

From Prompting to Pedagogical Design: Secondary Mathematics Teachers' Experiences with a Mixed-Model Prompt Engineering Framework for AI-Assisted Exemplar Development

Adriano V. Patac, Jr.¹, Louida P. Patac², Fernando T. Herrera³

¹Program Chair, Master of Arts in Mathematics Education, Surigao Del Norte State University ,
apatac@ssct.edu.ph

²Dean, College of Arts and Sciences, Surigao Del Norte State University , lpatac@ssct.edu.ph

³Professor, Caraga State University, fherrera@carsu.edu.ph

HOW TO CITE:

Adriano V. Patac, Jr., Louida P. Patac, Fernando T. Herrera (2026). From Prompting to Pedagogical Design: Secondary Mathematics Teachers' Experiences with a Mixed-Model Prompt Engineering Framework for AI-Assisted Exemplar Development. International Journal of Special Education, 41(18s), 880-894

ABSTRACT:

Generative artificial intelligence is increasingly used in mathematics lesson planning, yet teachers often lack a structured process for designing, refining, and validating prompts. This study examined the experiences of 39 secondary mathematics teachers who participated in an AI-assisted exemplar-development training using the Mixed-Model Prompt Engineering framework. The model guided teachers through persona, task, and context definition; output structuring; validation and iterative refinement; curriculum alignment; and final human checking. A convergent mixed-methods program-evaluation design with qualitative priority was employed. Quantitative data were obtained from an 11-item post-training evaluation, while written responses from 36 participants were analyzed thematically. The program received a high overall rating ($M = 4.85$, $SD = 0.38$), with speaker effectiveness and content relevance obtaining the highest ratings ($M = 4.95$). Five themes emerged: prompt engineering as a professional practice, AI-assisted instructional design, learner-responsive mathematics teaching, teacher empowerment, and conditions for sustainable implementation. Participants valued structured prompting, lesson-exemplar development, and reduced preparation demands, but identified the need for additional practice, reliable connectivity, continued training, and human verification. The study contributes to mathematics education by proposing pedagogical prompt literacy as an emerging teacher competence that integrates mathematical knowledge, instructional judgment, contextual awareness, and critical evaluation of AI-generated materials.

COPYRIGHT STATEMENT:

Copyright: © 2026 Authors.
Open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Keywords: Generative Artificial Intelligence, Mathematics Education, Pedagogical Prompt Literacy, Prompt Engineering, Teacher Professional Development

1. Introduction

Lesson planning in mathematics is a form of pedagogical reasoning through which teachers sequence concepts, select tasks and representations, anticipate misconceptions, and determine evidence of learning. Effective planning extends beyond completing a prescribed format; it requires attention to curriculum expectations, learner diversity, equity, scaffolding, available resources, and the purposeful use of technology (Ayalon et al., 2021; Calleja et al., 2024; Dingman et al., 2021; Jackson et al., 2023; Yang et al., 2023; Zaragoza et al., 2023).

Generative AI expands this process by producing lesson outlines, examples, assessments, explanations, and differentiated materials. However, fluent responses may still be generic, mathematically inaccurate, or poorly aligned with curriculum and classroom conditions (Kasneci et al., 2023; Lo, 2023; Ogunleye et al., 2024). This risk is especially important in mathematics, where subtle errors may appear plausible; some preservice teachers have even interpreted incorrect AI-generated solutions as alternative methods (Gurl et al., 2025). Evidence nevertheless suggests that AI can support richer examples, feedback, personalization, and curriculum-aligned materials when teachers actively review and adapt its outputs (Rizos et al., 2024; Rutherford et al., 2025). Effective AI-assisted planning is therefore iterative: teachers generate, evaluate, revise, contextualize, and validate outputs rather than accept them as finished lessons (ElSayary et al., 2025).

1.1 Prompt engineering as an emerging teacher competency

Prompt quality depends on how clearly users communicate the task, context, constraints, and expected output. For teachers, prompt engineering is therefore more than technical interaction with an AI system; it is an emerging professional competence that combines prompt structure, method selection, critical evaluation, and contextual judgment (Federiakin et al., 2024). A vague request may yield a polished but unusable lesson, whereas a context-rich prompt

can specify learners, curriculum goals, misconceptions, resources, differentiation, assessment, and format. In this sense, the prompt functions as a pedagogical specification. Robertson et al. (2024) frame this process as human–AI co-construction through context setting, prompt structuring, validation, and refinement, with final responsibility remaining with the teacher. Similarly, ElSayary et al. (2025) found that useful lesson materials emerged through repeated localization and revision. These findings suggest that effective prompting is an expression of pedagogical content knowledge: teachers must understand what a sound lesson requires before they can direct AI to support its design.

1.2 The Mixed-Model Prompt Engineering framework

The professional-development intervention examined in the present study used the *Mixed-Model Prompt Engineering for Mathematics Education Teaching and Learning Activities* developed for classroom-oriented mathematics planning. The module was designed from a mathematics education rather than a software-development perspective. It focuses on classroom applications of few-shot prompting, reasoning-oriented prompting, alternative-response generation, self-fact-checking, persona prompting, reflection, revision, localization, and ethical AI use. It explicitly limits its scope to practical and predominantly low-technology mathematics settings and recognizes that generated outputs remain subject to teacher verification.

In this study, the term mixed-model does not refer to a mixed-methods research design. Instead, it describes the integration of several complementary prompt-engineering frameworks and established lesson-planning practices into a single, classroom-oriented process. The model combines the prompt-engineering framework proposed by Federiakin et al. (2024), the Context–Structure–Evaluate protocol developed by Robertson et al. (2024), and the iterative lesson-planning practices identified by ElSayary et al. (2025).

