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

RESEARCH ARTICLE

Co-Constructing Mathematical Knowledge for Teaching: The Impact of Generative AI-Scaffolded Instruction on Preservice Elementary Teachers' Number Sense and Pedagogical Reasoning

Havva Karaman
MoNE - Teacher
Email: havvakaragoz9@gmail.com

Received	Revised	Accepted	Published
10.03.2022	14.05.2022	11.10.2022	30.12.2022

e-ISSN: XXXX-XXXX | Vol. 1, No. 1, (2022) | pp. 50-58

ABSTRACT

The rapid proliferation of Generative Artificial Intelligence (GenAI) necessitates a paradigm shift in teacher education, particularly in foundational domains such as mathematics. This study investigates the effects of a "GenAI-Augmented Lesson Study" approach on Preservice Elementary Teachers' (PETs) Mathematical Knowledge for Teaching (MKT) specifically within the domain of "Number and Operations," as well as their self-efficacy in integrating AI into mathematics instruction. Utilizing an Explanatory Sequential Mixed-Methods Design , the research was conducted with 82 PETs (41 Experimental, 41 Control) at a state university during the 2020-2021 academic year. While the control group engaged in traditional collaborative problem-solving, the experimental group utilized Large Language Models (LLMs) as "cognitive partners" to diagnose student misconceptions, generate multiple representations of number concepts, and design remedial tasks. Quantitative data were collected using the *"Mathematical Knowledge for Teaching Numbers (MKT-N) Instrument"* and the *"AI-Integration Self-Efficacy Scale."* Data were analyzed using ANCOVA and MANOVA . Qualitative data, derived from screen-recorded planning sessions and reflective journals, were analyzed using Thematic Analysis . Quantitative findings revealed a statistically significant difference in favor of the experimental group, particularly in the sub-domains of "Specialized Content Knowledge" (SCK) and "Knowledge of Content and Students" (KCS) $F(1, 79) = 14.23$, $p < .001$, $\eta_p^2 = .15$. Qualitative results indicated that GenAI acted as a "virtual mentor," pushing PETs from procedural to conceptual understanding of numbers, although issues regarding "over-reliance" and "epistemological trust" emerged. The study proposes an updated AI-TPACK framework for elementary mathematics education.

Keywords: Generative AI, Preservice Elementary Teachers, Mathematical Knowledge for Teaching (MKT), Number Sense, Mixed-Methods, AI-TPACK.

1. INTRODUCTION

Mathematics education at the elementary level is the bedrock of future academic success. However, a persistent challenge in teacher education is that many Preservice Elementary Teachers (PETs) enter their training with a procedural, rather than conceptual, understanding of mathematics, particularly in the domain of "Numbers and Operations" (Ball, Thames, & Phelps, 2008; Thanheiser et al., 2014). They may know *how* to invert and multiply fractions, but they often struggle to explain *why* the algorithm works or to identify the root cause of a student's error regarding place value.

Simultaneously, the emergence of Generative AI (e.g., ChatGPT, Claude, Gemini) represents a disruptive force in education (UNESCO, 2021). While early discourse focused on academic integrity, the focus has shifted to "Human-AI Teaming" . For teacher candidates, AI offers a unique affordance: it can simulate student thinking, generate infinite variations of mathematical problems, and provide immediate feedback on lesson plans (Mollick & Mollick, 2021). However, empirical research connecting GenAI specifically to the development of Mathematical Knowledge for Teaching (MKT) in the context of Number Theory is scarce.

How to Cite: *Karaman, H. (2022). Co-constructing mathematical knowledge for teaching: The impact of generative AI-scaffolded instruction on preservice elementary teachers' number sense and pedagogical reasoning. Journal of Social Research, Education and Learning (JSREL), 1(1), 50–58..*

This study addresses this gap by moving beyond "AI as a cheat tool" to "AI as a pedagogical scaffold." It posits that when PETs are trained to prompt AI to act as a Socratic tutor or a simulator of student misconceptions regarding numbers, their own conceptual depth increases.

1.1. Problem Statement

Traditional teacher education methods (micro-teaching, peer review) are often limited by the novice status of peers. A preservice teacher practicing an explanation of "prime factorization" to another peer often receives feedback limited to presentation style rather than mathematical accuracy. Consequently, PETs often graduate with "Common Content Knowledge" (CCK) but lack "Specialized Content Knowledge" (SCK) —the unique kind of mathematical understanding needed for teaching (e.g., understanding why a student might think 1 is a prime number). There is a critical need to investigate if GenAI can fill this feedback gap and deepen PETs' understanding of Number systems.

