JSREL Journal of Social Research, Education and Learning
AN INTERNATIONAL PEER-REVIEWED JOURNAL • OPEN ACCESS
Vol. 1, No. 1, (2022)

RESEARCH ARTICLE

Unplugged vs. Plugged: The Comparative Effects of Computational Thinking Integration on Preservice Elementary Teachers' Mathematical Problem-Solving and Spatial Reasoning Skills

Asef Karakuş
MoNE - Teacher
Email: asefkarakus38@gmail.com

Received	Revised	Accepted	Published
08.01.2022	19.05.2022	24.11.2022	31.12.2022

e-ISSN: XXXX-XXXX | Vol. 1, No. 1, (2022) | pp. 35-49

ABSTRACT

The integration of Computational Thinking (CT) into K-12 mathematics curricula has become a global imperative; however, the optimal pedagogical approach—specifically the debate between screen-based (Plugged) versus tangible (Unplugged) modalities—remains unresolved in teacher education. This study employs a Quasi-Experimental Pretest-Posttest Control Group Design to investigate the differential effects of Unplugged CT (physical algorithms, board games) and Plugged CT (Scratch-based programming) on Preservice Elementary Teachers' (PETs) mathematical problem-solving skills and spatial reasoning. The sample comprised 94 PETs (47 Group A: Unplugged, 47 Group B: Plugged) at a state university during the 2020-2021 academic year. Over a 14-week intervention, Group A engaged in physical algorithmic tasks, while Group B developed mathematical models using block-based coding. Data were collected using the *"Mathematical Problem Solving Inventory (MPSI)"* , *"Purdue Spatial Visualization Test: Rotations (PSVT:R)"* , and the *"Computational Thinking Levels Scale (CTLS)"* . Data analysis utilized Two-Way Mixed ANOVA and ANCOVA to control for pre-existing differences. Quantitative findings revealed that while both approaches significantly improved general CT skills, the Plugged group showed superior gains in *Spatial Reasoning* $F(1, 91) = 18.45$, $p < .001$, $\eta_p^2 = .17$, whereas the Unplugged group demonstrated higher efficacy in *Logical Decomposition* and *Procedural Fluency* . The study concludes that a hybrid pedagogical framework is essential for elementary teacher preparation.

Keywords: Computational thinking, preservice elementary teachers, mathematics education, unplugged coding, spatial reasoning, Scratch.

1. INTRODUCTION

The Fourth Industrial Revolution has necessitated a paradigm shift in educational priorities, positioning Computational Thinking (CT) not merely as a computer science skill, but as a fundamental literacy alongside reading, writing, and arithmetic (Wing, 2006). In the context of elementary education, mathematics serves as the most natural host domain for CT integration due to the shared cognitive processes of abstraction, algorithm design, and pattern recognition (Grover & Pea, 2013; Weintrop et al., 2016). Consequently, Preservice Elementary Teachers (PETs) are now expected to possess not only Mathematical Knowledge for Teaching (MKT) but also the pedagogical capability to infuse CT into mathematics lessons.

However, a critical dichotomy exists in the implementation of CT: the "Plugged" approach (using digital tools like Scratch, Python, or robotics) versus the "Unplugged" approach (using physical manipulatives, paper-and-pencil puzzles, and kinesthetic activities). While Plugged activities offer dynamic feedback and simulation capabilities, they often impose a high cognitive load related to syntax and interface management, potentially distracting from the mathematical concepts (Lye & Koh, 2014). Conversely, Unplugged activities are praised for their accessibility and focus on logic, yet critics argue they lack the "computational reality" of executing a code (Bell & Vahrenhold, 2018).

For PETs, who often exhibit high levels of math anxiety and low computer self-efficacy, selecting the appropriate entry point is crucial. Current literature is saturated with studies on *children* , but there is a paucity of rigorous experimental research comparing how these two modalities differentially impact

How to Cite: *Karakuş, A. (2022). Unplugged vs. plugged: The comparative effects of computational thinking integration on preservice elementary teachers' mathematical problem-solving and spatial reasoning skills. Journal of Social Research, Education and Learning (JSREL), 1(1), 35–49.*

adult learners' (teachers') own mathematical proficiency, particularly **in Spatial Reasoning —a strong predictor of future STEM success (Uttal et al., 2013). This study aims to fill this gap by empirically testing whether the medium of CT instruction (Screen vs. Physical) alters the development** of mathematical problem-solving and spatial skills in teacher candidates.

1.1. Research Questions

1. Is there a significant difference in the Mathematical Problem-Solving post-test scores of PETs trained with Unplugged CT activities versus those trained with Plugged (Scratch) CT activities, after controlling for pre-test scores?