Figure 1 presents the Context–Structure–Evaluate protocol, which provides the iterative co-construction component incorporated into

the Mixed-Model Prompt Engineering framework.

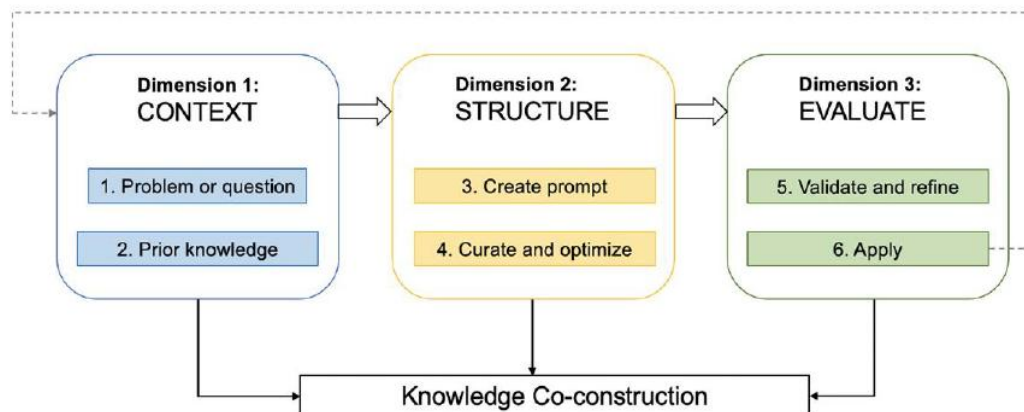


Figure 1. Context–Structure–Evaluate Cycle for Human–AI Knowledge Co-construction

These components were reorganized into a flexible sequence that can support mathematics teachers across different educational systems. Rather than being tied to one national curriculum, the model can be adapted to various grade levels, curriculum standards, classroom contexts, learner profiles, and levels of technological access. Its central purpose is to help teachers move from general or one-step prompting toward a more deliberate process of defining the instructional task, providing relevant classroom context, specifying the expected lesson structure, refining the generated response, and validating the final output for mathematical accuracy, pedagogical suitability, inclusiveness, and curriculum alignment.

The model contains four interrelated layers. The first layer establishes the persona, task, and context. It asks the teacher to identify the professional role the AI should assume, define the instructional product to be created, and provide information the system cannot reasonably infer, such as class size, learner readiness, common misconceptions, available materials, language conditions, and local examples. The second layer specifies the structure of the intended output, including objectives or outcomes, content, resources, teaching procedures, assessment, remarks, and reflection. The third layer requires validation, refinement,

and iteration, through which the initial output is contextualized, aligned, expanded, and corrected. The fourth layer checks curriculum alignment, instructional feasibility, student-centeredness, and consistency with the expected competencies.

The model's master-prompt sequence guides teachers through a deliberate process of AI-supported lesson design. Teachers begin by assigning the AI the role of an experienced mathematics educator and identifying a clearly defined mathematical topic. They then describe the learners, available resources, instructional constraints, curriculum expectations, and preferred lesson structure. The prompt is progressively refined to include collaborative learning, differentiation, common misconceptions, culturally and locally meaningful examples, assessment strategies, and inclusive practices. Before the lesson is used, the generated plan is reviewed for curricular alignment, classroom feasibility, mathematical accuracy, and the realistic allocation of instructional time.

Importantly, the process does not end when the AI produces a lesson plan. The generated material is treated as an initial draft that remains subject to professional review. The teacher decides whether the mathematical explanations are accurate, whether the activities are appropriate for the learners, whether the

assessment tasks reflect the intended outcomes, and whether the material complies with ethical and institutional expectations. This final stage preserves the teacher's role as the principal instructional decision-maker. Within the framework, generative AI supports planning, but it does not replace pedagogical judgment, subject-matter expertise, or responsibility for what ultimately takes place in the classroom.

The model also connects prompt structure with recognizable mathematics pedagogies. It includes the Concrete–Representational–Abstract sequence, the Enactive–Iconic–Symbolic progression, Singapore bar modelling, multiple-representation instruction, and the Concrete–Virtual Manipulative–Abstract sequence. These approaches provide teachers with language for asking AI systems to organize lessons around established instructional progressions rather than produce collections of unrelated activities. The model therefore attempts to connect three domains that are often treated separately: prompt construction, mathematics pedagogy, and curriculum alignment.

Ethical and professional safeguards are likewise incorporated. The module requires disclosure of AI assistance, verification of learning objectives and mathematical content, alignment of activities and assessments, removal of personally identifiable information, and substantial teacher revision of generated outputs. It describes AI responses as drafts and assigns final responsibility for errors and instructional decisions to the human author. These provisions are consistent with a human-centered approach, which emphasizes privacy, human agency, inclusion, ethical validation, and pedagogical appropriateness in the educational use of generative AI (Miao & Holmes, 2023).

1.3 Global Curriculum and Professional-Development Context

Generative AI is reshaping lesson planning, assessment, differentiation, and instructional support across education systems. Its value, however, depends less on access than on teachers' ability to use it critically, ethically, and in alignment with curriculum goals and learner

needs. A human-centered approach therefore remains essential to protect privacy, ensure equitable access, and preserve professional judgment (Miao & Holmes, 2023). Because effective use requires more than basic familiarity with ChatGPT, teachers need sustained opportunities to refine prompts, evaluate inaccurate outputs, and adapt generated materials to authentic classroom contexts. This need is reinforced by evidence that demand for AI integration is outpacing relevant professional development and that prompt engineering is emerging as a distinct professional competence (Federiakin et al., 2024; Tan et al., 2025). In mathematics education, AI may support explanation, planning, assessment, feedback, and personalization, but its benefits depend on purposeful use, careful validation, and continued teacher responsibility for final instructional decisions (Pepin et al., 2025).