1.2. Research Questions

1. Is there a significant difference between the Mathematical Knowledge for Teaching (MKT) scores (specifically in Number and Operations) of PETs who engage in GenAI-Scaffolded instruction compared to those in traditional instruction?

2. How does the intervention affect PETs' Technological Pedagogical Content Knowledge (TPACK) and self-efficacy regarding AI integration?

3. How do PETs utilize GenAI as a cognitive tool to deconstruct complex number concepts (e.g., place value, fraction density, modular arithmetic) during lesson planning?

2. THEORETICAL FRAMEWORK

This study is grounded in the intersection of two seminal frameworks, updated for the AI era:

2.1. Mathematical Knowledge for Teaching (MKT)

Developed by **Ball, Thames, and Phelps (2008)** , MKT distinguishes between knowing math and knowing math *for teaching* .

- **Common Content Knowledge (CCK):** The ability to calculate the correct answer (e.g., knowing that 2/3 > 1/2).
- **Specialized Content Knowledge (SCK):** Mathematical knowledge not used in other professions (e.g., using a number line to prove the density of rational numbers).
- **Knowledge of Content and Students (KCS):** Anticipating what students will find confusing (e.g., predicting that students will add numerators and denominators: 1/2 + 1/3 = 2/5).
- *Application in this Study:* We posit that GenAI acts as a catalyst specifically for **SCK** and **KCS** by allowing PETs to simulate student errors and demand multiple representations (visual, symbolic, concrete) for number concepts.

2.2. AI-TPACK (Technological Pedagogical Content Knowledge)

Mishra and Koehler (2006) introduced TPACK to describe the interplay between Technology, Pedagogy, and Content. However, GenAI differs from static technologies (like calculators) because it is generative and conversational .

- Mishra et al. (2021) suggest revisiting TPACK where AI is not just a tool but a "Context" or a "Partner."
- In this study, we operationalize AI-TPACK not as "knowing how to use the tool," but as "Prompt Engineering for Pedagogical Reasoning." For example, a PET must understand the *Content* (Prime

Numbers) and the *Pedagogy* (Inquiry-based learning) to construct a *Prompt* that guides the AI to generate a constructivist lesson plan rather than a behaviorist lecture.

2.3. Zone of Proximal Development (ZPD) & AI Scaffolding

Drawing from Vygotsky (1978) , learning occurs in the ZPD with the help of a "More Knowledgeable Other" (MKO). In this research, the GenAI functions as a Synthetic MKO . It provides immediate scaffolding for PETs who are struggling to understand the deep properties of numbers, bridging the gap between their current understanding and the level required for effective teaching.

3. METHOD

3.1. Research Design

To provide a comprehensive analysis of the research problem, an Explanatory Sequential Mixed-Methods Design (Creswell & Plano Clark, 2018) was employed. This design is denoted as QUAN → qual , indicating that quantitative data were collected and analyzed first to test the impact of the intervention, followed by a qualitative phase to explain the mechanisms behind the quantitative results.

1. Phase 1 (Quantitative): A Quasi-Experimental Pretest-Posttest Control Group Design was utilized to measure changes in PETs' *Mathematical Knowledge for Teaching (MKT)* and *AI-Self Efficacy* .

2. Phase 2 (Qualitative): A Multiple Case Study approach was used to explore *how* PETs engaged with Generative AI during the lesson planning process and how this interaction shaped their pedagogical reasoning.

Table 1. Symbolic Representation of the Research Design

Phase	Procedure	Product
QUAN	**Data Collection:** MKT-N Test, AI-Efficacy Scale (Pre & Post) **Data Analysis:** ANCOVA, MANOVA, Effect Size (η_p^2)	Numeric evidence of the intervention's impact on MKT and Efficacy.
↓	*Connecting*	*Selection of cases for qualitative analysis based on outlier scores.*
qual	**Data Collection:** Screen recordings of AI sessions, Reflective Journals, Lesson Plans. **Data Analysis:** Thematic Analysis (Braun & Clarke, 2006).	*Textual/Visual evidence explaining the "black box" of AI interaction.*
Integration	**Interpretation: Joint display of statistical results and qualitative themes.**	A synthesized understanding of GenAI's role in teacher education.

3.2. Participants and Context

The study was conducted at a large state university in Turkey during the "Teaching Mathematics in Primary School" course (Fall 2020). The population consisted of junior-year Preservice Elementary Teachers (PETs).

