International Journal of Learning, Teaching and Educational Research
Vol. 25, No. 3, pp. 463-493 March 2026
https://doi.org/10.26803/ijlter.25.3.20
Received Dec 17, 2025; Revised Feb 23, 2026; Accepted Feb 26, 2026

Enhancing Pre-Service Mathematics Teachers' Digital Pedagogy through a Two-Rotation AI-Integrated Blended Learning Model for Critical and Reflective AI Use

Kuralay Atirbek*, Roza Kadirbayeva
Znanibekov University
Shymkent, Republic of Kazakhstan

Şenol DOST
Hacettepe Üniversitesi
Ankara, Turkey

Zhannat Saduakassova
Znanibekov University
Shymkent, Republic of Kazakhstan

Abstract. The rapid growth of artificial intelligence (AI) in mathematics education places new demands on teachers' digital-pedagogical competence and their ability to use AI critically. Although research highlights both the potential and risks of AI tools, there is limited evidence on how structured instructional models help future teachers move beyond uncritical reliance on AI-generated explanations. This study examined a two-rotation AI-integrated blended learning model implemented with 94 pre-service mathematics teachers in Kazakhstan. A quasi-experimental pre–post design with a control group was utilised. The experimental group ($n = 47$) completed an eight-week intervention consisting of guided AI-supported construction (visualisation, modelling, micro-teaching) followed by a systematic analysis of AI outputs (error identification, human–AI comparison, modelling with real data). Quantitative analyses (t-tests, ANCOVA) and thematic analysis of reflections and artefacts revealed significant improvements in the experimental group. Digital pedagogy showed a large effect ($d = 0.93$), while professional skills and AI readiness demonstrated medium effects ($d = 0.56$). Qualitative findings indicate a shift from initial acceptance of AI outputs towards systematic verification and more informed pedagogical decision-making. Overall, the findings suggest that combining guided AI use with structured analytical reflection supports

*Corresponding author: Kuralay Atirbek; *atirbek979797@gmail.com*

responsible and reflective AI integration in mathematics teacher education. The findings also suggest that teacher education programmes should systematically incorporate structured AI-integrated models that combine guided application with critical analysis to foster reflective, responsible, and pedagogically sound AI use.

Keywords: artificial intelligence; blended learning; digital pedagogy; pre-service teachers; mathematics education

1. Introduction

The rapid expansion of artificial intelligence (AI) in education is reshaping expectations for teachers' technological and pedagogical competence. Pre-service mathematics teachers increasingly use generative and mathematical AI tools (e.g., ChatGPT and AI-enhanced dynamic mathematics environments) to obtain explanations, visualisations, solution strategies, and modelling support (Pllana et al., 2025; Urhan et al., 2024; Vintere et al., 2024). However, effective AI integration requires more than technical proficiency; it demands digital-pedagogical competence, critical AI literacy, and evidence-based professional judgement (Kohnke et al., 2025).

Research indicates that students often struggle to formulate precise prompts and critically evaluate AI-generated reasoning. In combined ChatGPT–GeoGebra tasks, learners may accept incomplete or incorrect explanations despite recognising the support provided by the tools (Galiç et al., 2025; Yunianto et al., 2024). While specialised systems such as GeoGebra Discovery provide more reliable verification, generative models may produce logically inconsistent arguments (Botana et al., 2024). These findings highlight the need for structured instructional approaches that foster reflective engagement with AI outputs.

Although AI adoption is accelerating, teacher education programmes frequently offer fragmented exposure rather than developmentally sequenced learning experiences. As a result, pre-service teachers often lack systematic opportunities to evaluate AI-generated reasoning or integrate AI meaningfully into pedagogical decision-making (Lee & Bryan, 2025). Existing rotation-based instructional designs provide a framework for structuring learning phases (Aksak Kömür et al., 2023) yet rarely incorporate explicit AI-error analysis or staged development of critical AI literacy.

To address this gap, the present study proposes a two-rotation AI-integrated instructional model aimed at developing digital-pedagogical competence, reflective AI literacy, and professional mathematical judgement. Rotation 1 emphasises guided AI-supported construction, including explanation building, modelling, and micro-teaching tasks. Rotation 2 introduces systematic analytical deconstruction, requiring students to evaluate AI-generated reasoning, diagnose modelling errors, compare human and AI solutions, and reconstruct mathematically valid explanations. By positioning AI-generated inaccuracies as didactic resources, the model frames AI as an object of professional scrutiny rather than an authoritative source of knowledge.

Accordingly, this study examines how competence develops across two sequential rotations representing a progression from AI-assisted exploration to analytically grounded professional use. The research questions are:

RQ1. How does digital-pedagogical competence develop in a two-rotation AI-integrated environment?

RQ2. How does reflective and critical AI use evolve across instructional stages?

RQ3. How is professional mathematical and pedagogical competence shaped through engagement with AI-generated reasoning and modelling tasks?

2. Literature Review

2.1 AI in Mathematics Education: Opportunities and Challenges

The integration of AI into mathematics education has generated both pedagogical opportunities and methodological concerns. Generative and mathematical AI tools can enhance conceptual understanding through instant explanations, visualisations, and modelling support (Memiş, 2025). However, research has consistently shown that students often accept AI-generated outputs without systematic verification and struggle to evaluate automated reasoning (Galiç et al., 2025; Pitts et al., 2025; Yunianto et al., 2024). While specialised systems such as GeoGebra Discovery demonstrate higher formal reliability, generative models may produce logically inconsistent arguments (Botana et al., 2024). Beyond issues of mathematical correctness, scholars highlight concerns related to ethics, equity, and responsible implementation (Gqoli & Baidoo, 2025). These findings indicate that effective AI integration requires structured pedagogical frameworks that support critical evaluation rather than mere tool usage.

2.2 Digital-Pedagogical Competence and AI Literacy

Digital pedagogy emphasises alignment between instructional goals, teaching strategies, and technological tools (Väätäjä & Ruokamo, 2021; Voicu, 2025). Frameworks such as DigCompEdu conceptualise competence as developmental, progressing from operational use to reflective and analytical practice (Ng et al., 2023). However, in teacher education contexts, exposure to AI often remains fragmented. Pre-service teachers may experiment with tools without systematic opportunities to evaluate AI-generated reasoning or integrate AI meaningfully into pedagogical decision-making (Abualrob, 2025). Although AI literacy is increasingly recognised as essential, existing frameworks provide limited guidance on how reflective AI judgement develops within disciplinary contexts such as mathematics education.

2.3 Professional Mathematical Competence and Error-Based Learning

Professional mathematical competence involves iterative validation, comparison of alternative reasoning pathways, and systematic error diagnosis (Kumar & Behera, 2023; Toxanov et al., 2024). Emerging research suggests that comparing human and AI-generated solutions may deepen disciplinary understanding (McGalliard & Otten, 2025). Nevertheless, few instructional models explicitly position AI-generated inaccuracies as structured learning resources for developing diagnostic reasoning and pedagogical judgement in mathematics teacher education.

2.4 Rotation-Based Instructional Designs

Rotation-based instructional models organise learning through sequenced phases of theory, practice, and reflection (Aksak Kömür et al., 2023). Although effective for structuring learning, existing rotation designs rarely integrate systematic AI-error analysis or staged development of critical AI literacy.

2.5 Research Gap

Current literature addresses digital competence (Loureiro et al., 2024), AI literacy, and professional mathematical reasoning (Manunure & Leung, 2024) largely in isolation. There remains a lack of developmentally sequenced instructional models that intentionally combine AI-supported construction with analytical deconstruction of AI-generated reasoning. This study addresses this gap by proposing and empirically examining a two-rotation AI-integrated instructional model designed to support the co-development of digital-pedagogical competence, critical AI literacy, and professional mathematical and pedagogical judgement.

3. Methodology

3.1 Research Context and Participants

This study was conducted in October and November 2025 at Znanibekov University in Shymkent, Kazakhstan. The research involved third-year pre-service mathematics teachers enrolled in the same institutional mathematics education programme. Two intact academic groups were available and were assigned as the experimental and control groups, respectively, based on scheduling feasibility and comparable baseline performance in diagnostic pre-tests. Random assignment was not feasible due to institutional constraints; therefore, the study employed a quasi-experimental design with intact groups. Pre-test comparisons indicated no statistically significant differences between the groups across all measured variables prior to the intervention ($p > .05$).

All enrolled students in both groups participated in the study. Only complete pre–post datasets were included in the analysis. In all, 94 eligible students enrolled in the course participated in the study (N = 94). Two intact groups were allocated to experimental (n = 47) and control (n = 47) conditions based on scheduling feasibility and comparable pre-test performance. No participants were excluded, no attrition occurred during the eight-week intervention, and no data were removed from analysis. Thus, the final analytical sample comprised 94 complete pre–post datasets. Both groups followed the regular curriculum; however, only the experimental group received the two-rotation AI-integrated blended learning intervention. Pre-test comparisons supported baseline equivalence between groups ($p > .05$). All instructional sessions were taught by the same instructor to minimise variability in teaching style.