1.4 Research Problem and Gap

Research on generative AI in education has expanded rapidly, yet limited attention has been given to practicing secondary mathematics teachers who receive structured training in prompt engineering. Existing studies commonly examine broad perceptions of ChatGPT, preservice teachers, or individual applications such as lesson planning and assessment. Moreover, studies of teachers' AI use do not always explain how participants were taught to formulate, refine, and validate prompts, making it difficult to identify a transferable process for AI-assisted instructional design (ElSayary et al., 2025; Pepin et al., 2025).

The relevance of AI-supported lessons also depends on curriculum expectations, learner diversity, language, class size, connectivity, and available resources. Consequently, evidence from one educational setting may not apply directly to another. Teachers need professional development that connects prompt construction with actual pedagogical decisions rather than merely demonstrating an AI tool. Recent reviews show that demand for AI integration has grown faster than the availability of sustained, practice-

based teacher development (Federiakin et al., 2024; Tan et al., 2025).

The Mixed-Model Prompt Engineering framework was developed as a practical training model and had not previously undergone controlled empirical testing. The present study therefore examines how 39 secondary mathematics teachers experienced the framework, what they learned from structured prompting, how they perceived its usefulness for lesson-exemplar development, and what conditions may support or hinder its continued use. The study evaluates teachers' experiences and professional learning rather than claiming direct effects on student achievement.

1.5 Research objectives

The study examined how secondary mathematics teachers experienced the Mixed-Model Prompt Engineering framework during an AI-assisted lesson-exemplar development program. It assessed their evaluation of the training, identified their significant learning, and explored how structured prompting supported lesson planning, contextualization, differentiation, instructional efficiency, and technology integration. It also documented teachers' experiences in constructing, refining, and validating prompts, together with the technological, logistical, pedagogical, and ethical concerns that may influence the sustained implementation of AI-focused professional development.

2. Methodology

2.1 Research Design

The study adopted a convergent mixed-methods program-evaluation design with qualitative priority. Quantitative ratings and written responses were collected through the same post-training questionnaire, analyzed separately, and brought together during the interpretation of the findings. This design was appropriate because numerical ratings provided an overview of the participants' assessment of the training, while their written accounts explained what they learned, how they experienced the prompting process, and what difficulties they encountered.

Mixed-methods integration was therefore used to develop a more complete account than either data form could provide independently (Creswell & Plano Clark, 2018; Fetters et al., 2013).

The study was evaluative and descriptive rather than experimental. It did not include a pretest, control group, classroom observation, or direct measure of student achievement. Consequently, the findings were used to describe participants' experiences and perceptions, not to establish that the training caused improvements in teaching performance or mathematics learning.

The term *mixed-methods* refers to the research design of the present study. It should be distinguished from the *Mixed-Model Prompt Engineering framework*, which was the instructional model introduced during the training.

2.2 Research Setting and Participants

The study was conducted following the *Enhancing Mathematics Education Through AI-Assisted Exemplar Development: MathGPT Extension Project*. The program involved secondary mathematics teachers from a public-school division and focused on the use of generative AI in lesson planning, exemplar development, assessment preparation, localization, and instructional presentation.

The study involved 39 secondary mathematics teachers who attended the MathGPT training and completed the post-activity evaluation. Criterion-based purposive sampling was used because only teachers with direct experience of the program could provide relevant feedback, while total enumeration allowed all eligible responses to be included. Participants were required to be actively engaged in secondary mathematics teaching, complete the post-training questionnaire, and provide voluntary consent under the data-privacy notice. All 39 responses were included in the quantitative analysis, while 36 participants contributed substantive written responses for the qualitative analysis. Entries such as "NA," "N/A," or a single punctuation mark were excluded as non-substantive, whereas "none" was retained when it clearly indicated that no improvement was suggested.

2.3 The Training Intervention

The training was organized around the Mixed-Model Prompt Engineering for Mathematics Education Teaching and Learning Activities. The framework was developed for classroom use and integrates prompt structure, prompting methods, curriculum alignment, mathematics pedagogy, and human verification. Its objectives include introducing prompt engineering from a mathematics-education perspective, developing lesson plans with generative AI, and applying ethical practices in educational AI use.

The framework contains four connected layers:

1. **Persona, task, and context** – defining the role assigned to the AI, the instructional product to be created, and the relevant classroom conditions;
2. **Structure** – specifying the required lesson-plan format and instructional components;
3. **Validation, refinement, and iteration** – reviewing, localizing, correcting, and extending the generated response; and
4. **Curriculum alignment** – checking whether the activities, assessments, instructional resources, and intended outcomes remain consistent with curriculum requirements.

During the training, teachers were introduced to prompt strategies such as persona prompting, few-shot prompting, alternative-response generation, self-fact-checking, and iterative revision. They were guided to provide information that an AI system could not reasonably infer, including learner readiness, class size, available resources, anticipated misconceptions, lesson duration, and expected output format.

