https://doi.org/10.37420/j.cer.2026.095

Generative AI-Assisted Exercise Design in Primary Mathematics

Jun Wang*
Chengguan Town Central Primary School, Huoqiu County, Lu'an, Anhui 237400, China
***Corresponding author. E-mail: wangjun_huoqiu@163.com**

Abstract
The instructional value of generative artificial intelligence in primary mathematics depends on how generated exercises connect mathematical content, representation, and evidence of understanding. This conceptual article develops a framework for that connection through a focused narrative synthesis of research on AI-supported instructional materials, mathematics pedagogy, fraction instruction, and formative assessment. Five design principles specify goal alignment, mathematical completeness, cognitive demand, representational accessibility, and diagnostic potential. A preparation workflow translates these principles into task specification, candidate generation, mathematical verification, pedagogical revision, and sequencing. Three worked examples address equal partitioning, fraction magnitude, and the reference whole. Each links a design problem to a complete exercise, a solution rationale, and an instructional response. The analysis identifies representational precision as a central condition of exercise quality: the whole, partition, scale, and comparison conditions must support the intended inference. The framework connects AI assistance to decisions that teachers can articulate and inspect, providing a practical basis for developing conceptually coherent exercises. The examples are constructed illustrations; the framework's usability and educational effects require empirical evaluation.

Keywords:
generative artificial intelligence; primary mathematics; exercise design; mathematical reasoning; teacher judgment; formative assessment

1 Introduction

The quality of a primary mathematics exercise rests on the relationship between its learning objective, mathematical structure, and anticipated student response. Numerical correctness is essential, but the instructional purpose also depends on the reasoning elicited by the question and the representations through which quantities are expressed. An exercise on fractions, for example, may require learners to identify a shaded part, explain equal partitioning, or distinguish a fraction number from an amount. Although these tasks share familiar notation, they provide different evidence of understanding.

Generative artificial intelligence introduces new possibilities for producing and varying instructional materials. Teachers can request alternative contexts, representations, and explanations within a conversational

© 2026 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

interface. The educational significance of this capacity lies in the subsequent selection and development of exercises. NCTM's position on artificial intelligence emphasizes the continuing importance of teachers' mathematical and pedagogical expertise (National Council of Teachers of Mathematics [NCTM], 2024). Miao and Holmes (2023) similarly locate educational AI use within a human-centered approach. For exercise design, these positions raise a practical question: how should professional judgment be expressed in the preparation of a task?

Recent studies locate this question at the level of instructional materials. Sawyer and Aga (2026) analyzed 191 ChatGPT-generated elementary mathematics tasks, identifying concerns about repetition and grade-level appropriateness. Iida et al. (2026) examined practicing teachers' evaluations of AI-generated worksheets and highlighted the role of modification. These studies establish a basis for examining the relationship between generation and instructional suitability. An account of that relationship requires criteria specific enough to guide decisions about individual questions, diagrams, and solutions.

This article develops such an account for primary mathematics exercises. Its central question is how AI-supported preparation can maintain coherence among learning goals, task representations, and evidence of mathematical understanding. The contribution comprises five design principles, an associated preparation workflow, and worked examples that make the underlying decisions explicit. Introductory fractions provide the illustrative domain because equal partitioning, magnitude, and the reference whole expose consequential relationships between language, drawings, and mathematical inference. The framework concerns teacher preparation of classroom or homework exercises and can be implemented through a standard conversational interface.

2 Conceptual approach and theoretical basis

2.1 Analytical approach

The article adopts a conceptual design approach informed by a focused narrative synthesis. Sources were selected for their relevance to AI-supported preparation of instructional materials, mathematics task design, and the teaching and assessment of fractions. Peer-reviewed studies inform the account of AI assistance; professional and curriculum guidance establish pedagogical priorities; fraction instruction and formative assessment provide the basis for analyzing exercise structure and intended responses. The selection is purposive rather than systematic.