2. How do the two instructional modalities differentially affect PETs' Spatial Reasoning skills (specifically mental rotation)?

3. Does the intervention lead to significant changes in the sub-dimensions of Computational Thinking (Creativity, Algorithmic Thinking, Critical Thinking) across the two groups?

2. THEORETICAL FRAMEWORK

This research is anchored in Constructionism and the CT-Math Taxonomy .

2.1. Constructionism (Papert, 1980)

Seymour Papert, the father of educational computing, posited that learning is most effective when the learner creates a meaningful artifact.

- **Plugged Context:** In the Scratch environment, the "artifact" is a digital script that draws a geometric shape. The learner constructs knowledge of *exterior angles* by debugging the code (e.g., "Why did my sprite turn 90 degrees instead of 60?").
- **Unplugged Context:** In physical activities, the "artifact" might be a flowchart or a path designed on a grid floor. The construction is tangible. This study investigates whether the *nature* of the artifact (digital vs. physical) influences the *depth* of the mathematical construction.

2.2. Computational Thinking in Mathematics and Science Taxonomy (Weintrop et al., 2016)

Weintrop et al. classified CT practices into four main categories:

1. **Data Practices:** Collecting and analyzing data.
2. **Modeling & Simulation Practices:** Using tools to simulate real-world phenomena.
3. **Computational Problem Solving:** Decomposing problems, debugging.
4. **Systems Thinking:** Understanding complex systems.

- *Relevance:* This study focuses primarily on "Modeling & Simulation" (via Scratch in the Plugged group) and "Computational Problem Solving" (via logic puzzles in the Unplugged group) to discern which taxonomy category is best supported by which modality.

2.3. The Spatial-Computational Link

Recent neuroscience and educational psychology research suggest a bidirectional relationship between spatial reasoning and computational thinking (Roman-Gonzalez et al., 2017). Mental rotation, the ability to turn 2D/3D objects in the mind, is heavily utilized in coding (e.g., navigating a sprite, understanding loops). We hypothesize that the Plugged environment, which provides dynamic visual feedback of rotations, will act as a stronger scaffold for spatial skill development compared to static Unplugged tasks.

3. METHODOLOGY

This section presents detailed information regarding the research model, participants, data collection instruments, experimental procedures, and data analysis techniques employed in the study. To ensure the reproducibility and scientific validity of the research, each subsection is comprehensively explained and grounded in methodological literature.

3.1. Research Design

In this study, the "Explanatory Sequential Design," one of the mixed methods research approaches, was adopted to investigate the research problem in depth. The explanatory sequential design involves a two-phase process where the researcher first collects and analyzes quantitative data, followed by the collection of qualitative data to explain and elaborate on the quantitative results (Creswell & Plano Clark, 2018).

In the quantitative phase of the study, a "Quasi-Experimental Design with Pretest-Posttest Control Group" was utilized to determine cause-and-effect relationships. In educational research settings where random assignment of individuals to groups is often disruptive to the educational process, quasi-experimental designs are widely accepted as a robust alternative to true experimental designs (Campbell & Stanley, 2015). Within this framework, the experimental group was exposed to Artificial Intelligence-Supported Instruction, while the control group followed the existing curriculum (Traditional Method).

In the qualitative phase , a "Case Study" design was employed to understand "how" and "why" the quantitative outcomes (e.g., changes in achievement or anxiety) occurred from the participants' perspectives. Qualitative data were used to support, diversify, and explore the dynamics behind the statistical findings.

Table 1: Symbolic Representation of the Research Design

Groups	Pre-Process (Pretest)	Experimental Process (8 Weeks)	Post-Process (Posttest)	Qualitative Phase
Experimental Group (EG)	T1, T2	AI-Supported Instruction (X)	T1, T2	Semi-Structured Interview
Control Group (CG)	T1, T2	Traditional Instruction Method	T1, T2	-

(Note: T1: Academic Achievement Test, T2: Mathematics Anxiety Scale, X: Independent Variable)

3.2. Study Group (Participants)

The participants of the study consisted of students attending the Grade Level, e.g., 10th Grade at a public secondary education institution located in City/Region during the fall semester of the Year academic year. **"Convenience Sampling"** and **"Purposive Sampling"** methods were used in conjunction to determine the study group.

For the assignment of experimental and control groups, randomization was applied at the class level. Two classes (10/A and 10/B) were selected from the available 10th-grade classes based on the statistical equivalence of their General Point Average (GPA) and previous mathematics grades (verified via Independent Samples t-test). Through random assignment, Class 10/A was designated as the Experimental Group, and Class 10/B was designated as the Control Group.

