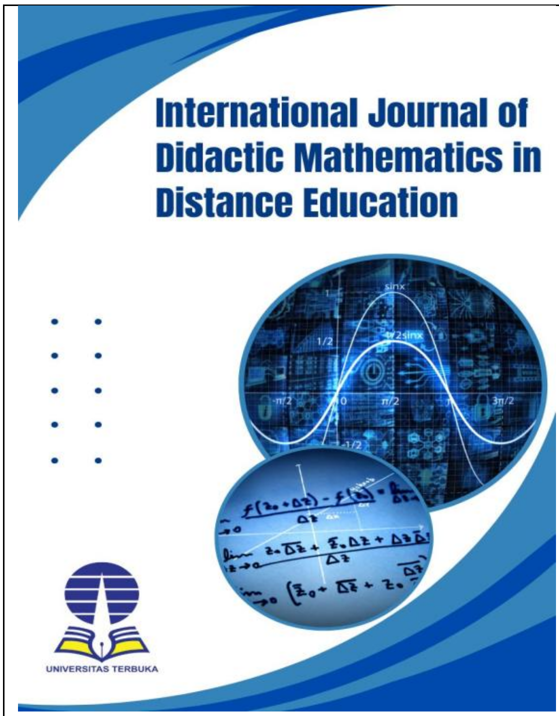
INTERNATIONAL JOURNAL OF
DIDACTIC MATHEMATICS IN DISTANCE EDUCATION

Journal homepage: https://jurnal.ut.ac.id/index.php/ijdmde



From calculation to adaptive mathematical competence: Reconceptualizing mathematical operations for the generative AI era in school mathematics

Jiajie Yan
Chengdu Academy of Educational Sciences, Chengdu, China, 201931130001@mail.bnu.edu.cn

To cite this article:
Yan, J. (2026). From calculation to adaptive mathematical competence: Reconceptualizing mathematical operations for the generative AI era in school mathematics. *International Journal of Didactic Mathematics in Distance Education*, *3*(2), 210-232.
To link to this article:
https://jurnal.ut.ac.id/index.php/ijdmde
Published by:
Universitas Terbuka
Jl. Pd. Cabe Raya, Pd. Cabe Udik, Kec. Pamulang, Kota Tangerang Selatan, Banten 15437

International Journal of Didactic Mathematics in Distance Education
https://jurnal.ut.ac.id/index.php/ijdmde

From calculation to adaptive mathematical competence: Reconceptualizing mathematical operations for the generative AI era in school mathematics

Jiajie Yan
Chengdu Academy of Educational Sciences, Chengdu, China, 201931130001@mail.bnu.edu.cn

Abstract
Mathematical operations are often treated as routine calculation, yet generative artificial intelligence (AI) challenges this view because procedures can now be executed instantly by digital tools. This structured narrative review reconceptualizes mathematical operation competence for the generative AI era by synthesizing 44 English-language sources published between 1991 and June 2026 on procedural fluency, conceptual knowledge, strategy flexibility, adaptive expertise, mathematical errors, instruction, and AI-mediated learning. The synthesis shows that operation competence is multidimensional, integrating accuracy, efficiency, conceptual understanding, flexibility, metacognitive monitoring, verification, and transfer. Conceptual and procedural knowledge develop iteratively, while effective strategy selection requires knowledge of multiple procedures and sensitivity to problem characteristics. Persistent operational errors arise from interacting conceptual, cognitive, affective, and instructional factors, including misconceptions, working-memory demands, mathematics anxiety, and teaching focused primarily on answer production. Generative AI introduces an additional risk of procedural outsourcing, whereby AI-generated solutions replace students' independent reasoning, execution, and monitoring. Effective instruction should therefore integrate structural and conceptual understanding, comparison of solution strategies, deliberate and spaced practice, analysis of errors, productive struggle, formative feedback, and systematic verification of AI- and CAS-generated outputs. The review proposes an integrative definition of mathematical operation competence and identifies priorities for construct validation, longitudinal and cross-cultural research, classroom interventions, teacher knowledge, and human-AI interaction, positioning structured AI verification as a central component of future mathematical competence.

Article History
Received:
July 12, 2026
Revised:
August 27, 2026
Accepted:
August 29, 2026
Published Online:
September 3, 2026

Keywords:
Adaptive expertise; Generative artificial intelligence; Mathematical operation competence; Procedural fluency;

1. Introduction

Operations are among the most visible features of school mathematics. Students add and multiply numbers, simplify expressions, transform equations, manipulate functions, calculate limits, and execute statistical procedures. Because these activities produce observable answers, operational competence has often been equated with speed and

International Journal of
Didactic Mathematics in Distance Education
https://jurnal.ut.ac.id/index.php/ijdmde
LPPM UT
LEMBAGA PENELITIAN DAN PENGABDIAN KEPADA MASYARAKAT
UNIVERSITAS TERBUKA

accuracy. This narrow interpretation is educationally consequential. It encourages repetitive practice detached from meaning, treats errors as carelessness rather than evidence of reasoning, and overlooks the strategic decisions that precede and accompany a calculation. The arrival of generative artificial intelligence, capable of producing a fluent, complete solution to almost any school-mathematics procedure on request, has made this narrow interpretation newly untenable: if operation competence is reduced to obtaining a correct answer, a chatbot already possesses it, and the construct itself must be reconceptualized around what remains distinctively the learner's to know, decide, and verify. This equation of competence with speed and accuracy echoes recurring critiques in the field. Star (2005) observed that procedural knowledge has often been assessed by the ability to reproduce steps rather than to understand or adapt them, and National Council of Teachers of Mathematics (NCTM, 2014) explicitly warns against treating fluency as rapid, accurate recall in isolation from understanding. The pattern is not only theoretical: international assessment frameworks such as PISA continue to report items that reward computational speed and accuracy as core indicators of mathematical proficiency in many education systems, even as the assessment's own framework argues for a broader, reasoning-oriented construct (OECD, 2023).

International frameworks define mathematical proficiency more broadly. The five-strand model in Adding It Up positions procedural fluency alongside conceptual understanding, strategic competence, adaptive reasoning, and productive disposition, emphasizing that the strands develop together rather than in isolation (National Research Council, 2001). The National Council of Teachers of Mathematics similarly argues that procedural fluency should be built from conceptual understanding and should include flexible, accurate, and efficient use of procedures (NCTM, 2014). The OECD's PISA 2022 framework further places mathematical procedures within reasoning, problem solving, modeling, and computational thinking in technology-rich environments (OECD, 2023). These perspectives challenge the assumption that an operationally competent student is merely one who calculates quickly.

The issue is particularly important in curriculum systems that explicitly identify mathematical operations as a core competency. In the Chinese senior-secondary curriculum, mathematical operation competence refers to understanding operation objects, applying rules, exploring operation routes, selecting methods, designing procedures, and obtaining results (Ministry of Education of the People's Republic of China, 2020). This formulation is conceptually rich, but the research base associated with it has often remained fragmented across studies of calculation ability, procedural knowledge, error diagnosis, algebraic manipulation, and classroom teaching. A direct translation of the term mathematical operation also creates difficulty in international publication because English-language scholarship more commonly uses procedural fluency, procedural knowledge, computational fluency, procedural flexibility, and adaptive expertise. A publishable review therefore needs to connect the curriculum-specific construct to these established research traditions. This claim of fragmentation is illustrated by how the original literature on mathematical operation competence is

International Journal of
Didactic Mathematics in Distance Education
https://jurnal.ut.ac.id/index.php/ijdmde

typically disseminated: often as classroom-implementation reports for specific algorithms or grade-level error analyses in domestic pedagogical outlets, rather than as a cumulative body of work that builds on itself or engages directly with the international procedural-fluency, procedural-flexibility, or adaptive-expertise traditions synthesized here. It should also be stated explicitly that this review interprets the Chinese curriculum construct of mathematical operation competence through the lens of the English-language research traditions identified in Section 2.2, rather than through a systematic search of Chinese-language sources; the characterization of earlier work as descriptive and practitioner-based therefore refers to English-language and translated commentaries on the construct, not to the Chinese-language empirical literature as a whole, whose full scope lies beyond what this review can evaluate. This boundary is acknowledged as a limitation in Section 4.4.

