Design and Feasibility of a Modular Prompt Framework for Polya-Based Mathematical Instructional Scaffolding

Hsuan-Yung Fang[1], Regina Juchun Chu[2*]
[1,2] Institute of Learning Sciences and Technologies, Tsing Hua University, Taiwan
[*]chu@mx.nthu.edu.tw

Abstract: *This study addresses the lack of real-time heuristic guidance in mathematics self-study by designing an AI prompt framework grounded in Polya's four-stage model. We developed a five-module structure—Persona, Rules, Instruction, Reflection, and Practice—to provide digital scaffolding simulating teacher-like guidance. Testing with Gemini 2.5 and teacher interviews demonstrates that the framework delivers coherent reasoning without hallucinations. Findings suggest the Reflection module effectively supports metacognitive awareness. While limited for students with low prior knowledge, the study confirms the feasibility of prompt engineering as instructional scaffolding in mathematics.*

Keywords: Structured Prompt Patterns, Polya's Problem Solving, AI Math Learning

1. Introduction

Generative Artificial Intelligence (GenAI) has emerged as a vital tool for fostering student engagement, self-efficacy, and accessibility (Chohan & Khan, 2024; Dmitrenko et al., 2025). Its pedagogical utility extends to automating instructional design, enriching content knowledge, and diagnosing error patterns for targeted remediation (Getenet, 2024; Oh, 2023). Nevertheless, mathematical applications remain constrained by "hallucinations," fragmented reasoning, and premature answer disclosure that hinders conceptual depth (Drushlyak et al., 2025; Schorcht et al., 2024; Wu, 2024). These limitations are frequently exacerbated by poorly structured prompting, leading to logical fallacies and diminished rigor (Drushlyak et al., 2025; White et al., 2023).

While prior inquiries emphasize terminal accuracy, few address the nuances of process-oriented interactions in autonomous learning where students struggle with strategic reasoning without heuristic scaffolding (Getenet, 2024; White et al., 2023). Addressing this gap, this study integrates Polya's four-stage model with modular prompt engineering to examine the feasibility of an interactive framework designed as a process-oriented learning partner.

2. Literature Review

2.1. The Mathematical Problem-Solving Process

Mathematical problem-solving is a central proficiency in national curriculum frameworks. Within this domain, Polya's four-stage model—*understanding* the problem, *planning* (devising a plan), *execution* (carrying out the plan), and *reflection* (looking back)—remains the dominant framework for analyzing student behaviors and designing instructional support. Students must transform verbal descriptions into mathematical representations during understanding, while planning and execution require flexible strategy selection. Finally, reflection involves metacognitive strategies for evaluating reasoning and revising solutions.

Despite this structure, students often lack real-time scaffolding for strategy formulation in out-of-class contexts. To bridge this gap, this study operationalizes these stages as explicit AI dialogue rules. Rather than providing direct answers, the framework structures prompt to guide students through meaning-making and reflective thinking, acting as a heuristic partner for self-directed study.

2.2. AI Applications in Mathematical Problem Solving

Current research indicates that while AI models often exhibit fragmented reasoning and "hallucinations" (Drushlyak et al., 2025; Oh, 2023), their performance remains highly sensitive to prompt quality. Beyond simple model upgrades, prompt engineering has emerged as a decisive factor in improving mathematical coherence and strategy use (Schorcht et al., 2024). Within this landscape, White et al. (2023) introduced a Prompt Pattern Catalog—comprising patterns such as *Persona*, *Flipped Interaction*, *Cognitive Verifier*, *Fact-Check*, and *Reflection*. These patterns enable natural-language instructions to express complex pedagogical functions, including role assignment, task decomposition, and error verification. Such patterns are inherently composable, allowing researchers to design structured prompt sequences that simulate the heuristic scaffolding required for mathematical problem-solving (Oh, 2023).

Empirical studies demonstrate that such structured frameworks significantly boost accuracy and instructional quality. For instance, incorporating roles and process guidance based on Polya's model has led to accuracy rates exceeding 90% (Oh, 2023; Chohan & Khan, 2024). Moreover, interactive AI tools have shown potential in fostering metacognitive skills, such as error detection and reflective thinking (Dmitrenko et al., 2025; Drushlyak et al., 2025). However, limitations persist in strategic diversity, as models often struggle to move beyond algebraic reasoning to graphical or heuristic approaches (Getenet, 2024). Synthesizing these findings, the present study adopts Polya's process as a structural core, integrating multiple prompt patterns to facilitate coherent, process-oriented interactions that support students' strategy refinement and self-reflection.

