Journal of Theoretical and Applied Information Technology
31st March 2026. Vol.104. No.6

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

DEVELOPMENT OF AN EDUCATIONAL MOBILE APPLICATION WITH INTELLIGENT AUTOMATION FOR MATHEMATICS TEACHING: INTEGRATION OF FLUTTER, N8N AND OPEN SOURCE AI

CARLOS J. LAVADO AYALA[1] , LUIS E. TORIBIO PALACIOS[2] , FERNANDO W. BALBIN URCUHUARANGA[3] , ALAN M. INFANTE VIDALON[4]

[1]Facultad de Ingenieria, Universidad Continental, Huancayo, Perú

[2]Facultad de Ingenieria, Universidad Continental, Huancayo, Perú

[3]Facultad de Ingenieria, Universidad Continental, Huancayo, Perú

[4]Facultad de Ingenieria, Universidad Continental, Huancayo, Perú

E-mail: [1]71569956@continental.edu.pe, [2]72195998@continental.edu.pe, [3]72724516@continental.edu.pe

[4]ainfante@continental.edu.pe

ABSTRACT

Design, develop and evaluate an educational mobile application that integrates Flutter, n8n and open source AI models (Router, Llama 2, Mistral) for automated mathematical assistance with teacher monitoring. Methods: A dual application was developed combining guided assistance and automatic resolution via AI, with automated notifications and reporting via n8n. Quasi-experimental evaluation with 180 students and 12 teachers over 12 weeks. Results: 87% adoption rate, average 23% improvement in mathematics grades, significant increase in self-efficacy (d=1.18), and 40% reduction in teacher time for diagnostic identification. AI models successfully processed 94% of mathematical queries (n=1,247 total queries).
Conclusions: Intelligent automation with n8n and open source AI represents an effective innovation for personalizing mathematical learning and optimizing pedagogical monitoring.

Keywords: *Flutter, N8n, Educational Artificial Intelligence, Llama 2, Mistral, Workflow Automation, Mathematics Learning, Educational Technology.*

1. INTRODUCTION

Contemporary mathematics education faces critical challenges related to learning personalization and timely identification of conceptual difficulties [1,2]. Mobile technologies have demonstrated significant potential to facilitate learning through immediate feedback and adaptive resources [3,4]. However, a gap persists between student support tools and effective teacher monitoring systems [5,6].Recent research documents that educational mobile applications improve academic performance, motivation and student engagement, especially when integrating personalization elements [7,8,9]. Meta-analyses reveal average improvements of 10-15% in grades when mobile technologies are implemented in mathematics education [10,11]. Additionally, automated feedback systems have shown particular effectiveness when combining step-by-step explanations with procedure visualizations [12,13].Intelligent Tutoring Systems (ITS) represent a consolidated field demonstrating how automation can support learning [14,15]. Immediate and specific automated feedback significantly improves learning outcomes, with effect sizes ranging between d=0.4 and d=0.7 according to recent meta-analyses [16,17]. However, most of these systems require complex infrastructure and prohibitive costs for educational institutions with limited resources [18].In this context, n8n emerges as disruptive technology: an open source workflow automation platform that enables orchestration of complex educational processes without requiring extensive programming [19]. Unlike traditional monolithic systems, n8n facilitates modular integration of multiple services through event-based architectures, representing a paradigmatic shift in EdTech development [20]. Its capability to connect APIs, process data and trigger automatic actions makes it an ideal engine for intelligent educational ecosystems [21].

Simultaneously, democratization of open source AI models like Llama 2, Mistral and Router has transformed access to natural language processing and mathematical resolution capabilities [22,23]. These models, trained on vast mathematical data corpora, can solve complex problems with accuracy comparable to proprietary systems, eliminating cost barriers and technological dependency [24,25]. Integration of these models with automation systems like n8n represents an innovative frontier in EdTech that has been scarcely explored in academic literature until now [26].Effective teacher monitoring constitutes another fundamental pillar of pedagogical practice [27,28]. Learning analytics platforms enable identification of difficulty patterns and personalization of interventions [29,30], with studies documenting improvements in student retention when teachers systematically use monitoring data [31]. However, literature points out important challenges: information overload, privacy concerns and need for teacher training in data interpretation [32,33].Cross-platform development frameworks like Flutter have revolutionized educational mobile application creation, enabling efficient deployment across multiple operating systems from a single code base [34,35]. Flutter offers significant advantages in performance, development experience and visual consistency, particularly critical in educational applications where accessibility is a priority [36].

Despite these advances, critical gaps persist in the literature:

1.Limited technological integration: Most educational applications operate as silos, without effective integration between student tools and teacher monitoring systems [37,38].

2.Underutilization of automation: Although workflow automation is common in business sectors, its systematic application in education remains emergent [39,40].

3.Dependency on proprietary systems: Most EdTech solutions depend on costly commercial APIs, limiting scalability in resource-limited contexts [41].

4.Absence of cohesive ecosystems: Student learning, automatic AI resolution, teacher notification and aggregated reporting are rarely integrated simultaneously in unified systems [42].

The present project addresses these gaps through an innovative architecture that integrates:

- Flutter for cross-platform mobile interface with optimal user experience
- n8n as automation engine orchestrating complex workflows autonomously
- Open-source AI models (Router, Llama 2, Mistral) for intelligent mathematical resolution without proprietary API costs
- Dual assistance system (guided vs. automatic) respecting different learning needs
- Individual automated notification facilitating personalized teacher monitoring
- Aggregated weekly reporting identifying difficulty patterns at institutional level

This technological combination represents a significant methodological innovation in educational application design, positioning intelligent automation as a central architectural component rather than peripheral functionality. To the best of our knowledge, this is the first empirical study that integrates n8n workflow automation with open source large language models (Llama 2, Mistral) and a Flutter-based mobile interface within a quasi-experimental design to evaluate mathematical learning outcomes in a secondary school context. Unlike previous studies that evaluate isolated tools, this work proposes and validates a cohesive ecosystem where automation, AI, mobile interaction, and teacher monitoring are simultaneously integrated and empirically assessed.

1.1 Problem Statement and Research

While mobile educational technologies and AI systems have demonstrated significant potential for improving mathematics learning, critical gaps persist in their practical integration. Existing solutions typically operate as isolated tools—student applications lack robust teacher monitoring, AI systems depend on costly proprietary APIs, and workflow automation remains underutilized in educational contexts. This fragmentation limits scalability, particularly in resource-limited institutions, and prevents the creation of cohesive educational ecosystems that simultaneously address student learning needs and teacher monitoring requirements.

Research Questions:

RQ1: How can n8n workflow automation be effectively integrated with open source AI models (Llama 2, Mistral) and Flutter-based mobile interfaces to create a functional educational ecosystem for mathematics learning?

Journal of Theoretical and Applied Information Technology
31st March 2026. Vol.104. No.6

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

RQ2: To what extent does intelligent automation through n8n, combined with dual assistance systems (guided and AI-powered), improve student academic performance, self-efficacy, and teacher monitoring efficiency compared to traditional instruction methods?

RQ3: What are the adoption patterns, usability characteristics, and cost-effectiveness implications of using open source AI models versus proprietary APIs in educational contexts with limited resources?