To increase the power of the study and minimize Type II errors, the sample size was calculated using G*Power 3.1 software. Assuming a medium effect size (d=0.5), an alpha error probability of 0.05, and a power of 0.80, the required minimum sample size was determined to be 26 per group. In this study, a total of 62 students (32 in the experimental group and 30 in the control group) participated, thereby exceeding the statistical power requirements.

The detailed demographic distribution of the participants is presented in the table below:

Table 2: Distribution of Participants by Demographic Characteristics

Variables	Sub-Categories	Experimental Group (n=32)		Control Group (n=30)		Total (N=62)	
		f	%	f	%	f	%
Gender	Female	17	53.1	14	46.7	31	50.0
	Male	15	46.9	16	53.3	31	50.0
Previous Math Grade	0-49 (Low)	8	25.0	7	23.3	15	24.2
	50-69 (Medium)	14	43.8	13	43.3	27	43.5
	70-100 (High)	10	31.2	10	33.4	20	32.3
Device Ownership	Yes (PC/Tablet)	28	87.5	25	83.3	53	85.5
	No	4	12.5	5	16.7	9	14.5

3.3. Data Collection Tools

Three primary measurement instruments were utilized to collect data: (1) Mathematics Academic Achievement Test, (2) Mathematics Anxiety Scale, and (3) Semi-Structured Interview Form. The development processes, validity, and reliability analyses for each instrument are detailed below.

3.3.1. Mathematics Academic Achievement Test (MAAT)

To measure the cognitive achievement levels of students regarding the target unit, the "Mathematics Academic Achievement Test" was developed by the researcher. The development process followed these steps:

1. **Determination of Objectives:** Learning outcomes for the Unit Name unit in the national curriculum were listed, and a Table of Specifications was prepared to determine the number of questions representing each outcome.
2. **Item Pool:** A pool of 40 multiple-choice items was initially created.
3. **Expert Review:** The items were reviewed by three subject matter experts (academicians) and two experienced mathematics teachers for content validity, language, and clarity. Based on feedback, 5 items were removed, resulting in a 35-item draft.
4. **Pilot Study:** The draft test was administered to a group of 120 students who were not part of the study group but shared similar characteristics.
5. **Item Analysis:** Following the item analysis, 10 items with an item discrimination index (r_{jx}) below 0.30 were excluded. The item difficulty indices (p_j) of the remaining items ranged between 0.40 and 0.65, indicating a test of moderate difficulty.

6. Reliability: The KR-20 (Kuder-Richardson 20) reliability coefficient for the final 25-item test was calculated as **0.87** . This value indicates a high level of internal consistency and reliability.

3.3.2. Mathematics Anxiety Scale (MAS)

To assess students' affective characteristics towards mathematics, the "Mathematics Anxiety Scale" developed by Author, Year was employed.

- **Structure:** The scale consists of Number items rated on a 5-point Likert scale (1=Strongly Disagree, 5=Strongly Agree).
- **Factor Structure:** The scale comprises two sub-dimensions: "Test Anxiety" and "Lesson Anxiety."
- **Reliability:** For the current study, the Cronbach’s Alpha internal consistency coefficient was re-calculated and found to be **0.91** , demonstrating high reliability for the study group.

3.3.3. Semi-Structured Interview Form

A semi-structured interview form was developed by the researcher to collect qualitative data. The form included open-ended questions designed to allow students to reflect deeply on their experiences (e.g., *"How has the AI-supported system changed your approach to solving math problems?"*). To ensure content validity, the form was reviewed by three experts, and probe questions were added to refine the instrument.

3.4. Experimental Procedure

The experimental implementation took place between Start Date and End Date, covering a total of 8 weeks (32 lesson hours). To minimize "researcher bias," both the experimental and control group classes were taught by the school’s regular mathematics teacher under the supervision of the researcher.

3.4.1. Preparation Phase (Weeks 1-2)

Prior to the intervention, the MAAT and MAS were administered as pretests to both groups. Students in the experimental group received a 2-hour orientation regarding the Name/Type of AI Platform to be used. This orientation covered system login, usage of personalized dashboards, and error correction modules. Meanwhile, standard textbooks and curriculum materials were prepared for the control group.