3. Method

Following Polya's four-stage problem-solving framework, this study designed an AI-based interactive mathematical learning tool that supports the stages of understanding, planning, execution, and reflection, enabling students to conduct personalized after-school learning and problem practice. To ensure response quality and maintain a natural interaction environment, a structured prompt architecture was implemented, consisting of five modules—Persona, Rules, Instruction, Reflection, and Practice. The dynamic interaction workflow and module coordination are visualized in Figure 1.

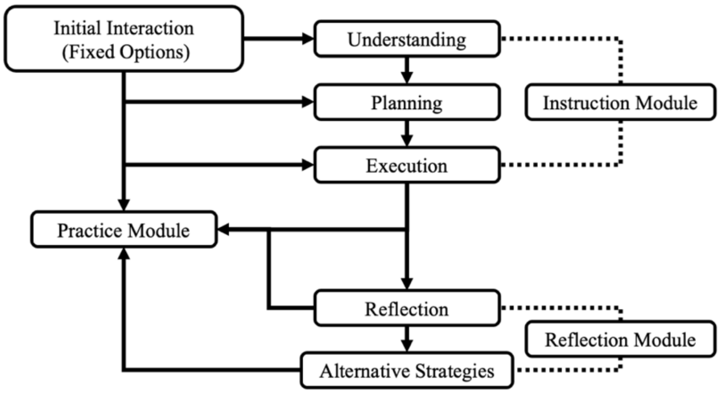

Figure 1. Dynamic interaction workflow and module transition logic

As illustrated in Figure 1, the Rules Module governs the overall interaction flow; students begin by selecting from fixed options (e.g., "stuck midway") that explicitly express their needs, which then triggers the corresponding instructional stage to clarify intent and reduce cognitive load. The Persona Module defines the AI as a facilitator responsible for eliciting student thinking, employing an "affirm-then-correct" strategy to maintain an encouraging tone. The Rules Module further implements "cognitive load management" (e.g., pausing after three consecutive equations) and ensures that all interactions strictly follow Ask-Me-Anything (AMA) and Chain-of-Thought (CoT) principles to prevent proactive disclosure of final answers. The Instruction Module operationalizes Polya's first three stages; instead of explicitly naming prerequisite concepts, it uses questioning and condition decomposition to guide students toward independently recognizing relevant methods, utilizing the Cognitive Verifier Pattern. During the planning stage, the Fact-Check Pattern helps evaluate strategy feasibility, while the execution stage provides reasoning via CoT, with scaffolding gradually reduced as student independence increases. The Reflection Module corresponds to Polya's "looking back" stage, enforcing three core questions (identifying challenges, critical steps, and logic explanation) and offering alternative strategies to foster strategic flexibility. Finally, the Practice Module facilitates learning transfer by generating tiered tasks (Basic, Advanced, and Challenge problems) based on the Gameplay and Infinite-generation patterns, concluding with an automated learning report for long-term self-regulation.

To examine its feasibility and practical effectiveness, this study implemented full interaction tests using one number theory problem and one geometry problem, followed by demonstration sessions and semi-structured interviews with three middle school mathematics teachers, anonymized as TA, TB, and TC. Interview excerpts are indicated using the prefix (I), labeled according to question numbers 1 through 5. The interview data were verified by the interviewees to ensure the accuracy of interpretation.

4. Result

To address educational accessibility and equity, this study employed Gemini 2.5 to examine whether the proposed framework could effectively reproduce Polya's four-stage process and maintain reasoning consistency without additional fine-tuning. The model's capacity for unrestricted long-form dialogue facilitates authentic autonomous learning, helping to reduce disparities among students with varying educational resources when seeking support outside the classroom.

4.1. Model Performance and Reasoning Stability

To verify the framework's stability, this study implemented interaction tests using two distinct tasks: a number theory problem (*(x+1) / (x^2-3x+3) is an integer*) and a geometry folding problem.

In the number theory test, the Rules Module correctly identified the "stuck midway" intent and triggered the Instruction Module. The AI utilized Cognitive Verifier and Fact-Check patterns to decompose conditions without proactive disclosure of final answers. Chain-of-Thought (CoT) reasoning was activated only when computations became complex, maintaining logical consistency without skipped steps. Upon requesting alternative approaches, the system transitioned smoothly to the Reflection Module to guide comparison of logical structures. During the Practice Module, when the researcher adopted an inappropriate approach, the AI redirected reasoning through counter-questioning rather than providing the solution, demonstrating scaffolded instructional behavior.

In the geometry test, the system demonstrated robust representational transformation capabilities. When the learner expressed unfamiliarity with trigonometric ratios, the AI reformulated the solution using more fundamental approaches. It subsequently activated CoT reasoning, breaking calculations into micro-steps while maintaining logical coherence. Throughout both tasks, no skipped logical steps, computational errors, or fabricated conditions (hallucinations) were observed, confirming the framework's instructional robustness and stability in complex mathematical inference.