The original literature on mathematical operations has offered useful descriptions of the construct, its relation to other competencies, common student errors, and instructional recommendations. However, much of that work is descriptive, locally published, and based on practitioner experience rather than theoretically grounded empirical evidence. It also predates recent developments in strategy-flexibility research, mathematics-anxiety meta-analysis, digital mathematics education, and generative artificial intelligence. An updated synthesis should not simply translate earlier claims; it should reorganize them around internationally recognizable constructs, evaluate the strength of the evidence, and specify a future research program.

This review addresses four research questions: Despite growing international attention to procedural fluency, procedural flexibility, and adaptive expertise, no prior synthesis has systematically connected these traditions to the curriculum-specific construct of mathematical operation competence or proposed an integrative, measurable definition that spans both research communities. This gap motivates the four research questions addressed below.

RQ1. How should mathematical operation competence be conceptualized in relation to procedural fluency, conceptual and procedural knowledge, procedural flexibility, and adaptive expertise?

RQ2. How does operation competence interact with broader mathematical competencies such as reasoning, representation, problem solving, modeling, and computational thinking?

RQ3. Which conceptual, cognitive, affective, and instructional factors, including uncritical reliance on generative AI, explain students' persistent operational errors?

RQ4. Which instructional approaches, including verification routines for generative-AI-assisted work, are supported by theory and evidence, and what research gaps remain?

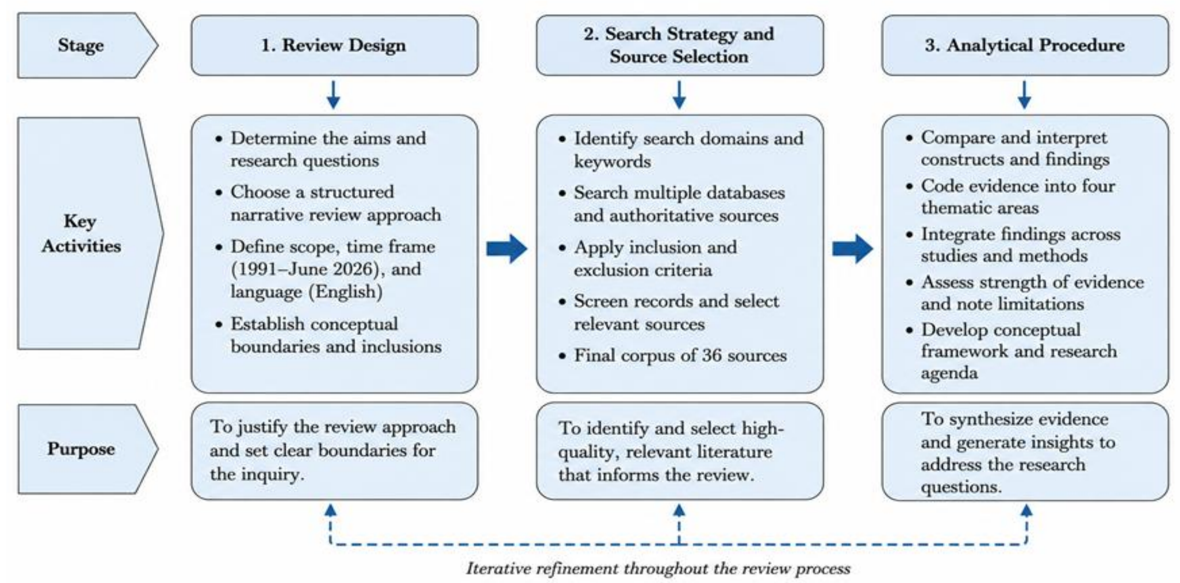
Building directly on the gap identified above, the review makes three contributions. First, it proposes a multidimensional definition that is sufficiently broad to include numerical and symbolic work but sufficiently precise to guide measurement and instruction. Second, it reframes operation errors and instruction within an integrated model linking mathematical knowledge, cognitive resources, affect, classroom tasks, teacher knowledge, and technological tools. Third, it treats generative AI not as a peripheral instructional tool but as a condition that redefines what operation competence must now demonstrate, and on that basis develops verification-oriented instructional principles for operation-focused teaching in an AI-saturated learning environment.

2. Methods

The methods employed in this study were designed to provide a transparent and structured process for identifying, selecting, analyzing, and synthesizing relevant literature on mathematical operation competence. The review process consisted of three interconnected components: review design, search strategy and source selection, and analytical procedure. Figure 1 provides an overview of these methodological stages and illustrates how the selected literature was progressively organized and synthesized to address the research questions.

Figure 1

Methodological Framework for the Structured Narrative Review Process



2.1 Review design

A structured narrative review was selected because the literature is theoretically heterogeneous and uses overlapping but non-equivalent constructs. Narrative synthesis is appropriate when the objective is to clarify concepts, connect research traditions, and develop an interpretive framework rather than estimate a single pooled effect (Grant & Booth, 2009; Snyder, 2019). The review is structured in the sense that the search domains, selection principles, and thematic synthesis are made explicit; it is not presented as an exhaustive systematic review or meta-analysis. Structured approach follows established guidance on narrative synthesis and review typologies (Grant & Booth, 2009; Snyder, 2019); Section 2.2 specifies the search domains, source-selection criteria, and screening rationale in place of a PRISMA-style flow diagram, consistent with the review's declared narrative, rather than systematic, scope.

2.2 Search domains and source selection

English-language literature published from 1991 through June 2026 was identified through targeted searches of ERIC, PubMed and PubMed Central, the National Academies Press, NCTM publications, OECD publications, and major journal-publisher platforms. Search terms were combined across five domains: (a) procedural fluency, computational fluency, mathematical operations, calculation, and symbolic manipulation; (b) conceptual knowledge, procedural knowledge, and mathematical equivalence; (c) procedural flexibility, strategy flexibility, and adaptive expertise; (d) arithmetic or algebra errors, misconceptions, working memory, and mathematics anxiety; and (e) comparison, worked examples, incorrect examples, productive failure, feedback, digital tools, and artificial intelligence in mathematics education. Reference lists of foundational reviews and recent articles were used to locate additional studies.

Sources were included when they made a clear conceptual, empirical, methodological, or policy contribution to understanding mathematical procedures in school mathematics. Priority was given to peer-reviewed journal articles, research syntheses, major scholarly books, and authoritative curriculum or policy frameworks. Studies were excluded when computation referred mainly to computer science rather than mathematical learning, when they did not address learning or teaching, or when they offered only unsupported classroom opinion. Foundational publications were retained where they established constructs still used in current research, while recent work was added to capture developments through 2026. In total, 44 sources met the inclusion criteria described above and form the reference list for this review. Because the review followed a narrative rather than a systematic search protocol, records identified and excluded at each screening stage were not tallied in a PRISMA-style count; this is acknowledged as a methodological limitation in Section 4.4. All searching, screening, and inclusion decisions in this review were carried out by the author alone; no second screener was involved, and full search strings, per-database record counts, and duplicate-removal figures were not retained in a form suitable for tabular reporting. This single-screener process is acknowledged as a limitation in Section 4.4, and future systematic updates of this synthesis should log search strings, per-source yields, and inter-screener agreement in line with reporting standards for transparent evidence synthesis.