One of the researchers participated as a non-intrusive observer during instructional sessions, ensuring that the natural flow of classroom activities was preserved and that the learning environment in both groups remained authentic and aligned with institutional practices. Participation in the study was voluntary, and all students provided informed consent for the use of anonymised assessment data and learning artefacts. The instructional intervention was designed as an eight-week AI-integrated blended learning programme organised into two sequential rotations. The structure of the intervention, including assessment points, instructional phases, and learning activities, is presented in Figure 1.

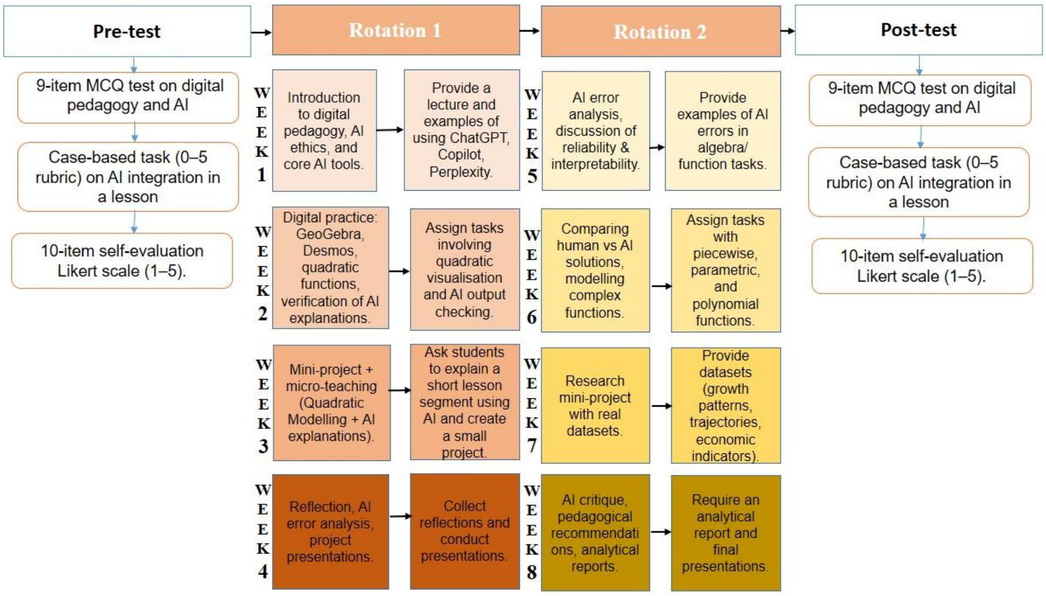


Figure 1: Structure of the two-rotation AI-integrated blended learning intervention

Although students in the control group had potential access to publicly available AI tools outside formal instruction, no structured AI-supported activities were incorporated into their coursework. The instructional tasks, modelling activities, and analytical error-based components were implemented exclusively within the experimental group. While AI use in the control group was not formally measured, the absence of guided AI integration and structured analytical tasks reduced the likelihood of systematic contamination across conditions.

3.2 Data Collection Procedures

Data were collected over an eight-week period during regularly scheduled university coursework sessions in both experimental and control groups (50 minutes per session). In the experimental group, data were gathered continuously throughout the two-rotation AI-integrated intervention to document students' engagement with AI tools, mathematical tasks, and reflective activities across progressive stages of competence development. A detailed week-by-week overview of learning objectives, activities, AI tools, outputs, and assessment points is provided in Appendix 1.

Multiple data sources ensured the comprehensive documentation of learning processes. Quantitative data included the following pre- and post-tests: a digital pedagogy knowledge test, a rubric-based instructional case task, and a 10-item AI readiness scale. Qualitative data comprised weekly artefacts (AI prompts and outputs, GeoGebra and Desmos constructions, analytical reports), micro-teaching recordings, observation notes, and written reflections.

All materials were compiled and anonymised prior to analysis. Rotation 1 focused on AI-supported exploration and micro-teaching, whereas Rotation 2 emphasised analytical engagement, including human–AI comparisons, error diagnosis, and modelling tasks. These artefacts provided evidence of developmental shifts in reflective AI use and professional mathematical judgement. Although digital platforms supported data capture, all interpretive procedures (episode selection, modelling categorisation, and thematic coding) were conducted following established mixed-methods protocols to ensure alignment with the research questions.

3.3 Instructional Model

The intervention was structured as a two-rotation AI-integrated instructional model designed to guide pre-service mathematics teachers through a developmental progression from initial AI familiarisation to analytical, evidence-based, and pedagogically grounded AI use (Tasarib et al., 2025). Conceptually, the model functions as a pedagogical framework rather than an analytical coding system. It organises learning experiences, scaffolds competence development, and specifies the instructional logic through which digital-pedagogical, reflective, and professional mathematical skills are formed.

The intervention was implemented by the course instructor, who had prior experience in digital pedagogy and AI-supported teaching. Before the intervention, a structured implementation plan was developed, including weekly learning objectives, task descriptions, AI tools to be used, and expected student outputs. Session plans were aligned with the rotation model framework to ensure consistency across weeks. To maintain standardisation, all instructional sessions followed a predefined sequence of activities (introduction, guided AI practice, modelling task, reflective discussion). The same instructor taught both the experimental and control groups to minimise variability in teaching style and content delivery.

The “blended” format consisted of in-class interactive sessions (approximately 100 minutes weekly) combined with guided independent tasks completed outside the classroom. In-class sessions focused on collaborative modelling, AI analysis, and discussion of error patterns. Out-of-class components included AI-supported mini-projects, analytical reports, and reflective logs submitted through the university’s learning management system. AI tools used included ChatGPT, GeoGebra, and Desmos. Fidelity of implementation was monitored through structured session planning and post-session documentation. An implementation log was maintained after each session to document adherence to the planned instructional sequence and completion of core activities.

To justify the design of the two-rotation AI-integrated instructional model, the study draws on and critically extends existing digital pedagogy and AI competence frameworks, including DigCompEdu, SIPE-AI, and rotation models in blended learning. As summarised in Table 1, while these frameworks address digital competencies, stages of AI-supported reasoning, or instructional sequencing, they lack structured mechanisms for developing critical AI literacy through systematic error analysis, pedagogical reflection, and diagnostic reasoning.

Table 1: Positioning the two-rotation AI model within existing frameworks

Framework	Main focus	Identified gap	Extension offered by the two-rotation AI model	Key source
DigCompEdu	Digital competences for educators	No mechanism for developing critical AI literacy or structured error-based reasoning	Phase 2 provides a systematic scaffold for evaluating AI outputs and diagnosing reasoning errors	European Commission (2017); Ng et al. (2023)
SIPE-AI	Stages of AI-supported reasoning	No pedagogical interpretation or competence-development pathway	The two-rotation AI model integrates reasoning stages with pedagogical tasks, reflection, and modelling	Yoon et al. (2024)
Rotation models in blended learning	Learning optimisation and task sequencing	Lack of integration of AI-error-based learning and diagnostic reflection	The two-rotation AI model introduces analytical deconstruction, positioning AI errors as a central didactic resource	Aksak Kömür et al. (2023); Kaidieva et al. (2025)

The proposed two-rotation AI model addresses these gaps by integrating analytical deconstruction, human–AI comparison, and error-based reflection as central didactic components. In contrast to optimisation-oriented rotation models, AI errors are positioned as purposeful learning resources, enabling the development of reflective pedagogical judgement and professional mathematical competence. This theoretical grounding informed the design of the two instructional rotations described below.

Across both rotations, the model integrates four recurring instructional components:

1. AI-assisted exploration – guided engagement with generative and mathematical AI tools.
2. Mathematical modelling and validation – constructing, interpreting, and evaluating models with and without AI support.
3. Comparison of human and AI reasoning – analysing AI-generated solutions relative to mathematically correct reasoning.
4. Structured reflective analysis – examining AI limitations, correcting reasoning chains, and formulating pedagogical justifications.

Together, these components operationalise the study's theoretical constructs by transitioning students from using AI as a supportive digital tool (Rotation 1) to treating it as an object of professional analysis and critique (Rotation 2). The complete structure of the two rotation models is presented in Figure 2.

