financial Education Affects Financial Knowledge," 255–72; Kaiser and Menkhoff, "Does Financial Education Impact Financial Literacy," 611–30; Urban et al., "Effects of High School Personal Financial Education Policies," 101786.
[15] Ma et al., "Intelligent Tutoring Systems and Learning Outcomes," 901–18; Kulik and Fletcher, "Effectiveness of Intelligent Tutoring Systems," 42–78.
[16] Amagir et al., "Review of Financial-Literacy Education Programs," 56–80.

scenarios—mechanisms that align closely with experimental evidence showing that AI-enabled personalization strengthens learner engagement and academic outcomes.[17] From a cognitive-processing and self-regulated learning perspective, immediate corrective scaffolding reduces extraneous cognitive load during multi-step financial problem solving, prevents misconceptions from consolidating, and prompts learners to monitor and revise flawed choices in real time rather than waiting for delayed summative grading.

The domain-level divergence reinforces this cognitive interpretation and refines existing curriculum theory. AI-supported instruction generated its largest differential advantage in budgeting (+0.94 points, $g = 0.48$)—the sole domain where conventional instruction failed to produce statistically significant progress (+0.10, $p = .503$)—followed by meaningful gains in saving ($g = 0.28$) and digital financial safety ($g = 0.25$). In contrast, between-group differences vanished for needs versus wants (+0.15 points, $g = 0.08$, $p = .444$), where teacher-led instruction achieved comparable improvement. This asymmetry demonstrates that AI efficacy is domain-contingent rather than uniform: budgeting, saving, and digital financial safety are procedural, computational, and risk-evaluation competencies requiring trade-off calculations under resource constraints and pattern recognition against digital threats, making them highly responsive to interactive simulations and immediate feedback. Conversely, distinguishing needs from wants is primarily a conceptual and value-laden classification skill that responds well to teacher-facilitated dialogue and everyday classroom examples.

2. Modality Parity, Infrastructural Equity, and Enabling Conditions

A central insight from the multi-tool comparison is the complete absence of statistically significant performance or satisfaction differences across the AI educational chatbot, interactive financial learning application, AI-generated case studies, and adaptive quiz platform ($F(3,163) = 0.53$, $p = .662$; all tools $g = 1.20$–1.53 vs. conventional instruction). Although the adaptive quiz platform recorded a slightly lower descriptive mean gain (3.33), its smaller subgroup ($n = 27$) widened estimation uncertainty, rendering that difference inconclusive. This convergence extends earlier critiques by Zawacki-Richter and colleagues, Hwang and colleagues, Chen, Chen, and Lin, and Zhai and colleagues, who cautioned against technology-centric implementations divorced from pedagogical design.[18] The Balkh evidence establishes functional pedagogical equivalence:

[17] Huang, Lu, and Yang, "Effects of Artificial Intelligence–Enabled Personalized Recommendations," 104684.

learning gains are driven by the core instructional functions embedded across tools—structured practice, authentic scenarios, and immediate formative feedback—rather than by any single software interface or proprietary brand.

Equally consequential is the equity trajectory across urban general and rural vocational schools. Whereas conventionally taught students in rural vocational lycees recorded the weakest gains in the sample (1.28 vs. 1.79 in urban schools), rural students receiving AI support achieved learning gains virtually identical to urban AI-supported peers (3.68 vs. 3.63), generating a larger standardized AI advantage in rural settings ($g = 1.74$ vs. $g = 1.24$) and consistent positive differentials across all seven districts. Although the formal group-by-setting interaction remained marginal ($p = .073$), the pattern suggests that structured AI scaffolding functions as an instructional equalizer that compensates for rural pedagogical resource constraints. The covariate analysis further illuminates how institutional access operates. Under conventional teaching, lacking a personal device severely penalized learning (0.93 vs. 1.86 points, $p < .001$), whereas within the AI-supported group that deficit narrowed sharply to a non-significant 0.34 points (interaction $p = .066$), and household internet access showed no independent effect ($p = .183$). Tentatively, structured school-based AI sessions appear to have buffered students who lacked personal devices or home connectivity, even as overall device access remained an independent positive predictor across the full sample ($b = 0.64$, $p < .001$). Crucially, teacher AI training exerted a statistically significant independent effect after full adjustment ($b = 0.48$, $p = .003$), directly refuting narratives of technological teacher replacement and confirming that educator preparation remains indispensable to translating digital tools into classroom learning.[18]