The synthesis connects these strands through the exercise as the unit of design. A learning objective is translated into a mathematical relationship, that relationship is expressed through a task and its representation, and the expected response supplies evidence for subsequent instruction. The framework is developed by identifying the decisions required at each connection. Its internal coherence is examined through three constructed examples with complete task specifications and solution rationales. These examples are constructed analytical illustrations rather than an empirical sample of model outputs or classroom observations.

2.2 AI-supported instructional preparation

AI-supported preparation involves both the production of candidate materials and their pedagogical in-

© 2026 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

terpretation. Luo and Tahir (2025) investigated teacher–AI collaboration in children's STEAM arts lesson planning through comparison and evaluation of lesson plans. Their study illustrates the value of examining preparation as a process in which an initial output is developed into an instructional resource. Its disciplinary setting differs from primary mathematics, where the validity of quantitative relations and representations requires particular attention.

Sawyer and Aga (2026) bring that attention to generated mathematics tasks, connecting content analysis with cognitive demand and curricular suitability. Iida et al. (2026) extend the material-centered perspective through teachers' evaluations of worksheets. Together, these studies motivate a task-level account of professional judgment. The present framework develops this account by linking specific design criteria to observable properties of an exercise. Its focus is the organization of instructional decisions within AI-assisted preparation.

2.3 Mathematical proficiency and formative interpretation

The National Research Council (2001) conceptualizes mathematical proficiency as the integration of conceptual understanding, procedural fluency, strategic competence, adaptive reasoning, and productive disposition. Exercise design operationalizes selected aspects of this broader conception. A question intended to consolidate a procedure differs from one intended to expose the meaning of a representation or elicit justification. The relevant criterion is the correspondence between the intended learning and the activity required for completion.

NCTM (2014) emphasizes reasoning, representations, mathematical discourse, and evidence of student thinking. China's compulsory education mathematics curriculum also places mathematical expression and reasoning within the aims of mathematics education (Ministry of Education of the People's Republic of China, 2022). For AI-assisted preparation, these priorities direct attention to task structure: the operation performed, the relationship examined, and the explanation made available to the teacher.

Fraction instruction makes these relationships especially visible. Siegler et al. (2010) recommend developing the understanding of fractions as numbers and using number lines as a central representation. Black and Wiliam (2009) connect assessment evidence with pedagogical decisions. These perspectives support a design orientation in which the expected solution and the interpretation of alternative responses are considered during preparation. A pedagogically well-specified exercise therefore includes both a valid solution and an account of how possible responses could inform instruction.

3 Design principles

3.1 Alignment with an explicit learning objective

A task specification should identify the mathematical relationship to be understood and the action through which that understanding will be expressed. For introductory fractions, explaining why a part represents one quarter of a rectangle provides a more determinate objective than a general request for fraction practice. The former specifies equal partitioning and the relation between part and whole; it also indicates what a satisfactory explanation should address.

© 2026 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

Prerequisites and curricular boundaries give the specification further precision. A sequence may assume recognition of halves and quarters while reserving formal equivalence procedures for later instruction. Folding, shading, and oral explanation can then be selected as appropriate means of engagement. The teacher supplies these constraints before requesting candidate exercises, so that alignment can be judged against stated knowledge and intended activity.

3.2 Mathematical completeness and validity

Mathematical verification encompasses the question, its representation, and the proposed solution. Relevant quantities, units, and comparison conditions must be stated sufficiently to support the intended conclusion. A shaded-fraction question requires a defined whole and a defensible partition. A comparison of fractional amounts requires information about the wholes to which the fractions refer.

Independent solution is therefore a substantive design step. It includes checking whether alternative interpretations produce different answers and whether the actual diagram satisfies the wording. For elementary visual tasks, dimensions, grids, and scale conventions can make the necessary relations explicit. An AI-generated solution may assist this process, but mathematical validity is established through the reasoning that supports the answer.