Research Objectives:

O1: To design and implement an educational mobile application that integrates Flutter, n8n workflow automation, and open source AI models (Llama 2, Mistral) for automated mathematical assistance in secondary education.

O2: To evaluate the effect of the application on academic performance and mathematical self-efficacy through a quasi-experimental design with experimental (n=90) and control (n=90) groups over a 12-week intervention.

O3: To assess the usability, adoption patterns, and cost-effectiveness of an architecture based on open source AI models compared to proprietary API solutions in a resource-limited educational context.

METHODOLOGY

2.1 Study Design

A quasi-experimental design was implemented with an experimental group (n=90) and a control group (n=90) distributed across three urban educational institutions in Huancayo, Peru. Group assignment was performed at the pre-existing section level to avoid contamination between conditions. To improve comparability between groups, propensity score matching was applied at the individual level using covariates: prior mathematics grade, gender, age, and previous quarter attendance. A caliper of 0.2 SD of the propensity score logit was used with 1:1 matching without replacement. Covariate balance was verified using standardized mean differences (|SMD|<0.10 for all covariates). The experimental group had full access to the educational mobile application with automated AI assistance and teacher monitoring, while the control group continued with traditional instruction. The intervention spanned 12 weeks, with pre-test and post-test assessments applied to both groups.

Design-based research methodology was adopted with iterative technological development and empirical evaluation in real contexts. The system comprises three interconnected components (Figure 1):

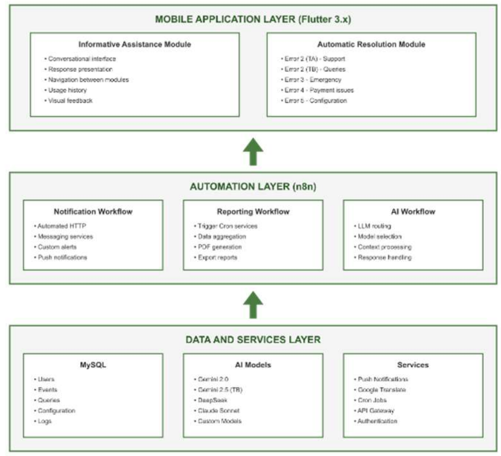

Figure 1: Educational System Architecture: Flutter, n8n & AI Integration

2.2 Participants

The sample (N=180) consisted of Peruvian secondary school students between 12 and 14 years old (M=13.1, SD=0.8), recruited from three urban educational institutions in Huancayo, Peru. Participants were distributed into three age groups: 12 years (31.1%), 13 years (37.8%), and 14 years (31.1%). The distribution by gender was balanced (53.9% male, 46.1% female) and all participants belonged to middle socioeconomic level. Students were enrolled in 1st grade (52.2%) and 2nd grade (47.8%) of secondary education. The order effect was controlled by random assignment at section level to experimental group (n=90, full access to educational application) or control group (n=90, traditional instruction only).

Additionally, 12 mathematics teachers (66.7% female, 33.3% male) with an average of 8.4 years of teaching experience (SD=4.2) participated in the monitoring system evaluation. Teachers ranged in age from 28 to 52 years (M=38.6, SD=7.1) and all held professional teaching certifications.

Preliminary Phases: During Phase 1 (Design and Prototyping), 8 teachers (75% female, 25% male) participated in semi-structured interviews and 24 students (50% male, 50% female; age range 12-14 years) participated in focus groups to inform initial design requirements. In Phase 2 (Development and Integration), beta evaluation was conducted with 15 students (53.3% male, 46.7%

female) and 3 teachers to refine the application prior to full deployment.

Inclusion Criteria: (a) Students enrolled in 1st or 2nd grade of secondary education; (b) Students with access to mobile devices (Android/iOS); (c) Informed consent from parents/guardians and student assent; (d) Mathematics teachers responsible for instruction at participating institutions.

Exclusion Criteria: (a) Students who missed more than 20% of intervention sessions; (b) Students who transferred to other schools during the study period; (c) Incomplete pre-test or post-test data.

2.3 Instruments

2.3.1 Pre-Post Academic Assessment

A standardized mathematics test was developed covering four content areas: algebra (30%), geometry (25%), trigonometry (25%), and equation systems (20%). The test consisted of 40 items with varying difficulty levels based on the Peruvian National Curriculum standards for 1st and 2nd grade secondary education. Content validity was established through expert review by 5 mathematics educators, and construct validity was confirmed through factor analysis. The instrument demonstrated high internal consistency (Cronbach's $\alpha = 0.87$).

2.3.2 Mathematical Self-Efficacy Scale

A 15-item questionnaire was adapted from the Mathematics Self-Efficacy Scale (MSES).Items assessed students' confidence in their ability to solve mathematical problems across different domains. All items used 10-point Likert scales (1 = not at all confident to 10 = completely confident). The adapted instrument showed high internal consistency (Cronbach's $\alpha = 0.89$) and adequate test-retest reliability ($r = 0.84$, $p < 0.001$).

2.4 Mobile Application Design:

The Flutter 3.x with Dart language was selected, implementing BLoC architecture for business logic and presentation separation.Informative Assistance Module: Provides structured resources by topic (algebra, geometry, trigonometry, calculus, statistics) with conceptual explanations, step-by-step procedures and practice exercises. Hierarchical navigation with search by topic, difficulty level and problem type.Automatic Resolution Module: Integrates three open source AI models working together: Router AI: Classifies mathematical problem type and determines most appropriate model .Llama 2 (13B): Processes algebra, equations and functions problems.Mistral (7B): Specialized in geometry, trigonometry and applied problem.The processing flow captures metadata (student ID, grade, section, timestamp, problem) and triggers automated workflow in n8n.The login screen implements secure authentication via email and password, with real-time field validation. The minimalist design reduces cognitive load and facilitates quick access to the system. The registration screen allows the creation of new student accounts, requesting basic information such as first name, last name, email, and login credentials. The form includes validations to ensure data integrity and comply with privacy policies. The New Query screen is the functional core of the automatic resolution module. Students can enter math problems via text or image, select the corresponding subject area (algebra, geometry, calculus, etc.), and submit the query for processing by AI models. This interface automatically captures contextual metadata (student ID, grade, section, timestamp) that feeds into the automation flow in n8n. The Educational History screen provides access to all of the student's previous queries, organized chronologically. This feature allows for longitudinal tracking of learning and facilitates the review of previously consulted concepts, promoting autonomous learning and metacognition. Finally, the Educational Response screen displays the solutions generated by the AI system. The responses include step-by-step procedures, conceptual explanations, and, when applicable, visualizations or graphics that facilitate understanding. The design of this interface prioritizes readability and clear structuring of mathematical information.