The participants then applied the framework to mathematics lesson-exemplar development. Initial AI-generated responses were treated as drafts. Teachers were encouraged to review what had been generated, identify missing or inappropriate content, revise their prompts, contextualize examples, add differentiation and

assessment tasks, and verify the final output for mathematical accuracy and classroom feasibility. This iterative structure reflects current research showing that context-rich and revisable prompts produce more relevant instructional materials than one-time, general requests (ElSayary et al., 2025; Federiakin et al., 2024).

The final stage required human validation. The model explicitly directs teachers to check curriculum alignment, realism, instructional time, and the mathematical and pedagogical quality of the completed lesson before use.

2.4 Research Instrument

Data were collected through a researcher-developed post-activity questionnaire administered via Google Forms. The instrument included a data-privacy notice, participant information, 11 activity-level items rated on a five-point scale, presenter-specific items, and four open-ended questions. The quantitative items assessed the clarity and attainment of the training objectives, applicability and relevance of the content, ease of technology use, speaker effectiveness, adequacy of time and duration, and suitability of the venue. Only the 11 activity-level items were included in the main analysis because the study focused on the overall training experience rather than comparisons among presenters. The qualitative questions elicited participants' views on the teaching of the technology, their most significant learning, areas for improvement, and their experiences in prompting ChatGPT during the training.

The fourth question used the expression *mixed-methods design* in the original form. For the analysis, this wording was understood as referring to the *Mixed-Model Prompt Engineering framework* demonstrated during the training, not to the mixed-methods research design used in the study. The questionnaire was intended as a program-evaluation instrument rather than a standardized psychological or educational scale. Accordingly, the numerical ratings were treated as descriptive indicators of participant evaluation and were not interpreted as measurements of a validated latent construct.

2.5 Data-Collection Procedure

The study followed five stages. First, the participants were introduced to generative AI and the role of prompt engineering in mathematics instructional planning. The facilitators explained how vague instructions commonly produce generalized responses and how additional context and constraints may improve the relevance of AI-generated materials. Second, the four layers of the Mixed-Model Prompt Engineering framework were presented. The teachers examined how persona, task, context, format, localization, curriculum alignment, iterative revision, and human validation could be incorporated into a sequence of prompts.

Third, the facilitators demonstrated the use of ChatGPT in developing mathematics lesson exemplars. The participants observed how an initial output could be revised to address misconceptions, learner diversity, classroom resources, instructional activities, formative assessment, and curriculum requirements. Fourth, the teachers engaged in guided application. They constructed prompts for selected mathematics topics, reviewed the resulting responses, and provided follow-up instructions when the generated material was incomplete, overly general, or unsuitable for the intended lesson.

Finally, the participants completed the electronic evaluation form. The responses were exported to a spreadsheet and prepared for analysis. Names, email addresses, residential information, and other direct identifiers were separated from the analytical dataset. Each participant was assigned a code from P1 to P39.

2.6 Data Analysis

Quantitative analysis

The 11 activity-level evaluation items were analyzed using descriptive statistics. Frequencies and percentages were calculated for each response category. The mean and standard deviation were also computed for each item and for the overall evaluation. Inferential statistical tests were not performed because the study did not involve comparison groups, random

assignment, or pretest–posttest measurement. The quantitative analysis was intended to describe the evaluations of the participating teachers rather than generalize statistically to all secondary mathematics teachers.

Qualitative analysis

The written responses were analyzed through inductive, semantic reflexive thematic analysis following Braun and Clarke's (2021) six-phase process. The researchers first read the responses repeatedly to become familiar with the data, then assigned descriptive codes to meaningful statements. Related codes were grouped into preliminary categories and developed into candidate themes and subthemes across the four open-ended questions. These themes were reviewed against the original responses for coherence, distinctiveness, and relevance to the research questions, after which representative statements were selected to illustrate each theme.

The coding was primarily semantic because it focused on what participants explicitly reported. Codes included structured prompting, persona construction, prompt refinement, lesson-exemplar development, workload reduction, technology integration, learner engagement, teacher confidence, limited training time, connectivity problems, access to premium tools, and responsible AI use. A participant could be assigned to more than one code or theme. However, repeated references to the same idea by one participant were counted only once within a subtheme. The reported frequencies therefore represented the number of participants mentioning an idea rather than the total number of times particular words appeared. These counts were used to indicate recurrence, not to determine the importance of a theme. Quotations were identified through participant codes. Minor corrections to spelling, capitalization, and punctuation could be made for readability, but the meaning of each statement was preserved.

Mixed-methods integration

The two strands were integrated during interpretation through side-by-side comparison. Quantitative ratings were compared with the qualitative themes to determine whether the

written responses clarified, supported, or qualified the numerical patterns. For example, ratings concerning training time and ease of technology use were examined alongside comments about additional hands-on practice, internet connectivity, and access to AI tools.

This form of integration allowed the numerical results to show the general pattern of evaluation while the written responses explained why particular aspects of the program were viewed favorably or identified for improvement. Such merging at the interpretation stage is a recognized strategy for achieving integration in convergent mixed-methods research (Fetters et al., 2013).

3. Results

This section presents the quantitative evaluation of the training and the themes generated from the

participants' written responses. The training introduced the four-layer Mixed-Model Prompt Engineering framework: persona, task, and context; output structure; validation, refinement, and iteration; and curriculum alignment.

3.1 Quantitative Evaluation of the Training

The participants evaluated 11 aspects of the program using a five-point scale. Across the 429 individual ratings, 368 responses, or 85.8%, received the highest rating of 5. Fifty-eight responses, or 13.5%, received a rating of 4, while three responses, or 0.7%, received a rating of 3. No rating below 3 was recorded. The overall mean evaluation was **4.85** ($SD = 0.38$). Of the 429 ratings, 426, or 99.3%, were at scale points 4 or 5.