Finally, two behavioral and governance signals merit rigorous scrutiny. First, session dosage exhibited only a weak, non-significant correlation with learning gains ($r = 0.13$, $p = .091$), with gains rising modestly from 5–6 sessions (3.32) to 7–8 sessions (3.90) before plateauing at 9–10 sessions (3.75). Within this short instructional window, even five to six well-structured sessions unlocked the bulk of achievable gains, indicating diminishing marginal returns beyond moderate exposure. Second, attitudinal perceptions were entirely decoupled from cognitive mastery: although 68.3% of AI-supported

[18] Zawacki-Richter et al., "Systematic Review of Research," 39; Hwang et al., "Vision, Challenges, Roles and Research Issues," 100001; Chen, Chen, and Lin, "Artificial Intelligence in Education," 75264–78; Zhai et al., "Review of Artificial Intelligence (AI) in Education," 8812542.

students reported high satisfaction ($M = 3.79$), neither satisfaction ($r = 0.04$, $p = .586$) nor financial literacy interest ($r = -0.01$ and -0.09) predicted actual score improvement. Consistent with emerging warnings from Kasneci and colleagues and Chan and Hu regarding generative AI in education, relying on learner satisfaction as a proxy for instructional quality risks masking superficial engagement, factual inaccuracies, student data privacy vulnerabilities, or cognitive over-reliance—risks that become particularly acute across multilingual populations (Dari, Uzbeki, Pashto, and Turkmen) where AI output fidelity in less widely resourced languages remains uneven.[19]

3. Global Relevance, Boundary Conditions, and Theoretical Contributions

Although empirical data were collected in Balkh Province, Afghanistan, the findings transcend a localized case study and speak directly to global educational research, learning analytics, and development agendas. International benchmarks from the OECD PISA financial literacy assessment and cross-national economic syntheses by Lusardi and Mitchell and by Lusardi underscore that adolescent financial vulnerability—especially among socioeconomically disadvantaged and rural youth—is a structural global challenge spanning both high-income and developing economies.[20] Whereas most prior AI-in-education research originates from university classrooms and digitally saturated school networks in high-income countries, this study demonstrates that the cognitive benefits of adaptive AI scaffolding transfer robustly to under-resourced secondary and vocational schools operating under severe connectivity constraints. In doing so, the evidence directly advances United Nations Sustainable Development Goal 4 (SDG 4)—specifically equitable quality secondary and technical-vocational education—and equips adolescents in emerging economies with foundational competencies required to navigate digital payments, household budgeting, and online financial risk within an expanding global digital economy.

At the same time, cross-national transferability depends on explicit boundary conditions. First, the compensatory rural and device-access effects observed here presuppose structured, school-facilitated access to AI sessions; deploying AI tools as self-directed homework in regions where fewer than half of students have home internet (48.9%) would likely widen rather than narrow equity gaps. Second, external validity is

[19] Kasneci et al., "ChatGPT for Good?" 102274; Chan and Hu, "Students' Voices on Generative AI," 43.
[20] OECD, *PISA 2018 Results (Volume IV)*; Lusardi and Mitchell, "Economic Importance of Financial Literacy," 5–44; Lusardi, "Financial Literacy and the Need for Financial Education," 1.

bounded by instructional orchestration and linguistic quality: consistent gains across Dari, Uzbeki, Pashto, and Turkmen speakers require ongoing verification that AI-generated financial explanations remain culturally contextualized, mathematically accurate, and free from translation distortion. Third, because the sample comprised male students in public general and vocational lycees, transferability to female learners, primary cohorts, and non-formal learning environments requires direct empirical validation.