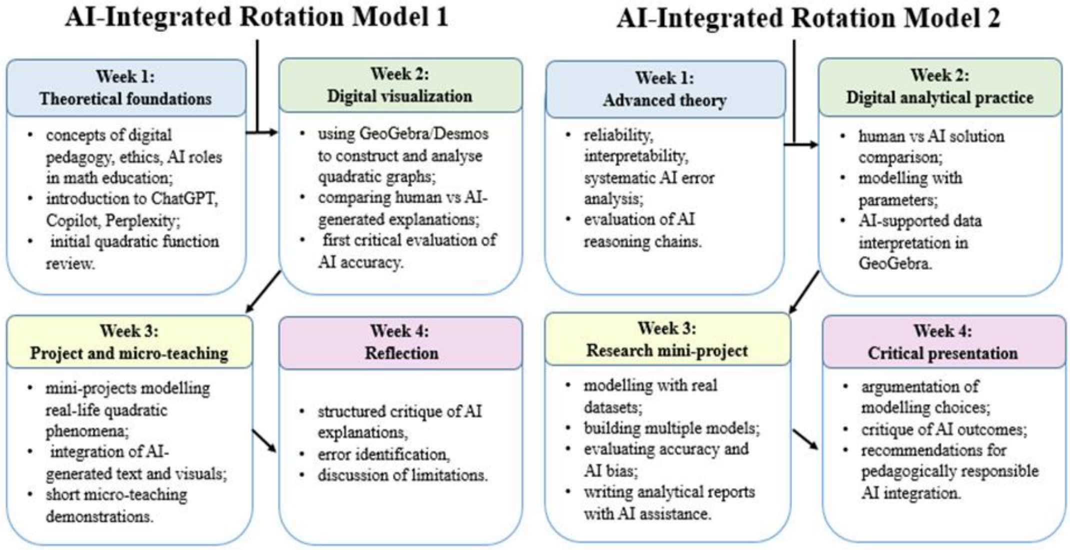


Figure 2: Structure and components of AI-integrated rotation models 1 and 2

3.3.1 Rotation 1: Foundational digital and pedagogical preparation (weeks 1–4)
Rotation 1 establishes foundational digital-pedagogical competence and introduces students to AI-supported mathematics teaching. It follows a structured movement from conceptual orientation to practical application and initial critical reflection. The four weeks in this rotation are structured as follows:

- Week 1 – Theoretical foundations: digital pedagogy, ethics, introduction to AI tools.
- Week 2 – Digital visualisation: GeoGebra/Desmos, quadratic functions, first evaluation of AI explanations.
- Week 3 – Mini-project and micro-teaching: real-world modelling, AI-supported explanations.
- Week 4 – Reflection: analysis of AI errors, discussion of limitations, pedagogical conclusions.

3.3.2 Rotation 2: Analytical and research-oriented phase (weeks 5–8)
Rotation 2 deepens analytical and professional competence by positioning AI outputs as objects of systematic mathematical and pedagogical evaluation. The four weeks in this rotation are structured as follows:

- Week 1 – Advanced theory: AI reliability, interpretability, error patterns.
- Week 2 – Analytical practice: comparison of human and AI solutions, modelling with parameters.
- Week 3 – Research project: datasets, multiple models, AI prediction error analysis, analytical writing.
- Week 4 – Critical presentation: critique of AI outcomes; pedagogical recommendations.

Through this progression, students move from guided AI use towards independent critical judgement, culminating in evidence-based decisions about responsible AI integration in mathematics teaching.

Both Rotation 1 and Rotation 2 were designed as sequential, development-oriented instructional models, each targeting a distinct layer of competence formation. Rotation 1 develops foundational digital-pedagogical competence through guided familiarisation with AI tools, introductory modelling tasks, and early critical awareness of AI-generated explanations. Its primary function is to enable students to work with AI confidently and meaningfully in mathematics teaching contexts.

Rotation 2 extends this foundation towards analytical, research-oriented, and professionally grounded engagement with AI. Through the systematic comparison of human and AI reasoning, modelling with real datasets, identification of error patterns, and pedagogical justification of instructional choices, Rotation 2 develops critical AI literacy, mathematical validation skills, and professional pedagogical judgement.

Taken together, the two rotations form a coherent developmental trajectory from initial AI-assisted exploration to advanced, evidence-based, and pedagogically responsible AI use. This directly supports the growth of digital-pedagogical competence (RQ1), reflective AI use (RQ2), and professional mathematical and pedagogical competence (RQ3). All planned instructional components were completed during the eight-week period. Informal observation notes were kept ensuring adherence to the rotation sequence and learning objectives.

3.4 Data Analysis Procedures

The study applied descriptive and inferential statistical techniques to evaluate the effectiveness of the AI-enhanced pedagogical intervention. The analytical strategy aligned with the research questions and the structure of the dataset, which consisted of pre–post measurements from an experimental and a control group across three constructs: Digital Pedagogy (Block A), Professional Teaching Skills (Block B), and AI Attitudes & Readiness (Block C). These blocks correspond to key competence domains of the DigCompEdu and P21 frameworks: Block A reflects digital teaching and learning competencies, Block B captures professional pedagogical and methodological skills, and Block C represents AI-related readiness, confidence, and dispositions towards responsible AI use. All analyses were performed using SPSS Statistics 28.

3.4.1 Block A (theoretical awareness test)

Block A consisted of nine multiple-choice items with one correct answer each. Items were scored dichotomously (correct = 1, incorrect = 0), and total scores were calculated as the proportion of correct responses (sum/9), yielding a continuous index ranging from 0 to 1. Content validity was established through expert review by two specialists in mathematics education and digital pedagogy, who evaluated item clarity, relevance, and alignment with the intended domains. Minor wording refinements were made prior to implementation based on their feedback.

Block A was constructed as a short, criterion-referenced knowledge index sampling multiple theoretically distinct but related domains of digital-pedagogical and AI-related competence. These included conceptual understanding, AI integration, ethical awareness, instructional roles, and critical evaluation. Item development was guided by key domains of the DigCompEdu and P21 frameworks to ensure theoretical alignment (see Appendix 2 for item-to-domain mapping). Given the heterogeneous content coverage and the limited number of items (n = 9), the instrument was not intended to function as a unidimensional latent scale. Accordingly, internal consistency was examined for transparency but was not treated as a primary indicator of measurement quality. The total score represents a proportion-correct index reflecting breadth of theoretical awareness rather than internal homogeneity of items.

Normality of pre–post difference scores was assessed using the Shapiro–Wilk test. Although the test indicated statistically significant deviations from normality for some variables (e.g., diff_BlockA: experimental group p = .016; control group p = .005), visual inspection of Q–Q plots and skewness/kurtosis values suggested only minor departures from normality. Given the sample size (n = 47 per group), paired-sample t-tests were considered robust to these deviations. Outlier analysis using boxplots revealed one extreme outlier in the experimental group for Digital Pedagogy. A sensitivity analysis excluding this case did not materially change the statistical significance or effect size of the results; therefore, all cases were retained in the final analyses.

For independent-samples comparisons, homogeneity of variances was assessed using Levene's test. In cases where variances were unequal, Welch's correction was examined; however, this did not alter the substantive conclusions. Accordingly, standard t-test results are reported. Because multiple dependent variables were analysed, Holm–Bonferroni corrections were applied to control Type I error inflation across the three primary outcomes. The main experimental effects remained statistically significant under the adjusted thresholds, supporting the robustness of the findings.

3.4.2 Block B (practical scenario case task)

Block B assessed applied digital-pedagogical competence through an open-ended instructional design scenario (Appendix 3). Participants were asked to describe how they would integrate AI tools into a Grade 10 lesson on quadratic functions, specifying tool choice, instructional stage, and expected student activity. Responses were evaluated using a four-criterion analytic rubric (1–5 scale per criterion) assessing:

1. Pedagogical alignment with lesson objectives.
2. Justification and appropriateness of AI tool selection.
3. Quality and feasibility of student activity design.
4. Critical and ethical consideration of AI use.

Each criterion was rated on a five-point performance scale (1 = vague or inappropriate integration; 3 = partially developed but general response; 5 = specific, pedagogically coherent, and critically justified integration). The overall Block B score was calculated as the mean of the four rubric dimensions (range 1–5).

The task was administered to both experimental and control groups and evaluated using anonymised scripts to minimise potential scoring bias. All responses were scored by a single trained rater following a standardised analytic rubric with predefined performance descriptors and anchor indicators. Prior to full-scale scoring, a brief calibration phase was conducted to ensure consistent interpretation of rubric criteria. Although interrater reliability coefficients were not calculated, the use of structured descriptors, anonymised scripts, and a standardised scoring protocol was intended to enhance scoring consistency.

3.4.3 Block C (AI attitudes & readiness)

Block C was calculated as the mean score of the 10 Likert-type items (range 1−5), with higher values indicating stronger perceived readiness and reflective AI use. The instrument functioned as an integrative index of perceived AI-related pedagogical readiness, combining dimensions of self-efficacy (e.g., confidence using AI tools), instructional integration, critical evaluation of AI-generated outputs, ethical awareness, and attitudes towards AI in education. Although the items address related but conceptually distinct aspects, they were analysed as a composite indicator of overall perceived AI-integrated teaching competence.