2.3 Analytical procedure

The synthesis proceeded in three stages. First, constructs were compared according to their definitions, indicators, and implied learning outcomes. Second, empirical findings were coded into four substantive themes corresponding to the research questions: conceptualization, relations with broader competence, sources of errors, and instruction. Third, the themes were integrated into a proposed framework and research agenda. Because the corpus includes theoretical papers, experiments, longitudinal studies, meta-analyses, policy documents, and classroom research, the review emphasizes convergence across methods rather than treating all evidence as methodologically identical. Convergence was judged qualitatively: a claim was treated as well supported when it was reported consistently across at least two methodologically distinct sources (for example, an experimental or quasi-experimental study together with a meta-analysis, theoretical account, or classroom-based study), and evidence from meta-analyses and multi-study syntheses was weighted more heavily than single qualitative or theoretical accounts when the two were in tension. Sections 5 and 6 flag claims resting on a single study or source so that readers can judge the strength of the underlying evidence. The synthesis in Sections 3 through 6 draws on 44 sources spanning theoretical papers, cognitive and classroom experiments, meta-analyses, longitudinal studies, and policy or curriculum documents; a full tabulated study-characteristics matrix (author, year, country, participants, grade level, mathematical topic, design, construct, and finding) was judged impractical given this methodological heterogeneity and is instead recommended as a next step for a follow-up systematic mapping review (Section 4.4, Table 3). Within the narrative, evidence strength is signaled qualitatively throughout: findings drawn from meta-analyses (e.g., Peng et al., 2016; Namkung et al., 2019; Barroso et al., 2021) are treated as more robust than findings drawn from single experimental, correlational, or theoretical sources, and the instructional claims in Section 3.4 are qualified accordingly.

3. Results

3.1 Conceptualizing Mathematical Operation Competence

3.1.1 Beyond calculation and algorithm execution

The first conceptual problem is terminological. Calculation usually refers to obtaining a numerical or symbolic result through a sequence of steps. Procedural knowledge refers to knowledge of actions and rules, whereas procedural fluency highlights accurate, efficient, and flexible execution. Procedural flexibility concerns knowledge of multiple procedures and the ability to select or adapt them according to problem characteristics. Adaptive expertise extends this idea by combining efficient performance in familiar situations with innovation and transfer in unfamiliar ones (Star, 2005; Verschaffel et al., 2009; Verschaffel, 2024). These constructs overlap, but they are not interchangeable.

A narrow view of operation competence treats the procedure as an externally supplied algorithm: identify the task type, recall the rule, and execute it without error. A deeper view treats the procedure as an object of understanding. Students should know why the procedure is valid, how it connects to representations and concepts, when it is efficient, when another method would be preferable, and how to verify the outcome. Star (2005) therefore distinguished superficial procedural knowledge from deep procedural knowledge. Deep procedural knowledge includes connectedness, generality, and flexibility, not merely the ability to reproduce steps. This distinction has been reinforced and refined in subsequent work: Schneider, Rittle-Johnson, and Star (2011) showed that the depth of procedural knowledge, indexed by its connectedness to conceptual knowledge and flexibility, varies systematically with learners' prior knowledge, providing empirical support for treating procedural depth as a measurable dimension rather than a purely theoretical distinction.

International Journal of
Didactic Mathematics in Distance Education
https://jurnal.ut.ac.id/index.php/ijdmde
IJDMDE
LPPM UT
LEMBAGA PENELITIAN DAN PENGABDIAN KEPADA MASYARAKAT
UNIVERSITAS TERBUKA

The process-object perspective also helps explain why operations cannot be separated from concepts. Sfard (1991) argued that mathematical ideas may be understood operationally as processes and structurally as objects. For example, an algebraic expression can be viewed as an instruction to calculate or as a mathematical object with properties that can be transformed and compared. Competent operation depends on coordinating these views. Students who see only the process may execute steps without recognizing structure; students who see only the structure may understand a relation but lack reliable means of working with it.

Table 1

Related constructs and their distinguishing features

Construct	Core meaning	Distinguishing feature
Calculation skill	Execution of numerical or symbolic steps to obtain an answer.	Usually emphasizes correctness and speed.
Procedural knowledge	Knowledge of rules, algorithms, and action sequences.	May be superficial or deeply connected to concepts.
Procedural fluency	Accurate, efficient, flexible, and appropriate use of procedures.	Balances reliability with understanding and choice.
Procedural flexibility	Knowledge of multiple procedures and capacity to choose among them.	Requires sensitivity to problem features and goals.
Adaptive expertise	Efficient routine performance combined with innovation and transfer.	Extends beyond familiar tasks to novel situations.
Mathematical operation competence	Understanding operation objects and structures; selecting, executing, monitoring, justifying, and evaluating procedures.	Integrates conceptual grounding, fluency, flexibility, monitoring, and transfer across mathematical domains.

Because mathematical operation competence is the integrative construct proposed by this review rather than an independent, pre-existing category, its inclusion in Table 1 above is illustrative rather than a claim of empirical equivalence with the other constructs listed: it shows how existing constructs map onto the six dimensions defined in Section 3.1.3, not that it has already been validated as a separately measured construct alongside them.

3.1.2 The iterative relation between conceptual and procedural knowledge

A persistent debate in mathematics education asks whether conceptual understanding should precede procedural competence or whether practice with procedures can generate conceptual insight. Evidence reviewed here, although not assessed through a systematic quality-appraisal procedure, consistently argues against a

International Journal of
Didactic Mathematics in Distance Education
https://jurnal.ut.ac.id/index.php/ijdmde

simple one-way sequence. Rittle-Johnson, Siegler, and Alibali (2001) showed that conceptual and procedural knowledge can develop iteratively: initial conceptual knowledge supports procedural gains, and procedural gains can subsequently support conceptual development. Schneider et al. (2011) similarly found that the relations among conceptual knowledge, procedural knowledge, and procedural flexibility depend on prior knowledge and become more integrated through learning. Rittle-Johnson et al. (2015) later formalized this pattern explicitly as a "bidirectional street," showing across multiple domains that conceptual and procedural knowledge each predict gains in the other rather than one uniformly preceding the other.

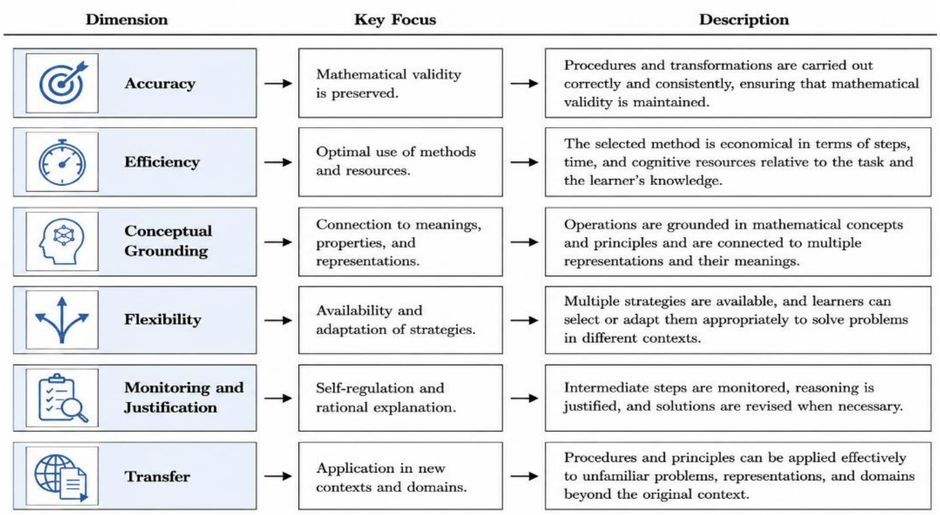
This iterative account has important implications. Conceptual explanations alone do not guarantee fluent performance, because procedures must be practiced, organized, and retrieved under task demands. Conversely, practice alone may stabilize errors or produce brittle knowledge if students do not understand the represented quantities and transformations. Baroody, Feil, and Johnson (2007) therefore argued against treating conceptual and procedural knowledge as independent or opposing categories. The educational target is coordinated knowledge in which meaning guides action and action refines meaning.

