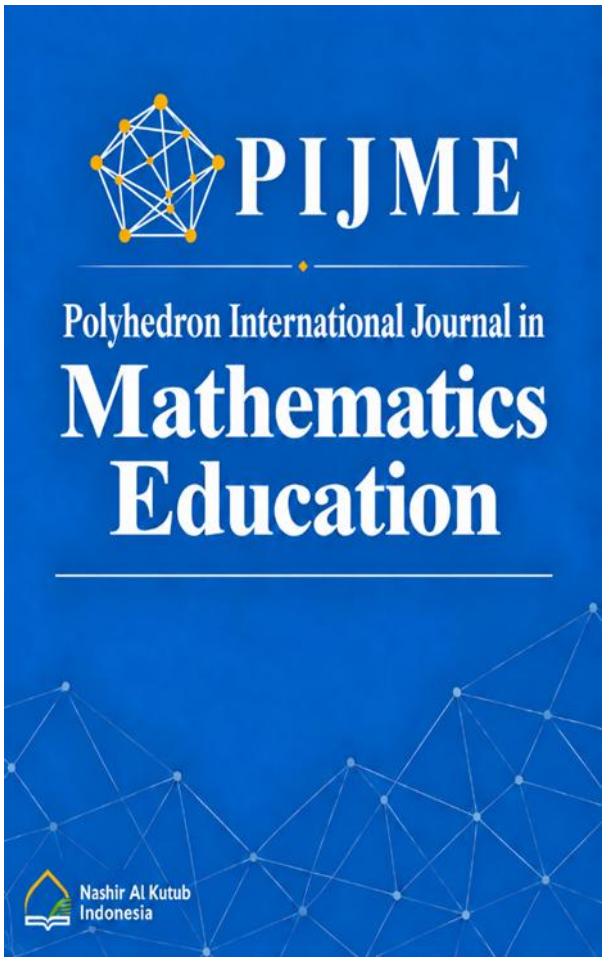
Polyhedron International Journal in Mathematics Education
https://nakiscience.com/index.php/pijme
PIJME
Nashir Al Kutub Indonesia



Prompting for products more than pedagogy: Prompt styles of elementary school teachers generating plane-figure questions with ChatGPT

Valeria Yekti Kwasaning Gusti
Department of Mathematics Education, Universitas Terbuka, Tangerang Selatan, Indonesia, valeria.gusti@ecampus.ut.ac.id

To cite this article:
Gusti, V. Y. K. (2027). Prompting for products more than pedagogy: Prompt styles of elementary school teachers generating plane-figure questions with ChatGPT. *Polyhedron International Journal in Mathematics Education*, *5*(1), 1–23.
To link to this article:
https://nakiscience.com/index.php/pijme
Published by:
Nashir Al-Kutub Indonesia
Residential Street Kila Rengganis, Block I, Number 11, Labuapi, Indonesia, 83361

Polyhedron International Journal in Mathematics Education
PIJME
https://nakiscience.com/index.php/pijme
Nashir Al Kutub Indonesia
E-ISSN: 2987-6540
Original Article

Prompting for products more than pedagogy: Prompt styles of elementary school teachers generating plane-figure questions with ChatGPT

Valeria Yekti Kwasaning Gusti
Department of Mathematics Education, Universitas Terbuka, Tangerang Selatan, Indonesia, valeria.gusti@ecampus.ut.ac.id

Abstract

Background: Teachers increasingly use ChatGPT to generate mathematics questions, yet their prompting and verification practices remain poorly understood.

Purpose: This study examined elementary teachers' prompting styles and verification practices when generating plane-figure questions with ChatGPT.

Method: Ten teachers from one public school in Depok, Indonesia, participated in a single session. Conversations underwent content analysis using four prompting strategies, eight specification elements, and follow-up classifications. Prompt profiles were derived through explicit rules, and generated items were reviewed for mathematical correctness.

Results: Nine opening prompts were analysable. All were brief, imperative, and zero-shot, with a median length of 14 words. Output customisation mainly concerned quantity, question type, and format. Only three prompts specified grade level and three specified cognitive level; none included learning objectives, student profiles, or examples. Four follow-up requests concerned additional items or formatting, without requesting verification or justification. One of ten fully reviewed multiple-choice items lacked a correct option. Four profiles emerged: level-labelled specification, format-focused, open delegation, and answer-key oriented.

Implications: Professional development should strengthen pedagogical specification and verification of AI-generated questions.

Contribution: The study identifies four prompting profiles and proposes a specify–generate–verify–revise cycle to support teachers' critical engagement with ChatGPT, although its effectiveness remains untested.

Article History
Received: September 2, 2026
Revised: October 28, 2026
Accepted: November 3, 2026
Published Online: November 11, 2026
Keywords:
ChatGPT;
Elementary school teachers;
Prompt engineering;
Question generation;
Teacher professional development

1. Introduction

Generative artificial intelligence (GenAI) chatbots such as ChatGPT are now available to teachers at no cost, and education researchers have quickly mapped both their opportunities and their risks (Kasneci et al., 2023; Lo, 2023; Tlili et al., 2023). Mathematics educators have begun to study how teachers use them to plan lessons, design tasks, and assess students (Almarashdi et al., 2024; Pepin et al., 2025; Su et al.,

This is an open-access article under the CC BY-SA 4.0 license.
Copyright © 2027 by Author

Polyhedron International Journal in Mathematics Education
PIJME
https://nakiscience.com/index.php/pijme
Nashir Al Kutub Indonesia
E-ISSN: 2987-6540
Original Article

2026; Wardat et al., 2023). A scoping survey reports that ChatGPT can help teachers adapt task formulations, for example by varying the level of cognitive demand, the link to the real world, and the rigour of mathematical language, while stressing that human oversight remains essential (Pepin et al., 2025). In a systematic review of 54 studies, the generation and planning of instructional materials was the second most frequent task supported by ChatGPT (22.2%), after mathematical problem solving (46.3%; Su et al., 2026), and a review of 20 studies on ChatGPT in school mathematics recommended structured teacher guidance because of limits in complex problem solving and student overreliance (Turmuzi et al., 2026).

For teachers, these tools change what professional knowledge is needed. Frameworks of teacher knowledge for the age of GenAI argue that technological, pedagogical, and content knowledge must now include knowing how to elicit, judge, and adapt machine-generated material (Celik, 2023; Mishra et al., 2023), and research on artificial intelligence for teachers identifies planning and assessment as the areas where teachers expect the most help but also face the greatest risk of uncritical adoption (Celik et al., 2022; Kim et al., 2022). Studies with pre-service teachers show that ChatGPT can generate lesson ideas and alternative solution strategies (Getenet, 2024; van den Berg & du Plessis, 2023; Yanar & Ergene, 2025), yet pre-service secondary mathematics teachers who critiqued ChatGPT's lesson plans often failed to recognise incorrect mathematical solutions and described them as alternative approaches (Gurl et al., 2025). Within the didactical tetrahedron, ChatGPT thus becomes a fourth vertex whose contributions the teacher has to mediate rather than simply accept (Dasari et al., 2024). In the present study, the tetrahedron serves as an interpretive lens rather than as a coding framework. The prompt is read as the teacher–artifact relation through which the teacher passes decisions about content and students to the chatbot, and verification is read as the teacher–content relation through which the output of the artifact is judged before it reaches students. The coding itself relies on the prompting and specification frames described in Section 2.4, and the tetrahedron is used in the Discussion to interpret what the teachers delegated and what they left unchecked.

What a chatbot returns depends on how it is asked. Prompt engineering is increasingly treated as part of AI literacy (Cain, 2024; Knoth et al., 2024; Ng et al., 2021; Walter, 2024). Studies of prompts in mathematics classify them by prompting strategy, namely zero-shot, few-shot, chain-of-thought, and output customisation (Bulut & Borromeo Ferri, 2026), and describe how pre-service teachers phrase prompts and detect errors when they solve mathematics tasks with ChatGPT (Noster et al., 2024). Prompts that explore a concept, seek feedback, or state a learning goal are treated as higher quality than prompts that only paste a task (Braskén et al., 2025), and single-prompt use that merely receives the output corresponds to the passive mode of engagement (Jaiswal & Nanda, 2026). How users prompt also matters for outcomes:

This is an open-access article under the CC BY-SA 4.0 license.
Copyright © 2027 by Author

Polyhedron International Journal in Mathematics Education
PIJME
https://nakiscience.com/index.php/pijme
Nashir Al Kutub Indonesia
E-ISSN: 2987-6540
Original Article

among 106 secondary students, a guided group used about twice as many adaptation prompts as an unguided group (Klar, 2025); prompt-writing training changed programming students' interactions (Güner & Er, 2025); follow-up questioning distinguished higher from lower performers (Yang et al., 2025); and students' use has been described as a cycle of stating an outcome, revising the prompt, and evaluating the response (Yoon et al., 2024).