3.3 Cognitive demand embedded in task structure

The intended reasoning should be necessary for completing the exercise. A request for explanation is more informative when it specifies a relationship to examine, a claim to assess, or a representation to construct. Comparing two partitions, for example, can require a learner to distinguish the number of pieces from the equality of their areas. The solution rationale should identify this distinction as part of the required reasoning.

Cognitive demand should be considered separately from numerical size and procedural length. Small unit fractions can support substantial reasoning about magnitude or the reference whole. Conversely, a lengthy calculation may require only a rehearsed procedure. Support should reduce incidental barriers while preserving the central mathematical decision. Providing a rectangle outline, for instance, can reduce drawing demands while retaining the need to construct equal parts.

3.4 Accessible language and precise representation

An exercise should express its mathematical conditions in language and representations accessible to the intended learners. Contextual details are useful when they clarify a relationship or motivate a question; unnecessary detail can obscure the relevant quantities. Drawings require corresponding precision. Learners should be able to determine the whole, the meaning of boundaries, and whether size or distance represents a quantity.

Accessibility may be supported through manipulatives, enlarged diagrams, oral responses, or partially prepared figures. These adaptations concern the means of expressing understanding and can retain a common mathematical objective. Within the proposed framework, the teacher selects support according to the demands of the task and learners' prior instruction, rather than treating difficulty as a fixed characteristic of a child.

© 2026 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

3.5 Diagnostic potential and instructional response

Preparation should anticipate how a response could inform instruction. Alongside an accepted solution, the teacher identifies an alternative interpretation worth investigating. Selecting one of four unequal pieces as one quarter, for example, may indicate attention to the number of pieces without attention to equal measure. A follow-up question can test that interpretation by directing attention to area or inviting a new partition.

This anticipatory analysis links exercise design to formative assessment (Black & Wiliam, 2009). The interpretation remains provisional until it is examined through student work and dialogue. Its design value is that the task has a clear instructional consequence: the teacher can identify what to ask, compare, or represent next. Table 1 summarizes the five principles and their associated design requirements.

Table 1. Design principles and corresponding task specifications

Principle	Design requirement	Evidence in the prepared material
Goal alignment	Specify the relationship and the activity through which it is expressed	Learning objective, prerequisites, and content boundaries
Mathematical validity	Establish sufficient and consistent conditions for the solution	Complete wording, diagram specification, and verified rationale
Cognitive demand	Make the intended reasoning necessary for completion	Comparison, justification, or construction requirement
Accessibility	Support participation while preserving the mathematical decision	Concise wording and appropriate representation or support
Diagnostic potential	Connect possible responses to subsequent instruction	Interpretive focus and a follow-up question

4 Preparation workflow

The workflow translates the five principles into a sequence of preparation decisions. It begins with a teacher-defined specification of the objective, prerequisites, mathematical boundaries, and intended response. A generation request then asks for a small set of candidates in a consistent format: complete wording, solution rationale, representation requirements, and optional support. The request can specify these requirements directly so that candidate outputs include the information needed for mathematical and pedagogical review.

Mathematical verification precedes pedagogical revision. The teacher solves each candidate, examines its assumptions, and checks the consistency of words, diagrams, and quantities. A substantive defect requires revision of the task conditions or exclusion of the candidate. This ordering prevents stylistic refinement from substituting for verification of the mathematical relationship.

Pedagogical revision addresses the correspondence between the exercise and its objective. The teacher considers the reasoning required, the accessibility of the representation, and the evidence available from a response. A candidate may be reformulated as a comparison, supplemented with an exact drawing, or provided with targeted support. Recording the reason for a substantive revision preserves the connection between the initial specification and the final exercise.

Sequencing organizes the selected exercises around a developing mathematical relationship. Representa-

© 2026 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

tion, comparison, and critique can offer complementary approaches to the same concept. Planned follow-up questions connect the sequence to evidence from student work. The preparation record can remain brief: the specification, selected candidate, verified solution, final task, and rationale for revision. A tool name and generation date can be retained for traceability. Table 2 distinguishes the contributions of AI assistance from the teacher's design decisions.