3.1.3 A multidimensional definition

Drawing on these traditions, this review defines mathematical operation competence as the capacity to identify the mathematical objects and structures involved in a task; select, execute, and adapt valid procedures; monitor and justify intermediate transformations; and interpret and evaluate the result in relation to the original problem. Six dimensions are embedded in this definition (See Figure 2).

Figure 2

Dimensions of Mathematical Operation Competence



Dimension	Key Focus	Description
Accuracy	Mathematical validity is preserved.	Procedures and transformations are carried out correctly and consistently, ensuring that mathematical validity is maintained.
Efficiency	Optimal use of methods and resources.	The selected method is economical in terms of steps, time, and cognitive resources relative to the task and the learner's knowledge.
Conceptual Grounding	Connection to meanings, properties, and representations.	Operations are grounded in mathematical concepts and principles and are connected to multiple representations and their meanings.
Flexibility	Availability and adaptation of strategies.	Multiple strategies are available, and learners can select or adapt them appropriately to solve problems in different contexts.
Monitoring and Justification	Self-regulation and rational explanation.	Intermediate steps are monitored, reasoning is justified, and solutions are revised when necessary.
Transfer	Application in new contexts and domains.	Procedures and principles can be applied effectively to unfamiliar problems, representations, and domains beyond the original context.

Each dimension can, in principle, be indexed by observable evidence: accuracy and efficiency through correctness and time- or step-count on comparable tasks; conceptual grounding through justification and representation-translation tasks; flexibility through multiple-strategy-use and strategy-choice measures; monitoring and justification through verification behavior, error-detection tasks, and think-aloud or written explanation; and transfer through performance on structurally related but superficially novel problems. Section 4.4 identifies the construction and validation of such a multidimensional assessment framework as a priority for future research, since no existing instrument currently operationalizes all six dimensions jointly.

This multidimensional definition avoids two extremes. It does not reduce operation competence to mechanical performance, but it also does not marginalize reliable execution. Mathematical thinking requires both. A student may have a sophisticated idea but fail because of unstable symbol manipulation; another may calculate accurately yet be unable to explain, choose, or transfer the method. Both profiles indicate incomplete competence. This raises a direct question about added value: what does mathematical operation competence contribute beyond existing constructs such as the five-strand proficiency model, NCTM's procedural fluency, adaptive expertise, and metacognitive self-regulation? The dimensions of accuracy, efficiency, and conceptual grounding are not new; they closely parallel procedural fluency as defined by NCTM (2014) and the fluency-understanding strands of Adding It Up (National Research Council, 2001). Flexibility and transfer are similarly borrowed directly from procedural-flexibility and adaptive-expertise research (Star, 2005; Verschaffel et al., 2009; Verschaffel, 2024). The contribution of mathematical operation competence is therefore not the invention of new dimensions but their integration into a single, curriculum-anchored construct: whereas the five-strand model treats procedural fluency, conceptual understanding, and strategic competence as co-equal but separately labeled strands, and adaptive-expertise research typically emphasizes strategy choice and innovation without formally incorporating monitoring and justification as a distinct dimension, the present definition packages accuracy, efficiency, conceptual grounding, flexibility, monitoring and justification, and transfer under a single measurable construct that corresponds directly to the operation-competence language used in curriculum documents such as the Chinese senior-secondary standards (Ministry of Education of the People's Republic of China, 2020). This integrative move is the review's central conceptual contribution, offered as a bridge between curriculum language and established research constructs rather than as a claim of an entirely new psychological process; Table 1 is intended to be read alongside this clarification.

3.2 Relations with Broader Mathematical Competencies

3.2.1 Reasoning, representation, and abstraction

Operations are inseparable from mathematical reasoning. Each legal transformation of an equation, expression, function, or representation depends on a warrant: an equality property, operation law, definition, theorem, or invariant. The written chain of equalities is therefore not merely a record of calculation; it is a compressed argument. When students cannot state why a transformation is valid, an apparently correct solution may reflect memorized pattern matching rather than reasoning.

Representation is equally central. Operations act on represented objects, and the form of representation influences which procedures become visible or efficient. Fractions, decimals, diagrams, tables, graphs, symbolic expressions, and digital representations highlight different structures. The process-object duality described by Sfard (1991) suggests that learners must move between acting on representations and treating the results as mathematical objects. Such coordination supports abstraction because students identify properties that remain invariant across specific calculations.

3.2.2 Problem solving, modeling, and strategic competence

In routine exercises, the operation is often specified implicitly by the task format. In genuine problem solving, selecting and organizing operations is part of the problem. Students must formulate a representation, decide which quantities and relations matter, choose an operation sequence, and interpret the output. The five-strand proficiency model therefore places procedural fluency in interaction with strategic competence and adaptive reasoning (National Research Council, 2001). Operation competence is a means of solving problems, but problem solving also develops operation competence by creating reasons to compare, adapt, and justify procedures.

The distinction between routine and non-routine tasks is important. Routine performance can be efficient without being adaptive. Non-routine tasks require students to reorganize known procedures, coordinate representations, or construct a new route. Recent work on teacher knowledge for mathematical problem solving emphasizes maintaining cognitive demand, eliciting students' ideas, and responding to emerging misunderstandings rather than reducing complex tasks to prescribed steps (Lit et al., 2026). These teacher practices are also necessary for operation-focused lessons when the aim is flexible competence rather than answer production.

3.2.3 Computational thinking and digital tools

Digital tools complicate the meaning of operational competence. Calculators, computer algebra systems, dynamic mathematics environments, and generative AI can execute procedures that previously occupied substantial instructional time. This does not make operation competence obsolete. Instead, it shifts emphasis toward choosing representations, specifying valid inputs, interpreting outputs, detecting implausible results, and deciding when manual work is mathematically or pedagogically valuable. The PISA 2022 framework explicitly recognizes computational thinking as part of mathematical literacy in a technology-rich world (OECD, 2023).Recent empirical work substantiates this policy-level claim: a 2026 meta-analysis of 22 empirical studies published between 2023 and 2025 found a moderate positive average effect of generative-AI-assisted instruction on students' mathematics learning outcomes ($g = 0.534$), with effects strongest when GenAI was used for creative transformation of tasks and in geometry learning (Liu et al., 2026).

Technology can support conceptual development by making multiple representations and patterns available, but its effects depend on task design, teacher orchestration, and the relationship between tool techniques and mathematical techniques (Drijvers, 2013; Kieran & Drijvers, 2006); a more recent survey of the field confirms that this orchestration challenge has only intensified as computer algebra systems, dynamic geometry packages, and now generative AI systems have converged in the same classroom (Engelbrecht & Borba, 2024). Uncritical use can conceal misconceptions or

International Journal of
Didactic Mathematics in Distance Education
https://jurnal.ut.ac.id/index.php/ijdmde
IJDMDE
LPPM UT
LEMBAGA PENELITIAN DAN PENGABDIAN KEPADA MASYARAKAT
UNIVERSITAS TERBUKA

encourage students to accept output without validation. The same concern applies more strongly to generative AI, which may produce plausible but invalid steps. UNESCO (2023) recommends human oversight, transparency, and age-appropriate pedagogical design. For operation competence, the central question is therefore not whether students should use tools, but how tool use can preserve reasoning, monitoring, and accountability. Direct experimental evidence supports this concern: in a large-scale field experiment with high-school mathematics students, unrestricted access to a generative-AI tutor improved practice-session performance but reduced subsequent unaided performance once AI access was removed, indicating weaker skill acquisition; prompting the tutor to give hints rather than final answers substantially mitigated this effect (Bastani et al., 2025).