Verification is the second concern. GenAI output in mathematics can contain errors that remain unnoticed unless the user checks them (McGalliard & Otten, 2025). ChatGPT-generated hints on mathematics problems failed quality checks on about one-third of problems before additional safeguards were applied (Pardos & Bhandari, 2024), and verification is described as a core requirement for classroom use (Su et al., 2026). School-based systems built on ChatGPT, such as a tool for fifth-grade word problems with automated test generation (Liu et al., 2025), and pilot studies reporting gains in engagement and self-regulated learning (Ng et al., 2024) show the potential of the technology, but they also presuppose that the generated content is correct and aligned with learning goals.

Most of this work concerns students who solve problems or pre-service teachers in university courses. Less is known about the prompts that in-service elementary school teachers write when they ask ChatGPT to produce questions for their own classes. In this situation, the prompt functions as an instructional specification: grade level, learning objective, cognitive demand, student characteristics, and format are parameters that the teacher can state explicitly. When they are not stated, the model does not leave them empty but fills them with its own defaults, for example by assuming a grade or a level of difficulty, and these assumptions may differ from the class the teacher has in mind. The distinction between explicit teacher specification and model inference therefore matters for instructional alignment, because only the former records the teacher's pedagogical intention. Teachers' design of assessment items with ChatGPT has been examined (Kwon & Ko, 2024, as cited in Pepin et al., 2025; the original study was not consulted, and only the information reported in the scoping survey is used here), but descriptions of prompt style are scarce, particularly for elementary school mathematics in Indonesia, where the use of GenAI by teachers is spreading faster than professional guidance. Studies of item generation mainly judge the items that a chatbot returns, and studies of prompting strategies classify how students or pre-service teachers phrase requests when they solve tasks (Bulut & Borromeo Ferri, 2026; Noster et al., 2024). Analysing the unguided prompts that in-service teachers actually write for their own classes adds a third view: it shows which instructional decisions teachers make explicit, which they leave to the model, and whether they check the product before it could reach students. The prompt-style profiles proposed here extend the strategy-based classification of Bulut and Borromeo Ferri (2026) by describing what a prompt specifies, not only how it is phrased. The

This is an open-access article under the CC BY-SA 4.0 license.
Copyright © 2027 by Author

Polyhedron International Journal in Mathematics Education
PIJME
https://nakiscience.com/index.php/pijme
Nashir Al Kutub Indonesia
E-ISSN: 2987-6540
Original Article

present study addresses this gap by describing how elementary school teachers prompted ChatGPT to generate questions on plane figures (*bangun datar*), how they continued the conversation, and what the generated items revealed about the need for verification. Its contribution is a descriptive vocabulary of prompt-style profiles that links the content of teachers' prompts to concrete targets for teacher professional development. Three research questions (RQs) guided the analysis:

RQ1. Which prompting strategies and specification elements appear in the teachers' opening prompts, and which prompt-style profiles do the combinations of these elements form?

RQ2. How do the teachers continue the conversation after the first response?

RQ3. What quality issues were identified in the generated items available in the saved conversations?

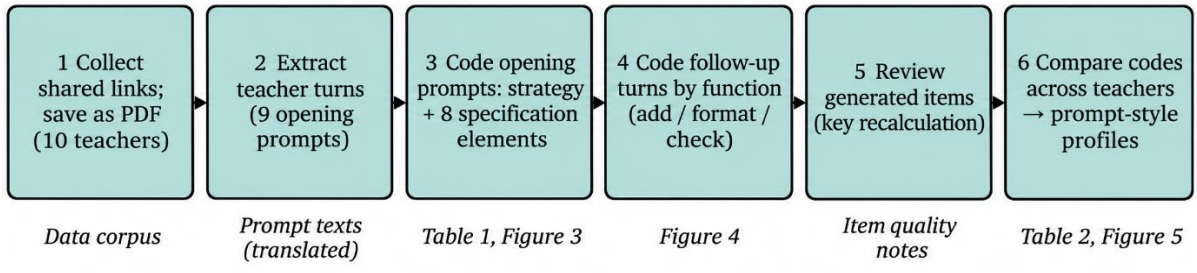
2. Methods

2.1 Research Design

This study used a qualitative descriptive design with content analysis of shared AI conversations. The coding of prompting strategies, specification elements, and follow-up functions was deductive, because the categories were defined in advance from the literature (Elo & Kyngäs, 2008). The prompt-style profiles were developed afterwards by comparing the coded combinations across teachers; this step was inductive in that the profiles were not fixed beforehand, and it was then formalised as explicit assignment rules (Section 2.4), following the logic of grouping coded data into higher-order categories described for inductive content analysis (Vears & Gillam, 2022). The design suits the aim because the prompts are written artefacts that can be analysed without influencing the teachers' wording. The analysis followed six steps (Figure 1): (1) collecting the shared conversation links and saving each conversation as a PDF file, (2) extracting the teachers' turns, (3) coding each opening prompt with a prompting-strategy frame and a specification checklist, (4) coding the follow-up turns by function, (5) reviewing the available items for mathematical correctness, with a full check of one item set, and (6) summarising prompt style in profiles by comparing the coded elements across teachers and applying the assignment rules.

Figure 1

Research Procedure



This is an open-access article under the CC BY-SA 4.0 license.
Copyright © 2027 by Author

Polyhedron International Journal in Mathematics Education
PIJME
https://nakiscience.com/index.php/pijme
Nashir Al Kutub Indonesia
E-ISSN: 2987-6540
Original Article

2.2 Participants

The participants were ten in-service teachers (coded A, B, C, E, F, H, I, J, K, and L) from one public elementary school in Depok, West Java, Indonesia. They were recruited by convenience after the principal agreed to host the session. The analysed corpus comprised conversations shared by teachers from the school who participated in the session; nine captured opening prompts met the requirements for prompt coding. The letter codes are anonymised identifiers retained from the study records. Their sequence is not used to infer attendance order, non-participation, or the total number of teachers approached; the reason for the gaps in the sequence was not documented in the study records. The available study records do not specify the teachers' gender or years of teaching experience. The study did not collect data on the teachers' prior experience with ChatGPT, which limits the interpretation of the profiles (Section 4.5). The distribution of teachers across the grades they taught was also not recorded. The target grades explicitly stated in the opening prompts were Grade 6 (Teacher F), Grade 5 (Teacher J), and Grade 2 (Teacher K); these prompt specifications should not be interpreted as the teachers' actual teaching assignments. All teachers used ChatGPT. Because the sample is a convenience sample from one school, it is described and not used to estimate proportions for a population.

2.3 Data Collection

The teachers were asked to generate mathematics questions on plane figures with a GenAI chatbot in a single session of about 15 minutes, without further instructions on how to write the prompt, so that their spontaneous prompting could be observed. The task can be summarised in Indonesian as "Buatlah soal matematika tentang bangun datar menggunakan ChatGPT" (English: "Use ChatGPT to generate mathematics questions on plane figures"). This wording is a reconstruction of the task, not a verbatim quotation of the oral instruction; the exact wording and any additional parameters supplied orally were not recorded. Consequently, the coding records what each teacher explicitly wrote; an unstated element cannot establish that the teacher had not received that information in the oral instruction. The teachers worked individually using ChatGPT. Device type, internet connection, account tier, and the model or version displayed in the interface were not recorded. The task is described as unguided with respect to prompt writing; this does not establish whether the teachers had received prompting training before the study. Each teacher then shared the link to the conversation, and the conversations were saved as PDF files. The session ran from 14:25 to 14:40 on 30 July 2026; this was the activity window of the session rather than a verified time range for every exported conversation. The opening prompt of Teacher E was not captured in the saved file, so nine opening prompts were coded. The reason for the missing opening prompt and the usability of any remaining turns could not be established. To keep the cross-case denominator consistent, Teacher E was excluded from all prompt, profile, and conversation-flow counts in Sections 3.1–3.3. Word

This is an open-access article under the CC BY-SA 4.0 license.
Copyright © 2027 by Author

Polyhedron International Journal in Mathematics Education
PIJME
https://nakiscience.com/index.php/pijme
Nashir Al Kutub Indonesia
E-ISSN: 2987-6540
Original Article

counts were taken from the original Indonesian prompts, not from the translations. Examples are the prompt of Teacher F, "Create grade 6 Mathematics questions. Topic: plane figures. Form: 10 multiple choice. HOTS," and the prompt of Teacher A, "Create interactive plane figure questions accompanied by systematic steps."