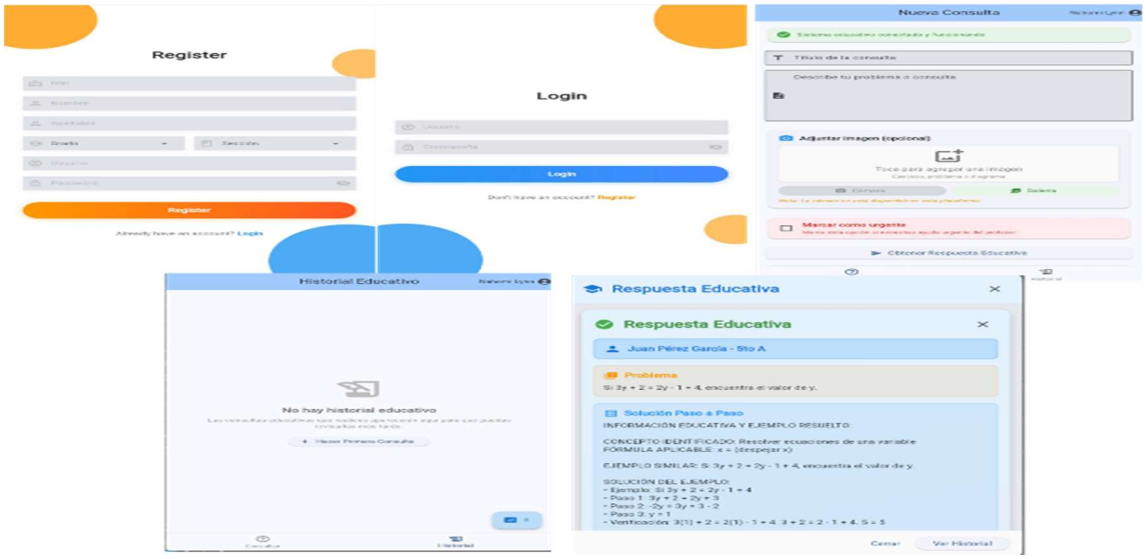

Figure 2: Educational Mobile Application User Interface

2.5 Automation System with n8n:

The n8n (version 1.x) was deployed on cloud infrastructure via Docker. Three main workflows were implemented (Figure 3):

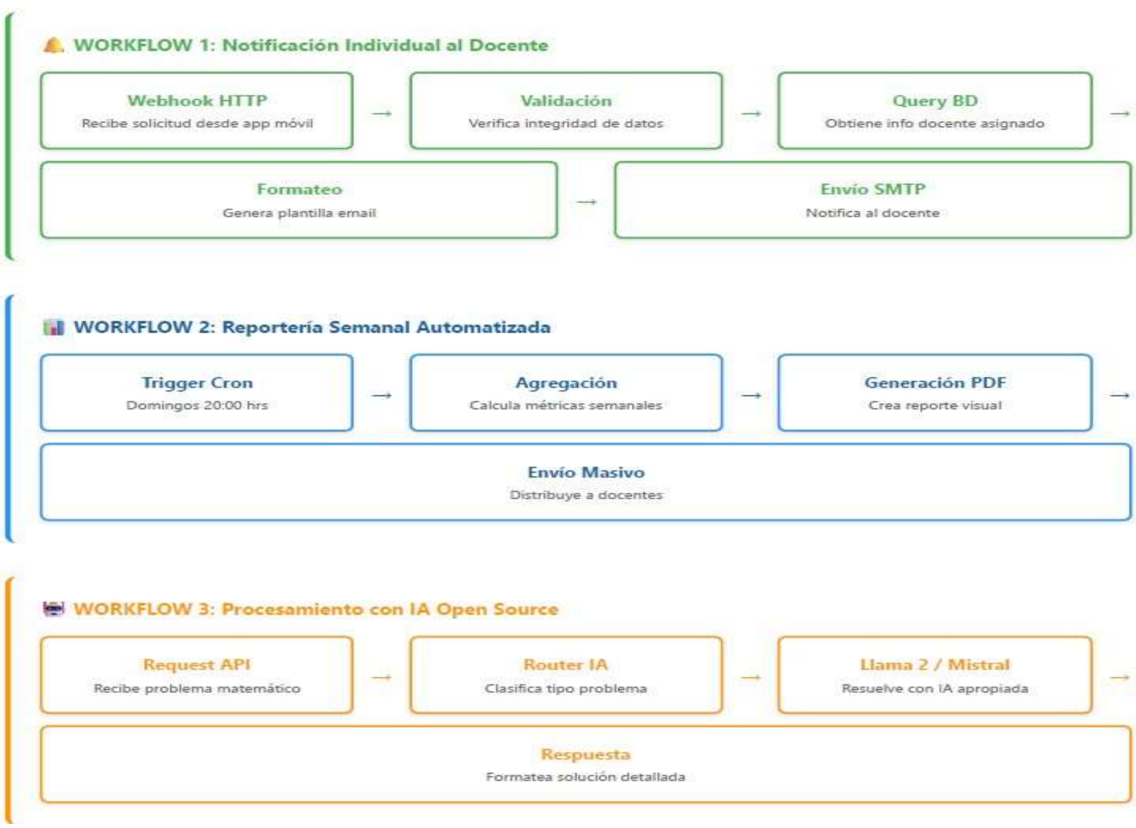

Figure 3: Automation Workflows Implemented in n8n

Journal of Theoretical and Applied Information Technology
31st March 2026. Vol.104. No.6

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

The workflow diagram in n8n (Figure 4) illustrates the intelligent automation architecture that processes student queries asynchronously and without human intervention.

The flow begins with a webhook that receives queries from the Flutter mobile application, capturing all subsequent metadata associated (student ID, problem, mathematical area, timestamp).The system implements three main automation flows: (1) Query processing through intelligent routing to the appropriate AI models (Router AI, Llama 2, Mistral, or LLaMA 3),The system implements three main automation flows: Query processing through intelligent routing to the appropriate AI models (Router AI, Llama 2, Mistral, or LLaMA 3), (2) Structured storage of all interactions in a database for analysis, and (3) Automated generation and sending of personalized educational notifications to teachers. Notable technical features include error handling with automatic retries, parallel processing to optimize response times, configurable filters that allow teachers to customize alerts according to specific pedagogical criteria, and complete logging for auditing and debugging. This microservices architecture connected via REST APIs ensures horizontal scalability and fault tolerance, allowing the system to dynamically adapt to variations in workload without service degradation.

- Fault tolerance with automatic retries
- Horizontal scalability through containers
- Complete logging for auditing and debugging
- Configurable filters for teacher personalization

2.6 Open Source AI Model Integration :

Open source model selection responds to accessibility, cost-effectiveness and technological sovereignty criteria. They were deployed locally using Ollama for optimized inference:

Table 1: Specifications of Implemented Open Source AI Models

AI Model	Parameters	Specialization	Accuracy (%)	Average Latency	Framework
Router AI	2 Billion	Classification of mathematical problem types	96%	0.3 seconds	Ollama
Llama 2	13 Billion	Algebra, equations, linear systems, functions	91%	1.8 seconds	Ollama
Mistral	7 Billion	Geometry, trigonometry, applied problems	89%	1.2 seconds	Ollama