The 10 items were developed for the present study based on constructs identified in prior research on digital-pedagogical competence and AI readiness in education. Item wording was informed by the DigCompEdu and P21 frameworks and refined for clarity and contextual relevance. The scale was designed to function as an integrative composite index of perceived AI-integrated teaching competence rather than as a multidimensional diagnostic instrument. Given the sample size and the applied research focus, formal factor analysis was not conducted. The items were analysed as a single composite score representing overall AI attitudes and readiness. Inspection of descriptive statistics indicated no substantial ceiling effects at post-test; mean scores remained below the upper scale boundary, and variance was retained, suggesting adequate sensitivity to intervention-related changes.

Internal consistency was assessed using Cronbach's alpha. The scale demonstrated high reliability at pre-test (a = .839) and acceptable reliability at post-test (a = .706), supporting its use for research purposes. The slight reduction at post-test may reflect increased differentiation in students' self-perceptions following the intervention rather than measurement instability. In addition, pre−post scores were strongly correlated (r = .780, p < .001), indicating substantial temporal stability of the construct and supporting its suitability for use in covariance-based analyses. All items are provided in Appendix 4. The analysis entailed first computing descriptive statistics (means and standard deviations) to summarise central tendencies and variability for each construct in both groups.

To examine within-group changes, paired-sample t-tests were conducted separately for the experimental and control groups. Effect sizes were calculated to evaluate practical significance. For paired comparisons, Cohen's d was computed as the mean difference divided by the standard deviation of the difference scores. For between-group comparisons, Cohen's d was calculated using pooled post-test standard deviations. Ninety-five per cent confidence intervals (95% CI) were reported for mean differences.

In addition, analysis of covariance (ANCOVA) tests were conducted for each outcome variable, with post-test scores as dependent variables, group (experimental vs control) as the fixed factor, and pre-test scores as covariates. This approach allowed adjustment for baseline differences in the quasi-experimental design. Prior to interpretation, the assumption of homogeneity of regression slopes was examined by testing the interaction between group and pre-test scores. Partial eta squared ($\eta^2 p$) values were reported as measures of effect size.

Qualitative data were analysed using a systematic thematic analysis approach (Braun & Clarke, 2019). All materials (reflective logs, classroom artefacts, micro-teaching episodes, and final presentations) were read iteratively and coded through a combination of deductive and inductive strategies. Initial codes were informed by Research Question 2 (critical AI evaluation, ethical reasoning, pedagogical justification), while additional patterns emerged directly from the data. Codes were subsequently grouped into broader categories and refined into overarching themes through constant comparison across data sources (Table 2).

Table 2: Example of coding process (raw data to themes)

Raw excerpt	Initial code	Category	Final theme
"AI made a mistake with the direction of the branches — I checked it manually and realised it was a visualisation error."	Error identification	Critical evaluation of AI output	Growth of critical AI evaluation
"I compared AI's reasoning with my own solution before accepting the answer."	Comparative verification	Reflective judgement	Development of pedagogically grounded AI use
"We discussed whether it is ethically appropriate to rely on AI explanations in assessment tasks."	Ethical reflection	Responsible AI integration	Advancement of mathematical modelling skills

To enhance credibility, coding decisions were documented throughout the process, and themes were reviewed for internal consistency and coherence. Triangulation across multiple qualitative sources strengthened the trustworthiness of interpretations. Qualitative findings were then compared with quantitative results to explain how and why observed changes occurred, particularly regarding the development of reflective and critical AI use.

3.5 Ethical Considerations

In accordance with the Znanibekov University Policy on Research Ethics for Educational Studies Conducted within Standard Instructional Practice (2022), minimal-risk pedagogical research embedded in regular coursework does not require separate institutional review board approval. Confirmation of exemption was obtained from the University Research Office.

Participants received written information about the study's aims, procedures, and data use protocols prior to participation. Informed consent was obtained for the use of anonymised reflections, learning artefacts, assessment data, and micro-teaching recordings. Participation was voluntary and had no impact on course grades. To mitigate potential bias associated with the instructor–researcher dual role, students were explicitly informed that participation or withdrawal would not influence evaluation outcomes, and standardised grading criteria were applied uniformly.

Micro-teaching sessions were recorded with prior consent and stored on password-protected institutional devices accessible only to the research team. Data were processed in accordance with institutional data protection regulations and retained for a maximum of three years before secure deletion. All materials were anonymised prior to analysis and reported in aggregate form. AI integration followed principles of responsible and critical use. Students were instructed to verify AI-generated outputs, reflect on potential limitations, and avoid uncritical reliance. The study adhered to principles of transparency, autonomy, data protection, and academic integrity.

4. Results and Findings

Assumption diagnostics reported in the methodology section indicated no violations requiring modification of the planned parametric analyses. Boxplot inspection identified one extreme outlier in Digital Pedagogy; sensitivity analysis confirmed that its exclusion did not alter substantive conclusions.

4.1 Descriptive statistics

Descriptive statistics were first used to summarise central tendencies and variability across the three measured constructs—Digital Pedagogy, Professional Teaching Skills, and AI Attitudes & Readiness—for both the experimental and control groups. Means and standard deviations were calculated for pre- and post-test scores, providing an initial indication of overall performance changes during the intervention. For instance, the experimental group demonstrated a notable increase in Digital Pedagogy (M = 0.6147 → 0.8298), whereas the control group showed only a minimal shift. These descriptive trends offered a preliminary view of the magnitude and direction of change prior to conducting inferential analyses.

Table 3 summarises the pre- and post-intervention means and standard deviations for the experimental and control groups across all three constructs. These descriptive indicators provide a clear overview of the initial group equivalence and the magnitude of change following the intervention.

Table 3: Pre- and post-intervention means (M) and standard deviations (SD) for experimental (EG) and control (CG) groups

Construct	Group	Pre-test M (SD)	Post-test M (SD)
Digital Pedagogy (Block A)	EG	0.61 (0.20)	0.83 (0.14)
	CG	0.66 (0.12)	0.70 (0.13)
Professional Skills (Block B)	EG	3.38 (1.33)	3.85 (1.00)
	CG	3.02 (1.34)	3.23 (1.18)
AI Attitudes & Readiness (Block C)	EG	3.78 (0.52)	4.17 (0.34)
	CG	3.90 (0.43)	3.97 (0.38)

4.2 Paired-sample t-tests

Assumption checks indicated that difference scores were approximately normally distributed; minor deviations detected via Shapiro–Wilk tests did not materially affect the robustness of parametric analyses. Paired-sample t-tests were conducted to examine pre–post changes within each group. The experimental group showed significant improvements across all three constructs: Digital Pedagogy ($t(46) = –7.14$, $p < .001$), Professional Skills ($t(46) = –3.64$, $p = .001$), and AI Attitudes & Readiness ($t(46) = –9.26$, $p < .001$) (Table 4).

Table 4: Paired-sample t-test results for experimental and control groups

Group	Measure	Mean difference	*t*	df	*p*
Experimental	Digital Pedagogy	–0.215	–7.144	46	<.001
Experimental	Professional Skills	–0.468	–3.643	46	.001
Experimental	Self-evaluation (AI)	–0.389	–9.257	46	<.001
Control	Digital Pedagogy	–0.040	–2.121	46	.039
Control	Professional Skills	–0.213	–1.700	46	.096
Control	Self-evaluation (AI)	–0.074	–3.287	46	.002

These results indicate that the AI-integrated instructional model substantially enhanced students' digital, pedagogical, and AI-related competencies. In contrast, the control group showed minimal or non-significant changes in Digital Pedagogy ($t(46) = –2.12$, $p = .039$), Professional Skills ($t(46) = –1.70$, $p = .096$), and AI Attitudes & Readiness ($t(46) = –3.29$, $p = .002$), suggesting limited improvement under standard instruction. Mean differences for paired-sample t-tests were calculated as pre-test minus post-test scores; therefore, negative values indicate improvement from pre- to post-test.

4.3 Independent-sample t-tests

Independent-sample t-tests indicated that there were no significant differences between the experimental and control groups at the pre-test stage across all three domains, including Digital Pedagogy, Professional Skills, and AI Attitudes & Readiness (all $p > .15$; effect sizes small, $|d| < 0.30$), confirming baseline equivalence of the groups prior to the intervention (Table 5).