2.4 Data Analysis

The coding frame had three parts (Figure 2). Prompting strategy was coded with a frame adapted from Bulut and Borromeo Ferri (2026): zero-shot (a direct request without examples or templates), few-shot (worked examples or a template to follow), chain-of-thought (an explicit request for step-by-step reasoning or justification), and output customisation (specification of style, format, or level of detail). The definitions were applied to the generation of questions instead of the solution of problems. Eight specification elements were added as a checklist and grouped into pedagogical elements (grade level, cognitive level, learning objective or curriculum, and student profile or context) and product elements (number of items, item type, visual or file format, and answer key or worked solution), following the distinction between pedagogical and technological decisions in teacher knowledge frameworks for GenAI (Mishra et al., 2023). Follow-up turns were coded as adding items, changing format, or checking content, a distinction informed by the contrast between receiving, verifying, and explanation-seeking use (Jaiswal & Nanda, 2026) and by the self-regulation and feedback-seeking dimension of prompting (Braskén et al., 2025). The functional coding of turns follows earlier analyses of chatbot dialogues that used socio-emotional, cognitive, metacognitive, and coordinative categories (Feng, 2025) and dialogue-act coding schemes (Wang et al., 2025).

The codes were not mutually exclusive, with one exception: a prompt was coded as either zero-shot or few-shot. A request for chain-of-thought-style reasoning and output customisation could be added to either, so a prompt could be zero-shot and at the same time specify the output and request steps. Because the teachers asked the chatbot to generate questions rather than to solve them, a request for step-by-step reasoning could only take the form of a request that solution steps be shown. Such requests (Teacher A's "systematic steps" and Teacher I's pembahasan) were coded as a chain-of-thought type request because they ask the model to make reasoning explicit in the output; the qualifier "type" is kept because these requests also work as output requirements and are not chain-of-thought prompting in the strict sense. Teacher I's request was also coded as the product element worked solution, because pembahasan unambiguously refers to solutions of the items. Teacher A's "systematic steps" was not coded as a worked solution, because the prompt does not make clear whether the steps refer to solutions of the items or to the steps of an interactive activity. When a feature fitted two codes, the narrower code was used, and the decision was recorded in the coding sheet.

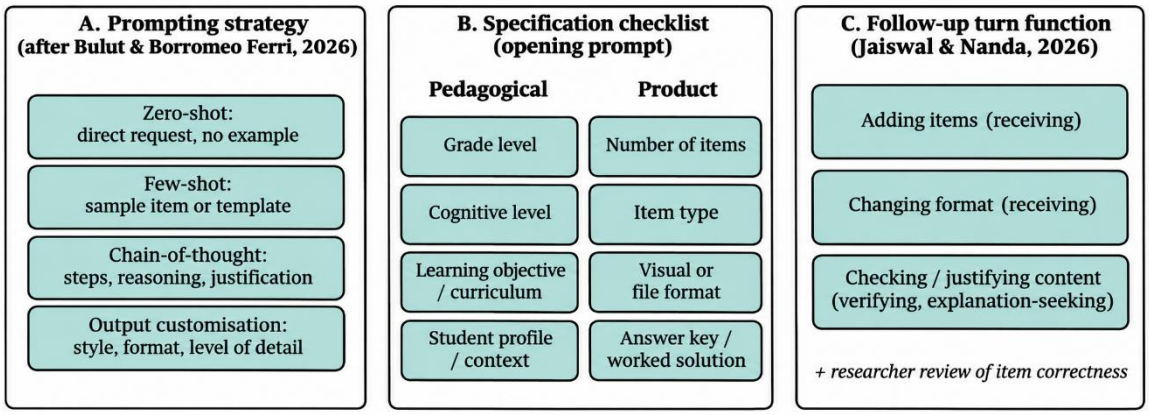
This is an open-access article under the CC BY-SA 4.0 license.
Copyright © 2027 by Author

Polyhedron International Journal in Mathematics Education
https://nakiscience.com/index.php/pijme
PIJME
Nashir Al Kutub Indonesia
E-ISSN: 2987-6540
Original Article

The eight specification elements were operationalised as follows. Grade level was coded only when a grade was named (e.g., "grade 6"); a school level such as "for an elementary school" (Teacher I) was not counted as grade level. Cognitive level was coded when the prompt contained any label for cognitive demand, such as HOTS or "level 1–3"; the label was recorded as written, and because the teachers did not define the labels, "level 1–3" (Teacher J) was not assumed to refer to the same framework as HOTS. Learning objective or curriculum reference was coded when the prompt stated what students should learn or be able to do or named a curriculum element; a topic description such as "about recognising plane figures" (Teacher K) was coded as the topic rather than as a learning objective, because it specified the content of the items but not an intended outcome. Student profile or context was coded when the prompt described the students (e.g., their ability or difficulties) or the classroom situation; general wording such as "attractive to children" (Teacher B) was treated as part of a format request. Assigning a role to the chatbot (e.g., "act as a grade 5 teacher") was recorded separately as a prompt feature and was not counted as a pedagogical element. Number of items, item type (e.g., multiple choice, essay), visual or file format (pictures, poster, interactive format, Word file), and answer key or worked solution were coded when they were stated explicitly. Features that the chatbot added on its own initiative, such as unrequested answer keys, were recorded separately and were not coded as teacher specifications.

Figure 2

Coding Frame Applied to Each Conversation



Prompt length was measured by counting words in the original Indonesian prompts. A word was defined as a string separated by spaces: numerals (e.g., "10") and abbreviations (e.g., HOTS) counted as one word each, hyphenated words and reduplications counted as one word, and punctuation was not counted. Length is reported only to describe the prompts and is not used as a measure of quality, because a short prompt can be well specified and a long one can be vague. To examine RQ3, mathematical correctness was checked by solving the items independently and comparing the solutions with the chatbot's answer key. A complete correctness check

This is an open-access article under the CC BY-SA 4.0 license.
Copyright © 2027 by Author

Polyhedron International Journal in Mathematics Education
PIJME
https://nakiscience.com/index.php/pijme
Nashir Al Kutub Indonesia
E-ISSN: 2987-6540
Original Article

was carried out for the ten multiple-choice items of Teacher F. No rationale for selecting this set was recorded, so it is treated as a case-specific review; the other sets were read once, and items that appeared mathematically ambiguous were flagged as unresolved. Each outcome was recorded as a confirmed error, an unresolved ambiguity, or not checkable because of export problems. Pedagogical quality (grade suitability, cognitive demand, clarity of wording, distractor quality, and alignment with the plane-figure topic) was not assessed systematically, so the review supports claims about correctness in one item set only and not an error rate for the corpus. Prompt-style profiles were formed by comparing the combination of pedagogical and product elements across teachers and were then fixed as assignment rules, applied in order to the opening prompt only: (1) level-labelled specification, a cognitive-level label together with at least two other elements; (2) format-focused, no cognitive-level label, and a visual or file format stated together with the number and type of items; (3) answer-key oriented, no cognitive-level label or visual format, and an answer key or worked solution requested; and (4) open delegation, no cognitive-level label and at most one product element, leaving most decisions to the model. Because the rules are applied in order, overlapping features (e.g., the answers requested by Teacher H in a poster prompt) do not lead to double assignment. Later turns were not used to assign profiles. All coding was done by one rater. To make the coding auditable, Table 2 presents the translated, abridged opening prompts and their coding, Appendix A gives the teacher-by-element matrix, and Supplementary Material S1 contains the full original Indonesian prompts, their English translations, the word counts, and the coding decision for each element, with identifying details removed.

2.5 Ethical Considerations

The school principal provided written permission for the session. Teachers were informed about the study's purpose, the analysis of their conversations, and their right to participate voluntarily without employment consequences. All teachers provided informed consent. No student data were collected. Teachers were identified using letter codes, and identifying school information was removed from the report. Saved conversations were stored on a password-protected device accessible only to the author. These safeguards applied to the downloaded files; the original conversations remained in the teachers' accounts and under their control.