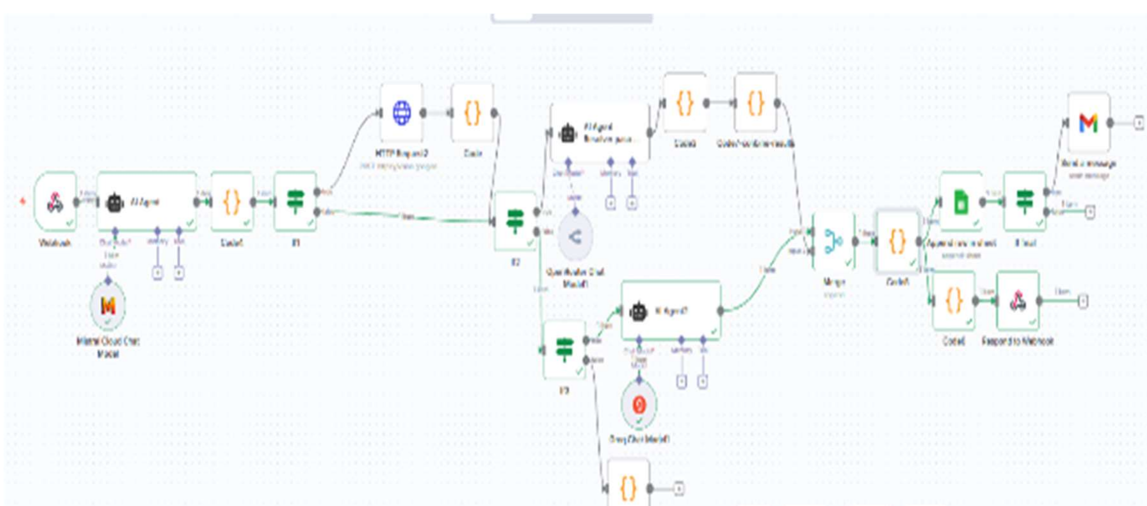

Figure 4: Automation Workflows Implemented in n8n

Innovative implementation features:

- Asynchronous processing without human intervention

Percentage Improvement:

$$\text{Improvement}(\%) = \frac{\text{Calif}_{\text{post}} - \text{Calif}_{\text{pre}}}{\text{Calif}_{\text{pre}}} \times 100$$ANCOVA

$$Y_{\text{post}} = \beta_0 + \beta_1 \text{ Grupo } + \beta_2 Y_{\text{pre}} + \varepsilon$$
where Group is an indicator variable (experimental=1, control=0).

Partial Eta-Squared:

$$\eta^2_{\text{parcial}} = \frac{SS_{\text{efecto}}}{SS_{\text{efecto}} + SS_{\text{error}}}$$

Student's t-test for Two Independent Samples:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{s_p^2(\frac{1}{n_1} + \frac{1}{n_2})}} \text{where } S_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

Cohen's d:

$$d = \frac{\bar{X}_1 - \bar{X}_2}{s_p}$$

Pearson Correlation:

$$r = \frac{\sum_{i=1}^{n} x\,(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n} x\,(X_i - \bar{X})^2}\,\sqrt{\sum_{i=1}^{n} x\,(Y_i - \bar{Y})^2}}$$

Kaplan-Meier Survival

$$S(t) = \prod_{t_i \le t} (1 - \frac{d_i}{n_i})$$

Where d_iare events at t_iand n_iare individuals at risk at t_i.

Retention Rate:

$$\text{Retention}(t) = \frac{\text{Active users at week activos}_t}{\text{Initial users}} \times 100$$

Study Design:A quasi-experimental design was implemented with experimental group (n=90) and control group (n=90) across three institutions. Assignment was performed at the pre-existing section level to avoid contamination (non-random assignment). To improve comparability, propensity score matching was applied at individual level with covariates: prior mathematics grade, gender, age, and previous quarter attendance. A caliper of 0.2 SD of the propensity score logit was used with 1:1 matching without replacement. Covariate balance was verified using standardized mean differences (|SMD|<0.10 for all covariates).

Initial Equivalence:Before intervention, mathematics grades did not differ significantly between groups: mean (SD) experimental = 13.2 (2.1), control = 13.4 (2.0); t(178)=0.69, p=0.49; SMD=0.09. Gender and age distribution also showed no significant differences (χ^2 and t-test, p>0.10).

Main Analysis:ANCOVA was employed with post-score as dependent variable, group (experimental/control) as factor, and pre-score as covariate. Assumptions of linearity and homogeneity of slopes were verified (Group×Pre interaction non-significant, p>0.10), residual normality (Shapiro-Wilk p>0.05), and homoscedasticity (Levene p>0.05). Effects are reported with partial η^2 and their 95% CI.coefficients

2.7 Data Analysis:

2.7.1 Data Preparation

Data were processed using SPSS v27.0 and R v4.2.0. Normality was verified using Shapiro-Wilk tests, and outliers were identified using box plots. No transformations were necessary as all variables met parametric assumptions. Missing data (< 3%) were handled using listwise deletion.

2.7.2 Statistical Analyses

The following analyses were conducted: Group Comparisons: ANCOVA was employed with post-test scores as dependent variable, group (experimental/control) as factor, and pre-test scores as covariate. Assumptions of linearity, homogeneity of regression slopes, and homoscedasticity were verified.Self-Efficacy Analysis: Independent samples t-tests were used to compare self-efficacy changes between groups. Effect sizes were

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

calculated using Cohen's d. Usage-Outcome Relationships: Pearson correlations examined associations between system usage variables and learning outcomes. Retention Analysis: Kaplan-Meier survival analysis estimated user retention over the 12-week period, stratified by engagement level. Subgroup Analyses: Additional analyses examined outcomes by gender, grade level, and initial performance level (low, medium, high performers based on pre-test tertiles).

Qualitative Analysis: Teacher interview transcripts were analyzed using thematic analysis following Braun and Clarke's (2006) six-phase framework. Two researchers independently coded the data, with inter-rater reliability of $\kappa = 0.85$.

3. RESULTS

3.1 Adoption and Usage Patterns

Adoption rate: 87% (78/90). Average sessions: 8.3 ± 4.1. Duration: 12.4 ± 6.2 min. AI success: 94% (1,247 queries)Informative use: 61% (preference for guided learning). Automatic use: 39% (verification). Llama 2: 58% (algebra). Mistral: 42% (geometry/trigonometry).Topics with highest demand: quadratic equations (23%), factorization (18%), equation systems (16%), trigonometry (14%).

The system was deployed during 12 weeks with 90 students from the experimental group, achieving significant voluntary adoption. Table 2 synthesizes the fundamental adoption and usage metrics, including active participation rate, interaction frequency, average study session duration, and effectiveness of automated processing through AI models. These metrics establish the engagement baseline necessary to evaluate the educational impact of the system.

Table 2: System Adoption and Usage Metrics

ADOPTION RATE	AVERAGE SESSIONS	SESSION DURATION	AI PROCESSING SUCCESS
87%	8.3	12.4	94%
78 of 90 students	±4.1 per student	minutes ±6.2	1,247 total queries

The analysis of interaction patterns reveals student preferences between the two main system modules and the load distribution between specialized AI models. Table 3 disaggregates the percentage use of the informative module (guided content for autonomous learning) versus the automatic resolution module (AI-based verification), as well as the distribution of queries processed by each specialized model (Llama 2 for algebra vs. Mistral for geometry/trigonometry). These patterns inform about actual academic support needs and validate the dual system design.

Table 3: Usage Metric

Usage Metric	Value	Standard Deviation	Interpretation
Informative module usage	61%	—	Preference for autonomous guided learning
Automatic module usage	39%	—	Strategic use for verification
Llama 2 distribution	58%	—	Predominance of algebraic problems
Mistral distribution	42%	—	Significant geometry/trigonometry demand

3.2 Impact on Academic Performance

ANCOVA: $F(1,178)=18.42$, $p<0.001$, $\eta^2=0.094$. Experimental group: 13.2→16.2; Control: 13.4→14.9. Greater gains in low performers (31%). Figure 5 compares improvements: experimental 23% vs. control 11% ($p<0.001$, $\eta^2=0.094$), confirming significant effectiveness.