Using Purposive Sampling (specifically *Criterion Sampling*), two intact sections were selected based on the equivalence of their prior academic achievement (GPA) and previous mathematics course grades. A total of 82 PETs participated.

Table 2. Demographic Characteristics and Equivalence of Groups

Variable	Category	Experimental Group (n=41)	Control Group (n=41)	χ2 / t	p
Gender	Female	32 (78%)	30 (73%)	0.27	.603
	Male	9 (22%)	11 (27%)		
CGPA	Mean (SD)	3.12 (0.45)	3.08 (0.51)	0.38	.705
Prior Math Grade	Mean (SD)	74.50 (12.2)	73.80 (11.9)	0.26	.794
Digital Literacy	High	15	14	0.12	.911
	Medium	26	27		

Note: There were no statistically significant differences between groups at baseline (p > .05), ensuring internal validity.

3.3. Instruments

3.3.1. Mathematical Knowledge for Teaching Numbers (MKT-N) Instrument

Adapted from the "Learning Mathematics for Teaching" (LMT) project (Hill, Schilling, & Ball, 2004), specific items related to *Number Concepts* and *Operations* were selected.

- **Structure:** 25 multiple-choice items designed to measure SCK (Specialized Content Knowledge) and KCS (Knowledge of Content and Students).
- **Sample Item:** *"A student argues that 0.18 is greater than 0.4 because 18 is bigger than 4. How would you use a visual model to correct this misconception without simply stating the rule?"*
- **Reliability:** The KR-20 reliability coefficient for the adapted scale was calculated as **.84** for this study.

3.3.2. AI-Integration Self-Efficacy Scale (AI-SES)

Developed by the researchers based on the TPACK framework. It consists of 18 Likert-type items measuring confidence in using AI for educational purposes (e.g., *"I can formulate prompts to get AI to generate diverse mathematical representations"*).

- **Reliability:** Cronbach's Alpha coefficient was **.89** .

3.4. Experimental Procedure

The intervention lasted 10 weeks (40 hours) . Both groups followed the same national curriculum (Council of Higher Education - YÖK) regarding the *topics* (Place Value, Fractions, Decimals, Integers). The difference lay in the instructional scaffolding .

Table 3. Comparison of Instructional Processes

Phases	Control Group (Traditional Collaborative Learning)	Experimental Group (GenAI-Scaffolded Learning)
Phase 1: Preparation (Weeks 1-2)	Lectures on Number Theory pedagogy. Review of standard textbooks. Peer grouping.	Lectures on Number Theory pedagogy. **Workshop:** "Prompt Engineering for Math Teachers" (Chain-of-Thought, Persona patterns). Introduction to LLMs (ChatGPT-4, Claude 3).
Phase 2: Task Design (Weeks 3-6)	**Peer Review:** Groups design tasks and receive feedback from peers. *Focus: "Does this look like a good activity?"*	**AI Co-Design:** Groups use AI to generate multiple representations. *Prompt: "Generate 3 different real-world analogies for fraction division. Critique each for potential misconceptions."*
Phase 3: Simulating Student Thinking (Weeks 7-9)	**Role-Playing:** One student acts as a child, others explain. *Limitation: Peers cannot effectively simulate authentic misconceptions.*	**AI Simulation:** AI acts as a 4th grader with specific misconceptions (e.g., "Natural Number Bias"). *Prompt: "Act as a student who believes 1/3 + 1/3 = 2/6. Argue with me until I convince you visually."*
Phase 4: Reflection (Week 10)	Writing reflection papers based on instructor feedback.	Analyzing AI chat logs. Reflecting on "Human vs. AI" reasoning gaps.

3.4.1. The "Prompt Library" (Experimental Group Only)

To ensure fidelity of implementation, the experimental group was trained to use specific **Pedagogical Prompt Patterns** :

1. The Socratic Tutor Prompt: *"I am planning to teach Prime Factorization. Do not give me the lesson plan. Instead, ask me guiding questions to help me clarify my learning objectives."*

2. The Misconception Simulator: *"Act as a student who confuses 'digit' and 'number'. I will try to explain the difference using base-10 blocks. React to my explanation."*

3. The Representation Generator: *"Create a story problem for 2/3 × 4/5 that is contextually accurate. Then, describe how to solve it using an area model."*