3. Results

3.1 Prompting Strategies and Specification Elements (RQ1)

Nine of the ten opening prompts could be coded. All nine were written in ChatGPT, in informal Indonesian, as short imperative requests, with a median length of 14 words (mean 14.8; range 8–22). Table 1 summarises the elements that the opening prompts specified, and Figure 3 shows the pattern for each teacher.

This is an open-access article under the CC BY-SA 4.0 license.
Copyright © 2027 by Author

Polyhedron International Journal in Mathematics Education
https://nakiscience.com/index.php/pijme
PIJME
Nashir Al Kutub Indonesia
E-ISSN: 2987-6540
Original Article

Table 1

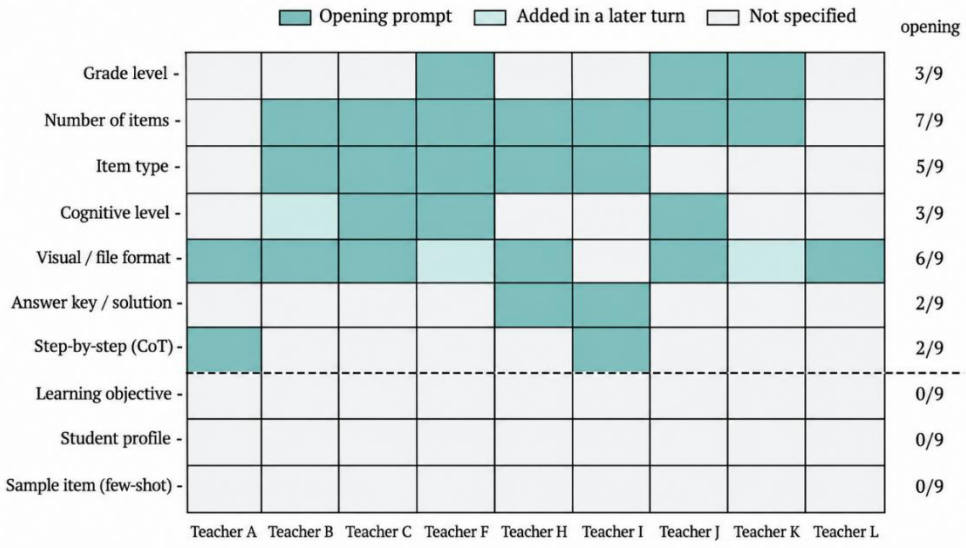
Prompting Strategies and Specification Elements in the Opening Prompts (n = 9)

Category	Description	Prompts (*n*)
Length	Median (mean; range) in words	14 (14.8; 8–22)
Prompting strategy	Zero-shot	9
	Few-shot (sample item or template)	0
	Chain-of-thought type request	2
	Output customisation	9
Product elements	Number of items stated	7
	Visual, interactive, or file format	6
	Item type stated	5
	Answer key or worked solution requested	2
Pedagogical elements	Grade level stated	3
	Cognitive level stated	3
	Learning objective or curriculum reference	0
	Student profile or classroom context	0

Note. Counts refer to the opening prompt only. Teacher E's prompt was not captured. No prompt assigned a role to the chatbot; role assignment was recorded separately and is not counted as a pedagogical element.

Figure 3

Specification Elements by Teacher in the Opening Prompt and Later Turns



Note. Each cell shows whether a teacher specified an element: dark cells mark elements stated in the opening prompt, light cells mark elements first stated in a later turn, and grey cells mark elements not stated in any saved turn. Values on the right are the

This is an open-access article under the CC BY-SA 4.0 license.
Copyright © 2027 by Author

Polyhedron International Journal in Mathematics Education
PIJME
https://nakiscience.com/index.php/pijme
Nashir Al Kutub Indonesia
E-ISSN: 2987-6540
Original Article

number of opening prompts (of 9) that contained each element; later turns are not included in these counts. Elements below the dashed line were never specified. Teacher E is not shown because the opening prompt was not captured.

Every opening prompt was zero-shot. None contained a sample item, an item grid, a curriculum reference, or a template to imitate, and none assigned a role to the chatbot or described the students. Output customisation was present in all nine prompts, and it concerned the product more than the pedagogy. The most frequent specifications were the number of items (7 prompts), visual, interactive, or file format (6), and item type (5). Six prompts requested a visual, interactive, or file format: pictures (Teachers B, J, and L), a poster (Teachers C and H), and an interactive format (Teacher A), with a Microsoft Word file added by Teacher J. Teachers F and K asked for a diagram or poster in later turns, so eight of the nine teachers requested a visual or interactive product at some point. Cognitive level was stated in three opening prompts (Teachers C, F, and J) and in a second turn by Teacher B, as the label HOTS (Teachers B, C, and F) or as "level 1–3" (Teacher J), in each case without a definition. The labels were recorded as written; because the prompt does not indicate which framework "level 1–3" refers to, it is not treated as equivalent to HOTS. Grade level was stated by three teachers (F, J, and K). Teacher I named the school level ("for an elementary school") but not a grade. Two prompts included a chain-of-thought type request: Teacher A asked for "systematic steps," and Teacher I asked for worked solutions (*pembahasan*). Teacher L asked for learning material on plane figures and their characteristics rather than for questions. This was a departure from the assigned task, so Teacher L's prompt is not comparable with the others on item type, number of items, or answer key. When Teacher L was excluded (n = 8), the pattern was unchanged: all eight prompts were zero-shot and contained output customisation, the number of items was stated in seven, item type in five, a visual or file format in five, grade level and cognitive level were each specified in three prompts, and no prompt stated a learning objective, a student profile, or a sample item.

3.2 Prompt-Style Profiles (Synthesis of RQ1)

Comparing the coded elements across teachers produced four prompt-style profiles (Table 2 and Figure 4). Level-labelled specification (Teachers C, F, and J) combined a cognitive-level label with the number of items, the item type or format, and, for F and J, the grade. The label refers to the presence of a cognitive-level label, not to completeness: none of these prompts stated a learning objective or a student profile, and Teacher C did not state a grade. Format-focused prompts (Teachers B and H) put the product first, stating a visual format together with the number and type of items: B requested pictures, and H requested a poster with answers. Open delegation (Teachers A, K, and L) stated at most one product element and left most decisions to the model: A and L named a format only, and K named the number of items and the grade but no item type, format, or level. Teacher K was assigned on the basis of the opening prompt;

This is an open-access article under the CC BY-SA 4.0 license.
Copyright © 2027 by Author

Polyhedron International Journal in Mathematics Education
PIJME
https://nakiscience.com/index.php/pijme
Nashir Al Kutub Indonesia
E-ISSN: 2987-6540
Original Article

the poster requested in the second turn is reported with the conversation flow (Section 3.3). The answer-key oriented profile (Teacher I) requested five essay questions with worked solutions and named the school but not the grade. Teacher E could not be classified. As Figure 4 shows, no teacher stated more than two of the four pedagogical elements. The number of product elements varied more: one element was stated by Teachers A, K, and L, two by F and J, three by B, C, and I, and four by H (median 2), so four of the nine opening prompts stated three or four product elements. Appendix A lists the elements coded for each teacher.

Table 2

Coding of the Opening Prompts by Teacher

Teacher	Opening prompt (translated, abridged)	Strategy	Profile
A	Create interactive plane figure questions accompanied by systematic steps	ZS, CoT, OC	Open delegation
B	Help me create 10 multiple-choice questions about plane figures, attractive to children using pictures (later: 5 HOTS items; poster without key)	ZS, OC	Format-focused
C	Create HOTS plane figure questions as 5 multiple choice, 5 short answer, and 5 essay in a poster image with a small title	ZS, OC	Level-labelled specification
E	Not captured in the saved file	Not coded	Not classified
F	Create grade 6 Mathematics questions. Topic: plane figures. Form: 10 multiple choice. HOTS (later: simple colourful diagram)	ZS, OC	Level-labelled specification
H	Create 10 multiple-choice questions about plane figures in poster form along with the answers	ZS, OC	Format-focused
I	Create 5 essay questions about plane figures for an elementary school, complete with worked solutions	ZS, CoT, OC	Answer-key oriented
J	Create 10 plane figure questions for grade 5 with pictures and question types level 1–3 in Microsoft Word	ZS, OC	Level-labelled specification
K	Create 5 grade 2 mathematics questions about recognising plane figures (later: in poster form)	ZS, OC	Open delegation
L	Create material on plane figures and their characteristics, with colourful pictures	ZS, OC	Open delegation