At post-test, however, the experimental group scored significantly higher than the control group on all outcome measures. Digital Pedagogy showed the strongest improvement ($t(92) = 4.49$, $p < .001$), with a large effect size ($d = 0.93$). Professional Skills also improved more in the experimental group ($t(92) = 2.73$, $p = .008$), showing a medium effect ($d = 0.56$). Similarly, AI Attitudes & Readiness demonstrated a significant medium-sized effect in favour of the experimental group ($t(92) = 2.71$, $p = .008$, $d = 0.56$).

Overall, these findings indicate that the AI-integrated rotation model produced statistically significant and practically meaningful improvements in all three measured domains.

Table 5: Results of the independent-sample t-tests between groups with Cohen's *d*

Variable	*t*	df	*p*	Mean difference	95% CI	Cohen's *d*	Interpretation of *d*
preDigPed	−1.45	92	.152	−0.0497	[−0.118, 0.019]	−0.30	Small
preProfSkil	1.31	92	.192	0.362	[−0.185, 0.909]	0.27	Small
preSelfEst	−1.17	92	.244	−0.115	[−0.309, 0.080]	−0.24	Small
postDigPed	4.49	92	<.001	0.125	[0.070, 0.181]	0.93	Large
postProfSkil	2.73	92	.008	0.617	[0.168, 1.066]	0.56	Medium
postSelfEst	2.71	92	.008	0.200	[0.054, 0.346]	0.56	Medium

To further strengthen the quasi-experimental design and control for potential baseline differences, additional ANCOVA tests were conducted using post-test scores as dependent variables, group as the fixed factor, and corresponding pre-test scores as covariates.

4.4 ANCOVA tests

For Digital Pedagogy (Block A), the assumption of homogeneity of variances was met (Levene's F (1, 92) = 1.43, p = .234) (Table 6). After adjusting for pre-test scores, a significant group effect was observed (F (1, 91) = 27.97, p < .001, partial η^2 = .235). This indicates a large intervention effect, with the experimental group significantly outperforming the control group. Item-level inspection showed that improvement occurred across multiple competence domains rather than in a single narrow skill, supporting the interpretation of Block A as a broad knowledge index.

Table 6: ANCOVA results for experimental vs control groups

Construct	Levene's *F* (*p*)	*F* (1, 91)	*p*	Partial η^2	Effect size interpretation
Digital Pedagogy (Block A)	1.43 (.234)	27.97	<.001	.235	Large
Professional Skills (Block B)	2.87 (.094)	5.16	.026	.054	Small–medium
AI Attitudes & Readiness (Block C)	1.32 (.254)	70.51	<.001	.437	Very large

For Professional Skills (Block B), homogeneity of variances was also satisfied (Levene's F(1, 92) = 2.87, p = .094). Controlling for baseline scores, a significant group effect was found (F(1, 91) = 5.16, p = .026, partial η^2 = .054), indicating a small-to-moderate effect size.

For AI Attitudes & Readiness (Block C), the assumption of homogeneity of regression slopes was examined by testing the interaction between group and pre-test scores. The interaction term was statistically significant (F(1, 90) = 13.65, p < .001), indicating that the relationship between baseline and post-test scores differed across groups. Although the adjusted group effect remained statistically significant ($F(1, 91) = 70.51$, $p < .001$, partial $\eta^2 = .437$), the significant interaction suggests that the intervention effect varied depending on initial readiness levels. Therefore, results for this variable should be interpreted as reflecting differential gains across baseline competence levels rather than a uniform treatment effect.

Pre–post scores were strongly correlated ($r = .780$, $p < .001$), indicating substantial temporal stability of AI readiness. This strong association contributed to the high proportion of explained variance observed in the ANCOVA model. Overall, the ANCOVA results confirm that the AI-integrated rotation model produced statistically significant improvements across all three domains, even after controlling for initial baseline differences. The strongest effects were observed in AI Attitudes & Readiness and Digital Pedagogy, while Professional Skills demonstrated a smaller but meaningful effect.

To control for potential Type, I error inflation due to multiple outcome testing, a Holm–Bonferroni correction was applied. All primary group effects remained statistically significant under the adjusted thresholds.

4.5 Qualitative Analysis Overview

Analysis of 376 reflective logs, 91 activity artefacts, 47 micro-teaching episodes, and 8 final presentations revealed three interrelated themes demonstrating a consistent progression of students' professional competencies. At the start of Rotation 1, most students approached AI-generated explanations with considerable trust, often assuming their correctness without verification. By the end of Rotation 2, however, students demonstrated systematic checking of AI outputs, scrutiny of reasoning steps, comparison of human and AI solutions, and recognition of typical AI-error patterns. These qualitative developments align closely with the quantitative improvements observed across all measured constructs.

4.5.1 Theme 1: Growth of critical AI evaluation

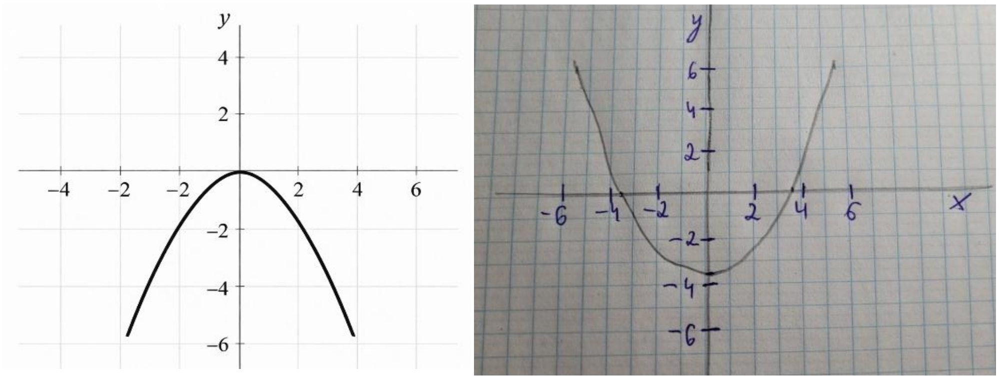
a) Sub-theme 1.1: Shift from acceptance to analytical verification

Early reflections frequently demonstrated uncritical acceptance of AI-generated answers:

> *"AI generated the graph correctly; everything is clear."* (Student 11, Week 2)

By Rotation 2, systematic verification strategies were evident (Figure 3):

> *"AI made a mistake with the direction of the branches – it used a* = 2, *but the graph looked as if a* < 0. *I checked it manually – it was a visualisation error."* (Student 11, Week 6)



Note. The black curve depicts an AI-generated graph that does not correspond to the specified coefficient (a = 2). The hand-drawn graph on the right represents the correct student-constructed parabola based on the given coefficient. The images are anonymised reproductions of classroom artefacts and are presented without identifying information. Publication of these materials complies with institutional ethical guidelines and participant consent.

Figure 3: Example of a typical AI visualisation error

With the coefficient $a = 2$, the parabola should open upwards; however, the AI-generated graph displayed the branches opening downward, which corresponds to $a < 0$. This type of visualisation error was commonly identified by students when comparing AI-generated graphs with manual constructions. This shift reflects the transition from passive reliance on AI towards analytical verification of mathematical correctness.

b) Sub-theme 1.2: Evaluation of AI reasoning chains

Students increasingly required AI systems to provide intermediate solution steps and cross-checked these using formal mathematical rules:

> *"ChatGPT gave an answer but skipped an important transformation. Now I always ask for detailed steps."* (Student 5, Week 7)

This behaviour was observed in 29 of the 47 experimental-group participants.

c) Sub-theme 1.3: Evidence-based comparison of AI and human models

During modelling tasks with real data, students demonstrated the ability to compare competing models using statistical and conceptual criteria. For example, in a plant-growth modelling task, students identified:

- incorrect sign of coefficient *b* in the AI-generated model;
- failure of the AI model to pass through key data points; and
- lower r^2 compared with a student-generated quadratic model.

> *"The AI suggested a cubic model, but the simpler quadratic model fit better."* (Students 18, 23, 24)

These observations indicate a developing capacity for evidence-based model selection grounded in mathematical validity.

4.5.2 Theme 2: Development of pedagogically grounded AI use

a) Sub-theme 2.1: The teacher as the moderator of AI use

Students increasingly recognised that the responsibility for conceptual accuracy and instructional coherence remains with the teacher rather than with AI systems:

> *"I used to think AI replaces explanation. Now I understand that the teacher checks, corrects, and explains AI's mistakes."* (Student 26, Week 8)

This shift reflects the emergence of pedagogical control over AI-supported explanations.

b) Sub-theme 2.2: Pedagogical redesign of lesson structure with AI

Students began to integrate AI outputs not as authoritative answers but as pedagogical prompts to stimulate critical thinking and classroom discussion:

> *"I would give students a task to find an error in an AI explanation — this develops their thinking. AI becomes a provocateur, not the teacher."* (Student 17, Week 7)

Overall, the qualitative and quantitative findings converge to show that students moved from technical use of AI towards pedagogically grounded AI use, marked by a clearer understanding of the teacher's professional role and the instructional implications of AI-generated solutions.