3.3 Why Students Make Persistent Operational Errors

3.3.1 Conceptual and structural sources

Many operational errors are systematic rather than random. Students apply an available rule to an inappropriate structure, overgeneralize a familiar pattern, or treat symbols according to surface appearance. Algebraic errors such as combining unlike terms, distributing incompletely, canceling across addition, or altering an equation asymmetrically reveal underlying conceptions of variables, fractions, operations, and equivalence (Booth et al., 2014). These errors may persist because a procedure appears locally reasonable and occasionally produces correct answers.

Understanding the equal sign illustrates the connection between concept and operation. Students who interpret ‘=‘ only as a signal to calculate rather than a relation between equivalent quantities are less successful in equation solving (Knuth et al., 2006). The operational error is therefore rooted in a relational misconception, and this predictive link is durable: a longitudinal study found that equal-sign knowledge measured in second grade predicted algebra competence two years later, net of general mathematics ability (Matthews & Fuchs, 2020). Recent reviews of algebra education likewise identify weak arithmetic foundations, limited symbolic fluency, abstraction difficulties, and shifts between operational and structural thinking as interconnected sources of difficulty (Klee-Schramm et al., 2026).

3.3.2 Cognitive resources and task demands

Even when students possess relevant knowledge, performance depends on cognitive resources. Multi-step procedures require learners to hold intermediate results, coordinate goals, inhibit irrelevant responses, and monitor symbols. A meta-analysis found a moderate positive relation between working memory and mathematics performance across mathematical domains and learner groups (Peng et al., 2016). High element interactivity can overload working memory, especially when notation is unfamiliar or multiple transformations must be coordinated. Cognitive load theory therefore predicts that instruction should reduce unnecessary load while preserving the mathematical relationships students need to learn (Sweller et al., 2019).

Operational automaticity can free working-memory resources, but speed should not become the sole goal. Overemphasis on timed performance may increase pressure and encourage guessing or premature rule application. Fluency is better understood as reliable access to facts and procedures combined with strategic control. Students need enough practice for efficient execution and enough variation to prevent procedures from becoming tied to one superficial format.

3.3.3 Affective and motivational sources

Mathematics anxiety is consistently associated with lower mathematics achievement. Meta-analyses show a small-to-moderate negative relation across age groups and assessment types, with stronger relations for complex and multi-step tasks in some studies (Barroso et al., 2021; Namkung et al., 2019). Anxiety can reduce available working-memory capacity, increase avoidance, and narrow attention toward threat or time pressure (Ashcraft & Kirk, 2001); a more recent meta-analysis confirms that working memory mediates this relationship, with tasks demanding higher executive control showing the strongest anxiety-related impairment (Finell et al., 2022). Repeated operational failure can also produce anxiety, creating a reciprocal cycle in which weak performance and negative affect reinforce each other (Dowker et al., 2016).

Affective factors should not be used to explain away conceptual problems. A student may be anxious because prior instruction has made mathematics unpredictable, error-intolerant, or excessively timed. Conversely, a supportive environment will not correct misconceptions without mathematically focused teaching. Effective diagnosis considers affect, cognition, and knowledge together.

3.3.4 Instructional sources

Instruction contributes to errors when it emphasizes imitation without meaning, presents only one canonical method, moves quickly past foundational notation, or treats correct answers as sufficient evidence of understanding. Teachers may also inadvertently reinforce misconceptions by using language such as ‘move it to the other side and change the sign’ without connecting the action to equivalent transformations. When students’ written steps are not examined, persistent errors remain invisible until high-stakes assessment.

Teacher knowledge is therefore central. Mathematical knowledge for teaching includes the capacity to interpret nonstandard methods, identify the mathematical source of an error, select representations, and explain why procedures work (Ball et al., 2008). Empirical evidence links teachers’ mathematical knowledge for teaching to student achievement gains (Hill et al., 2005), a relationship a recent meta-analysis of mathematics and science teachers confirms at a broader scale, finding pedagogical content knowledge positively associated with teaching quality and student achievement across 56 studies (Fukaya et al., 2025). Operation-focused instruction requires teachers to notice not only whether a student is wrong, but what mathematical conception could have produced the response.

Table 2

A diagnostic framework for operational errors

Source of difficulty	Typical evidence	Instructional response
Conceptual/structural	Systematic misuse of equality, variables, fractions, signs, or operation properties.	Elicit explanations; connect procedures to definitions, properties, and multiple representations.
Procedural	Omitted steps, unstable algorithms, inaccurate fact retrieval, or inefficient method selection.	Use guided practice, worked examples, retrieval, spacing, and gradual variation.
Strategic/metacognitive	Inability to choose a method, check reasonableness, or revise after an error.	Compare methods; require estimation, self-explanation, checking, and reflection.
Cognitive-load related	Errors increase with notation density, long sequences, or simultaneous representations.	Segment tasks, make structure visible, fade scaffolds, and build prerequisite fluency.
Affective/motivational	Avoidance, rushing, freezing, or reduced persistence under pressure.	Reduce unnecessary time pressure; normalize error analysis; build attainable success and agency.
Instructional/technological	Dependence on templates or tools; inability to explain or validate output.	Use formative assessment, teacher noticing, and explicit verification of calculator, CAS, or AI output. (Kieran & Drijvers, 2006; UNESCO, 2023; Bastani et al., 2025)

3.4 Instruction for Mathematical Operation Competence

3.4.1 Build fluency from meaning, not instead of meaning

The most defensible instructional principle is integration. Conceptual understanding and procedural fluency should be developed together, with the balance changing according to the learner's prior knowledge and the mathematical topic (National Research Council, 2001; NCTM, 2014). Teachers should make explicit the objects being operated on, the properties that authorize transformations, and the relations preserved by a procedure. Explanations should then be followed by sufficient practice for students to achieve reliable execution. Throughout this review, evaluative language such as 'the strongest evidence' or 'the most strongly supported approach' should be read as summarizing the general direction and consistency of the cited studies rather than as the output of a formal, systematic quality-appraisal or grading procedure, since the review did not conduct such an assessment (Section 2.1). A structured evidence-strength appraisal, for example classifying each instructional recommendation in Sections 3.4.1 through 6.6 by design type (experimental, quasi-experimental, correlational, qualitative, theoretical, or meta-analytic) and consistency across studies, is identified as a priority methodological extension for future systematic updates of this synthesis (Section 4.4).

This principle rejects both discovery without consolidation and practice without sense making. Some procedures benefit from explicit modeling, particularly when novice learners lack prerequisite knowledge. However, modeling should reveal decision points and mathematical warrants rather than present an opaque sequence. Scaffolds can be gradually faded as students assume responsibility for selecting, explaining, and checking methods.

3.4.2 Compare procedures and cultivate flexibility

Comparison is an approach for which experimental evidence is comparatively well replicated, though this review did not apply a systematic quality-appraisal procedure to grade that evidence. When students compare different solution methods for the same problem, they are more likely to notice structural similarities and differences, evaluate efficiency, and connect procedures to concepts. Experimental studies of equation solving found benefits for conceptual knowledge and procedural flexibility when students compared solution methods, although effects depend on prior knowledge and the design of the comparison (Rittle-Johnson & Star, 2007, 2009; Rittle-Johnson et al., 2009).