Theoretically, this study contributes four distinct propositions to learning analytics, educational technology, and financial education scholarship:

a. Pedagogy-Led Functional Equivalence: Across heterogeneous AI platforms, shared instructional affordances (immediate feedback, adaptive practice, authentic decision scenarios) account for learning gains more powerfully than interface-specific technological features, challenging tool-deterministic models in educational technology.
b. Domain-Contingent Scaffolding Theory: Adaptive AI systems yield disproportionate cognitive advantages in procedural, quantitative, and risk-assessment domains (budgeting, saving, digital financial safety) relative to conceptual or normative domains (needs versus wants), refining how instructional design frameworks map digital affordances onto cognitive task structures.
c. Compensatory Digital Equity Mechanism: When deployed through structured school-level orchestration paired with trained educators, AI scaffolding narrows institutional urban–rural learning disparities that otherwise widen under conventional lecture instruction.
d. Affective–Cognitive Decoupling in EdTech Analytics: Objective achievement metrics operate independently of subjective student satisfaction and topic interest, underscoring the theoretical necessity of pairing perceptual survey analytics with validated performance diagnostics.

4. Managerial and Policy Implications

Derived directly from the verified statistical estimates, four stakeholder groups can translate these findings into actionable educational interventions:

a. Educators and Curriculum Designers: Adopt a strategic hybrid allocation of instructional time. Concentrate AI-supported interactive scenarios and

automated feedback on procedural, calculation-intensive, and risk-sensitive financial literacy units—particularly budgeting, saving, and digital financial safety—where conventional lectures struggle to shift mastery, while retaining teacher-led classroom discussion for value-based topics such as needs versus wants. Structuring interventions into focused 5-to-8-session blocks captures peak learning gains without inducing curriculum fatigue.

b. School Administrators and Institutional Managers: Prioritize human capacity and shared access over costly proprietary software procurement. Because all four AI tools delivered equivalent gains while teacher AI training ($b = 0.48$) and device access ($b = 0.64$) independently boosted achievement, school leaders should invest in structured teacher professional development and rotational school-based device laboratories rather than assuming students possess home internet or personal devices.

c. EdTech Developers and Learning Analytics Architects: Design lightweight, low-bandwidth, and offline-capable educational platforms tailored for digitally uneven secondary and vocational schools. Developers must embed domain-verified financial content, audit linguistic accuracy across underrepresented languages (such as Dari, Uzbeki, Pashto, and Turkmen), and build educator dashboards grounded in objective domain-mastery diagnostics rather than vanity metrics such as student satisfaction ratings.

d. Education Policymakers and Development Partners: Integrate structured AI-supported financial education into national secondary and vocational curricula, explicitly targeting rural agricultural and multi-disciplinary vocational schools where conventional achievement lags furthest behind. Parallel governance policies should establish mandatory safeguards for student data privacy, content accuracy auditing, and pedagogical guidelines that prevent passive student over-reliance on automated financial advice.

5. Limitations and Directions for Future Research

Several methodological and contextual limitations warrant candid consideration when interpreting these findings. First, although baseline equivalence was confirmed across all measured demographic, infrastructural, and prior-knowledge indicators, assignment to instructional conditions was non-random; unmeasured differences in student motivation, teacher enthusiasm, or classroom micro-climate cannot be entirely

ruled out.[21] Second, the sample consisted exclusively of male secondary students across 14 public schools in a single province, preventing empirical generalization to female students or to other geographical and institutional contexts. Third, because both AI-supported and conventional groups operated within every participating school, informal peer interaction across conditions may have introduced modest spillover effects. Fourth, financial literacy was measured once immediately after instruction using a cognitive knowledge-and-skills assessment rather than real-world financial behavior, leaving open whether gains persist over time, whether novelty effects temporarily elevated effort, or whether improved test scores translate into actual budgeting, saving, and safe digital transaction habits. Fifth, the narrow dosage window (5–10 sessions) was not experimentally manipulated, and granular process logs capturing teacher mediation quality, exact device-sharing arrangements, and AI output accuracy across home languages were not recorded.