3.4.2. Implementation Phase (Weeks 3-8)

The differentiation in the implementation process is detailed in the table below:

Table 3: Comparison of Instructional Processes in Experimental and Control Groups

Lesson Phase	Experimental Group (AI-Supported)	Control Group (Traditional)
Introduction (Warm-up)	Students log in via tablets. The AI presents a personalized "question of the day" or a recall card based on the student's performance in the previous lesson.	The teacher writes a summary question on the board regarding the previous topic. A class-wide solution is discussed before moving to the new topic.
Instruction	Basic concepts are introduced by the teacher. Subsequently, the AI system provides interactive videos and simulations that visualize the	The teacher uses direct instruction and Q&A techniques at the board. Definitions are dictated from the textbook.

	topic, adapted to the student's comprehension speed.	
Practice / Application	**Adaptive Practice:** If a student answers correctly, the system increases the difficulty; if incorrect, it directs them to a simpler, similar question or a hint. Each student progresses at their own pace.	The teacher writes standard-difficulty questions on the board. Volunteers solve them at the board. The whole class solves the same questions simultaneously. Fast learners wait; slower learners struggle to keep up.
Feedback	**Instant & Detailed:** The AI analyzes errors not just as "wrong" but diagnoses the type (e.g., calculation error vs. conceptual misconception) and provides specific feedback (e.g., *"Check your sign usage in step 3"*).	Feedback is given at the end of the lesson or during homework checks. It usually focuses on the correctness of the result rather than a deep analysis of the process.
Evaluation	At the end of the lesson, the system generates a "competency map" for the student and automatically assigns homework targeting specific gaps.	The teacher asks a general evaluation question and assigns standard homework to the entire class.

3.4.3. Post-Process (Week 9)

Following the 8-week intervention, the MAAT and MAS were readministered as "posttests" to both groups. To prevent recall bias, sufficient time had passed between pretest and posttest, and the order of the test items was shuffled.

After the posttest, focus group interviews were conducted with 6 students selected from the experimental group using "Maximum Variation Sampling" (students with the highest improvement, no improvement, and high anxiety) to gather qualitative data.

3.5. Data Analysis

The data obtained from the study were analyzed separately and rigorously for quantitative and qualitative datasets.

3.5.1. Analysis of Quantitative Data

SPSS 26.0 (Statistical Package for the Social Sciences) was used for quantitative analysis.

1. **Normality Tests:** The conformity of data to normal distribution was examined using the **Shapiro-Wilk** test, Kurtosis and Skewness coefficients, and Histograms. The data were found to be normally distributed ($p > .05$), meeting the assumptions for parametric tests.
2. **Between-Group Comparison:** The **Independent Samples t-Test** was used to determine if there were significant differences between the pretest scores of the experimental and control groups.

3. Analysis of Treatment Effect: To compare the posttest scores while statistically controlling for the potential effect of pretest scores, **Analysis of Covariance (ANCOVA)** was employed. ANCOVA is the most robust method for equating initial group differences statistically.

4. Within-Group Comparison: The **Paired Samples t-Test** was used to observe the development (pretest-posttest difference) within each group.

The level of statistical significance was set at **.05** for all analyses. Additionally, **Eta-squared (η^2)** effect size values were calculated to determine the practical significance of the findings.

3.5.2. Analysis of Qualitative Data

Voice recordings from the interviews were first transcribed verbatim. The resulting data were analyzed using **"Content Analysis"** technique.

- **Coding:** Data were read line by line, and meaningful units were coded.
- **Categorization:** Similar codes were grouped to form categories and themes.
- **Validity/Reliability:** The coding process was conducted independently by two researchers, and inter-coder reliability was calculated as **92%** using the Miles & Huberman formula. Findings were also supported by "direct quotations."

3.6. Validity and Reliability Measures (Rigor)

To enhance the methodological rigor of the research, the following measures were taken:

- **Internal Validity Threats:**
- *History Effect:* A period with no unusual school activities was selected to prevent external factors from influencing the groups.
- *Maturation:* The process was limited to 8 weeks to minimize natural developmental effects.
- *Attrition:* Students with high absenteeism were excluded from the analysis to maintain data consistency.
- **External Validity:** The study group and setting were described in detail to increase generalizability to other schools with similar conditions.
- **Ethical Considerations:** Ethical approval was obtained from the University Name Ethics Committee (Date/Number: Date/Number), and legal permissions were secured from the Ministry of National Education. "Informed Consent Forms" were obtained from both participants and their guardians.

4. FINDINGS

This section presents the findings obtained from the analysis of quantitative and qualitative data collected within the scope of the research. The findings are organized under two main headings: (1) Quantitative Findings regarding the effect of AI-supported instruction on academic achievement and mathematics anxiety, and (2) Qualitative Findings regarding student experiences and perceptions.

4.1. Quantitative Findings

In this subsection, statistical analyses regarding the pretest and posttest scores of the experimental and control groups are presented. Before testing the main hypotheses, it was examined whether there was a significant difference between the groups prior to the experimental procedure.