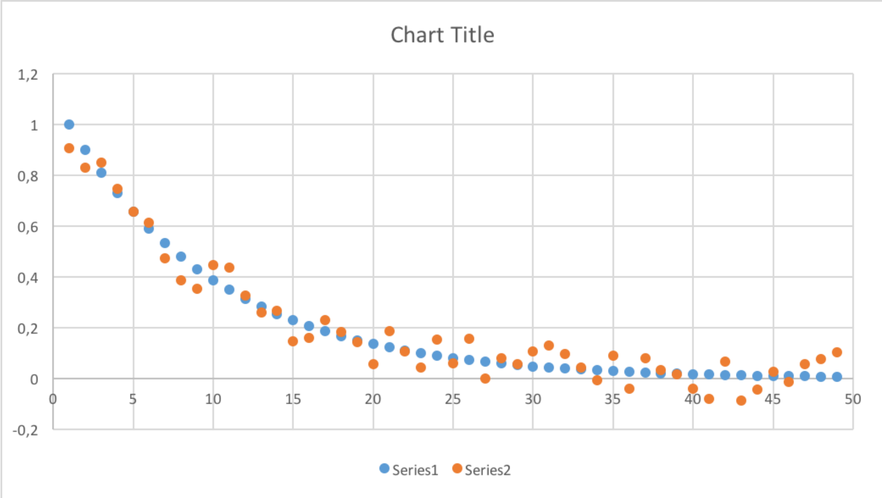
4.5.3 Theme 3: Advancement of mathematical modelling skills

a) Sub-theme 3.1: Working with real data

Students increasingly recognised the limitations of AI-generated models when applied to noisy empirical datasets:

> *"AI smoothed the data too much. Real data are noisy, and this must be considered."* (Student 20, Week 7)

This reflects growing awareness of the distinction between idealised mathematical representations and authentic data behaviour. Figure 4 illustrates a typical modelling situation in which AI tools produced a smooth model capturing the overall trend but suppressing data variability. Students noted that such over-smoothing failed to adequately represent noisy empirical data, prompting critical evaluation and alternative model selection.



Note. Blue: AI-generated model (smoothed); orange: empirical data (noisy). The visualisation is presented in anonymised form for illustrative purposes. No identifying information is included, and materials are published in accordance with participant consent and institutional ethical standards.

Figure 4: Comparison of AI-generated model output and empirical data identified by students during the second rotation

b) Sub-theme 3.1: Comparison of competing models

Students demonstrated enhanced analytical judgement when evaluating alternative models:

> *"The AI suggested an exponential model, but it didn't fit the lower points. Our polynomial model had a better r^2."* (Student 13, Week 7)

Model selection was increasingly based on statistical fit and conceptual consistency rather than on AI preference.

c) Sub-theme 3.3: Recognition of conceptual constraints

Students became more sensitive to domain and contextual restrictions in modelling:

> *"AI created a model that ignored the domain. A model must match the meaning of the task."* (Student 5, Week 6)

Across these tasks, students progressed from basic graph construction towards more sophisticated modelling practices involving validation of assumptions, comparison of competing representations, systematic error analysis, and interpretation of results in relation to a real-world context.

These qualitative developments correspond directly with the observed quantitative growth across all three measured constructs. Across Themes 1 to 3, the qualitative evidence consistently reinforces the quantitative findings:

- Digital-Pedagogical Competence: large improvement (d = 0.93)
- Professional Skills: medium improvement (d = 0.56)
- AI Attitudes & Readiness: strong gain (t(46) = –9.26, p < .001)

The intervention enabled students to evolve from passive AI users into critical evaluators, reflective practitioners, and analytically capable future mathematics teachers. The two-rotation structure was essential in supporting this developmental trajectory.

4.5.4 Rotation 1 → Rotation 2: Development of competencies

Table 7 illustrates the progression from foundational digital–mathematical skills in Rotation 1 to analytical, research-oriented, and pedagogically grounded competencies in Rotation 2. Quantitative indicators demonstrate the effectiveness of this developmental trajectory.

Table 7: Developmental progression across instructional rotations and corresponding quantitative indicators

Component	Rotation 1	Rotation 2	Quantitative evidence
Perception of AI	Trust, acceptance	Critical verification	Block C (AI Attitudes & Readiness): $F(1, 91) = 70.51$, $p < .001$, $\eta^2 = .437$
Mathematical activity	Quadratic tasks	Advanced modelling	Block B (Professional Skills): $F(1, 91) = 5.16$, $p = .026$, $\eta^2 = .054$
Modelling	Basic graphs	Model comparison (r^2)	Block A (Digital Pedagogy): $F(1, 91) = 27.97$, $p < .001$, $\eta^2 = .235$
Pedagogical thinking	AI as explainer	AI as pedagogical object	Block C pre–post increase: $t(46) = -9.26$, $p < .001$
Argumentation	Descriptive	Evidence-based	Block B pre–post increase: $t(46) = -3.64$, $p = .001$
The teacher's role	Intermediary	Moderator	Reflected in Block C growth ($\eta^2 = .437$)

4.6 Integrated Summary of Qualitative and Quantitative Evidence

Qualitative analysis revealed a clear developmental trajectory across the two instructional rotations. Rotation 1 established a foundational level of competence, including basic understanding of digital tools, initial visualisation skills, and the use of AI primarily as a tool for explanation. Rotation 2 fostered more advanced competencies, including data analysis, comparison of competing mathematical models, systematic critique of AI-generated outputs, and pedagogically grounded argumentation. As summarised in Table 7, the qualitative progression across rotations corresponded closely with quantitative gains observed in Blocks A to C.

The quantitative results strongly corroborated these patterns:

- Digital Pedagogy (Block A): large effect of improvement
- Professional Skills (Block B): medium effect of improvement
- AI Attitudes & Readiness (Block C): strong and consistent growth

The control group showed no comparable improvement, confirming that the observed growth was attributable to the two-rotation AI-integrated instructional intervention.

4.6.1 RQ1: Development of digital-pedagogical competence
Students demonstrated improvement in theoretical understanding of AI in education, evaluation of AI-generated explanations, integration of AI into lesson elements, use of dynamic visualisation tools, and confidence in digital pedagogy. Rotation 1 contributed primarily to foundational growth, whereas Rotation 2 supported more advanced evaluative and methodological competence.

4.6.2 RQ2: Change in conscious and reflective AI use
Students progressed from basic awareness towards reflective and critical AI use, including systematic evaluation of AI outputs, justification of pedagogical appropriateness of AI, metacognitive reflection on AI reliability, and transition from solution-seeking to analytical critique.

4.6.3 RQ3: Influence on professional mathematical and pedagogical competence
Students demonstrated growth in mathematical modelling, analytical reasoning, pedagogical design, reflective justification of instructional decisions, and error analysis of both mathematical and AI-generated solutions. Micro-teaching artefacts and research projects revealed increasing methodological and pedagogical sophistication across the two rotations.

5. Discussion

This study investigated whether a two-rotation AI-integrated instructional model that is developmentally sequenced contributes to the development of pre-service mathematics teachers' digital-pedagogical competence, professional skills, and reflective AI use. The findings provide converging quantitative and qualitative evidence suggesting that the intervention supported meaningful growth across all three domains, with the strongest effects observed in digital pedagogy.

The large between-group effect in Digital Pedagogy (d = 0.93) indicates that structured AI integration was associated with substantial gains in instructional competence. Importantly, this development appears to reflect more than increased familiarity with digital tools. Post-intervention artefacts and case-task responses revealed clearer alignment between lesson objectives and AI-supported activities, more deliberate tool selection, and systematic incorporation of verification-oriented tasks. Participants increasingly differentiated between generative AI used for explanation and dynamic mathematical environments used for modelling and validation.

Rather than positioning AI outputs as authoritative answers, they designed instructional scenarios in which students were expected to analyse, compare, and critically evaluate AI-generated reasoning. These shifts indicate development across key dimensions of digital-pedagogical competence, including instructional planning, tool–task alignment, formative assessment design, facilitation of student verification processes, and critical mediation of technology within mathematics instruction.

Moderate improvements in Professional Skills (d = 0.56) indicate development in modelling competence and pedagogical judgement. Qualitative data help clarify how this progression unfolded. During Rotation 1, students primarily engaged in guided AI-supported construction, often focusing on procedural modelling and visualisation. Rotation 2 introduced systematic analytical deconstruction, including comparison of human and AI reasoning chains, identification of modelling inconsistencies, and evaluation of underlying assumptions. Work with real datasets further enhanced awareness of data variability, contextual constraints, and the limitations of algorithmic smoothing. This shift from procedural engagement towards evidence-based evaluation corresponds closely with the observed quantitative gains in professional competence.