Table 1. Descriptive Evaluation of the MathGPT Training

Evaluation statement	Rating 5, n	Rating 4, n	Rating 3, n	M	SD
Training goals and objectives were clearly defined.	33	6	0	4.85	0.37
Training objectives were attained.	31	8	0	4.79	0.41
Information was helpful and applicable to participants' work.	36	3	0	4.92	0.27
The technology was easy to study and apply.	29	10	0	4.74	0.44
The speakers explained the technology effectively.	37	2	0	4.95	0.22
The discussion content was useful and relevant.	37	2	0	4.95	0.22
Sufficient time was allocated for learning the technology.	29	9	1	4.72	0.51
The duration of the activity was appropriate.	30	8	1	4.74	0.5
The venue was well ventilated.	35	4	0	4.9	0.31
The venue was comfortable.	35	3	1	4.87	0.41
The venue was accessible.	36	3	0	4.92	0.27
Overall	368	58	3	4.85	0.38

Note. $N = 39$. M = mean; SD = standard deviation.

The highest mean ratings were recorded for speaker effectiveness and content relevance, both with a mean of 4.95. These were followed by the applicability of the training information and venue accessibility, both with a mean of 4.92. The lowest mean rating concerned the sufficiency of the training time, with a mean of 4.72. The ease of learning and applying the technology and the appropriateness of the activity duration both obtained a mean of 4.74.

3.2 Qualitative Findings

Analysis of the 36 substantive responses yielded five major themes: prompt engineering as a professional practice, AI-assisted mathematics instructional design, engaging and learner-responsive instruction, professional learning and teacher agency, and conditions for continued implementation. Frequencies indicate the number of participants who mentioned each subtheme at least once; because individual responses could reflect several ideas, the categories were not mutually exclusive.

Table 2. Themes and Subthemes Identified in the Teachers' Responses

Major theme	Subtheme	n	Percentage
Prompt engineering as a professional practice	Structured and purposeful prompting	19	52.8
	Persona, localization, and classroom context	4	11.1
	Prompt refinement and output specificity	6	16.7
AI-assisted mathematics instructional design	Development of lesson plans, exemplars, and DLLs	18	50
	Efficiency and reduced instructional workload	15	41.7
	PowerPoint, software, and digital-tool integration	15	41.7
Engaging and learner-responsive instruction	Interactive and engaging learning activities	10	27.8
	Clearer, adaptive, and learner-centered materials	4	11.1
	Twenty-first-century teaching competence	4	11.1
Professional learning and teacher agency	Clear, practical, and relevant facilitation	14	38.9
	Confidence, empowerment, and further exploration	15	41.7
	Requests for additional training or detailed guidance	7	19.4

Major theme	Subtheme	n	Percentage
Conditions for continued implementation	More time and hands-on practice	9	25
	Internet connectivity and access to premium tools	6	16.7
	Venue and logistical concerns	2	5.6
	Human oversight, originality, and plagiarism	2	5.6

Note. $N = 36$ participants with substantive written responses. Participants could be represented in more than one subtheme.

3.3 Prompt Engineering as a Professional Practice

Structured prompting was the most frequently reported learning, appearing in the responses of 19 participants. Teachers referred to “proper prompting,” “prompt engineering,” “prompting structure,” and a “systematic way of prompting.”

One teacher stated:

“The most significant learning is how to prompt an AI in such a way that I could get the best possible output.”—P3

Another participant described the process as follows:

“A systematic way of prompting is crucial in creating the desired lesson plan.”—P32

Four participants specifically referred to persona, localization, and contextual information. These teachers reported that defining the role of the AI and providing local classroom conditions resulted in responses that were more closely connected to the intended lesson.

“I had no idea how to make prompts so that the output from AI would be localized and related to the persona.”—P4

Six participants described the need to refine prompts or provide additional instructions when the first response lacked sufficient detail. Participant 8 explained:

“AI gave me a brief description only, so I needed to prompt more so that it would suit the information I needed about my topic.”—P8

Participant 24 similarly reported that the prompting sequence was “organized and in order” and enabled the AI to produce more exact details.

3.4 AI-Assisted Mathematics Instructional Design

Eighteen participants referred to the preparation of lesson plans, lesson exemplars, daily lesson logs, or instructional activities. The use of AI in lesson preparation was the most frequently reported instructional application.

“My most significant learning from this training session was understanding how AI can efficiently support and enhance the creation of mathematical exemplars.”—P13

“I learned a lot about the proper use of AI in crafting a lesson exemplar.”—P38

Fifteen participants reported that AI reduced the time or effort required for instructional preparation. Their responses included references to easier lesson planning, reduced workload, faster access to teaching resources, and more efficient exemplar development.

“The application of AI in making DLLs makes my task easy.”—P36

“It lessens the work of teachers, such as grading, communication, and access to teaching-material resources.”—P34

Fifteen participants also mentioned the use of PowerPoint, online platforms, digital applications, or other technology-based resources. Participant 29 reported learning about AI assistance in preparing “lesson plans,

exemplars, and PowerPoint presentations.” Other participants requested further instruction in developing interactive slides, using software applications, and integrating answer or score buttons into presentations.

3.5 Engaging and Learner-Responsive Instruction

Ten participants associated technology integration with more interactive or engaging mathematics lessons. The responses referred to learner participation, classroom interaction, motivation, and technology-supported presentations.