Comparison should be carefully orchestrated. Presenting many methods without a clear purpose may overload novices or communicate that all approaches are equally useful in every situation. Teachers need to select contrasting methods, ask students to identify the mathematical idea underlying each, and discuss conditions under which one method is more efficient. Flexibility involves both potential flexibility, knowing alternatives, and practical flexibility, using them adaptively. Cross-national evidence indicates that students' procedural flexibility varies by opportunity to learn and classroom norms, not only by individual ability (Star et al., 2022).

3.4.3 Use deliberate, spaced, and varied practice

Practice is necessary for fluency, but its design matters. Repetition of one problem type in a single block can improve immediate performance while encouraging superficial cue-based method selection. Spaced and interleaved practice asks students to retrieve procedures after delays and discriminate among problem types. Such practice should begin after students possess enough understanding to avoid stabilizing misconceptions. Variation should change mathematically relevant features while preserving opportunities to recognize invariant structure.

Practice should also include estimation and reasonableness checks. A student who can execute an algorithm but cannot judge whether the answer is plausible lacks monitoring competence. Requiring students to predict answer size, identify likely sign, or use an alternative representation can convert checking from an optional final step into part of the procedure itself. This aligns with evidence that structured self-monitoring and error-locating prompts improve diagnostic habits and learning gains more than outcome-only feedback (Hattie & Timperley, 2007).

3.4.4 Learn from errors and productive struggle

Incorrect examples can be productive when students are asked to explain the error, compare it with a correct solution, and revise the reasoning. Durkin and Rittle-Johnson (2012) found that comparing incorrect and correct examples supported learning about decimal magnitude. Error analysis directs attention to boundaries of applicability and can expose misconceptions that ordinary correct-example practice leaves untouched. A subsequent meta-analysis of worked-examples research corroborates this pattern at a broader scale, reporting a medium overall effect on mathematics performance when correct examples are paired with erroneous ones, though correct examples alone yield the largest gains (Barbieri et al., 2023).

Productive failure extends this idea by allowing students to attempt complex problems before formal instruction, followed by carefully designed consolidation. Kapur (2008) showed that initial failure can prepare learners to understand canonical solutions more deeply when the subsequent teaching connects student-generated ideas to target concepts. Productive struggle is not equivalent to leaving students unsupported. Its effectiveness depends on task accessibility, teacher monitoring, and a consolidation phase that makes the relevant mathematical structures explicit.

3.4.5 Provide formative feedback and develop self-monitoring

Feedback is most useful when it reduces the gap between current reasoning and a clear learning goal, rather than simply labeling an answer correct or incorrect (Hattie & Timperley, 2007). For mathematical operations, feedback can address the selected method, the validity of a transformation, notation, efficiency, or checking. Students should be prompted to locate the first invalid step, explain why it is invalid, and repair the solution. This approach preserves agency and develops diagnostic habits.

Teachers should also collect evidence from intermediate work. Final answers hide strategy and may conceal compensating errors. Think-aloud explanations, annotated solutions, comparison tasks, and short diagnostic prompts reveal how students interpret operation objects and rules. Such evidence supports teacher noticing and enables instruction to respond to emerging misunderstandings (Lit et al., 2026).

3.4.6 Integrate digital tools and generative AI with verification routines

Digital tools can reduce routine burden and create opportunities for exploration, but they should be used to extend rather than replace mathematical judgment. Productive tasks may ask students to predict a result before using a tool, compare manual and digital methods, explain discrepancies, identify hidden assumptions, or test whether a generated solution remains valid under changed conditions. Computer algebra systems can make structure visible, yet students need to connect machine techniques with paper-and-pencil techniques and theoretical reasoning (Kieran & Drijvers, 2006).

Generative AI adds a new form of procedural outsourcing. Because AI systems can produce complete solutions with fluent explanations, students may appear competent without selecting, executing, or monitoring any operation. Instruction should therefore include verification protocols: restate the problem independently, inspect each transformation, test the result, request an alternative method, and explain why the final answer is reasonable. Research should examine when AI feedback supports metacognition and when it produces dependency or disengagement. Human oversight and transparent pedagogical purpose remain essential (UNESCO, 2023). Direct evidence of this dynamic comes from a randomized field experiment in which high-school students who used an unguarded generative-AI tutor performed worse on unaided follow-up assessments than peers who did not, despite equal or better performance while AI assistance was available, consistent with procedural outsourcing rather than genuine skill acquisition (Bastani et al., 2025).

4. Discussion

This integrative model carries an epistemic trade-off common to multi-system explanatory frameworks: because it allows almost any observed error to be redescribed as a misalignment among the four systems, it functions better as a diagnostic heuristic for practice than as a strictly falsifiable causal theory. The alternative of treating operational errors as a single-cause phenomenon, attributable only to carelessness, insufficient drill, or low ability, has, however, demonstrably worse instructional consequences, since it directs teachers toward a narrow remedy of additional repetition regardless of the actual source of difficulty. The model proposed here is therefore defended on practical rather than purely explanatory grounds, and Section 4.4 calls explicitly for confirmatory measurement work capable of testing its structure empirically.

4.1 An integrative model

The reviewed literature supports an integrative model in which mathematical operation competence emerges from interactions among four systems. The mathematical-knowledge system includes concepts, properties, representations, facts, and procedures. The strategic-control system includes method selection, monitoring, explanation, and revision. The learner-resource system includes working memory, prior knowledge, anxiety, motivation, and beliefs. The instructional-ecological system includes tasks, teacher knowledge, classroom norms, feedback, assessment, and technological tools. Operational performance at any moment reflects the alignment—or misalignment, among these systems.

This model explains why the same observable error can have different causes. An incorrect sign may result from a misconception, a working-memory lapse, rushed responding under anxiety, or an unexamined rule taught through shorthand. It also explains why a single intervention rarely addresses all students. Additional practice may help a learner with unstable fact retrieval but reinforce the problem for a learner who holds a structural misconception. Effective teaching begins with diagnosis and then selects an intervention matched to the source of difficulty.

4.2 Reframing the relation between skill and thinking

The evidence does not support a choice between operational skill and mathematical thinking. Reliable procedures can support higher-order reasoning by reducing cognitive burden, while conceptual understanding and reasoning make procedures more memorable, adaptable, and resistant to error. The central educational distinction is not skills versus understanding, but routine expertise versus adaptive competence. Routine expertise is valuable for familiar tasks; adaptive competence adds flexibility, explanation, monitoring, and transfer.

This reframing is important for assessment. Conventional tests often reward the correct final answer and provide little evidence about method choice or reasoning. A fuller assessment of operation competence should sample multiple dimensions through tasks that ask students to execute, compare, explain, diagnose, estimate, adapt, and verify. Measures should distinguish knowledge of alternatives from actual adaptive use and should include numerical, algebraic, graphical, and technology-mediated contexts.

4.3 Implications for curriculum and teacher education

Curriculum documents should define operation competence with observable dimensions and learning progressions. Lists of verbs such as understand, select, design, and calculate are useful but insufficient unless accompanied by examples of increasing sophistication. Progressions might move from accurate execution of a supported procedure, to explanation of its basis, to comparison of alternatives, to independent selection and adaptation in unfamiliar contexts. In Indonesia, the Kurikulum Merdeka's learning outcomes (Capaian Pembelajaran) for mathematics likewise foreground a numeracy competency (numerasi) that combines computational accuracy with reasoning and problem solving (Badan Standar, Kurikulum, dan Asesmen Pendidikan [BSKAP], Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia, 2022); the multidimensional definition proposed in Section 3.1.3 could equally inform how observable indicators of this competency are specified in Indonesian curriculum and teacher-training materials, extending the review's relevance for IJDMDE's Indonesia-based readership beyond the Chinese case discussed above.