The comparison with the control group strengthens the interpretation of these findings. While the control group followed the standard mathematics education curriculum (lectures, problem-solving seminars, and limited use of conventional digital tools such as presentation software), it did not engage in structured AI-supported modelling, human–AI comparison, or systematic error-based analytical tasks. The absence of comparable growth in reflective AI use suggests that the staged analytical structure of the two-rotation model—particularly the deconstruction phase—played a central role in competence development.

With respect to AI Attitudes & Readiness, the results indicate significant gains in the experimental group. However, the ANCOVA revealed a significant interaction between baseline readiness and group membership, suggesting that the magnitude of improvement varied depending on initial competence levels. Thus, the observed effect reflects differential growth rather than a uniform impact across all participants. Qualitative reflections nonetheless indicate a shift in professional orientation: AI was increasingly perceived not as a replacement for teacher explanation, but as a tool requiring verification, contextualisation, and pedagogical framing. Students emphasised teacher responsibility for evaluating AI outputs, identifying limitations, and guiding learners in critical engagement.

Conceptually, the model operationalises AI literacy as situated professional reasoning embedded in disciplinary practice. Rather than treating AI competence as tool proficiency or ethical awareness in isolation, the two-rotation design integrates AI-supported construction with structured analytical deconstruction. This sequencing aligns with theories of reflective practice and diagnostic expertise, positioning AI-generated outputs as objects of professional scrutiny. By framing AI inaccuracies as epistemic resources, the intervention appears to support the development of diagnostic reasoning and evidence-based pedagogical judgement.

The model also differs from conventional blended rotation approaches, which typically emphasise instructional sequencing or format variation. In contrast, the present design embeds systematic human–AI comparison and error analysis within authentic mathematical modelling tasks. This integration may explain why improvements were observed not only in perceived readiness but also in demonstrated pedagogical and modelling competence.

Although the study was conducted within a university course context, the competencies developed—verification of AI outputs, comparison of alternative modelling approaches, and pedagogically justified integration of AI—are directly relevant to classroom practice. Future research should examine longitudinal transfer through practicum-based observations, delayed post-tests, or analysis of independently implemented AI-supported lessons in school settings. Such work would clarify the sustainability, transferability, and authentic classroom enactment of reflective AI competence beyond the immediate instructional environment.

Overall, the findings suggest that a developmentally sequenced, two-rotation AI-integrated approach can support pre-service teachers in moving beyond technical AI use towards analytically grounded, ethically responsible, and critically aware professional practice, including the recognition of AI limitations, potential biases, and the teacher's role in safeguarding pedagogical judgement. By combining guided construction with systematic deconstruction, the intervention contributes to the formation of critical AI literacy embedded in disciplinary and pedagogical reasoning.

6. Conclusion

This study examined the impact of a two-rotation AI-integrated instructional model on pre-service mathematics teachers' digital-pedagogical competence, professional skills, and reflective AI use. The findings indicate that structured, staged AI integration is associated with significant improvements across all three domains, with the strongest gains observed in digital pedagogy and AI-related readiness. Participants demonstrated a shift from procedural AI use towards more systematic verification, modelling evaluation, and pedagogically justified decision-making. These results suggest that combining guided AI-supported construction with structured analytical deconstruction can contribute to the development of reflective and responsible AI practices in mathematics teacher education.

The study contributes to the field by proposing a developmentally sequenced framework that embeds critical AI evaluation within authentic modelling and instructional design tasks. Rather than treating AI as a standalone digital tool, the model positions it within disciplinary reasoning and pedagogical judgement. Although conducted within a single institutional context, the framework offers a potentially transferable approach to structured AI integration in teacher preparation. Future research should explore long-term transfer, classroom enactment, and adaptation across diverse educational settings.

7. Limitations and Implications for Research and Practice

This study has several limitations that indicate directions for future research. First, the sample was limited to pre-service mathematics teachers from a single institution, which restricts the generalisability of the findings to other educational contexts, disciplines, and in-service-teacher populations. Replication across institutions and educational systems is needed to examine contextual robustness.

Second, the intervention focused on short-term learning outcomes. Although significant improvements were observed, the long-term sustainability and classroom transfer of reflective AI use remain open questions. Longitudinal designs, delayed post-tests, and practicum-based observations would help determine whether these competencies persist and translate into authentic teaching practice.

Third, qualitative evidence was primarily derived from reflective logs and learning artefacts. While these sources provided insight into students' reasoning, they do not fully capture real-time cognitive processes or interactional dynamics. Future studies could incorporate classroom discourse analysis, screen recordings, or learning analytics to triangulate findings.

Fourth, responses to the rubric-based case task were scored by a single trained rater using structured analytic criteria. Although standardised descriptors were applied, future research should incorporate independent double-scoring and report interrater reliability coefficients (e.g., ICC) to further strengthen measurement robustness.

A further limitation concerns potential informal AI use by students in the control group outside structured instruction. Although no guided AI activities were implemented in the control condition, independent access to publicly available tools cannot be entirely excluded. Future research may incorporate monitoring procedures (e.g., self-report logs or usage tracking) to minimise possible cross-condition contamination.

From a practical perspective, the findings underscore the importance of staged and scaffolded AI integration in teacher education. Rather than treating AI as a standalone digital tool, the two-rotation model embeds AI within pedagogically coherent instructional sequences. In this way, AI integration promotes systematic verification, modelling evaluation, and reflective judgement. For mathematics education in particular, AI can function not only as a support for explanation and visualisation but also as a didactic resource for fostering diagnostic reasoning and critical thinking.

More broadly, the results suggest that AI-related competencies—such as systematic output verification, model interpretation, and responsible pedagogical integration—should be explicitly incorporated into teacher education curricula and professional standards.

Conflict of Interest
The authors declare no conflict of interest.

8. References

Abualrob, M. (2025). From reflection to re-reflection: ChatGPT and transformative learning in preservice elementary science education. *International Journal of Learning, Teaching and Educational Research, 24*(10), 769–793. https://doi.org/10.26803/ijlter.24.10.37

Aksak Kömür, İ., Kılınç, H., & Okur, M. R. (2023). The rotation model in blended learning. *Asian Journal of Distance Education, 18*(2), 63–74. https://doi.org/10.5281/zenodo.8197798

Botana, F., Recio, T., & Vélez, M. P. (2024). On using GeoGebra and ChatGPT for geometric discovery. *Computers, 13*(8), Article 187. https://doi.org/10.3390/computers13080187

Braun, V., & Clarke, V. (2019). Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health, 11*(4), 589–597. https://doi.org/10.1080/2159676X.2019.1628806

European Commission. (2017). European framework for the digital competence of educators: DigCompEdu. *Publications Office of the European Union.* https://doi.org/10.2760/159770

Galiç, S., Urhan, S., Dost, Ş., & Lavicza, Z. (2025). Examining mathematics teachers noticing the rationality: Scenario-based training with AI chatbot. *Science & Education, 34*, 2759–2790. https://doi.org/10.1007/s11191-025-00618-3

Gqoli, N., & Baidoo, J. (2025). The integration of artificial intelligence in mathematics education: Towards developing a conceptual framework for understanding student perspectives in higher education. *E-Journal of Humanities, Arts and Social Sciences, 6*(10), Article 31. https://doi.org/10.38159/ehass.202561031

Kaidieva, N. K., Kozhalieva, D. Zh., & Turgunbaeva, T. Sh. (2025). Implementing practice-oriented mathematics education for students within the Sustainable Development Goals framework. In E. G. Popkova (Ed.), *Sustainable development of the green entrepreneurial economy* (pp. 265–276). World Scientific. https://doi.org/10.1142/9789811290770_0023

Kohnke, L., Zou, D., Ou, A. W., & Gu, M. M. (2025). Preparing future educators for AI-enhanced classrooms: Insights into AI literacy and integration. *Computers and Education: Artificial Intelligence, 8*, Article 100398. https://doi.org/10.1016/j.caeai.2025.100398

Kumar, M., & Behera, B. (2023). Mathematical modelling: A pedagogical tool for competency-based learning. In M. Ramesh, M. J. Arokia Marie, & M. T. V. Nagaraju (Eds.), *Perspectives in higher education* (pp. 241–256). Manglam Publications.