Table 2. Distribution of functions in AI-assisted preparation

Stage	AI assistance	Teacher decision
Specification	Organize supplied constraints	Define objectives, prerequisites, and mathematical boundaries
Generation	Produce candidate formulations and variations	Select candidates for development
Verification	Suggest solutions and identify assumptions	Establish mathematical validity through independent reasoning
Revision	Offer wording and representation alternatives	Preserve cognitive demand and improve accessibility
Sequencing	Suggest connections and follow-up questions	Relate task order and instructional responses to the learning objective

5 Illustrative analysis of fraction exercises

5.1 Analytical scope

The examples address equal partitioning, fraction magnitude, and the reference whole. They assume familiarity with halves and quarters; the number-line example additionally assumes experience with a zero-to-one interval. Their placement within a teaching sequence depends on the local curriculum and prior instruction. Each example develops an initial question by making its mathematical conditions, representational demands, or opportunities for explanation more explicit. The student-facing wording remains direct, while the accompanying analysis specifies its instructional function. Halves and quarters are used to keep computational demands low while making the effects of partition, scale, and reference whole explicit. The analytical focus is the reasoning required by the exercise and the conditions supporting that reasoning.

5.2 Equal partitioning

Consider the question, "A rectangle is divided into four pieces. One piece is shaded. What fraction is shaded?" The number of pieces alone is insufficient to establish the fraction represented. Equal area is a necessary condition for identifying each piece as one quarter of the rectangle. The design problem concerns the relation between the partition and the proposed answer.

The developed exercise states: "Draw a rectangle 8 cm long and 2 cm high on square grid paper. Draw full-height vertical lines 1 cm, 2 cm, and 3 cm from the left edge. Shade the leftmost strip. Mei says, 'One of the four pieces is shaded, so one quarter of the rectangle is shaded.' Is Mei correct? Explain. In another rectangle of the same size, draw three full-height vertical lines to make four equal quarters."

The original strips have widths 1, 1, 1, and 5 cm, and therefore unequal areas. Mei's claim is invalid. Lines

© 2026 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

2, 4, and 6 cm from the left edge produce four equal-area strips in the second rectangle. Identifying the first shaded strip as one eighth is a possible extension; the central requirement is to explain why four pieces do not necessarily constitute four quarters.

The exercise connects critique with construction. The first component exposes the condition omitted from the initial question; the second requires a representation satisfying that condition. Providing the rectangle outlines can reduce measurement demands. A follow-up question such as "How can you show that the four parts cover equal amounts?" directs attention to the mathematical justification of the partition.

5.3 Magnitude on a number line

The question "Which is greater, one half or one quarter?" permits an answer without making the basis of comparison visible. A representation can strengthen its diagnostic potential by connecting the fractions to positions within a shared unit. This is consistent with the emphasis on fractions as numbers in Siegler et al. (2010).

The developed exercise states: "Draw a line segment 8 cm long. Label its left endpoint 0 and its right endpoint 1. The whole segment represents one unit. Mark one half and one quarter. Jun says, 'One quarter is greater because four is greater than two.' Do you agree? Explain using your number line."

One half lies 4 cm from zero and one quarter lies 2 cm from zero. Their positions establish that one half is greater. Dividing a fixed unit into four equal intervals produces smaller intervals than dividing it into two. The explanation links denominator meaning to magnitude, addressing the whole-number comparison embedded in the stated claim.

The task requires consistency among fraction labels, positions, and justification. A learner who identifies the greater fraction correctly but reverses the points produces evidence that warrants further interpretation. Asking how the drawing represents distance from zero can clarify that inconsistency. A blank line or foldable paper strip may support construction while preserving the placement decision. The mathematical demand resides in coordinating partition and magnitude, rather than recalling an isolated comparison rule.