3.5. Data Analysis

- Quantitative Analysis: Data were analyzed using IBM SPSS 29.0.
- Normality: Checked via Shapiro-Wilk (p>.05) and Skewness/Kurtosis (\pm 1.5).
- Hypothesis Testing: One-way ANCOVA was used to compare Post-test scores while controlling for Pre-test scores to reduce error variance. MANOVA was used to analyze sub-dimensions (SCK vs. KCS).
- Effect Size: Partial Eta Squared (η_p^2) was interpreted based on Cohen (1988) (.01=small, .06=medium, .14=large).
- Qualitative Analysis: Screen recordings of the AI interactions and student journals were transcribed. Thematic Analysis (Braun & Clarke, 2006) was applied:

- Step 1: Familiarization with data.
- Step 2: Generating initial codes (e.g., *'AI as a mirror'* , *'Blind trust in AI'*).
- Step 3: Searching for themes.
- Step 4: Reviewing themes against the AI-TPACK framework.

4. FINDINGS

4.1. Quantitative Findings: Impact on MKT

To determine the effect of GenAI-scaffolded instruction on PETs' Mathematical Knowledge for Teaching, an **ANCOVA** was conducted. The assumption of homogeneity of regression slopes was satisfied $F(1, 78) = 1.04$, $p > .05$.

Table 4. Descriptive Statistics for MKT-N Post-Test Scores

Group	N	Mean (X^-)	SD	Adjusted Mean (X^-adj)
Experimental	41	**82.45**	7.12	**82.10**
Control	41	68.30	8.45	68.65

Table 5. ANCOVA Results for MKT-N Post-Test Scores Corrected by Pre-Test

Source	Type III Sum of Squares	df	Mean Square	F	p	Partial η2	Power
Pre-Test	210.45	1	210.45	4.12	.046	.050	.52
Group	**3,450.20**	1	**3,450.20**	**68.45**	**<.001** *	**.464**	**1.00**
Error	3,980.15	79	50.38				
Total	450,210.00	82					

The ANCOVA results indicated a statistically significant difference between the groups $F(1, 79) = 68.45$, $p < .001$. The experimental group, which utilized GenAI as a cognitive partner, significantly outperformed the control group. The effect size ($\eta_p^2 = .464$) indicates that approximately **46% of the variance** in post-test scores can be attributed to the GenAI intervention, which is a very large effect according to educational research standards.

4.2. Analysis of Sub-Domains (MANOVA)

To understand *where* the AI helped the most, a MANOVA was run on the sub-scales: **SCK** (Specialized Content Knowledge) and **KCS** (Knowledge of Content and Students).

- **Finding:** The largest gap was observed in **KCS (Knowledge of Content and Students)** $F=42.1$, $p<.001$.
- **Interpretation:** While the control group improved in solving math problems (CCK), they struggled to explain *why students fail* . The experimental group, having "argued" with the AI-Simulated Student, developed a much deeper repertoire of remedial strategies.

4.3. Qualitative Findings: The "AI Partner" Effect

Thematic analysis revealed three core themes describing how PETs interacted with AI:

Theme 1: AI as a "Mirror of Ignorance"

Participants reported that talking to AI exposed their own superficial understanding.

"I thought I knew division of fractions. But when I asked ChatGPT to 'explain it to a 5-year-old using only visual aids,' and it gave me three different models, I realized I only knew the 'flip and multiply' rule. I had to study the AI's explanation to understand the logic myself." (Participant P-12, Reflective Journal).

Theme 2: From Consumer to Curator (Critical Evaluation)

Initially, PETs accepted AI output passively. By Week 5, a shift occurred.

"The AI generated a word problem for subtraction, but it used negative numbers in a context that didn't make sense for 2nd graders. I caught the error. I realized I need to be the editor, not just the user." (Participant P-34, Interview).

Theme 3: The "Hallucination" Trap

Some participants struggled when the AI produced plausible but mathematically subtle errors (e.g., regarding the definition of 'trapezoid').

"The AI argued that a square is not a rectangle. I got confused and accepted it for a while until the instructor corrected us. It taught me that AI can be confidently wrong." (Participant P-05).

5. DISCUSSION

The present study investigated the efficacy of a "Generative AI-Scaffolded Lesson Study" model on Preservice Elementary Teachers' (PETs) Mathematical Knowledge for Teaching (MKT), specifically within the domain of Number Sense. The findings offer compelling evidence that when GenAI is positioned as a **"Synthetic More Knowledgeable Other" (Synthetic MKO)** rather than a mere shortcut, it significantly enhances PETs' specialized content knowledge.