Lee, H., & Bryan, L. M. (2025). Integrating AI in teacher education: Exploring the impact on preservice teacher competencies. *Professional Development in Education, 51*(3), 478−494. https://doi.org/10.1080/19415257.2025.2490000

Loureiro, A. C., Santos, A. I., & Meirinhos, M. (2024). Digital competence for pedagogical integration: A study with elementary school teachers in the Azores. *Education Sciences, 14*(12), Article 1293. https://doi.org/10.3390/educsci14121293

Manunure, K., & Leung, A. (2024). Integrating inquiry and mathematical modeling when teaching a common topic in lower secondary school: An iSTEM approach. *Frontiers in Education, 9*, Article 1376951. https://doi.org/10.3389/feduc.2024.1376951

McGalliard, W., & Otten, S. (2025). AI responses to challenging problems and educator responses to AI availability. *Digital Experiences in Mathematics Education, 11*, 319−332. https://doi.org/10.1007/s40751-024-00167-4

Memiş, Y. (2025). Examining the potential and pitfalls of AI in problem solving. *Journal of Pedagogical Research, 9*(2), 177–190. https://doi.org/10.33902/JPR.202532099

Ng, D. T. K., Leung, J. K. L., Su, J., Zheng, C., Li, S., & Chan, C. K. (2023). Teachers' AI digital competencies and twenty-first century skills in the post-pandemic world. *Educational Technology Research and Development, 71*, 137–161. https://doi.org/10.1007/s11423-023-10203-6

Pitts, G., Rani, N., Mildort, W., & Cook, E.-M. (2025). Students' reliance on AI in higher education: Identifying contributing factors. *arXiv*. https://doi.org/10.48550/arXiv.2506.13845

Pllana, D., Baez, R., Lama, A. V., Sanchez, H., Sain, Z. H., & Sandeep, N. (2025). Integrating AI platforms in project-based learning. *Proceedings of the 14th Annual International Conference on Business, Technology and Innovation* (Vol. 4). UBT – Higher Education Institution. https://knowledgecenter.ubt-uni.net/conference/2025UBTIC/ED/4

Tasarib, A., Rosli, R., & Rambely, A. S. (2025). Impacts and challenges of mathematical modelling activities on students' learning development: A systematic literature review. *EURASIA Journal of Mathematics, Science and Technology Education, 21*(5), em2641. https://doi.org/10.29333/ejmste/16398

Toxanov, S., Abzhanova, D., Mukhatayev, A., Biloshchytskyi, A., & Biloshchytska, S. (2024). The development of a mathematical model for constructing an individual educational trajectory for teachers. *Education Sciences, 14*(7), Article 748. https://doi.org/10.3390/educsci14070748

Urhan, S., Gençaslan, O., & Dost, Ş. (2024). An argumentation experience regarding concepts of calculus with ChatGPT. *Interactive Learning Environments, 32*(10), 7186–7211. https://doi.org/10.1080/10494820.2024.2308093

Väätäjä, J., & Ruokamo, H. (2021). Conceptualizing dimensions and a model for digital pedagogy. *Journal of Pacific Rim Psychology, 15*. https://doi.org/10.1177/1834490921995395

Vintere, A., Safiulina, E., & Panova, O. (2024). AI-based mathematics learning platforms in undergraduate engineering studies: Analyses of user experiences [Conference session]. *Engineering for Rural Development,* 1042–1047. https://doi.org/10.22616/ERDev.2024.23.TF216

Voicu, C.-G. (2025). Integrating artificial intelligence into digital pedagogy: Challenges and opportunities. *Revista de Pedagogie Digitala, 4*(1), 31–38. https://doi.org/10.61071/RPD.2543

Yoon, H., Hwang, J., Lee, K., Roh, K. H., & Kwon, O. N. (2024). Students' use of generative artificial intelligence for proving mathematical statements. *ZDM Mathematics Education, 56*, 1531–1551. https://doi.org/10.1007/s11858-024-01629-0

Yunianto, W., Galic, S., & Lavicza, Z. (2024). Exploring computational thinking in mathematics education: Integrating ChatGPT with GeoGebra. *International Journal of Education in Mathematics, Science, and Technology, 12*(6), 1451–1470. https://doi.org/10.46328/ijemst.4437

Appendix 1

Week-by-Week Structure of the AI Two-Rotation Model

Week	Learning Objective	Core Activities	AI Tools Used	Outputs Collected	Assessment Point
1	Develop foundational AI literacy	Intro to AI pedagogy and ethics	ChatGPT, Copilot	Reflection notes	Formative
2	Verify AI-generated explanations	GeoGebra tasks, error checking	GeoGebra, ChatGPT	Annotated solutions	Rubric-based feedback
3	Apply AI in modelling tasks	Mini-project (quadratic modelling)	ChatGPT, Desmos	Micro-teaching artefact	Case-based rubric
4	Reflect on AI limitations	AI-error analysis	ChatGPT	Reflective log	Reflection coding
5	Analyse AI reliability	AI error comparison	ChatGPT, GeoGebra	Error-detection report	Analytical rubric
6	Compare human vs AI models	Complex function modelling	AI + manual modelling	Model comparison table	Professional skills rubric
7	Work with real datasets	Research mini-project	AI tools + datasets	Data modelling report	Block B assessment
8	Develop pedagogical AI stance	Final report + presentation	AI (critical use)	Analytical report	Final rubric

Appendix 2

Theoretical Awareness Test (9 items, one correct answer each)

ITEM	Question	Answer Options	Correct Answer	What it Measures	Domain
A1	What does the term digital pedagogy mean?	a) Using computers in a lesson b) Using technologies to enhance learning and thinking c) Replacing the teacher with artificial intelligence d) Creating online tests	**b**	Conceptual understanding of digital pedagogy	Conceptual digital pedagogy
A2	What is the teacher's role when using AI in the classroom?	a) Controlling the work of the AI b) Delegating all work to the AI c) Facilitating, guiding, and organising learning d) Not intervening in the process	**c**	Teacher mediation in AI-supported learning	Pedagogical mediation
A3	Which tool is most suitable for function visualisation?	a) ChatGPT b) GeoGebra c) Canva d) Grammarly	**b**	Appropriate tool selection for mathematical visualisation	Tool–task alignment
A4	Which example is not an ethical use of AI?	a) Using ChatGPT to explain a topic to a student b) Giving students an AI-generated answer without verification c) Creating tasks with the help of AI d) Checking errors in a solution using AI	**b**	Recognition of unethical AI use	Ethical AI awareness
A5	What is important when using AI in the classroom?	a) Always trusting its answers b) Verifying the accuracy of information and explaining it to students c) Avoiding AI completely d) Using it only for entertainment	**b**	Need for verification of AI outputs	Critical AI evaluation
A6	How can AI support future	a) Only by replacing the teacher	**b**	Functional integration of AI in teaching	Functional AI integration

	mathematics teachers?	b) By helping with planning, explaining, and analyzing c) Exclusively by creating presentations d) In no meaningful way			
A7	What does a project-rotation learning model include?	a) Continuous frontal instruction b) Alternation of formats: theory, practice, project, reflection c) Only independent student work d) Fully online learning only	**b**	Understanding of rotation-based instructional design	Instructional design literacy
A8	Which skill belongs to digital pedagogical competence?	a) Knowing all school programmes b) Using technologies for explanation and assessment c) Solving complex problems without tools d) Typing texts quickly	**b**	Definition of digital pedagogical competence	Technology-supported teaching
A9	Why is it important to develop critical thinking when working with AI?	a) To receive answers faster b) To distinguish between accurate and incorrect information c) To avoid using AI d) To memorise more information	**b**	Importance of critical thinking in AI contexts	Critical AI literacy

Appendix 3

Practical Scenario (Case Task)

Imagine you are preparing a lesson on "Quadratic Functions" for Grade 10 students. Your goal is to help learners understand how the graph changes when the coefficients are modified.

Question:
Describe how you can use AI to make this lesson more engaging and easier to understand.
- Which tools would you choose?
- At what stage of the lesson would they be useful?
- What will students do?

(Open-ended response, 5–10 sentences)

Appendix 4

Self-Assessment of Digital and Pedagogical Competencies

(rate each statement from 1 to 5, where 1 = "strongly disagree", 5 = "strongly agree")

№	Statement	1	2	3	4	5
1	I can use digital tools to explain mathematical ideas.	☐	☐	☐	☐	☐
2	I understand how to integrate AI into the learning process.	☐	☐	☐	☐	☐
3	I feel confident working with ChatGPT, GeoGebra, and other AI tools.	☐	☐	☐	☐	☐
4	I am able to create learning tasks that incorporate AI.	☐	☐	☐	☐	☐
5	I can explain to students how to use AI for learning.	☐	☐	☐	☐	☐
6	I understand the risks and limitations of AI.	☐	☐	☐	☐	☐
7	I believe that AI can improve the quality of learning.	☐	☐	☐	☐	☐
8	I am ready to use AI in my future teaching practice.	☐	☐	☐	☐	☐
9	I can critically evaluate results generated by AI.	☐	☐	☐	☐	☐
10	I feel inspired and interested in working with AI.	☐	☐	☐	☐	☐