Teacher education should include analysis of student work, common misconceptions, multiple procedures, and technology-generated solutions. Teachers need specialized knowledge to determine whether a nonstandard method is valid, why a misconception is attractive, and which representation can expose the underlying relation. Professional learning should therefore connect content knowledge, pedagogical content knowledge, noticing, and formative assessment rather than treating operation teaching as elementary classroom management.

4.4 Limitations and Future Research

This review has limitations. It is a structured narrative review rather than a database-exhaustive systematic review, and its synthesis is shaped by the terminology used in English-language research. The umbrella construct mathematical operation competence is not standardized internationally, so some relevant studies may be located under topic-specific labels such as arithmetic, algebraic manipulation, procedural fluency, or strategy use. The evidence base also contains more research on elementary arithmetic and equation solving than on advanced secondary topics such as functions, trigonometry, calculus, probability, and statistics. In total, 44 sources met the inclusion criteria and are cited throughout this review. Because selection and screening were conducted by a single author rather than by two independent reviewers, the corpus may reflect unintentional selection bias despite the explicit search domains and criteria specified in Section 2.2; this constraint is acknowledged in line with the transparency practices recommended by DOAJ and the Committee on Publication Ethics (COPE), and future updates of this synthesis should incorporate independent dual screening and documented inter-rater agreement. This concentration should be read as a boundary condition on the review's claims rather than as evidence that the six proposed dimensions transfer uniformly across mathematical domains. Operation competence plausibly takes a different shape in, for instance, geometric proof and construction, where operations act on spatial and deductive structures rather than symbolic expressions, or in statistical reasoning, where procedures often involve probabilistic judgment under uncertainty rather than deterministic transformation. The definition and dimensions proposed in Section 3.1.3 should therefore be treated as a starting point requiring domain-specific validation, a priority explicitly identified in the research agenda (Table 3).

These limitations point to a research agenda that can make the construct empirically stronger.

Table 3
Priority research agenda

Research gap	Recommended design	Expected contribution
Construct ambiguity	Delphi studies, cognitive interviews, and confirmatory measurement models across topics and grades.	A validated multidimensional definition and assessment framework.
Development over time	Longitudinal studies linking conceptual knowledge, fluency, flexibility, affect, and transfer.	Evidence on reciprocal and developmental relations.
Limited causal evidence	Classroom randomized or strong quasi-experimental studies of comparison, error analysis, spacing, and feedback.	Identification of effective instructional components and moderators.
Cross-cultural variation	Common tasks and measures across curriculum systems, languages, and achievement levels.	Understanding of how opportunity to learn and classroom norms shape flexibility.
Teacher enactment	Video-based studies combining teacher knowledge, noticing, task design, and student outcomes.	Mechanisms linking teacher expertise to operation competence.

Research gap	Recommended design	Expected contribution
Technology and AI	Experiments comparing unrestricted, scaffolded, and verification-oriented tool use.	Evidence on learning, monitoring, dependency, and transfer in human-AI activity.
Equity and learner diversity	Inclusive designs involving multilingual learners, students with mathematics difficulties, and diverse socioeconomic contexts.	More equitable theory, assessment, and intervention.

5. Conclusion

Mathematical operation competence should no longer be treated as a synonym for fast and accurate calculation, and the generative AI era makes that reduction untenable rather than merely outdated: a tool that can already execute the calculation forces the construct to be redefined around what a student must still know, decide, and verify. It is an adaptive form of mathematical competence that integrates conceptual meaning, procedural reliability, strategic choice, monitoring, justification, and transfer. Research consistently indicates that conceptual and procedural knowledge develop in interaction, that flexibility requires both a repertoire of methods and sensitivity to task features, and that operational errors reflect more than carelessness. They emerge from the interaction of misconceptions, cognitive demands, affective experiences, instructional practices, and technological environments, including the newly consequential risk of procedural outsourcing to generative AI.

For instruction, the implication is not to reduce practice but to redesign it. Students need explicit attention to mathematical structure, sufficient practice for fluency, comparison of methods, opportunities to analyze errors, productive struggle followed by consolidation, formative feedback, and routines for verifying digital or AI-generated output. For research, the next step is to move from broad descriptions toward validated constructs, longitudinal models, causal classroom studies, cross-cultural comparisons, and evidence on teacher and AI mediation. Such work can position mathematical operations not as the mechanical residue of school mathematics, but as a central site where knowledge, reasoning, and adaptive problem solving are coordinated.

Acknowledgments

The author thanks the editors and the anonymous reviewers of the International Journal of Didactic Mathematics in Distance Education for their careful reading and constructive comments, which substantially improved the conceptual framing and clarity of this manuscript.

Author Contributions

Jiajie Yan is the sole author of this manuscript and is solely responsible for its conceptualization, literature search and selection, analysis and synthesis, writing of the original draft, and revision. The author has read and approved the final version of the manuscript.

Funding Statement
This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflict of Interest
The author declares no conflict of interest related to this manuscript.

Data Availability Statement
No new empirical data were generated or analyzed in this study. As a structured narrative review, this manuscript synthesizes previously published sources, all of which are identified in the reference list and are publicly accessible through their respective publishers or repositories.

AI Use Declaration
Generative artificial intelligence tools were used during manuscript preparation to assist with language editing, reference formatting, and identification of candidate literature for verification. All AI-assisted output was independently reviewed, fact-checked against original sources, and edited by the author, who takes full responsibility for the accuracy, originality, and academic integrity of the final manuscript. No generative AI tool is credited as an author, and no AI system was used to fabricate data, findings, or citations.

6. References

Ashcraft, M. H., & Kirk, E. P. (2001). The relationships among working memory, math anxiety, and performance. *Journal of Experimental Psychology: General, 130*(2), 224–237. https://doi.org/10.1037/0096-3445.130.2.224

Badan Standar, Kurikulum, dan Asesmen Pendidikan [BSKAP], Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia. (2022). Keputusan Kepala BSKAP No. 008/H/KR/2022 tentang Capaian Pembelajaran pada Pendidikan Anak Usia Dini, Jenjang Pendidikan Dasar, dan Pendidikan Menengah pada Kurikulum Merdeka. Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi.

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

Barbieri, C. A., Miller-Cotto, D., Clerjuste, S. N., & Chawla, K. (2023). A meta-analysis of the worked examples effect on mathematics performance. *Educational Psychology Review, 35*(1), 1–33. https://doi.org/10.1007/s10648-023-09745-1

Baroody, A. J., Feil, Y., & Johnson, A. R. (2007). An alternative reconceptualization of procedural and conceptual knowledge. *Journal for Research in Mathematics Education, 38*(2), 115–131.

Barroso, C., Ganley, C. M., McGraw, A. L., Geer, E. A., Hart, S. A., & Daucourt, M. C. (2021). A meta-analysis of the relation between math anxiety and math achievement. *Psychological Bulletin, 147*(2), 134–168. https://doi.org/10.1037/bul0000307

Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, Ö., & Mariman, R. (2025). Generative AI without guardrails can harm learning: Evidence from high school mathematics. *Proceedings of the National Academy of Sciences, 122*(26), e2422633122. https://doi.org/10.1073/pnas.2422633122

Booth, J. L., Barbieri, C., Eyer, F., & Paré-Blagoev, E. J. (2014). Persistent and pernicious errors in algebraic problem solving. *The Journal of Problem Solving, 7*(1), 10–23. https://doi.org/10.7771/1932-6246.1161

Dowker, A., Sarkar, A., & Looi, C. Y. (2016). Mathematics anxiety: What have we learned in 60 years? *Frontiers in Psychology*, 7, Article 508. https://doi.org/10.3389/fpsyg.2016.00508

Drijvers, P. (2013). Digital technology in mathematics education: Why it works (or doesn't). PNA, 8(1), 1–20.