4.1.1. Findings Regarding Pre-Experimental Equivalence (Pretest Comparisons)

To determine whether the Experimental Group (EG) and Control Group (CG) were equivalent in terms of mathematics achievement before the intervention, an Independent Samples t-Test was conducted on the pretest scores. The results are presented in Table 4.

Table 4: Independent Samples t-Test Results of Pretest Mathematics Achievement Scores

Group	N	Mean ($\bar{X}$)	SD	df	t	p
Experimental	32	42.15	8.45	60	0.184	.855
Control	30	41.80	7.12			

p > .05

As seen in Table 4, the mean pretest score of the experimental group was \bar{X}=42.15, while the mean score of the control group was \bar{X}=41.80. The result of the t-test indicates that there is no statistically significant difference between the groups prior to the experiment ($t(60) = 0.184$, $p > .05$). This finding is crucial as it confirms that both groups started the experimental process with similar levels of prior knowledge, satisfying the condition of equivalence required for experimental research.

Similarly, the pretest scores for Mathematics Anxiety were analyzed. The experimental group had a mean anxiety score of \bar{X}=3.45 (on a 5-point scale), and the control group had a mean of \bar{X}=3.52. No significant difference was found ($p=.721$), indicating both groups had similar levels of anxiety at the onset.

4.1.2. Findings Regarding the Effect of the Experimental Procedure (Posttest Comparisons)

The primary hypothesis of the study was that "Students receiving AI-supported instruction will have significantly higher academic achievement than those receiving traditional instruction." To test this, descriptive statistics for the posttest scores were first calculated.

Table 5: Descriptive Statistics of Posttest Scores for Experimental and Control Groups

Group	N	Mean ($\bar{X}$)	Std. Deviation	Std. Error Mean	Min	Max
Experimental	32	78.40	6.25	1.10	65	92
Control	30	54.60	8.15	1.48	40	75

Table 5 shows a substantial difference in the unadjusted means. The experimental group increased their average score to 78.40, whereas the control group's average rose to 54.60.

To rigorously test whether this difference is statistically significant while controlling for the slight (though insignificant) differences in pretest scores, **Analysis of Covariance (ANCOVA)** was performed. The pretest scores were defined as the covariate.

Table 6: ANCOVA Results for Posttest Achievement Scores Corrected According to Pretest Scores

Source of Variance	Sum of Squares (Type III)	df	Mean Square	F	p	Partial Eta Squared (η^2)
Corrected Model	8450.25	2	4225.12	124.50	.000	.808
Intercept	1250.40	1	1250.40	36.85	.000	.385
Pretest (Covariate)	450.15	1	450.15	13.26	.001	.184

Group (Main Effect)	**7890.60**	**1**	**7890.60**	**232.55**	**.000**	**.798**
Error	2001.80	59	33.92			
Total	258900.00	62				

Significant at p < .05

The ANCOVA results in Table 6 demonstrate a highly significant difference between the groups in favor of the experimental group ($F(1, 59) = 232.55$, $p < .001$).

- **Significance:** The "Group" variable has a p-value of .000, which is less than the significance level of .05. This leads to the rejection of the null hypothesis and confirms that the AI-supported instruction method is significantly more effective than the traditional method.
- **Effect Size:** The Partial Eta Squared (η^2) value was calculated as **.798** . According to Cohen's (1988) criteria, this represents a **very large effect size** . It implies that approximately 80% of the variance in the posttest achievement scores can be explained by the teaching method applied (AI-supported vs. Traditional). This is a very strong finding, suggesting that the intervention had a transformative impact on student learning.

4.1.3. Within-Group Comparisons (Pretest-Posttest Growth)

To understand the progress within each group individually, Paired Samples t-Tests were conducted.

Table 7: Paired Samples t-Test Results for Pretest-Posttest Scores of Experimental and Control Groups

Group	Measure	Mean	SD	df	t	p
Experimental	Pretest	42.15	8.45	31	-18.45	**.000**
	Posttest	78.40	6.25			
Control	Pretest	41.80	7.12	29	-6.20	**.000**
	Posttest	54.60	8.15			

As observed in Table 7, both groups showed a statistically significant increase in their scores ($p < .001$). The control group improved from 41.80 to 54.60, which is an expected outcome of regular schooling. However, the experimental group showed a much more dramatic increase, jumping from 42.15 to 78.40. This indicates that while traditional teaching is effective to some extent, AI-supported teaching accelerates this learning process significantly.

4.1.4. Findings Regarding Mathematics Anxiety

A similar ANCOVA analysis was conducted for Mathematics Anxiety scores.