“Teaching with technology makes learning interactive.”—P5

“Technology integration in lesson delivery creates an engaging and interactive teaching-learning process.”—P31

Four participants described AI-assisted instructional materials as clearer, more adaptive, or more responsive to learners.

“AI can meaningfully transform lesson design by helping teachers create clearer, more engaging, and more adaptive mathematics exemplars.”—P14

Participant 35 similarly referred to AI-supported materials as “clearer, more engaging, and more learner-centered mathematics lessons.” Four participants connected the training with twenty-first-century teaching competence. Their responses referred to digital readiness, technology integration, and the need for teachers to develop skills appropriate to changing educational environments.

3.6 Professional Learning and Teacher Agency

Fourteen participants described the delivery of the training as clear, useful, practical, relevant, comprehensive, or well explained.

“The teaching of the technology was clear, practical, and showed how AI can genuinely enhance teachers’ efficiency in developing high-quality mathematics exemplars.”—P35

Participant 38 reported that the prompting methods were easy to apply, while Participant 30 described the discussions as informative and a

source of new learning. Fifteen participants used terms indicating increased confidence, empowerment, curiosity, or willingness to explore AI further. Their descriptions included “insightful,” “empowering,” “amazing,” “enlightened,” and “mind-blowing.”

“I was amazed, enlightened, and ignited by the knowledge on prompting ChatGPT.”—P31

“I was able to explore a lot of things using AI.”—P34

Seven participants requested additional training, more detailed demonstrations, individualized guidance, or further discussion of AI applications in mathematics teaching.

“I hope this is not the last, but the beginning of more trainings to be conducted.”—P11

3.7 Conditions for Continued Implementation

Nine participants requested more time for hands-on application. These comments referred to PowerPoint development, software exploration, guided practice, and the application of AI to specific mathematics topics.

“Actual hands-on experience on how to apply it to a specific topic in classroom instruction.”—P39

Six participants reported concerns involving internet connectivity or limited access to premium AI tools.

“We cannot access the premium version of ChatGPT and only have free accounts, so our prompting is limited.”—P26

Two participants identified concerns about venue size, comfort, and access to facilities. Two participants raised issues related to responsible AI use. Participant 17 emphasized that AI still required human intervention, while Participant 33 recommended including guidance on plagiarism in future training.

“AI is helpful and powerful; however, it still requires human intervention.”—P17

“Include ways on how to avoid plagiarism since the outputs are AI-assisted.”—P33

4. Discussion

This study examined 39 secondary mathematics teachers' experiences with the Mixed-Model Prompt Engineering framework. Participants reported strong acceptance of the training, growing competence in structured prompting, and recognition of AI as a support for instructional design rather than a source of ready-made lessons. Sustained use, however, depended on adequate practice time, reliable internet access, hands-on support, and continued human verification.

4.1 Overall assessment of the training

The program received a strong overall evaluation ($M = 4.85$, $SD = 0.38$), with 99.3% of ratings at levels 4 or 5. Speaker effectiveness and content relevance obtained the highest means ($M = 4.95$), indicating that participants viewed the training as clear, relevant, and closely aligned with their instructional work. Its practical focus on lesson planning, exemplar development, assessment, contextualization, differentiation, and presentation design likely contributed to this positive response, consistent with Tan et al. (2025), who emphasized the value of practice-based, subject-specific AI professional development. Although ratings for training time ($M = 4.72$) and ease of technology use ($M = 4.74$) remained high, participants requested more guided practice and application time, suggesting that the main concern was not relevance but insufficient opportunity to develop practical fluency.

4.2 Prompt Engineering as an Emerging Professional Practice

Structured and purposeful prompting was the most frequently reported learning, mentioned by 19 of the 36 participants who submitted substantive written responses. Teachers referred to defining a persona, specifying a task, describing classroom conditions, identifying a desired format, and revising prompts when the initial response was incomplete. These accounts show a shift from treating ChatGPT as a question-and-answer tool toward using it through a more deliberate instructional-design process. This finding supports Federiakin et al.'s (2024)

conceptualization of prompt engineering as an emerging twenty-first-century competence. Their framework defines prompt engineering as the ability to communicate a problem, its context, and the constraints of the desired solution, together with the capacity to examine the resulting output critically. In the present study, teachers' descriptions of "organized," "proper," and "systematic" prompting illustrate how this competence can take a subject-specific form. The teachers were not simply learning to obtain longer AI responses; they were learning to communicate pedagogical requirements.

The Mixed-Model framework provided a visible structure for this process. Its first layer establishes the persona, task, and classroom context; the second defines the required output structure; the third involves validation, refinement, and iteration; and the fourth checks curriculum alignment and classroom feasibility. These layers appear to have helped participants recognize that the usefulness of an AI-generated lesson depends on the quality of the instructional specifications provided by the teacher. The teachers' accounts of revising prompts also align with Robertson et al.'s (2024) Context–Structure–Evaluate protocol, which treats human–AI interaction as an iterative process of framing, constructing, validating, and refining outputs. Similar patterns were reported by ElSayary et al. (2025), whose participating teachers found that context-rich and repeatedly refined prompts produced more relevant lesson-planning materials than general, one-time requests.

An important theoretical implication is that prompt engineering for teachers may be understood as a form of pedagogical prompt literacy. Effective prompts require teachers to identify intended learning outcomes, common misconceptions, learner differences, representations, assessment evidence, and classroom constraints. Prompt quality therefore depends partly on the teacher's pedagogical and mathematical knowledge. A teacher cannot clearly instruct an AI to produce a suitable mathematics lesson without first understanding what makes the lesson mathematically accurate and pedagogically appropriate.