4.2. Field Interviews with Teachers

Three teachers (TA, TB, TC) affirmed the system's instructional potential while noting discrepancies with actual classroom use. Key findings include: (1) Competence: AI responses partially reflected the researcher's *"too teacher-like"* (TC-I-1) style. The system is most effective for students with basic competence, as lower-achieving learners *"do not even know how to ask questions"* (TA-I-5). (2) Adaptation: Teachers suggested adding functions that *"recommends which unit students should review"* (TA-I-5) for students lacking fundamental concepts. (3) Over-reliance: Concerns were raised that rapid AI feedback might allow students to *"bypass genuine reasoning"* (TB-I-2). (4) Barriers: AI supplements support but cannot replace pedagogical judgment. Success remains contingent on student self-discipline and the tension between scaffolding and *"sometimes letting students struggle independently is part of their growth"* (TC-I-4).

5. Discussion and Conclusion

The findings demonstrate that the proposed five-module framework effectively reproduces Polya's four-stage process, functioning as a tool for metacognitive support through seamless transitions and question-driven scaffolding (Oh, 2023; Schorcht et al., 2024; White et al., 2023). However, consistent with pedagogical concerns identified during testing, challenges remain regarding instructional depth and learner over-reliance, as AI currently struggles to adapt its pathway for students lacking prior knowledge (Wu, 2024). Despite these limitations, the framework addresses critical gaps in after-school learning by providing heuristic guidance and promoting educational equity through Gemini 2.5.

Regarding limitations, since interaction tests relied on researcher simulation within the specific domains of number theory and geometry, future work should involve direct student participation to enhance real-world applicability. Furthermore, as the system relies exclusively on prompt engineering, its performance reflects the base model's reasoning rather than domain-specific instructional expertise. Future iterations should incorporate student-level diagnostic mechanisms and broader curriculum alignment to improve precision. Overall, this framework offers a transferable model for integrating AI into mathematics education, fostering a more personalized and accessible learning environment.

References

Chohan, I., & Khan, I. (2024). Enhancing mathematics education with ChatGPT-4 personalized problem-solving and consistent learning. 2024 2nd International Conference on Computing and Data Analytics (ICCDA). https://doi.org/10.1109/ICCDA64887.2024.10867401

Dmitrenko, N., Shakhina, I., Fedyk, T., Podzygun, O., & Franchuk, N. (2025). Leveraging ChatGPT to develop critical thinking through mathematical problem-solving in ESP classes. Environment. Technology. Resources. Proceedings of the 16th International Scientific and Practical Conference, 3, 91–98. https://doi.org/10.17770/etr2025vol3.8540

Drushlyak, M., Lukashova, T., Sabadosh, Y., Melnikov, I., & Semenikhina, O. (2024). Using ChatGPT for the development of critical thinking in youth: Example of inequality proof. 2024 47th MIPRO ICT and Electronics Convention (MIPRO). https://doi.org/10.1109/MIPRO60963.2024.10569759

Drushlyak, M., Lukashova, T., Shamonia, V., & Semenikhina, O. (2025). ChatGPT-based simulation helps to develop the pre-service mathematics teachers' critical thinking. International Journal of Instruction, 18(1), 153–172. https://doi.org/10.29333/iji.2025.1819a

Getenet, S. (2024). Pre-service teachers and ChatGPT in multistrategy problem-solving: Implications for mathematics teaching in primary schools. International Electronic Journal of Mathematics Education, 19(1), em0766. https://doi.org/10.29333/iejme/14141

Oh, S. J. (2023). Effective ChatGPT prompts in mathematical problem solving: Focusing on quadratic equations and quadratic functions. Communications of Mathematical Education, 37(3), 545–567. https://doi.org/10.7468/jksmee.2023.37.3.545

Schorcht, S., Buchholtz, N., & Baumanns, L. (2024). Prompt the problem – investigating the mathematics educational quality of AI-supported problem solving by comparing prompt techniques. Frontiers in Education, 9, 1386075. https://doi.org/10.3389/feduc.2024.1386075

White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., Elnashar, A., Spencer-Smith, J., & Schmidt, D. C. (2023). A prompt pattern catalog to enhance prompt engineering with ChatGPT. arXiv. https://doi.org/10.48550/arXiv.2302.11382

Wu, Y. (2024). Critical thinking pedagogics design in an era of ChatGPT and other AI tools—Shifting from teaching "what" to teaching "why" and "how". Journal of Education and Development, 8(1). https://doi.org/10.20849/jed.v8i1.1404