5.4 The reference whole

The question "Lea eats one half of a cake and Bo eats one quarter of a cake; who eats more?" leaves the relationship between the cakes unspecified. The fraction numbers have a determinate order, but the physical amounts depend on the reference wholes. The developed exercise makes this distinction explicit through lengths and equal widths.

It states: "Two paper strips are each 2 cm wide. Strip A is 8 cm long and strip B is 16 cm long. Shade one half of strip A and one quarter of strip B. Compare the shaded areas and explain your answer. If both strips had the same length, would shading one half or one quarter produce the larger shaded area? Explain."

Both shaded portions have length 4 cm and width 2 cm, giving equal areas of 8 square centimeters. Learners can establish equality through direct comparison of the rectangles without formal area calculation. For equal whole strips, one half covers more area than one quarter. The contrast separates the order of fraction

© 2026 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

numbers from the amounts obtained by applying them to differently sized wholes.

Specifying width is essential because equal shaded lengths alone would not establish equal areas. Providing accurately dimensioned strips also prevents visual appearance from substituting for the stated quantitative relationship. The follow-up question, "What changes when the whole strips have equal lengths?", directs attention to the condition governing the comparison. Table 3 summarizes the representational decisions across the examples.

Table 3. Representational conditions and intended mathematical evidence

Example	Condition made explicit	Intended evidence of understanding
Equal partitioning	The areas of parts within one whole	Explanation and construction of equal-area quarters
Fraction magnitude	A shared zero-to-one unit and equal intervals	Consistency among positions, fraction labels, and comparison
Reference whole	The lengths and equal widths of two wholes	Distinction between fraction order and represented amount

6 Discussion

6.1 Instructional coherence as a design criterion

The framework links the specification of an exercise to the mathematical reasoning and instructional interpretation it supports. Its task-level focus gives operational content to teacher judgment: the teacher defines a relationship, establishes valid conditions for representing it, and identifies evidence that would inform subsequent teaching. These connections offer a criterion for evaluating the instructional coherence of a candidate exercise.

The examples show why representation is central to that criterion. In equal partitioning, the boundaries must establish equal area. On a number line, the unit interval and positions must support the magnitude comparison. In the reference-whole example, the dimensions determine the relationship between fractions and amounts. Drawing functions here as part of the mathematical specification. It enables the learner to construct or inspect the relationship and gives the teacher a basis for interpreting the response.

This perspective also distinguishes structural variation from contextual substitution. Replacing names or objects may leave the mathematical problem unchanged. Varying partition conditions or the size of the whole changes the inference a learner must make. AI assistance can be directed toward such variations when the teacher specifies which relationship should remain invariant and which condition should change. The pedagogical rationale then remains visible across the original brief, candidate task, and final exercise. Although these principles also apply to conventional exercise design, their role in AI-assisted preparation is to turn pedagogical intentions into generation constraints and criteria for selecting and revising outputs. For example, a request can vary the size of the whole while preserving the fractions, require explicit dimensions, and ask for a solution rationale. Verification then checks whether each candidate preserves those conditions, while the revision record explains why an output was accepted, changed, or excluded.

© 2026 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

6.2 Implications for primary mathematics preparation

The framework positions AI within the teacher's preparation process. A conversational interface can supply candidates for printed or classroom materials, without requiring student accounts. Generic descriptions of prerequisites provide sufficient context for the illustrated requests. This arrangement accords with the emphasis on agency and privacy in Miao and Holmes (2023).

The practical unit of work is a small set of exercises organized around a common objective. Each candidate requires mathematical verification, so the volume generated should reflect the time available for substantive inspection. A brief revision record can support continuity between lessons by preserving the reason a diagram, wording, or comparison condition was selected. The value of AI assistance depends on whether its candidates contribute useful alternatives within this process.