4.3 AI as a co-designer of mathematics instruction

Eighteen participants identified lesson plans, exemplars, and daily lesson logs as key AI applications, while 15 reported greater efficiency. Teachers therefore viewed generative AI mainly as a planning assistant that could produce initial structures, activities, and assessments for later revision. This aligns with Pepin et al. (2025) and ElSayary et al. (2025), who found that AI can support lesson design when prompts are context-rich and teachers remain involved in refining outputs. However, perceived workload reduction should be treated cautiously because the study did not measure actual time savings, and verification may introduce additional work.

This caution is particularly important in mathematics, where plausible outputs may still contain conceptual or procedural errors. Gurl et al. (2025) showed that some preservice teachers failed to recognize incorrect AI-generated solutions, reinforcing the Mixed-Model framework's requirement for final checks of accuracy, alignment, feasibility, and instructional time. The findings therefore position AI as a **co-designer**: it generates drafts and alternatives, while teachers retain authority over mathematical correctness, pedagogical sequencing, learner suitability, and classroom use.

4.4 Learner-responsive instruction and teacher agency

Participants viewed AI-assisted planning as a means of producing more interactive, differentiated, and learner-responsive mathematics lessons. Ten teachers referred to engagement, while others noted clearer materials and digital presentations, echoing findings on varied examples, motivational support, feedback, and personalization in AI-supported mathematics instruction (Rutherford et al., 2025; Pepin et al., 2025). These claims remain perceptual, however, because no classroom or student-level outcomes were measured. Fifteen participants also reported greater confidence and willingness to explore AI, suggesting that the framework strengthened teacher agency by positioning educators as critical reviewers and

final decision-makers. Its contribution lies in extending instructional choice without transferring pedagogical responsibility to the AI system.

4.5 Conditions for Sustained Implementation

Positive attitudes toward AI did not remove practical barriers. Participants requested more time for guided practice, while others cited unstable connectivity and limitations of free accounts. Sustained competence therefore requires repeated application across mathematical topics, opportunities to compare and revise prompts, and practice in identifying inaccurate or unsuitable outputs. This supports Tan et al.'s (2025) view that AI professional development should be continuous, situated, and connected to teachers' actual work. Unequal access to reliable internet, devices, and paid tools also raises equity concerns. Future training should explicitly address plagiarism, privacy, disclosure, and human verification. These safeguards are consistent with international guidance promoting human-centered, equitable, and ethically validated educational AI use (Miao & Holmes, 2023).

4.6 Implications for Theory

The findings position prompt engineering as a link between technological competence and pedagogical reasoning. Within the Mixed-Model framework, teachers translate professional knowledge into explicit prompt elements: persona and task define purpose, context identifies learners and constraints, structure specifies the intended output, and iteration supports evaluation and revision. Prompt competence should therefore be judged not only by prompt wording or output quality, but also by the teacher's ability to identify deficiencies, justify revisions, and verify mathematical and pedagogical alignment. The results further support a human–AI co-design perspective in which AI provides provisional content, while teachers retain contextual judgment, mathematical verification, and professional responsibility.

4.7 Implications for Practice

AI-focused professional development should center on authentic instructional products, including lessons, exemplars, assessments, and presentations. Training should provide sufficient time for prompt construction, revision, validation, peer review, and follow-up coaching. Participants should examine weak, improved, and misleading AI responses, use mathematical verification checklists, document revisions, and work with low-connectivity or free-access alternatives. Applying the framework across algebra, geometry, statistics, modelling, and problem posing would also test whether teachers can transfer the approach beyond the examples used during training.

4.8 Implications for Policy

Educational policies should move beyond unrestricted adoption or complete prohibition of generative AI. Schools and teacher-education institutions need clear standards for privacy, disclosure, verification, acceptable use, and professional accountability. AI integration also requires adequate infrastructure, mentoring, time, and subject-specific training; otherwise, unequal access may deepen existing disparities. Program evaluation should distinguish participant satisfaction from teacher learning, classroom implementation, and student outcomes, since favorable ratings alone do not demonstrate improved teaching or learning.

4.9 Limitations

The study involved 39 purposively selected teachers from one program, limiting statistical generalization. Its post-training design lacked a pretest, comparison group, and delayed follow-up, so perceived changes cannot be attributed

solely to the intervention. The evaluation instrument was not a validated scale, and the high ratings may reflect ceiling or social-desirability effects. Qualitative data were limited to brief written responses, while lesson quality, classroom implementation, and student outcomes were not independently assessed. The framework was also newly developed, and the model developer's involvement may have introduced confirmation bias. Future studies should use larger multi-site samples, comparison designs, independent assessment of teacher-produced materials, classroom observations, student-level measures, and longitudinal follow-up.

5. Conclusion

The Mixed-Model Prompt Engineering framework helped 39 secondary mathematics teachers move from general AI use toward more deliberate instructional co-design. Participants valued its support for structured prompting, lesson-exemplar development, workload reduction, and responsive mathematics teaching, while noting that sustained use requires adequate practice, reliable access, continued professional learning, and human verification. The study's main contribution is the concept of *pedagogical prompt literacy*: the teacher's ability to translate mathematical goals, learner needs, classroom constraints, and curriculum expectations into prompts, then critically refine and validate AI-generated outputs. This positions generative AI as a co-design tool rather than a substitute for professional judgment. Future studies should test the framework across settings, compare structured and unstructured prompting, and examine effects on lesson quality, classroom practice, student reasoning, engagement, and achievement.