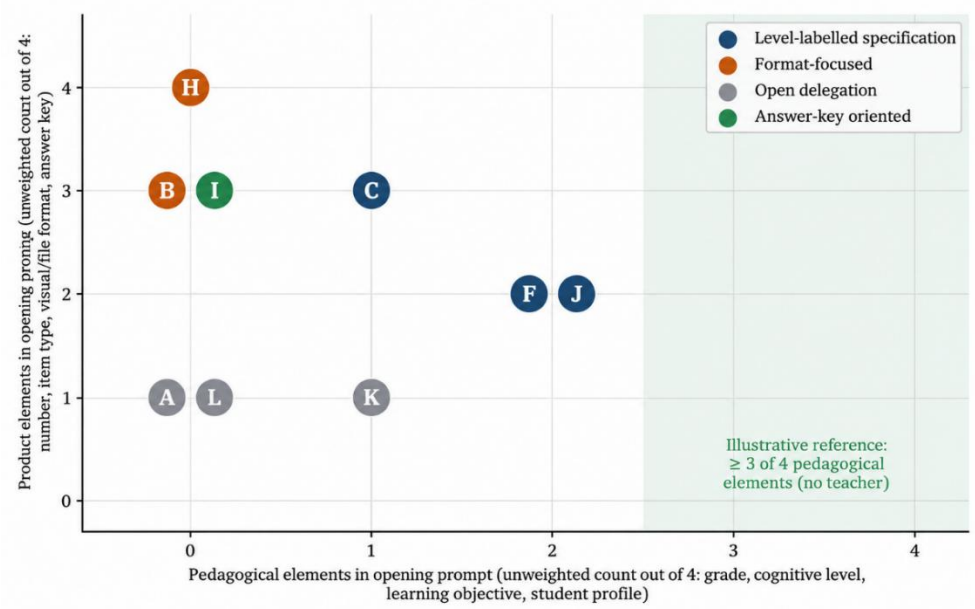
Note. ZS = zero-shot; CoT = chain-of-thought type request; OC = output customisation.

This is an open-access article under the CC BY-SA 4.0 license.
Copyright © 2027 by Author

Polyhedron International Journal in Mathematics Education
PIJME
https://nakiscience.com/index.php/pijme
Nashir Al Kutub Indonesia
E-ISSN: 2987-6540
Original Article

Figure 4

Prompt-Style Profiles by Pedagogical and Product Specification in the Opening Prompt



Note. Each circle is one teacher; overlapping positions are offset slightly. Both axes are unweighted counts of the elements stated in the opening prompt; they are descriptive and are not validated measures of prompt quality, and the elements on each axis may differ in importance. The shaded region (three or more of the four pedagogical elements) is an illustrative reference only, marking prompts that would state most of the pedagogical parameters; no teacher reached it.

3.3 Conversation Flow After the First Response (RQ2)

Figure 5 traces the flow of the conversations. The denominator is the nine teachers whose opening prompt was captured. Teacher E was excluded to keep the same nine-case denominator as in the opening-prompt and profile analyses; this exclusion does not imply that no later turns existed. Six of the nine teachers wrote a single prompt, with no follow-up recorded in the saved conversation. Teachers B, F, and K continued the conversation with four follow-up turns: Teacher B added five HOTS items and requested a poster without an answer key, Teacher F requested a simple colourful diagram, and Teacher K requested a poster. One turn therefore added items and three changed the format. No turn visible in the saved conversations pointed to an error, asked why an answer was correct, asked the chatbot to check its key, or asked for a change in difficulty. This does not show that the teachers did not check the items: they may have checked them independently, outside the chat, or after the session, which the saved conversations cannot reveal. In the terms of Jaiswal and Nanda (2026), the single-turn conversations correspond to receiving output without probing, and the multi-turn conversations stayed within a request–response pattern. The chatbot also added elements that the teachers had not requested, such as answer

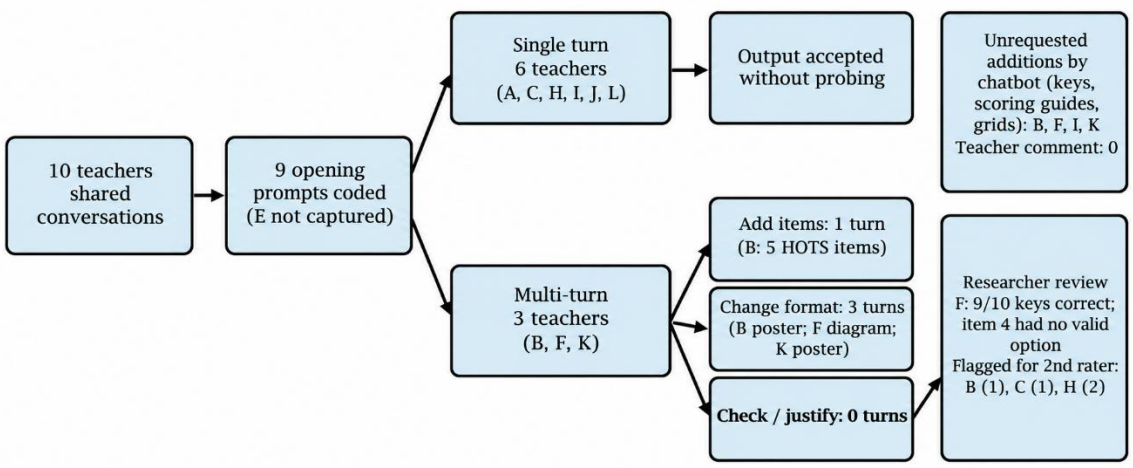
This is an open-access article under the CC BY-SA 4.0 license.
Copyright © 2027 by Author

Polyhedron International Journal in Mathematics Education
https://nakiscience.com/index.php/pijme
PIJME
Nashir Al Kutub Indonesia
E-ISSN: 2987-6540
Original Article

keys, scoring guides, or item grids (Teachers B, F, I, and K), and the saved conversations show no teacher comment on them. These model-added features were not counted as teacher specifications in Table 1, Figure 3, Figure 4, or Appendix A. They show that some parameters of task design were supplied by the model on its own initiative: the output of these conversations was more fully specified than the prompts, but those parts reflect the defaults of the model rather than decisions of the teachers.

Figure 5

Conversation Flow from the Opening Prompt to Follow-Up Turns and Item Review



3.4 Quality of the Generated Items (RQ3)

The saved conversations contained no teacher turns explicitly verifying the mathematical correctness of the generated items. Correctness was therefore examined through an author review of the available outputs; pedagogical quality was not assessed systematically (Section 2.4). The review distinguished confirmed mathematical defects from cases requiring further verification and products that could not be adequately inspected.

Teacher F's set of ten multiple-choice items was the only set recalculated in full. Nine items had correct answer keys, whereas item 4 had no mathematically valid answer option. The chatbot itself acknowledged this problem in the same response. This case therefore represents a confirmed defect in the generated item accompanied by a chatbot warning, rather than an error that went unacknowledged within the response. The available evidence does not establish whether the defect was subsequently corrected or whether the item was used. No teacher turn followed the warning in the saved record, but this absence does not establish whether the teacher read the warning or took action outside the conversation.

Four additional items were flagged for independent review: one potentially ambiguous item from Teacher B, one essay item from Teacher C, and two items from Teacher H. Their correctness remained unresolved, so they were not counted as confirmed errors or confirmed correct items. Inspection of the visual products was also

This is an open-access article under the CC BY-SA 4.0 license.
Copyright © 2027 by Author

Polyhedron International Journal in Mathematics Education
PIJME
https://nakiscience.com/index.php/pijme
Nashir Al Kutub Indonesia
E-ISSN: 2987-6540
Original Article

limited. Teacher B's requested poster was not available as a usable image in the saved conversation, and some saved poster outputs were truncated. These limitations prevented a complete assessment of their mathematical content and do not, by themselves, indicate errors in the generated products.

Overall, the review identified one confirmed item defect, four unresolved cases, and visual outputs that could not be fully assessed. Because only Teacher F's set was checked in full, these findings do not support an estimate of the error rate across all generated items. They describe the outcomes of the available review rather than establish the overall mathematical or pedagogical quality of the outputs.