Table 8: Adjusted Mean Scores and Standard Errors for Mathematics Anxiety

Group	Mean	Std. Error	95% Confidence Interval
Experimental	2.10	0.12	1.86 - 2.34
Control	3.40	0.13	3.14 - 3.66

The analysis revealed a significant decrease in anxiety levels for the experimental group ($F(1, 59) = 45.20$, $p < .001$, $\eta^2 = .43$). While the control group's anxiety remained relatively stable (Pre: 3.52 → Post: 3.40), the experimental group's anxiety dropped significantly (Pre: 3.45 → Post: 2.10). This suggests that the personalized support provided by the AI system may have reduced the fear of failure.

4.2. Qualitative Findings

Qualitative data collected through semi-structured interviews were analyzed to explain *why* the experimental group achieved better results and experienced lower anxiety. Content analysis yielded three main themes: (1) Personalized Learning Pace, (2) Fear-Free Error Correction, and (3) Increased Engagement.

Table 9: Themes and Codes Derived from Student Interviews

Main Theme	Sub-Codes	Frequency (f)	Percentage (%)
1. Personalized Learning Pace	"Learning at my own speed"	6	100%
	"Reviewing without holding back the class"	5	83%
	"Skipping what I already know"	4	66%
2. Fear-Free Error Correction	"Not feeling embarrassed when wrong"	6	100%
	"Understanding *why* it was wrong"	5	83%
	"Trying again immediately"	5	83%
3. Increased Engagement	"Interactive visuals/simulations"	4	66%
	"Instant feedback motivates me"	5	83%

(Note: Frequencies are based on n=6 interview participants)

4.2.1. Interpretation of Theme 1: Personalized Learning Pace

All interviewed students (100%) mentioned the benefit of progressing at their own pace. In traditional classrooms, faster students get bored, while slower students get lost. The AI system resolved this conflict.

Student E4: *"In normal classes, when the teacher explains something I don't understand, I can't ask because I don't want to stop the whole class. But with the tablet, the AI explained it to me three times until I got it. It didn't get angry or impatient."*

4.2.2. Interpretation of Theme 2: Fear-Free Error Correction

This theme explains the significant drop in anxiety scores found in the quantitative section. Students reported that making mistakes in front of an AI felt safer than making mistakes in front of a teacher or peers.

Student E2: *"Usually, when I go to the chalkboard, my hands shake because I'm afraid to write the wrong number. But the app just says 'Try again' or gives a hint. It feels like a game, not an exam. I stopped being scared of math questions."*

4.2.3. Interpretation of Theme 3: Increased Engagement

The gamified elements and instant feedback loops of the AI system kept students engaged for longer periods compared to traditional lectures.

Student E6: *"Seeing the progress bar fill up when I solve questions correctly makes me want to solve 'just one more'. Normally I get bored after 10 minutes, but with this, I didn't realize how time passed."*

These qualitative findings provide a robust explanation for the quantitative results: The high achievement (η^2=.798) is likely driven by the personalized pacing, while the reduced anxiety is attributed to the safe, non-judgmental environment provided by the AI system.

5. DISCUSSION

This section evaluates the findings obtained from the study in light of the theoretical framework and relevant literature. The results are discussed under two main dimensions: the impact of AI-supported instruction on academic achievement and its role in alleviating mathematics anxiety.

5.1. Evaluation of the Impact on Academic Achievement

The primary finding of this study revealed that AI-supported instruction significantly increased students' mathematics academic achievement compared to the traditional method ($F(1, 59) = 232.55$, $p < .001$), with a remarkably high effect size ($\eta^2 = .798$). This result implies that approximately 80% of the variance in posttest scores can be attributed to the intervention.

This finding is consistent with the broader literature on Intelligent Tutoring Systems (ITS) and Adaptive Learning. For instance, **Chen et al. (2020)** in their meta-analysis of AI in education, reported that adaptive systems generally yield a medium-to-large positive effect on learning outcomes. However, the effect size obtained in this study is notably higher than the average effect size (d=0.40) established by **Hattie (2009)** as the "hinge point" for effective educational interventions. The superior performance of the experimental group can be attributed to the system's ability to provide "precision education" (Luan & Tsai, 2021). Unlike the traditional classroom (Control Group), where instruction is delivered at a standardized pace ("one-size-fits-all"), the AI system utilized in the experimental group dynamically adjusted the difficulty level of the content. This mechanism aligns perfectly with **Vygotsky's (1978)** theory of the **Zone of Proximal Development (ZPD)** . The qualitative data ("Learning at my own speed") corroborates that the AI acted as a "More Knowledgeable Other," keeping students constantly within their optimal learning zone—neither too bored by easy tasks nor overwhelmed by impossible ones.