Durkin, K., & Rittle-Johnson, B. (2012). The effectiveness of using incorrect examples to support learning about decimal magnitude. *Learning and Instruction, 22*(3), 206–214. https://doi.org/10.1016/j.learninstruc.2011.11.001

Engelbrecht, J., & Borba, M. C. (2024). Recent developments in using digital technology in mathematics education. *ZDM – Mathematics Education, 56*(2), 281–292. https://doi.org/10.1007/s11858-023-01530-2

Finell, J., Sammallahti, E., Korhonen, J., Eklöf, H., & Jonsson, B. (2022). Working memory and its mediating role on the relationship of math anxiety and math performance: A meta-analysis. *Frontiers in Psychology, 12*, Article 798090. https://doi.org/10.3389/fpsyg.2021.798090

Fukaya, T., Nakamura, D., Kitayama, Y., & Nakagoshi, T. (2025). A systematic review and meta-analysis of research on mathematics and science pedagogical content knowledge: Exploring its associations with teacher and student variables. *Teaching and Teacher Education, 155*, Article 104881. https://doi.org/10.1016/j.tate.2024.104881

Grant, M. J., & Booth, A. (2009). A typology of reviews: An analysis of 14 review types and associated methodologies. *Health Information & Libraries Journal, 26*(2), 91–108. https://doi.org/10.1111/j.1471-1842.2009.00848.x

Hattie, J., & Timperley, H. (2007). The power of feedback. Review of Educational Research, 77(1), 81–112. https://doi.org/10.3102/003465430298487

Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal, 42*(2), 371–406. https://doi.org/10.3102/00028312042002371

Kapur, M. (2008). Productive failure. *Cognition and Instruction, 26*(3), 379–424. https://doi.org/10.1080/07370000802212669

Kieran, C., & Drijvers, P. (2006). The co-emergence of machine techniques, paper-and-pencil techniques, and theoretical reflection: A study of CAS use in secondary school algebra. *International Journal of Computers for Mathematical Learning*, 11, 205–263. https://doi.org/10.1007/s10758-006-0006-7

Klee-Schramm, I., Peters, A., & Rott, B. (2026). Algebra education for secondary students with difficulties: A review of the literature from 2019 to 2023. *International Journal of Science and Mathematics Education, 24*, Article 55. https://doi.org/10.1007/s10763-026-10676-w

Knuth, E. J., Stephens, A. C., McNeil, N. M., & Alibali, M. W. (2006). Does understanding the equal sign matter? Evidence from solving equations. *Journal for Research in Mathematics Education, 37*(4), 297–312. https://doi.org/10.2307/30034852

Lit, S., Kool, M., & Drijvers, P. (2026). Teacher knowledge and skills for mathematical problem solving: A systematic literature review. Journal of Mathematics Teacher Education. Advance online publication. https://doi.org/10.1007/s10857-025-09737-8

Liu, B., Zhang, W., & Wang, F. (2026). Can generative artificial intelligence effectively enhance students' mathematics learning outcomes? A meta-analysis of empirical studies from 2023 to 2025. *Education Sciences, 16*(1), Article 140. https://doi.org/10.3390/educsci16010140

Matthews, P. G., & Fuchs, L. S. (2020). Keys to the gate? Equal sign knowledge at second grade predicts fourth-grade algebra competence. *Child Development, 91*(1), 14–28. https://doi.org/10.1111/cdev.13144

Ministry of Education of the People's Republic of China. (2020). General high school mathematics curriculum standards (2017 edition, 2020 revision). People's Education Press.

Namkung, J. M., Peng, P., & Lin, X. (2019). The relation between mathematics anxiety and mathematics performance among school-aged students: A meta-analysis. *Review of Educational Research, 89*(3), 459–496. https://doi.org/10.3102/0034654319843494

National Council of Teachers of Mathematics. (2014). Principles to actions: Ensuring mathematical success for all. Author.

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

OECD. (2023). PISA 2022 assessment and analytical framework. OECD Publishing. https://doi.org/10.1787/dfe0bf9c-en

Peng, P., Namkung, J., Barnes, M., & Sun, C. (2016). A meta-analysis of mathematics and working memory: Moderating effects of working memory domain, type of mathematics skill, and sample characteristics. *Journal of Educational Psychology, 108*(4), 455–473. https://doi.org/10.1037/edu0000079

Rittle-Johnson, B., Siegler, R. S., & Alibali, M. W. (2001). Developing conceptual understanding and procedural skill in mathematics: An iterative process. *Journal of Educational Psychology, 93*(2), 346–362. https://doi.org/10.1037/0022-0663.93.2.346

Rittle-Johnson, B., & Star, J. R. (2007). Does comparing solution methods facilitate conceptual and procedural knowledge? An experimental study on learning to solve equations. *Journal of Educational Psychology, 99*(3), 561–574. https://doi.org/10.1037/0022-0663.99.3.561

Rittle-Johnson, B., & Star, J. R. (2009). Compared with what? The effects of different comparisons on conceptual knowledge and procedural flexibility for equation solving. *Journal of Educational Psychology, 101*(3), 529–544. https://doi.org/10.1037/a0014224

Rittle-Johnson, B., Star, J. R., & Durkin, K. (2009). The importance of prior knowledge when comparing examples: Influences on conceptual and procedural knowledge of equation solving. *Journal of Educational Psychology, 101*(4), 836–852. https://doi.org/10.1037/a0016026

Rittle-Johnson, B., Schneider, M., & Star, J. R. (2015). Not a one-way street: Bidirectional relations between procedural and conceptual knowledge of mathematics. *Educational Psychology Review, 27*(4), 587–597. https://doi.org/10.1007/s10648-015-9302-x

Schneider, M., Rittle-Johnson, B., & Star, J. R. (2011). Relations among conceptual knowledge, procedural knowledge, and procedural flexibility in two samples differing in prior knowledge. Developmental Psychology, 47(6), 1525–1538. https://doi.org/10.1037/a0024997

Sfard, A. (1991). On the dual nature of mathematical conceptions: Reflections on processes and objects as different sides of the same coin. *Educational Studies in Mathematics, 22*(1), 1–36. https://doi.org/10.1007/BF00302715

Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. Journal of Business Research, 104, 333–339. https://doi.org/10.1016/j.jbusres.2019.07.039

Star, J. R. (2005). Reconceptualizing procedural knowledge. *Journal for Research in Mathematics Education, 36*(5), 404–411. https://doi.org/10.2307/30034943

Star, J. R., Tuomela, D., Joglar-Prieto, N., Hästö, P., Palkki, R., Abánades, M. Á., Pejlare, J., Jiang, R. H., Li, L., & Liu, R.-D. (2022). Exploring students' procedural flexibility in three countries. *International Journal of STEM Education*, 9, Article 4. https://doi.org/10.1186/s40594-021-00322-y

Sweller, J., van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review, 31*, 261–292. https://doi.org/10.1007/s10648-019-09465-5

UNESCO. (2023). Guidance for generative AI in education and research. UNESCO.

Verschaffel, L. (2024). Strategy flexibility in mathematics. *ZDM–Mathematics Education, 56*, 115–126. https://doi.org/10.1007/s11858-023-01491-6

Verschaffel, L., Luwel, K., Torbeyns, J., & Van Dooren, W. (2009). Conceptualizing, investigating, and enhancing adaptive expertise in elementary mathematics education. *European Journal of Psychology of Education, 24*(3), 335–359. https://doi.org/10.1007/BF03174765