A descriptive cross-case comparison provides context for these findings. The three teachers who continued their conversations represented different prompting profiles: Teacher B used a format-focused approach, Teacher F used level-labelled specification, and Teacher K used open delegation. The single confirmed defect was identified in Teacher F's set, which also received the most complete correctness review. Differences in review coverage limit comparisons across profiles: the absence of confirmed defects in other cases cannot be interpreted as evidence of higher quality. Accordingly, this comparison does not establish an association between prompting profile, conversational continuation, and item correctness.

4. Discussion

4.1 Prompting for Products Rather Than Pedagogy

The main finding is that, in what they typed during this short task, the teachers positioned the chatbot mainly as a producer of finished products. Their prompts were brief, zero-shot, and aimed at the form of the output, whereas the parameters that make a question fit a class, namely the learning objective, a defined cognitive level, the student profile, and the context, were mostly absent. This pattern resembles the student prompts reported by Braskén et al. (2025), many of which were copied tasks with little goal setting or feedback seeking, and the single-turn use that Jaiswal and Nanda (2026) classify as receiving output without transforming it. The teachers differed from these students in one respect: output customisation appeared in every prompt. Teachers specified quantity, item type, and format, which suggests that the practical side of prompting, obtaining a usable product, is developed, while the pedagogical side, setting cognitive demand and checking validity, is not. Pre-service teachers have been reported to use output customisation only rarely (Gallart et al., 2026, as cited in Bulut & Borromeo Ferri, 2026). The original study was not consulted, and it is not certain that its definition of output customisation matches the definition used here, so the comparison is offered only as a possible contrast; the samples, tasks, and grade levels of the two studies also differ. In terms of teacher knowledge frameworks, the typed prompts made the product side of the request explicit more often than the pedagogical and content-specific decisions that Intelligent-TPACK and

This is an open-access article under the CC BY-SA 4.0 license.
Copyright © 2027 by Author

Polyhedron International Journal in Mathematics Education
PIJME
https://nakiscience.com/index.php/pijme
Nashir Al Kutub Indonesia
E-ISSN: 2987-6540
Original Article

GenAI-oriented TPACK describe (Celik, 2023; Mishra et al., 2023). This is a statement about the prompts, not about the teachers' knowledge or competence. Several alternative explanations are plausible: the teachers may have assumed that the chatbot would infer the context, the 15-minute session may have encouraged brief requests, and decisions about grade, objective, and students may have been made outside the chat, for example when the items were selected or adapted for a class. Read through the didactical tetrahedron (Dasari et al., 2024), the teacher–artifact relation was used to delegate the product, whereas the teacher–content relation, in which the teacher decides on the mathematical demand and judges the correctness of what the artifact returns, was rarely made visible in the conversations.

Delegating a finished product to the chatbot has also been reported in other fields. Most programming students who used a chatbot eventually prompted it for a full solution (Rahe & Maalej, 2025), and students interacted more with a chatbot than with peers (Feng, 2025), although they contributed less dialogue to it than to peers in a creative problem-solving task (Song et al., 2025). Students who asked sophisticated follow-up questions outperformed exploratory inquirers (Yang et al., 2025), whereas the follow-up turns of the teachers here only extended the content or changed the format. The short session and the absence of guidance probably encouraged this behaviour. Whether the same pattern occurs when teachers prepare materials in daily practice is a hypothesis that the present data cannot test: the session was hosted by the principal, the teachers knew that their conversations would be analysed, and both conditions may have shaped what they requested and how. Time pressure during everyday preparation (Kim et al., 2022) makes the hypothesis plausible, but testing it requires observation of routine practice.

4.2 Cognitive Demand Left to the Model

Cognitive level was named in only three opening prompts, with the label HOTS or "level 1–3" and without a definition. GenAI tends to produce tasks of low cognitive demand (McGalliard & Otten, 2025), so a bare label leaves the judgement of what counts as higher-order thinking to the model. Pepin et al. (2025) note that the cognitive requirement level, the reference to the real world, and the rigour of mathematical language can be varied through the prompt, and reviews of ChatGPT in mathematics education report that the quality of output depends strongly on the precision of the request (Almarashdi et al., 2024; Monike et al., 2026; Wardat et al., 2023). These controls were rarely used by the teachers, and none requested a real-world context, even though contextual problems are central to elementary mathematics.

4.3 The Absence of Verification

The absence of visible verification turns matters because an item defect was found. Su et al. (2026) identify plausible but incorrect output as a recurring concern in their review of 54 studies and recommend verification as a routine part of mathematical activity. The 46.3% reported in the Introduction refers to a different

This is an open-access article under the CC BY-SA 4.0 license.
Copyright © 2027 by Author

Polyhedron International Journal in Mathematics Education
PIJME
https://nakiscience.com/index.php/pijme
Nashir Al Kutub Indonesia
E-ISSN: 2987-6540
Original Article

category of that review (the share of studies concerned with mathematical problem solving) and is not an error rate. Error rates of about one-third have been reported for ChatGPT-generated mathematics help without safeguards (Pardos & Bhandari, 2024), and pre-service teachers often fail to recognise incorrect solutions in ChatGPT output (Gurl et al., 2025). The defective item in the set of Teacher F shows what could happen if a ready-made item set went to a class unchecked; in this case, the chatbot itself flagged the problem, but the saved conversation shows no response to the warning. Yoon et al. (2024) described students' use of a chatbot as a cycle of identifying the desired outcome, revising the prompt, evaluating the response, and deciding what to take. The follow-up turns visible in the present data did not enter this cycle, because none revised a request after an evaluation of the output had been expressed (Figure 6).

4.4 Implications for Teacher Professional Development

Two implications follow. First, workshops can model prompts that state the learning objective, the class profile, and the cognitive level with a short definition, and that include one sample item, since no teacher used few-shot prompting. Second, a checking routine can be built into the task (Pepin et al., 2025; Su et al., 2026). Asking the chatbot to solve or check its own items may expose some errors, but it does not replace the teacher's independent mathematical verification, because the same model can repeat or justify its own mistake. Figure 6 combines both implications in a specify–generate–verify–revise cycle. The cycle is a proposal derived from the findings; it was not implemented or evaluated in this study. For a plane-figure task, it could work as follows. In the specify step, the teacher states the grade, the objective (e.g., “students determine the perimeter of shapes composed of rectangles”), the cognitive demand with a short definition, the students' profile, and one sample item, which addresses the omissions found under RQ1. In the generate step, the chatbot produces the items. In the verify step, the teacher solves each item independently, checks that exactly one option is correct, and judges grade suitability and cognitive demand, which addresses the defect found under RQ3. In the revise step, the teacher returns to the chatbot with a specific correction or a changed parameter, the type of turn that was absent under RQ2. Future training studies should assess whether such a cycle changes the specification elements in teachers' prompts, the frequency of verification and revision turns, and the correctness and pedagogical quality of the items that teachers bring to class. Short sessions did not change students' spontaneous prompting in one study (Braskén et al., 2025), whereas experienced users reported that their prompts improved over months of use by adding context and using multistep approaches (Kock et al., 2025). Teacher training therefore probably needs repeated practice with feedback rather than a single demonstration. Guidance has helped in other settings: scaffolded ChatGPT use improved undergraduates' mathematical problem solving (Zhou et al., 2025), a guided group of secondary students used about twice as many adaptation prompts as an unguided group (Klar, 2025), and prompt-writing training

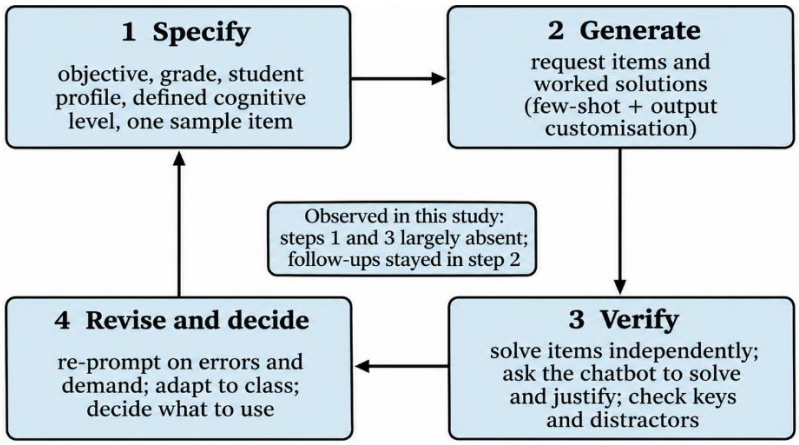
This is an open-access article under the CC BY-SA 4.0 license.
Copyright © 2027 by Author

Polyhedron International Journal in Mathematics Education
https://nakiscience.com/index.php/pijme
PIJME
Nashir Al Kutub Indonesia
E-ISSN: 2987-6540
Original Article

changed programming students' interactions (Güner & Er, 2025). Such training fits within broader efforts to build AI literacy and prompt engineering competence among educators (Cain, 2024; Knoth et al., 2024; Ng et al., 2021; Walter, 2024).