For teachers concerned with drawing strategies and mathematical problem solving, the framework offers a way to connect visual construction with reasoning. Diagrams can serve as objects of explanation, means of comparison, or indications of reasoning that warrant further investigation. Preparation accordingly includes deciding what learners should draw, what the drawing must preserve, and how its features relate to the expected justification. These decisions place mathematical meaning at the center of resource development.

6.3 Limitations and further research

The framework is conceptual and illustrated within introductory fractions. Its usability, preparation costs, and effects on classroom reasoning have not been evaluated. The purposive literature selection and limited content domain also constrain the scope of the synthesis. The design principles provide an organizing framework rather than a validated assessment scale.

Further research could examine preparation records linking teacher specifications, preserved AI outputs, revisions, and independent solutions. Such records would allow analysis of the decisions required across teachers and mathematical topics. Classroom evidence would then be needed to evaluate how the resulting exercises elicit reasoning and inform instruction. Distinguishing preparation quality from learning outcomes would clarify the contribution of each stage.

7 Conclusion

AI-assisted exercise design requires coherence among learning objectives, mathematical conditions, representations, and the interpretation of student responses. The five principles and preparation workflow developed here organize these relationships into explicit decisions. The fraction examples demonstrate how equal partitioning, magnitude, and the reference whole shape complete exercises and their solution rationales.

The framework provides a basis for using AI-generated candidates within purposeful primary mathematics preparation. Its central implication is that exercise quality depends on the mathematical relationships made available for learners to represent, examine, and explain. Empirical evaluation can establish how this design approach operates in teachers' preparation and classroom practice.

© 2026 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

References

[1]Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. Educational Assessment, Evaluation and Accountability, 21(1), 5–31. https://doi.org/10.1007/s11092-008-9068-5

[2]Iida, S., Nakagawa, S., & Sakakibara, N. (2026). Evaluations by practicing teachers of AI-generated educational materials and their potential for educational use. Japan Journal of Educational Technology, 50(1), 165–176. https://doi.org/10.15077/jjet.49081 [In Japanese; English title and abstract available.]

[3]Luo, Z., & Tahir, R. (2025). ChatGPT-assisted lesson planning for children's STEAM arts education: An experimental study on benefits, challenges, methods, and a prompt framework. International Journal of Artificial Intelligence in Education, 35, 3696–3745. https://doi.org/10.1007/s40593-025-00513-6

[4]Miao, F., & Holmes, W. (2023). Guidance for generative AI in education and research. UNESCO. https://doi.org/10.54675/EWZM9535

[5]Ministry of Education of the People's Republic of China. (2022). Yiwu jiaoyu shuxue kecheng biaozhun 2022 nian ban [Compulsory education mathematics curriculum standards 2022 edition]. Beijing Normal University Press. [In Chinese.]

[6]National Council of Teachers of Mathematics. (2014). Principles to actions: Ensuring mathematical success for all. NCTM.

[7]National Council of Teachers of Mathematics. (2024). Artificial intelligence and mathematics teaching: A position of the National Council of Teachers of Mathematics. https://www.nctm.org/uploadedFiles/Standards_and_Positions/Position_Statements/NCTM_AI_Position_Statement.pdf

[8]National Research Council. (2001). Adding it up: Helping children learn mathematics (J. Kilpatrick, J. Swafford, & B. Findell, Eds.). National Academy Press. https://doi.org/10.17226/9822

[9]Sawyer, A. G., & Aga, Z. G. (2026). A mile high and an inch deep: Exploring ChatGPT as a mathematics curriculum development tool. School Science and Mathematics, 126(1), 60–74. https://doi.org/10.1111/ssm.18348

[10]Siegler, R., Carpenter, T., Fennell, F., Geary, D., Lewis, J., Okamoto, Y., Thompson, L., & Wray, J. (2010). Developing effective fractions instruction for kindergarten through 8th grade: A practice guide (NCEE 2010-4039). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. https://ies.ed.gov/ncee/wwc/Docs/PracticeGuide/fractions_pg_093010.pdf

© 2026 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)