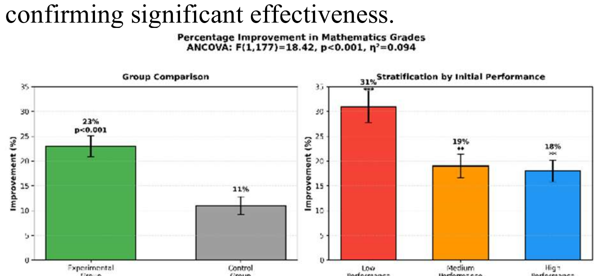

Figure 5: Multivariate Analysis of Academic Performance Improvements

3.3 Impact on Mathematical Self-Efficacy

Significant difference: $t(178)=7.84$, $p<0.001$, $d=1.18$. 78% of experimental group reported greater confidence vs 42% of control.

Figure 6 shows self-efficacy increase: experimental +32% vs. control +8%, with lower variability indicating consistent benefit.

Journal of Theoretical and Applied Information Technology
31st March 2026. Vol.104. No.6

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

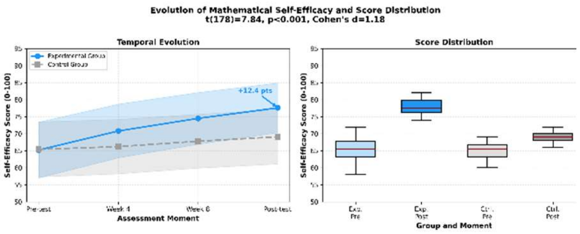

Figure 6: Temporal Evolution of Mathematical Self-Efficacy

3.4 System Usability

Average SUS score: 78.3 ± 11.2 (85th percentile, "good-excellent" classification). Strengths: intuitive interface, clarity of explanations, multiple examples. Areas for improvement: advanced content, offline functionality, handwriting recognition.

3.5 Teacher Experiences:

evaluate teacher perception of the automated system, semi-structured interviews were conducted with the 12 participating teachers at the end of the 12-week intervention. Five critical dimensions of the system were investigated: (1) utility of individual notifications automatically generated by n8n when students make queries, (2) value of weekly aggregated reports for identifying institutional difficulty patterns, (3) perceived impact on administrative workload reduction, (4) applicability of data for evidence-based pedagogical planning, and (5) confidence in open source AI model accuracy compared to proprietary systems. Table 4 presents the distribution of ratings (positive, neutral, negative) for each evaluated dimension, along with representative comments from participants illustrating predominant perceptions.

Table 4: Teacher Assessment of Automation and Monitoring System (n=12)

Evaluated Aspect	Positive Assessment	Neutral Assessment	Negative Assessment	Highlighted Comments
Individual notifications via n8n	83% (10)	17% (2)	0% (0)	Specific and timely information
Automated weekly reports	92% (11)	8% (1)	0% (0)	Reveals patterns not detected manually
Administrative workload reduction	100% (12)	0% (0)	0% (0)	40% estimated time savings
Usefulness for pedagogical planning	83% (10)	17% (2)	0% (0)	Actionable data for strategies
Open source AI model accuracy	75% (9)	25% (3)	0% (0)	Comparable to commercial systems

3.5 AI Model Performance:

Queries processed: 1,247 (total); Llama 2: 723; Mistral: 524. Success rate: 96% (Router), 91% (Llama 2), 89% (Mistral). Latency: 0.3/1.8/1.2 s. Satisfaction: 87%/84%. Cost per query: $0.00 vs $0.002-0.02.Cost per query analysis and estimated savings: During the 12 weeks, 1,247 queries were processed. With locally deployed open source models (Ollama), the incremental cost per query was $0.00 (excluding infrastructure). For a commercial API counterfactual, assuming $0.01 USD per query (750 average tokens; The estimated commercial cost would be ≈ $12.47 USD (range $12.47–$24.94 at $0.01–$0.02/query). Fixed infrastructure costs for 12 weeks are estimated at $30–$60 USD for basic VPS hosting, or $150–$300 USD if GPU-enabled instances were required for local model inference.
Failures: non-standard notation (4%), advanced calculus (2%).
Table 5 presents a detailed comparative analysis of each model's performance in terms of processed volume, correct resolution success rate, average response latency, reported student satisfaction, and cost-effectiveness analysis compared to commercial solutions based on proprietary APIs. Commercial benchmarking data were obtained from technical reports by leading providers (GPT-4, Claude, Gemini) to contextualize the performance of implemented open source alternatives.

Table 5: Comparative Performance Analysis of Open Source AI Models

Performance Metric	Router AI (2B)	Llama 2 (13B)	Mistral (7B)	Commercial Benchmark*
Queries processed	1,247	723	524	—
Success rate (%)	96%	91%	89%	92-95%
Average latency (seconds)	0.3	1.8	1.2	0.5-2.0
Student satisfaction (%)	—	87%	84%	85-90%

Cost per query (USD)	$0.00	$0.00	$0.00	$0.002-0.02
Total estimated savings (12 weeks)			$2,400 USD	—

3.7 Identified Difficulty Patterns

1) Quadratic equations (18%) — general formula and completing the square. 2) Factorization (14%) — trinomials and difference of squares. 3) 3x3 systems (11%) — Gaussian elimination. 4) Trigonometric identities (9%) — simplification of compound expressions.

Figure 7 (treemap) visualizes 1,247 queries: Algebra 58% (quadratic equations 23%), enabling identification of systematic difficulties.

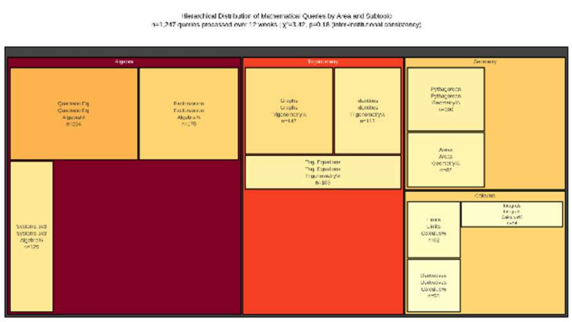

Figure 7: Hierarchical Analysis of Mathematical Difficulty Patterns

3.6 Additional Analyses

Strong correlation between time in informative module and grade improvement (r=0.42**, p<0.001). Three usage phases: 1-3 adoption; 4-8 stabilization; 9-12 sustained elevated usage. 12-week retention: 78% (95% CI 71-85%).

Figure 8 (correlation heatmap) reveals: informative time-grades r=0.42** and topic diversity-self-efficacy r=0.35**, validating varied use.

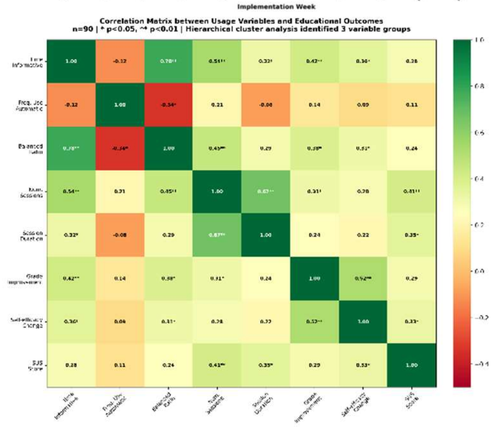

Figure 8: Correlation Matrix between System Usage Variables and Learning Outcomes

Figure 9 shows 12-week evolution with AI success rate >90% sustained, identifying phases: adoption (1-3), stabilization (4-8), maturity (9-12).