Figure 6

A Proposed Specify–Generate–Verify–Revise Cycle for Teacher Professional Development



Note. The cycle adapts the outcome–prompt–evaluation sequence described by Yoon et al. (2024) to item generation by teachers. The cycle was not implemented or evaluated in this study.

4.5 Limitations and Future Research

Five constraints affect how the findings should be interpreted. First, the sample is ten teachers from one school, recruited by convenience, so the four profiles describe this group and should not be generalised to other teachers, schools, or regions. Second, the prompts come from a single short session on one topic without guidance, and they may differ from the prompts these teachers write in routine practice. Third, all coding was done by one rater, so the reliability of the coding is not established (Appendix A and Supplementary Material S1 are provided so that readers can audit the coding decisions). Fourth, the teachers' prior experience with GenAI and their intentions were not collected, so the profiles describe the written prompts and not the reasons behind them. Fifth, the data show only what was visible in the saved conversations: checks that the teachers made outside the chat or after the session, and the effect of being observed in a session hosted by the principal, could not be captured. Future studies should involve teachers from several schools and examine prompting across multiple mathematics topics, use two independent coders, combine the analysis of conversations with interviews or think-aloud protocols, and examine whether training based on the cycle in Figure 6 changes teachers' prompts and the quality of the items they bring to class.

5. Conclusion

The elementary school teachers in this single-school study used brief, zero-shot prompts to ask ChatGPT for materials on plane figures. Their requests mainly specified the number, type, and visual format of the questions. Learning objectives and student

This is an open-access article under the CC BY-SA 4.0 license.
Copyright © 2027 by Author

profiles were absent, and the prompts that included cognitive-level labels did not define them. No teacher explicitly verified an item in the saved conversations. In the only set reviewed in full, one of ten multiple-choice items had no valid answer option, a defect that ChatGPT itself flagged.

The study contributes a descriptive framework for analysing what teachers specify when they use generative AI to design mathematics questions. The four profiles identified in the opening prompts were level-labelled specification, format-focused prompting, open delegation, and answer-key-oriented prompting. These profiles complement strategy-based classifications by showing that prompts using the same strategy can differ in which instructional decisions teachers make explicit and which they leave to the model. The profiles offer a basis for comparing prompting practices across schools, mathematics topics, and task settings.

For teacher professional development, the findings suggest pairing prompt-writing practice with independent mathematical verification. In the proposed specify–generate–verify–revise cycle, teachers articulate the learning objective, describe the learners, and define the intended cognitive demand before generating questions. They then solve each item independently, check that each multiple-choice item has exactly one valid answer, and assess whether the questions fit the learning objective and the learners. These checks guide revisions before classroom use. The cycle gives concrete form to calls for structured guidance and AI literacy in mathematics teacher education (Turmuzi et al., 2026). Future studies should test whether practising the cycle improves the pedagogical detail of teachers' prompts and the quality of the questions they select for classroom use.

6. Statements and Declarations

Funding:
This research received no external funding.

Competing Interests:
The author declares no competing interests.

Ethics Statement:
The school principal permitted the study. All teachers provided informed consent and participated voluntarily. No student data were collected, and teacher identities were protected using codes.

Data Availability:
Supplementary Material S1 contains the anonymised original Indonesian opening prompts, their English translations, the word counts, and the element-by-element coding decisions, and Appendix A gives the teacher-by-element matrix. The full coding sheet, which also records the follow-up turns, the features added by the model, and the outcomes of the item review, and the redacted generated items and follow-up turns are available from the author upon reasonable request for non-commercial research

This is an open-access article under the CC BY-SA 4.0 license.
Copyright © 2027 by Author

purposes; release is conditional on verification that the participants' consent permits the proposed sharing and that identifying information has been removed. The original conversations are not shared publicly because they may contain identifying information about the school.

Author Contributions:

Valeria Yekti Kwasaning Gusti: Conceptualization, Methodology, Investigation, Formal analysis, Visualization, Writing – original draft, Writing – review & editing.

Declaration of Generative AI Use:

ChatGPT was the object of study and was used by the participating teachers. In manuscript preparation, ChatGPT (model version not recorded) was used to assist with language editing, reference updating, and figure preparation. It was not used to code the data or to generate results. The author reviewed all AI-assisted content and takes full responsibility for the final text. All references were checked by the author against the publisher records (DOI, year, volume, issue, and pages or article number), and the labels and plotted values of every figure were checked against the coding sheet.

Acknowledgements:

The author thanks the principal and the teachers who participated in this study.

7. References

Almarashdi, H. S., Jarrah, A. M., Abu Khurma, O., & Gningue, S. M. (2024). Unveiling the potential: A systematic review of ChatGPT in transforming mathematics teaching and learning. *Eurasia Journal of Mathematics, Science and Technology Education, 20*(12), Article em2555. https://doi.org/10.29333/ejmste/15739

Braskén, M., Berts, K.-E., Wallin, S., & Frilund, S. (2025). Students' use of ChatGPT in an algebra class: A case study of prompts and attitudes. *LUMAT, 13*(1), Article 16. https://doi.org/10.31129/LUMAT.13.1.2877

Bulut, N., & Borromeo Ferri, R. (2026). AI-powered learning: Exploring secondary school students' use of ChatGPT for solving mathematical modelling problems. *ZDM – Mathematics Education, 58*, 993–1009. https://doi.org/10.1007/s11858-026-01803-6

Cain, W. (2024). Prompting change: Exploring prompt engineering in large language model AI and its potential to transform education. *TechTrends, 68*(1), 47–57. https://doi.org/10.1007/s11528-023-00896-0

Celik, I. (2023). Towards Intelligent-TPACK: An empirical study on teachers' professional knowledge to ethically integrate artificial intelligence (AI)-based tools into education. *Computers in Human Behavior, 138*, Article 107468. https://doi.org/10.1016/j.chb.2022.107468

Celik, I., Dindar, M., Muukkonen, H., & Järvelä, S. (2022). The promises and challenges of artificial intelligence for teachers: A systematic review of research. *TechTrends, 66*(4), 616–630. https://doi.org/10.1007/s11528-022-00715-y

This is an open-access article under the CC BY-SA 4.0 license.
Copyright © 2027 by Author

E-ISSN: 2987-6540
Original Article

Dasari, D., Hendriyanto, A., Sahara, S., Suryadi, D., Muhaimin, L. H., Chao, T., & Fitriana, L. (2024). ChatGPT in didactical tetrahedron, does it make an exception? A case study in mathematics teaching and learning. *Frontiers in Education, 8*, Article 1295413. https://doi.org/10.3389/feduc.2023.1295413

Elo, S., & Kyngäs, H. (2008). The qualitative content analysis process. *Journal of Advanced Nursing, 62*(1), 107–115. https://doi.org/10.1111/j.1365-2648.2007.04569.x

Feng, S. (2025). Group interaction patterns in generative AI-supported collaborative problem solving: Network analysis of the interactions among students and a GAI chatbot. *British Journal of Educational Technology, 56*, 2125–2145. https://doi.org/10.1111/bjet.13611

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), Article em0766. https://doi.org/10.29333/iejme/14141

Güner, H., & Er, E. (2025). AI in the classroom: Exploring students' interaction with ChatGPT in programming learning. *Education and Information Technologies, 30*, 12681–12707. https://doi.org/10.1007/s10639-025-13337-7

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*, 114–139. https://doi.org/10.1007/s40751-024-00162-9

Jaiswal, A., & Nanda, G. (2026). More than getting the answer: How students engage with generative AI in technical problem solving. *International Journal of STEM Education, 13*, Article 58. https://doi.org/10.1186/s40594-026-00646-7

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, 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

Kim, J., Lee, H., & Cho, Y. H. (2022). Learning design to support student-AI collaboration: Perspectives of leading teachers for AI in education. *Education and Information Technologies, 27*, 6069–6104. https://doi.org/10.1007/s10639-021-10831-6