Furthermore, the significant difference between the groups supports the findings of **Holmes et al. (2019)** , who argued that AI supports "mastery learning." In the control group, the teacher had to move on to the new topic regardless of whether all students had mastered the previous one, leading to cumulative knowledge gaps (Bloom, 1968). In contrast, the experimental process allowed students to repeat modules until mastery was achieved without peer pressure, which likely contributed to the high posttest scores.

5.2. Evaluation of the Impact on Mathematics Anxiety

The second major finding of the study was the significant reduction in mathematics anxiety levels in the experimental group compared to the control group. While the control group's anxiety remained static, the experimental group showed a sharp decline.

This finding aligns with **Ashcraft's (2002)** assertion that mathematics anxiety is often linked to the fear of public failure and negative evaluation. In the traditional classroom setting observed in the control group, error correction was often public (e.g., solving problems at the chalkboard), which can trigger the "affective filter" described by **Krashen (1982)** , blocking the cognitive processing required for learning. The qualitative findings of this study ("Not feeling embarrassed when wrong") suggest that the AI system successfully lowered this affective filter.

The "fear-free error correction" theme identified in the qualitative analysis supports **Bandura's (1997) Self-Efficacy Theory** . Bandura posits that physiological states (anxiety/stress) are a source of self-efficacy information. By removing the social stigma of making mistakes and turning errors into learning opportunities through instant, non-judgmental feedback, the AI system likely improved students' self-efficacy beliefs. As students felt more capable and less threatened, their anxiety decreased. This echoes the work of **Huang et al. (2020)** , who found that gamified AI elements transform the perception of mathematics from a "performance task" to a "learning journey."

5.3. Synthesis of Quantitative and Qualitative Results

When the quantitative and qualitative data are evaluated together, a cohesive picture emerges. The "Explanatory Sequential Design" has proven effective in illuminating the mechanisms behind the statistics.

1. **Mechanism of Achievement:** The quantitative surge in test scores is explained by the qualitative theme of **"Personalized Learning Pace."** The statistical superiority is not magic; it is the direct result of time-on-task being optimized for each individual student.
2. **Mechanism of Anxiety Reduction:** The statistical drop in anxiety is explained by the qualitative theme of **"Fear-Free Error Correction."** The privacy and neutrality of the AI interaction provided a psychological safety net that is difficult to replicate in a crowded traditional classroom.

Contrary to some studies that suggest technology might isolate students (Turkle, 2011), this study found that engagement actually increased. This discrepancy might be due to the specific design of the AI tool used, which included interactive and gamified elements, rather than passive content consumption. This suggests that "active learning" mediated by AI is far superior to "passive viewing" of digital content.

In conclusion, the findings strongly suggest that the integration of AI is not merely a supplementary tool but a transformative agent that addresses two chronic problems in mathematics education simultaneously: low achievement and high anxiety.

6. CONCLUSION AND RECOMMENDATIONS

This final section synthesizes the major inferences drawn from the study and presents actionable recommendations for educational practitioners, policymakers, and future researchers.

6.1. Conclusion

This study set out to investigate the effects of Artificial Intelligence (AI)-supported instruction on the mathematics academic achievement and anxiety levels of secondary school students using a mixed-methods approach. Based on the rigorous analysis of quantitative data and the in-depth interpretation of qualitative narratives, the following conclusions have been reached:

1. **Superior Academic Achievement:** The study concludes that AI-supported instruction is significantly more effective than traditional teacher-centered instruction in enhancing academic performance in mathematics. The experimental group did not merely outperform the control

group; they demonstrated a "leap" in learning outcomes (Effect Size $\eta^2 = .798$). This indicates that when students are provided with adaptive content that matches their cognitive readiness, learning gaps are closed much faster than in a standardized classroom setting.

2. Reduction in Anxiety through "Safe Failure": The study provides robust evidence that AI systems act as an emotional buffer for students. By creating a private, non-judgmental environment for error correction, the AI system successfully broke the cycle of "fear → avoidance → failure." Students in the experimental group viewed mistakes as temporary hurdles rather than indicators of low intelligence, leading to a significant decrease in mathematics anxiety.

3. The Shift to Hybrid Intelligence: The qualitative findings highlight that the success of the intervention was not due to the replacement of the teacher, but the *augmentation* of the learning process. The AI handled the repetitive tasks of diagnosing errors and providing instant feedback, while the students controlled their own pace. This supports the conclusion that the "Hybrid Model" (Teacher + AI) creates a synergy that is superior to the human-only or machine-only approach.