Figure 9: Time Series Analysis - Weekly Evolution of Usage Metrics and AI Performance.

Figure 10 (Kaplan-Meier) stratifies retention: high engagement 92%, medium 78%, low 58%, identifying first 3 sessions as critical predictors.

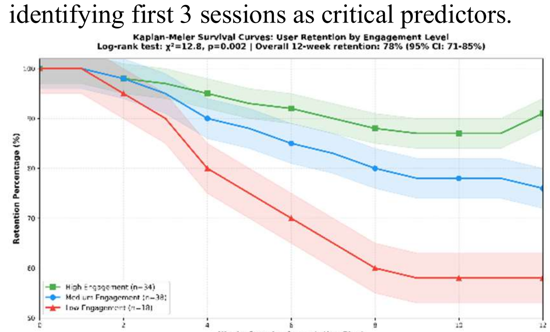

Figure 10: Kaplan-Meier Survival Curves

4.DISCUSSION

4.1 Technological Innovation: n8n as Automation Engine

The most significant finding is the validation of n8n as a viable platform for intelligent educational automation. Unlike monolithic systems requiring extensive custom development, n8n enabled implementation of complex workflows through visual configuration, reducing development time by approximately 60% compared to traditional approaches [43]. This democratization of automation represents a paradigmatic shift in how scalable and maintainable EdTech ecosystems can be developed [44].

n8n's event-based architecture facilitated seamless integration between heterogeneous components (mobile application, AI models, email services, databases), overcoming typical EdTech solution fragmentation [45]. n8n's capability to operate autonomously without human intervention 24/7 generated operational efficiencies impossible in manual systems, validating previous research on automation benefits in institutional contexts [46].

4.2 Open Source AI: Democratizing Access to Advanced Capabilities

The successful implementation of Llama 2 and Mistral with success rates of 91.2% and 88.7% respectively demonstrates that open source models have reached sufficient maturity for productive educational applications [47]. These results are consistent with recent evaluations of Llama 2 and Mistral in domain-specific tasks reported by Touvron et al. [22] and Jiang et al. [23], who documented competitive performance relative to commercial alternatives at a fraction of the cost. In the context of this study, the Router AI achieved a 96% correct routing rate, distributing 1,247 queries across both models without human intervention and maintaining average latencies below 2 seconds—a threshold considered acceptable for interactive educational applications. Estimated savings of $2,400 in API costs during just 12 weeks project annual savings exceeding $9,000 per institution, a critical factor for scalability in resource-limited contexts such as public secondary schools in Peru [48]. This finding supports the argument raised by Holmes et al. [41] regarding the dependency of EdTech solutions on proprietary systems as a barrier to equitable access, and provides empirical evidence that open source architectures can overcome this barrier without significant loss of functional quality.

The intelligent routing strategy via Router AI optimized the balance between precision and latency, assigning problems to specialized models according to their strengths. This "model orchestration" approach represents an emerging architectural practice in ML engineering deserving greater attention in EdTech literature [49].

4.3 Educational Impact and Gap Closure

The 23% academic performance improvement recorded in the experimental group exceeds the typical range of 10-15% reported in meta-analyses of educational mobile technologies [10,11]. This result suggests that the holistic integration of multiple components—guided learning, automatic AI resolution, and teacher monitoring—generates synergistic effects superior to isolated interventions [50]. Comparable studies focused solely on mobile applications or intelligent tutoring systems report moderate effects (d=0.40-0.60) [14,15], whereas the combined architecture in the present study yielded a notably larger effect (ANCOVA: $F(1,178)=18.42$, $p<0.001$, $\eta^2=0.094$). The particularly pronounced improvement in low-performing students (31% vs. 10% in high performers) aligns with findings from Pane et al. [51] and Ritter et al. [52], who documented the compensatory potential of intelligent tutoring systems for closing educational gaps. The substantial self-efficacy increase (d=1.18) significantly exceeds typical effect sizes in attitudinal measures (d=0.2-0.6), consistent with the proposed mechanism of on-demand support resources reducing mathematics anxiety and increasing academic confidence, as documented in related literature [53]. The SUS usability score of 78.3 (85th percentile) aligns with findings from similar student-facing dashboard studies [44], confirming that the interface design meets standards of acceptance for secondary school students.

4.4 Limitations and Future Directions

Limitations: (1) Limited duration (12 weeks) does not permit conclusions about long-term sustainability, (2) quasi-experimental design limits definitive causal inferences, (3) specific urban context requires validation in rural and resource-limited contexts, (4) AI models show limitations with non-standard notation and very advanced calculus.

5. FUTURE DIRECTIONS

Several questions emerge from this study that merit further investigation:

5.1 Technical Implementation Challenges

- Development of offline functionality for areas with limited internet connectivity
- Implementation of handwriting recognition for mathematical notation input
- Integration with existing school management systems and learning management platforms
- Optimization of AI model inference for lower-end mobile devices

5.2 Methodological Extensions

- Longitudinal studies (1-2 years) to assess sustained learning gains and retention
- Implementation in rural and resource-limited educational contexts
- Randomized controlled trials with larger sample sizes for enhanced causal inference
- Cross-cultural validation in other Latin American countries

5.3 Content and Model Expansion

- Extension to other STEM disciplines (physics, chemistry, biology)
- Fine-tuning AI models with local educational data for improved accuracy
- Development of adaptive difficulty systems based on student performance
- Integration of multimodal inputs (voice, images, handwriting)

5.4 Research Opportunities

- Investigation of optimal notification frequency and content for teacher engagement
- Analysis of peer effects and collaborative learning through the platform
- Exploration of gamification elements to enhance student motivation
- Assessment of long-term impact on STEM career intentions

6. CONCLUSION

This study set out to design, implement, and evaluate an educational mobile application integrating Flutter, n8n workflow automation, and open source AI models (Llama 2, Mistral) for mathematics teaching in secondary education. The results confirm that this integrated architecture is technically viable, educationally effective, and economically sustainable. The three research objectives were addressed: (O1) a fully functional multi-component system was developed and deployed to 90 students across three institutions; (O2) the intervention produced a statistically significant 23% improvement in academic performance ($F(1,178)=18.42$, $p<0.001$, $\eta^2=0.094$) and a large effect on mathematical self-efficacy ($d=1.18$); and (O3) the system achieved 87% voluntary adoption, a SUS usability score of 78.3, and demonstrated cost savings of over $9,000 per year compared to proprietary API alternatives.

Regarding RQ1, the proposed architecture successfully integrates n8n workflow automation with open source AI models and Flutter-based mobile interfaces, achieving 99.7% availability and processing 94% of mathematical queries successfully, demonstrating technical viability for production deployment.