To build upon this evidence base, future research should execute five concrete directions: (1) implement cluster- or school-level randomized controlled trials (RCTs) to strengthen causal identification and prevent within-school spillover; (2) expand sampling to include girls' secondary schools, additional provinces, and cross-national developing contexts to test gender and regional external validity; (3) incorporate longitudinal delayed posttests alongside behavioral financial indicators (such as simulated or tracked saving and budgeting decisions) to verify durability and behavioral transfer; (4) capture fine-grained learning analytics—such as platform interaction logs, error-correction trajectories, and structured observations of teacher scaffolding—to open the black box of how teachers and AI co-produce learning; and (5) systematically audit the factual accuracy, algorithmic fairness, cultural relevance, and linguistic reliability of AI-generated financial guidance across low-resource languages.

D. CONCLUSION

Structured artificial intelligence-supported instruction substantially strengthens secondary students' financial literacy in Balkh Province, delivering a covariate-adjusted advantage of 2.07 points over conventional teaching across baseline-equivalent cohorts in both urban general and rural vocational schools. These learning gains concentrate in applied procedural and risk-oriented domains—most notably budgeting, saving, and

[21] Shadish, Cook, and Campbell, *Experimental and Quasi-Experimental Designs*, 135–70.

digital financial safety—while all four AI tools perform equivalently, demonstrating that shared pedagogical functions of immediate feedback, repeated practice, and realistic scenarios matter far more than software selection. Although structured school-based AI delivery narrows urban–rural achievement disparities and operates independently of student satisfaction or household internet connectivity, device access and teacher AI training remain statistically decisive enabling conditions for equitable implementation. Future randomized, longitudinal, and gender-inclusive studies should verify knowledge retention, evaluate real-world financial behavior, and audit AI linguistic reliability across underrepresented local languages.

REQUIRED STATEMENTS

CRediT authorship contribution statement. S.S.: Conceptualization, Methodology, Investigation, Data curation, Formal analysis, Writing – original draft. A.K.A.: Conceptualization, Methodology, Validation, Supervision, Writing – review & editing.

Funding. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Declaration of competing interest. 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.

Data availability. The anonymized student assessment scores and survey datasets generated during the current study are not publicly available due to institutional secondary-school student privacy protections, but are available from the corresponding author upon reasonable request.

Ethics statement. Prior to data collection, formal institutional permission was obtained from the relevant educational authorities and participating public school administrations in Balkh Province, Afghanistan. Participation was strictly voluntary, informed consent was secured from all participating students and their parents or legal guardians, and all student assessment records and demographic responses were strictly anonymized prior to statistical analysis to ensure confidentiality and exclusive research use.

Declaration of generative AI and AI-assisted technologies in the writing process. The authors did not use generative AI or AI-assisted technologies in the writing, data

analysis, or preparation of this manuscript; artificial intelligence tools were utilized exclusively as the instructional intervention evaluated in the study.

REFERENCES

Amagir, Aisa, Wim Groot, Henriëtte Maassen van den Brink, and Arie Wilschut. "A Review of Financial-Literacy Education Programs for Children and Adolescents." *Citizenship, Social and Economics Education* 17, no. 1 (2018): 56–80. https://doi.org/10.1177/2047173417719555.

Chan, Cecilia Ka Yuk, and Wenjie Hu. "Students' Voices on Generative AI: Perceptions, Benefits, and Challenges in Higher Education." *International Journal of Educational Technology in Higher Education* 20 (2023): 43. https://doi.org/10.1186/s41239-023-00411-8.

Chen, Lijia, Pingping Chen, and Zhijian Lin. "Artificial Intelligence in Education: A Review." *IEEE Access* 8 (2020): 75264–78. https://doi.org/10.1109/ACCESS.2020.2988510.

Cohen, Jacob. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates, 1988, 24–27.

Hedges, Larry V. "Distribution Theory for Glass's Estimator of Effect Size and Related Estimators." *Journal of Educational Statistics* 6, no. 2 (1981): 107–28. https://doi.org/10.3102/10769986006002107.