Klar, M. (2025). Using ChatGPT is easy, using it effectively is tough? A mixed methods study on K-12 students' perceptions, interaction patterns, and support for learning with generative AI chatbots. *Smart Learning Environments, 12*, Article 32. https://doi.org/10.1186/s40561-025-00385-2

Knoth, N., Tolzin, A., Janson, A., & Leimeister, J. M. (2024). AI literacy and its implications for prompt engineering strategies. *Computers and Education:*

This is an open-access article under the CC BY-SA 4.0 license.
Copyright © 2027 by Author

Artificial Intelligence, 6, Article 100225. https://doi.org/10.1016/j.caeai.2024.100225

Kock, Z.-J., Salinas-Hernández, U., & Pepin, B. (2025). Engineering students' initial use schemes of ChatGPT as an instrument for learning. *Digital Experiences in Mathematics Education, 11*, 192–218. https://doi.org/10.1007/s40751-025-00169-w

Liu, J., Keane, T., Sun, D., Yang, Y., & Yang, Y. (2025). Design and evaluation of ChatGPT-MWPS: An AI-enhanced learning system for improving primary students' mathematical word problem solving. *Smart Learning Environments, 12*, Article 68. https://doi.org/10.1186/s40561-025-00419-9

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

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

Mishra, P., Warr, M., & Islam, R. (2023). TPACK in the age of ChatGPT and generative AI. *Journal of Digital Learning in Teacher Education, 39*(4), 235–251. https://doi.org/10.1080/21532974.2023.2247480

Monike, R. S., Sudirman, S., Kandaga, T., & Rodríguez-Nieto, C. A. (2026). Effectiveness of ChatGPT-integrated discovery learning on mathematical literacy in three-variable linear equation systems: A quasi-experimental study. *International Journal of Didactic Mathematics in Distance Education*, *3*(1), 64–82. https://doi.org/10.33830/ijdmde.v3i1.13157

Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence, 2*, Article 100041. https://doi.org/10.1016/j.caeai.2021.100041

Ng, D. T. K., Tan, C. W., & Leung, J. K. L. (2024). Empowering student self-regulated learning and science education through ChatGPT: A pioneering pilot study. *British Journal of Educational Technology, 55*, 1328–1353. https://doi.org/10.1111/bjet.13454

Noster, N., Gerber, S., & Siller, H.-S. (2024). Pre-service teachers' approaches in solving mathematics tasks with ChatGPT. *Digital Experiences in Mathematics Education, 10*, 543–567. https://doi.org/10.1007/s40751-024-00155-8

Pardos, Z. A., & Bhandari, S. (2024). ChatGPT-generated help produces learning gains equivalent to human tutor-authored help on mathematics skills. *PLOS ONE, 19*(5), Article e0304013. https://doi.org/10.1371/journal.pone.0304013

Pepin, B., Buchholtz, N., & Salinas-Hernández, U. (2025). A scoping survey of ChatGPT in mathematics education. *Digital Experiences in Mathematics Education, 11*, 9–41. https://doi.org/10.1007/s40751-025-00172-1

This is an open-access article under the CC BY-SA 4.0 license.
Copyright © 2027 by Author

Rahe, C., & Maalej, W. (2025). How do programming students use generative AI? *Proceedings of the ACM on Software Engineering, 2*(FSE), Article FSE045. https://doi.org/10.1145/3715762

Song, Y., Huang, L., Zheng, L., Fan, M., & Liu, Z. (2025). Interactions with generative AI chatbots: Unveiling dialogic dynamics, students' perceptions, and practical competencies in creative problem-solving. *International Journal of Educational Technology in Higher Education, 22*, Article 12. https://doi.org/10.1186/s41239-025-00508-2

Su, C. S., Hsu, C. C., & González Campos, J. A. (2026). ChatGPT as a resource to strengthen mathematics teaching and learning: A systematic review. *Journal of Computer Assisted Learning, 42*, Article e70241. https://doi.org/10.1002/jcal.70241

Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments, 10*, Article 15. https://doi.org/10.1186/s40561-023-00237-x

Turmuzi, M., Azmi, S., & Kertiyani, N. M. I. (2026). ChatGPT in school mathematics education: A systematic review of opportunities, challenges, and pedagogical implications. *Teaching and Teacher Education, 170*, Article 105286. https://doi.org/10.1016/j.tate.2025.105286

van den Berg, G., & du Plessis, E. (2023). ChatGPT and generative AI: Possibilities for its contribution to lesson planning, critical thinking and openness in teacher education. *Education Sciences, 13*(10), Article 998. https://doi.org/10.3390/educsci13100998

Vears, D. F., & Gillam, L. (2022). Inductive content analysis: A guide for beginning qualitative researchers. *Focus on Health Professional Education, 23*(1), 111–127. https://doi.org/10.11157/fohpe.v23i1.544

Walter, Y. (2024). Embracing the future of artificial intelligence in the classroom: The relevance of AI literacy, prompt engineering, and critical thinking in modern education. *International Journal of Educational Technology in Higher Education, 21*, Article 15. https://doi.org/10.1186/s41239-024-00448-3

Wang, D., Shan, D., Ju, R., Kao, B., Zhang, C., & Chen, G. (2025). Investigating dialogic interaction in K12 online one-on-one mathematics tutoring using AI and sequence mining techniques. *Education and Information Technologies, 30*, 9215–9240. https://doi.org/10.1007/s10639-024-13195-9

Wardat, Y., Tashtoush, M. A., AlAli, R., & Jarrah, A. M. (2023). ChatGPT: A revolutionary tool for teaching and learning mathematics. *Eurasia Journal of Mathematics, Science and Technology Education, 19*(7), Article em2286. https://doi.org/10.29333/ejmste/13272

Yanar, A. N., & Ergene, Ö. (2025). Integrating artificial intelligence in education: How pre-service mathematics teachers use ChatGPT for 5E lesson plan design.

This is an open-access article under the CC BY-SA 4.0 license.
Copyright © 2027 by Author

Polyhedron International Journal in Mathematics Education
https://nakiscience.com/index.php/pijme
PIJME
Nashir Al Kutub Indonesia
E-ISSN: 2987-6540
Original Article

Journal of Pedagogical Research, 9(2), 158–176. https://doi.org/10.33902/JPR.202533163

Yang, M., Jiang, S., Li, B., Herman, K., Luo, T., Chappell Moots, S., & Lovett, N. (2025). Analysing nontraditional students' ChatGPT interaction, engagement, self-efficacy and performance: A mixed-methods approach. *British Journal of Educational Technology, 56*, 1973–2000. https://doi.org/10.1111/bjet.13588

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

Zhou, R., He, X., Fan, Q., Li, Y., Li, Y., Xiao, X., & Fang, J. (2025). Exploring ChatGPT-facilitated scaffolding in undergraduates' mathematical problem solving. *Journal of Computer Assisted Learning, 41*, Article e70077. https://doi.org/10.1111/jcal.70077

Appendix A

Teacher-by-Element Matrix for the Opening Prompts (n = 9)

Teacher	Grade level	Cognitive level	Learning objective	Student profile	Ped. total	Number of items	Item type	Visual / file format	Answer key/ solution	Prod. total	Profile
A	–	–	–	–	0	–	–	✓	–	1	Open delegation
B	–	–	–	–	0	✓	✓	✓	–	3	Format-focused
C	–	✓	–	–	1	✓	✓	✓	–	3	Level-labelled specification
F	✓	✓	–	–	2	✓	✓	–	–	2	Level-labelled specification
H	–	–	–	–	0	✓	✓	✓	✓	4	Format-focused
I	–	–	–	–	0	✓	✓	–	✓	3	Answer-key oriented
J	✓	✓	–	–	2	✓	–	✓	–	2	Level-labelled specification
K	✓	–	–	–	1	✓	–	–	–	1	Open delegation
L	–	–	–	–	0	–	–	✓	–	1	Open delegation
Total	**3**	**3**	**0**	**0**		**7**	**5**	**6**	**2**		

Note. ✓ = stated in the opening prompt; – = not stated. Later turns and features added by the chatbot on its own initiative are not included. Teacher E is omitted because the opening prompt was not captured. Teacher A's "systematic steps" was coded as a chain-of-thought type request but not as a worked solution (Section 2.4). Profiles were assigned with the ordered rules in Section 2.4. Ped. = pedagogical; Prod. = product. The full original Indonesian prompts, translations, word counts, and coding decisions are given in Supplementary Material S1.

This is an open-access article under the CC BY-SA 4.0 license.
Copyright © 2027 by Author