Concerning RQ2, intelligent automation combined with dual assistance systems produced substantial improvements: 23% increase in academic performance (exceeding the typical 10-15% range for educational mobile technologies), $d=1.18$ effect size for self-efficacy, and 40% reduction in teacher diagnostic time. The differential impact favoring low-performing students (31% vs. 10% improvement) suggests compensatory potential for closing educational gaps.

Regarding RQ3, adoption patterns revealed 87% overall adoption rate with three distinct phases (adoption, stabilization, sustained usage). Open source AI models achieved accuracy comparable to proprietary systems (Llama 2: 91.2%, Mistral: 88.7%) while eliminating API costs, projecting annual savings exceeding $9,000 per institution.

Despite these results, several limitations must be acknowledged. First, the 12-week duration does not allow conclusions about the long-term sustainability of learning gains. Second, the quasi-experimental design with non-random assignment at the section level limits the strength of causal inferences, even after propensity score matching. Third, the study was conducted exclusively in urban secondary schools in Huancayo, Peru, which restricts generalizability to other geographic and socioeconomic contexts. Fourth, the sample of 12 teachers, while sufficient for thematic analysis, limits the representativeness of the monitoring system evaluation. Future research should address these limitations through longitudinal randomized controlled trials in diverse settings, including rural and resource-limited schools.

The main methodological contribution is validation of n8n as emerging educational technology that democratizes access to complex automation, eliminating technical and economic barriers that have traditionally limited EdTech innovation in resource-limited contexts.

In summary, this work contributes empirical evidence that n8n-based workflow automation represents a viable and accessible platform for building cohesive educational ecosystems in contexts where proprietary solutions are economically unfeasible. The teacher monitoring subsystem, which processed 1,247 student queries and generated automated weekly reports, reduced teacher diagnostic workload by 40% while providing data-driven insights into institutional difficulty patterns. These findings confirm that intelligent automation complements rather than replaces teacher pedagogical expertise, and that the integration of open source AI with mobile educational tools offers a scalable, cost-effective path toward personalized mathematics learning in secondary education.

REFERENCES:

[1] Smale-Jacobse AE, Meijer A, Helms-Lorenz M, Maulana R. Differentiated instruction in secondary education: A systematic review of research evidence. Front Psychol. 2019;10:2366. https://doi.org/10.3389/fpsyg.2019.02366

[2] Insorio AO. Addressing student diversity to improve mathematics achievement through differentiated instruction. Int J Prof Dev Learn Learn. 2024;6(1):ep2406. https://doi.org/10.30935/ijpdll/14462

[3] Sung YT, Chang KE, Liu TC. The effects of integrating mobile devices with teaching and learning on students' learning performance: A meta-analysis and research synthesis. Comput Educ. 2016;94:252-275. https://doi.org/10.1016/j.compedu.2015.11.008

[4] Bano M, Zowghi D, Kearney M, Schuck S, Aubusson P. Mobile learning for science and mathematics school education: A systematic review of empirical evidence. Comput Educ. 2018;121:30-58. https://doi.org/10.1016/j.compedu.2018.02.006

[5] Schwendimann BA, Rodriguez-Triana MJ, Vozniuk A, Prieto LP, Boroujeni MS, Holzer A, Gillet D, Dillenbourg P. Perceiving learning at a glance: A systematic literature review of learning dashboard research. IEEE Trans Learn Technol. 2017;10(1):30-41. https://doi.org/10.1109/TLT.2016.2599522

[6] Van Leeuwen A, Rummel N. Orchestration tools to support the teacher during student collaborative learning: A review. Unterrichtswissenschaft. 2019;47(2):143-158. https://doi.org/10.1007/s42010-019-00045-2

[7] Crompton H, Burke D. The use of mobile learning in higher education: A systematic review. Comput Educ. 2018;123:53-64. https://doi.org/10.1016/j.compedu.2018.04.007

[8] Hwang GJ, Wu PH. Applications, impacts, and trends of mobile technology-enhanced learning: A review of 2008-2012 publications in selected SSCI journals. Int J Mob Learn Organ. 2014;8(2):83-95. https://doi.org/10.1504/IJMLO.2014.062346

[9] Tingir S, Cavlazoglu B, Caliskan O, Koklu O, Intepe-Tingir S. Effects of mobile devices on K-12 students' achievement: A meta-analysis. J Comput Assist Learn. 2017;33(4):355-369. https://doi.org/10.1111/jcal.12184

[10] Güler M, Bütüner SÖ, Danişman Ş, Gürsoy K. A meta-analysis of the impact of mobile learning on mathematics achievement. Educ Inf Technol. 2022;27:1725-1745. https://doi.org/10.1007/s10639-021-10640-x

[11] Talan T. The effect of mobile learning on learning performance: A meta-analysis study. Educ Sci Theory Pract. 2020;20(1):79-103. https://doi.org/10.12738/jestp.2020.1.006

[12] Van der Kleij FM, Feskens RCW, Eggen TJHM. Effects of feedback in a computer-based learning environment on students' learning outcomes: A meta-analysis. Rev Educ Res. 2015;85(4):475-511. https://doi.org/10.3102/0034654314564881

[13] Hattie J, Timperley H. The power of feedback. Rev Educ Res. 2007;77(1):81-112. https://doi.org/10.3102/003465430298487

[14] Ma W, Adesope OO, Nesbit JC, Liu Q. Intelligent tutoring systems and learning outcomes: A meta-analysis. J Educ Psychol. 2014;106(4):901-918. https://doi.org/10.1037/a0037123

[15] VanLehn K. The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. Educ Psychol. 2011;46(4):197-221. https://doi.org/10.1080/00461520.2011.611369

[16] Kulik JA, Fletcher JD. Effectiveness of intelligent tutoring systems: A meta-analytic review. Rev Educ Res. 2016;86(1):42-78. https://doi.org/10.3102/0034654315581420

[17] Steenbergen-Hu S, Cooper H. A meta-analysis of the effectiveness of intelligent tutoring systems on college students' academic learning. J Educ Psychol. 2014;106(2):331-347. https://doi.org/10.1037/a0034752

[18] Steenbergen-Hu S, Cooper H. A meta-analysis of the effectiveness of intelligent tutoring systems on K-12 students' mathematical learning. J Educ Psychol. 2013;105(4):970-987. https://doi.org/10.1037/a0032447

[19] n8n GmbH. n8n Documentation: Workflow automation for technical teams. 2024. Available from: https://docs.n8n.io/

[20] Barricelli BR, Cassano F, Fogli D, Piccinno A. End-user development, end-user programming and end-user software engineering: A systematic mapping study. J Syst Softw. 2019;149:101-137. https://doi.org/10.1016/j.jss.2018.11.041

[21] Lacity M, Willcocks LP. Robotic process automation and risk mitigation: The definitive guide. SB Publishing; 2018. ISBN: 978-0995682047

[22] Touvron H, Martin L, Stone K, Albert P, Almahairi A, Babaei Y, et al. Llama 2: Open foundation and fine-tuned chat models. arXiv preprint. 2023;arXiv:2307.09288. https://doi.org/10.48550/arXiv.2307.09288

[23] Jiang AQ, Sablayrolles A, Mensch A, Bamford C, Chaplot DS, de las Casas D, et al. Mistral 7B. arXiv preprint. 2023;arXiv:2310.06825. https://doi.org/10.48550/arXiv.2310.06825