References

- Ayalon, M., Naftaliev, E., Levenson, E. S., & Levy, S. (2021). Prospective and in-service mathematics teachers' attention to a rich mathematics task while planning its implementation in the classroom. *International Journal of Science and Mathematics Education*, 19(8), 1695–1716. doi:10.1007/s880-894-880-894
- Braun, V., & Clarke, V. (2021). *Thematic analysis: A practical guide*. SAGE Publications.
- Calleja, J., Foster, C., & Hodgen, J. (2024). Teachers' structuring of mathematical inquiry lessons: Shifting from "task-first" to "scaffolded inquiry." *Research in Mathematics Education*, 26(3), 460–493. doi:10.1080/14794802.2023.2176915
- Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). SAGE Publications.

- Dingman, S., Teuscher, D., Olson, T. A., & Kasmer, L. A. (2021). Conceptualizing curricular reasoning: A framework for examining mathematics teachers' curricular decisions. *Investigations in Mathematics Learning*, 13(4), 267–286. doi:10.1080/19477503.2021.1981742
- ElSayary, A., Kuhail, M. A., & Hojeij, Z. (2025). Examining the role of prompt engineering in utilizing generative AI tools for lesson planning: Insights from teachers' experiences and perceptions. *Human Behavior and Emerging Technologies*, 2025, Article 9986139. <https://doi.org/10.1155/hbe2/9986139>
- Federiakina, D., Molerov, D., Zlatkin-Troitschanskaia, O., & Maur, A. (2024). Prompt engineering as a new 21st century skill. *Frontiers in Education*, 9, Article 1366434. <https://doi.org/10.3389/feduc.2024.1366434>
- Fetters, M. D., Curry, L. A., & Creswell, J. W. (2013). Achieving integration in mixed methods designs: Principles and practices. *Health Services Research*, 48(6, Pt. 2), 2134–2156. <https://doi.org/10.1111/880-894.12117>
- Gurl, T. J., Markinson, M. P., & Artzt, A. F. (2025). Using ChatGPT as a lesson planning assistant with preservice secondary mathematics teachers. *Digital Experiences in Mathematics Education*, 11(1), 114–139. <https://doi.org/10.1007/s880-894-880-894>
- Jackson, C., Woolford Buchheister, K., & Taylor, C. E. (2023). A planning framework foregrounding equity in mathematics teaching and learning. *Investigations in Mathematics Learning*, 15(2), 103–117. doi:10.1080/19477503.2022.2150462
- Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., ... Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, Article 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- Lo, C. K. (2023). What is the impact of ChatGPT on education? A rapid review of the literature. *Education Sciences*, 13(4), Article 410. <https://doi.org/10.3390/educsci13040410>
- Miao, F., & Holmes, W. (2023). *Guidance for generative AI in education and research*. UNESCO. <https://doi.org/10.54675/EWZM9535>
- Ogunleye, B., Zakariyyah, K. I., Ajao, O., Olayinka, O., & Sharma, H. (2024). A systematic review of generative AI for teaching and learning practice. *Education Sciences*, 14(6), Article 636. <https://doi.org/10.3390/educsci14060636>
- Pepin, B., Buchholtz, N., & Salinas-Hernández, U. (2025). A scoping survey of ChatGPT in mathematics education. *Digital Experiences in Mathematics Education*, 11(1), 9–41. <https://doi.org/10.1007/s880-894-880-894>
- Rizos, I., Foykas, E., & Georgakopoulos, S. V. (2024). Enhancing mathematics education for students with special educational needs through generative AI: A case study in Greece. *Contemporary Educational Technology*, 16(4), Article ep535. <https://doi.org/10.30935/cedtech/15487>
- Robertson, J., Ferreira, C., Botha, E., & Oosthuizen, K. (2024). Game changers: A generative AI prompt protocol to enhance human–AI knowledge co-construction. *Business Horizons*, 67(5), 499–510. <https://doi.org/10.1016/j.bushor.2024.04.008>
- Rutherford, T., Rodrigues, A., Duque-Baird, S., Veng, S., Mykyta-Chomsky, R., Cao, Y., Chisholm, K., & Bergwall, E. (2025). “I just think it is the way of the future”: Teachers' use of ChatGPT to develop motivationally-supportive math lessons. *Computers and Education: Artificial Intelligence*, 8, Article 100367. <https://doi.org/10.1016/j.caeai.2025.100367>
- Tan, X., Cheng, K. S., & Ling, M. H. A. (2025). Artificial intelligence in teaching and teacher professional development: A systematic review. *Computers and Education: Artificial Intelligence*, 8, Article 100355. <https://doi.org/10.1016/j.caeai.2024.100355>
- Yang, K.-L., Cheng, Y.-H., Wang, T.-Y., & Chen, J.-C. (2023). Preservice mathematics teachers' reasoning about their instructional design for using technology to teach mathematics. *Asia-Pacific Journal of Teacher Education*, 51(3), 248–265. doi:10.1080/1359866X.2023.2198116
- Zaragoza, A., Seidel, T., & Santagata, R. (2023). Lesson analysis and plan template: Scaffolding preservice teachers' application of professional knowledge to lesson planning. *Journal of Curriculum Studies*, 55(2), 138–152. doi:10.1080/00220272.2023.2182650

Author Declaration

Originality statement: The authors declare that this manuscript represents original work and has not been previously published, in whole or in substantial part, nor is it currently under consideration for publication elsewhere. All sources and contributions from other works have been appropriately acknowledged and cited.

Conflict of interest: The authors declare that there are no conflicts of interest, financial or otherwise, that could have influenced the conduct, analysis, interpretation, or reporting of this study.