Huang, Anna Y. Q., Owen H. T. Lu, and Stephen J. H. Yang. "Effects of Artificial Intelligence–Enabled Personalized Recommendations on Learners' Learning Engagement, Motivation, and Outcomes in a Flipped Classroom." *Computers & Education* 194 (2023): 104684. https://doi.org/10.1016/j.compedu.2022.104684.

Hwang, Gwo-Jen, Haoran Xie, Benjamin W. Wah, and Dragan Gašević. "Vision, Challenges, Roles and Research Issues of Artificial Intelligence in Education." *Computers and Education: Artificial Intelligence* 1 (2020): 100001. https://doi.org/10.1016/j.caeai.2020.100001.

Kaiser, Tim, Annamaria Lusardi, Lukas Menkhoff, and Carly Urban. "Financial Education Affects Financial Knowledge and Downstream Behaviors." *Journal of Financial Economics* 145, no. 2 (2022): 255–72. https://doi.org/10.1016/j.jfineco.2021.09.022.

Kaiser, Tim, and Lukas Menkhoff. "Does Financial Education Impact Financial Literacy and Financial Behavior, and If So, When?" *World Bank Economic Review* 31, no. 3 (2017): 611–30. https://doi.org/10.1093/wber/lhx018.

Kasneci, Enkelejda, Kathrin Seßler, Stefan Küchemann, Maria Bannert, Daryna Dementieva, Frank Fischer, Urs Gasser, et al. "ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education." *Learning and Individual Differences* 103 (2023): 102274. https://doi.org/10.1016/j.lindif.2023.102274.

Kulik, James A., and J. D. Fletcher. "Effectiveness of Intelligent Tutoring Systems: A Meta-Analytic Review." *Review of Educational Research* 86, no. 1 (2016): 42–78. https://doi.org/10.3102/0034654315581420.

Lakens, Daniël. "Calculating and Reporting Effect Sizes to Facilitate Cumulative Science: A Practical Primer for *t*-Tests and ANOVAs." *Frontiers in Psychology* 4 (2013): 863. https://doi.org/10.3389/fpsyg.2013.00863.

Lusardi, Annamaria. "Financial Literacy and the Need for Financial Education: Evidence and Implications." *Swiss Journal of Economics and Statistics* 155 (2019): 1. https://doi.org/10.1186/s41937-019-0027-5.

Lusardi, Annamaria, and Olivia S. Mitchell. "The Economic Importance of Financial Literacy: Theory and Evidence." *Journal of Economic Literature* 52, no. 1 (2014): 5–44. https://doi.org/10.1257/jel.52.1.5.

Ma, Wenting, Olusola O. Adesope, John C. Nesbit, and Qing Liu. "Intelligent Tutoring Systems and Learning Outcomes: A Meta-Analysis." *Journal of Educational Psychology* 106, no. 4 (2014): 901–18. https://doi.org/10.1037/a0037123.

OECD. *PISA 2018 Results (Volume IV): Are Students Smart About Money?* Paris: OECD Publishing, 2020, 1–248. https://doi.org/10.1787/48ebd1ba-en.

Shadish, William R., Thomas D. Cook, and Donald T. Campbell. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston: Houghton Mifflin, 2002, 135–70.

Urban, Carly, Maximilian Schmeiser, J. Michael Collins, and Alexandra Brown. "The Effects of High School Personal Financial Education Policies on Financial Behavior." *Economics of Education Review* 78 (2020): 101786. https://doi.org/10.1016/j.econedurev.2018.03.006.

Zawacki-Richter, Olaf, Victoria I. Marín, Melissa Bond, and Franziska Gouverneur. "Systematic Review of Research on Artificial Intelligence Applications in Higher Education—Where Are the Educators?" *International Journal of Educational Technology in Higher Education* 16 (2019): 39. https://doi.org/10.1186/s41239-019-0171-0.

Zhai, Xuesong, Xiaoyan Chu, Ching Sing Chai, Morris Siu Yung Jong, Andreja Istenic, Michael Spector, Jia-Bao Liu, Jing Yuan, and Yan Li. "A Review of Artificial Intelligence (AI) in Education from 2010 to 2020." *Complexity* 2021 (2021): 8812542. https://doi.org/10.1155/2021/8812542.