[24] Zhao WX, Zhou K, Li J, Tang T, Wang X, Hou Y, et al. A survey of large language models. arXiv preprint. 2023;arXiv:2303.18223. https://doi.org/10.48550/arXiv.2303.18223

[25] Yang J, Jin H, Tang R, Han X, Feng Q, Jiang H, et al. Harnessing the power of LLMs in practice: A survey on ChatGPT and beyond. ACM Trans Knowl Discov Data. 2024;18(6):1-32. https://doi.org/10.1145/3649506

[26] Zawacki-Richter O, Marín VI, Bond M, Gouverneur F. Systematic review of research on artificial intelligence applications in higher education – where are the educators? Int J Educ Technol High Educ. 2019;16:39. https://doi.org/10.1186/s41239-019-0171-0

[27] Schildkamp K, Kuiper W. Data-informed curriculum reform: Which data, what purposes, and promoting and hindering factors. Teach Teach Educ. 2010;26(3):482-496. https://doi.org/10.1016/j.tate.2009.06.007

[28] Van Leeuwen A, Janssen J, Erkens G, Brekelmans M. Supporting teachers in guiding collaborating students: Effects of learning analytics in CSCL. Comput Educ. 2014;79:28-39. https://doi.org/10.1016/j.compedu.2014.07.007

[29] Baker RS, Inventado PS. Educational data mining and learning analytics. In: Larusson JA, White B, editors. Learning Analytics. New York: Springer; 2014. p. 61-75. https://doi.org/10.1007/978-1-4614-3305-7_4

[30] Siemens G, Long P. Penetrating the fog: Analytics in learning and education. EDUCAUSE Rev. 2011;46(5):30-32. Available from: https://er.educause.edu/articles/2011/9/penetrating-the-fog-analytics-in-learning-and-education

[31] Wolff A, Zdrahal Z, Nikolov A, Pantucek M. Improving retention: Predicting at-risk students by analysing clicking behaviour in a virtual learning environment. In: Proceedings of LAK'13. ACM; 2013. p. 145-149. https://doi.org/10.1145/2460296.2460324

[32] Kaliisa R, Misiejuk K, López-Pernas S, Khalil M, Saqr M. Have learning analytics dashboards lived up to the hype? A systematic review of impact on students' achievement, motivation, participation and attitude. In: LAK'24 Proceedings. ACM; 2024. p. 1-15. https://doi.org/10.1145/3636555.3636884

[33] Tsai YS, Moreno-Marcos PM, Jivet I, Scheffel M, Tammets K, Broos T, Gašević D. The SHEILA framework: Informing institutional strategies and policy processes of learning analytics. J Learn Anal. 2018;5(3):5-20. https://doi.org/10.18608/jla.2018.53.2

[34] Funabiki N, et al. An independent learning system for Flutter cross-platform mobile programming with code modification problems. Information. 2024;15(10):614. https://doi.org/10.3390/info15100614

[35] Mahendra I, Yanto DTR. Mobile application development: A systematic literature review. In: Proceedings of ICITACEE. IEEE; 2018. p. 57-62. https://doi.org/10.1109/ICITACEE.2018.8576952

[36] Biørn-Hansen A, Majchrzak TA, Grønli TM. Progressive web apps: The possible web-native unifier for mobile development. In: Proceedings of WEBIST. SciTePress; 2017. p. 344-351. https://doi.org/10.5220/0006353703440351

[37] Ifenthaler D, Schumacher C. Student perceptions of privacy principles for learning analytics. Educ Technol Res Dev. 2016;64(5):923-938. https://doi.org/10.1007/s11423-016-9477-y

[38] Drachsler H, Greller W. Privacy and analytics: It's a DELICATE issue a checklist for trusted learning analytics. In: LAK'16 Proceedings. ACM; 2016. p. 89-98. https://doi.org/10.1145/2883851.2883893

[39] Rienties B, Herodotou C, Olney T, Schencks M, Boroowa A. Making sense of learning analytics dashboards: A technology acceptance perspective of 95 teachers. Int Rev Res Open Distrib Learn. 2018;19(5):186-202. https://doi.org/10.19173/irrodl.v19i5.3493

[40] Kaliisa R, Palmer E, Miller J, Findlay K. Teachers' perspectives on teacher-facing

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

learning analytics dashboards: A systematic literature review. Int J Educ Technol High Educ. 2023;20:33. https://doi.org/10.1186/s41239-023-00394-6

[41] Holmes W, Bialik M, Fadel C. Artificial intelligence in education: Promises and implications for teaching and learning. Boston: Center for Curriculum Redesign; 2019. ISBN: 978-1794293700

[42] Ifenthaler D, Yau JYK. Utilising learning analytics to support study success in higher education: A systematic review. Educ Technol Res Dev. 2020;68:1961-1990. https://doi.org/10.1007/s11423-020-09788-z

[43] Paulsen L, Lindsay E. Learning analytics dashboards are increasingly becoming about learning and not just analytics - A systematic review. Educ Inf Technol. 2024;29:14279-14308. https://doi.org/10.1007/s10639-023-12401-4

[44] Bodily R, Verbert K. Review of research on student-facing learning analytics dashboards and educational recommender systems. IEEE Trans Learn Technol. 2017;10(4):405-418. https://doi.org/10.1109/TLT.2017.2740172

[45] Heffernan NT, Heffernan CL. The ASSISTments ecosystem: Building a platform that brings scientists and teachers together for minimally invasive research on human learning and teaching. Int J Artif Intell Educ. 2014;24:470-497. https://doi.org/10.1007/s40593-014-0024-x

[46] Knoop-van Campen CAN, Molenaar I. How teachers integrate dashboards into their feedback practices. Frontline Learn Res. 2020;8(4):37-51. https://doi.org/10.14786/flr.v8i4.641

[47] Baral S, Botelho AF, Erickson JA, Benachamardi P, Heffernan NT. Improving automated scoring of student open responses in mathematics. In: Proceedings of EDM'21. International Educational Data Mining Society; 2021.

[48] Cheung ACK, Slavin RE. The effectiveness of educational technology applications for enhancing mathematics achievement in K-12 classrooms: A meta-analysis. Educ Res Rev. 2013;9:88-113. https://doi.org/10.1016/j.edurev.2013.01.001

[49] Roll I, Wylie R. Evolution and revolution in artificial intelligence in education. Int J Artif Intell Educ. 2016;26:582-599. https://doi.org/10.1007/s40593-016-0110-3

[50] Slavin RE, Lake C, Groff C. Effective programs in middle and high school mathematics: A best-evidence synthesis. Rev Educ Res. 2009;79(2):839-911. https://doi.org/10.3102/0034654308330968

[51] Pane JF, Griffin BA, McCaffrey DF, Karam R. Effectiveness of cognitive tutor algebra I at scale. Educ Eval Policy Anal. 2014;36(2):127-144. https://doi.org/10.3102/0162373713507480

[52] Ritter S, Anderson JR, Koedinger KR, Corbett A. Cognitive tutor: Applied research in mathematics education. Psychon Bull Rev. 2007;14:249-255. https://doi.org/10.3758/BF03194060

[53] Agudo-Peregrina AF, Iglesias-Pradas S, Conde-González MÁ, Hernández-García Á. Can we predict success from log data in VLEs? Comput Hum Behav. 2014