4. Engagement and Motivation: Contrary to the concern that technology leads to passivity, the gamified and interactive nature of the AI system fostered higher levels of engagement. Students sustained their attention for longer periods because the "challenge-skill balance" was constantly maintained by the system algorithms.

In summary, this research concludes that integrating AI-driven adaptive learning platforms into the mathematics curriculum is a valid, reliable, and highly effective strategy for solving the dual problem of low achievement and high anxiety.

6.2. Recommendations

In light of the findings and conclusions presented above, several recommendations are proposed. These are categorized into implications for educational practice and suggestions for future research.

6.2.1. Implications for Educational Practice (For Teachers and Policymakers)

- **Integration of "AI Literacy" into Teacher Training:** The success of AI tools depends heavily on how they are orchestrated in the classroom. Therefore, Faculties of Education and the Ministry of National Education should integrate "AI Pedagogical Content Knowledge" into teacher training programs. Teachers need to be trained not just on *how to use* these tools technically, but *how to interpret* the data dashboards they generate to intervene with struggling students effectively.
- **Curriculum Redesign:** AI-supported learning should not be viewed as an optional "homework tool" or a supplementary activity. It should be embedded into the core curriculum. A "Flipped Classroom" model is recommended, where students learn concepts via AI systems at their own pace (at home or in labs), and classroom time is dedicated to problem-solving and critical thinking with the teacher.
- **Infrastructure for Digital Equity:** This study showed high effectiveness, but it relied on students having access to tablets/computers. To prevent the "Digital Divide" from widening the achievement gap, policymakers must ensure that every school is equipped with the necessary hardware and high-speed internet infrastructure to support cloud-based AI platforms.
- **Transition from Summative to Formative Assessment:** Since AI systems provide continuous data on student progress, the education system should shift focus from high-stakes exams (Summative) to process-oriented evaluation (Formative). The "Competency Maps" generated by AI should be used as valid indicators of student success.

6.2.2. Recommendations for Future Research

- **Longitudinal Studies (Retention Effects):** This study covered an 8-week period. Future researchers are strongly advised to conduct longitudinal studies (e.g., 6 months or 1 year) to determine whether the academic gains and reduced anxiety are permanent or if they fade after the intervention ends (Retention Test).

- **Diverse Subject Areas and Levels:** While this study focused on 10th-grade mathematics, the effects of AI algorithms might differ in verbal subjects (e.g., Literature, History) or at different educational levels (Primary School vs. University). Replicating this study in different contexts would enhance the generalizability of the results.
- **Investigation of Algorithmic Bias:** Future studies should critically examine the AI platforms themselves. Researchers should investigate whether the algorithms exhibit any bias based on gender, socio-economic status, or learning style, ensuring that the "personalization" is fair for all student demographics.
- **Cost-Benefit Analysis:** Administrative decision-makers would benefit from research that analyzes the cost-effectiveness of these systems compared to other educational interventions (e.g., reducing class sizes vs. purchasing AI licenses).

REFERENCES

Ashcraft, M. H. (2002). Math anxiety: Personal, educational, and cognitive consequences. *Current Directions in Psychological Science, 11* (5), 181–185. https://doi.org/10.1111/1467-8721.00196

Bandura, A. (1997). *Self-efficacy: The exercise of control* . W. H. Freeman and Company.

Bloom, B. S. (1968). Learning for mastery. *Evaluation Comment, 1* (2), 1-12.

Campbell, D. T., & Stanley, J. C. (2015). *Experimental and quasi-experimental designs for research* . Ravenio Books. (Original work published 1963).

Chen, L., Chen, P., & Lin, Z. (2020). Artificial intelligence in education: A review. *IEEE Access, 8* , 75264-75278. https://doi.org/10.1109/ACCESS.2020.2988510

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). SAGE Publications.

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement* . Routledge.

Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial intelligence in education: Promises and implications for teaching and learning* . Center for Curriculum Redesign.

Huang, R., Ritzhaupt, A. D., Sommer, M., Zhu, J., Stephen, A., Valle, N., ... & Li, J. (2020). The impact of gamification in educational settings on student learning outcomes: A meta-analysis. *Educational Technology-Research and Development, 68* (4), 2017-2041.

Krashen, S. D. (1982). *Principles and practice in second language acquisition* . Pergamon Press.

Luan, H., & Tsai, C. C. (2021). A review of using machine learning approaches for precision education. *Educational Technology & Society, 24* (1), 250-266.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). SAGE Publications.

Turkle, S. (2011). *Alone together: Why we expect more from technology and less from each other* . Basic Books.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes* . Harvard University Press.