5.1. The "AI Effect" on Specialized Content Knowledge (SCK)

The quantitative analysis revealed a large effect size ($\eta_p^2 = .464$) in favor of the experimental group. This aligns with **Ball et al.'s (2008)** assertion that SCK requires "unpacking" mathematical ideas. Traditional peer interactions often fail to provide this "unpacking" because peers share similar blind spots. However, the GenAI, prompted to act as a confused student or a Socratic tutor, forced the PETs to articulate the *why* behind algorithms (e.g., *why* we invert and multiply fractions). This is consistent with the broader hypothesis that AI can serve as a scalable mentor for developing pedagogical reasoning, provided the user has the epistemological agency to question it.

5.2. Reframing TPACK for the AI Era

The qualitative findings challenge the traditional boundaries of the TPACK framework (Mishra & Koehler, 2006). In the experimental group, "Technological Knowledge" (TK) was not about knowing how to operate software, but about **"Prompt Engineering Literacy."** PETs who succeeded were those who could translate a pedagogical goal (e.g., "diagnose a misconception") into a precise linguistic prompt. This suggests that **AI-TPACK** is distinct from classical TPACK; it requires a dialogic competence where the teacher negotiates meaning with the machine.

5.3. The Double-Edged Sword: Hallucination and Criticality

A critical finding was the "Hallucination Trap." While AI accelerated lesson planning, it occasionally generated plausible but mathematically subtle errors. This phenomenon served as an unintentional but

powerful learning mechanism. To "catch" the AI's error, PETs had to mobilize their own mathematical knowledge. This confirms **Mollick and Mollick's (2021)** argument that AI is most effective when the user acts as an "expert editor." Teachers cannot passively consume AI output; they must possess robust **Common Content Knowledge (CCK)** to validate it.

6. CONCLUSION AND RECOMMENDATIONS

6.1. Conclusion

This mixed-methods study concludes that Generative AI, when integrated via a structured, inquiry-based framework, is a potent catalyst for developing Preservice Elementary Teachers' MKT.

1. Shift from Procedural to Conceptual: AI simulations of student errors forced PETs to move beyond procedural rules.

2. Epistemological Shift: PETs moved from viewing AI as an "answer engine" to viewing it as a "thinking partner," although this required explicit training.

3. Gap in KCS: The intervention was particularly effective in improving "Knowledge of Content and Students" (KCS), a domain notoriously difficult to teach in university settings without access to real children.

6.2. Recommendations

For Teacher Education Programs:

- **Curriculum Redesign:** "AI Literacy" should not be a standalone elective but embedded within domain-specific methods courses (e.g., "Math Methods with AI").
- **Prompt Engineering Workshops:** Preservice teachers must be trained in pedagogical prompting (e.g., Chain-of-Thought, Role-Playing) to unlock the tool's potential.
- **"Catch the Bot" Activities:** Instructors should intentionally use AI to generate erroneous math explanations and task PETs with identifying and correcting them to build critical awareness.

For Future Research:

- **Longitudinal Studies:** Follow these PETs into their first year of teaching to see if their "AI-Assisted" habits persist and if they translate to student achievement.
- **Domain Specificity:** Investigate if similar effects are observed in domains like Geometry or Probability, which rely more heavily on spatial reasoning than linguistic prompts.

REFERENCES

Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education, 59* (5), 389–407. https://doi.org/10.1177/0022487108324554

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3* (2), 77–101.

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.

Hill, H. C., Schilling, S. G., & Ball, D. L. (2004). Developing measures of teachers' mathematics knowledge for teaching. *The Elementary School Journal, 105* (1), 11–30. https://doi.org/10.1086/428763

Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record, 108* (6), 1017–1054. https://doi.org/10.1111/j.1467-9620.2006.00684.x

Mishra, P., Warr, M., & Islam, R. (2021). TPACK in the age of ChatGPT and Generative AI. *Journal of Digital Learning in Teacher Education, 39* (4), 235–251. https://doi.org/10.1080/21532974.2021.2247480

Mollick, E., & Mollick, L. (2021). Assigning AI: Seven approaches for students, with prompts. *SSRN Electronic Journal* . https://doi.org/10.2139/ssrn.4475995

Thanheiser, E., Whitacre, I., & Roy, G. J. (2014). Mathematical content knowledge for teaching elementary mathematics: A focus on whole-number concepts and operations. *Mathematics Enthusiast, 11* (2), 217–266.

UNESCO. (2021). *Guidance for generative AI in education and research* . UNESCO Publishing. https://unesdoc.unesco.org/ark:/48223/pf0000386693

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