Kalamatika: Jurnal Pendidikan Matematika

P-ISSN 2527-5615
E-ISSN 2527-5607

Journal homepage: https://kalamatika.matematika-uhamka.com/

Verification and reflection as metacognitive mechanisms in Polya-based problem-solving learning

Rahma Wahyu*, Tika Septia, Mazlini Adnan, and Supiat

To cite this entry:

Wahyu. R., Septia, T., Adnan, M., & Supiat. (2026). Verification and reflection as metacognitive mechanisms in Polya-based problem-solving learning. *Kalamatika: Jurnal Pendidikan Matematika*, *11*(2), 234-250. https://doi.org/10.22236/KALAMATIKA.vol11no2.2026pp234-250

Link to the article online:
https://kalamatika.matematika-uhamka.com/index.php/kmk/article/view/914

Submitted: *Jul 5, 2026* | Revised: *Sep 16, 2026* | Accepted: *Sep 18, 2026*
Published online: *November 22, 2026*

© 2026 The Author(s). Published by FKIP Universitas Muhammadiyah Prof. Dr. HAMKA in collaboration with Indonesian Mathematics Educators (IMES).

This work is an Open Access article licensed under a Creative Commons Attribution - ShareAlike 4.0 International License.

OPEN ACCESS

Kalamatika: Jurnal Pendidikan Matematika
Volume 11, No. 2, November 2026, pages 234-250
DOI: 10.22236/KALAMATIKA.vol11no2.2026pp234-250

Verification and reflection as metacognitive mechanisms in Polya-based problem-solving learning

Rahma Wahyu[1], Tika Septia[2], Mazlini Adnan[3], and Supiat[4]

[1]*Mathematics Education, Wisnuwardhana University, Malang, Indonesia*
[2]*Mechanical Engineering Technology Study, Surabaya Merchant Marine Polytechnic, Surabaya, Indonesia*
[3]*Mathematics Education, Universiti Pendidikan Sultan Idris, Perak, Malaysia*
[4]*Mathematics Education, Universitas Muhammadiyah Prof. DR. HAMKA, Jakarta, Indonesia*

ABSTRACT

Polya's "looking back" is a pivotal stage in mathematical problem solving. However, existing literature and classroom practices often reduce post-solution evaluation to an undifferentiated process of answer checking, failing to distinguish between verifying correctness and reflecting on problem-solving strategies. This conceptual article addresses this theoretical gap by operationalising the stage through two integrated metacognitive mechanisms: verification and reflection. Verification serves a product-oriented function, focusing on procedural accuracy, compliance with problem constraints, and contextual validity. In contrast, reflection serves a process-oriented function, involving the evaluation of strategy choices, comparison of alternative representations, and abstraction of principles for transfer to new situations. Drawing on 38 foundational and contemporary studies on mathematical problem solving, metacognition, and assessment, this article operationalises the proposed framework as a multi-tiered toolkit comprising observable indicators, an analytical rubric, and targeted instructional prompts. The proposed framework contributes to mathematics education by moving beyond binary correctness-based assessment and providing educators and researchers with explicit tools to foster and assess deeper metacognitive monitoring in problem-solving classrooms.

KEYWORDS

Metacognition; Polya framework; reflection; verification; look-back stage

CORRESPONDING AUTHOR

Rahma Wahyu ✉ rahmawahyu_pmat@wisnuwardhana.ac.id 🖃 Mathematics Education, Wisnuwardhana University, Malang, Indonesia

INTRODUCTION

In mathematical problem solving, producing a correct numerical answer represents only one dimension of the overall cognitive endeavour. A mathematical solution demonstrates greater instructional and conceptual depth when learners can articulate the underlying problem structure, justify the suitability of their chosen strategy, systematically test the result against the original constraints, and abstract the underlying principles for application to future problem-solving situations. Recent scholarship in mathematics education increasingly conceptualises problem solving not as a linear process of producing an answer, but as a complex, recursive activity involving reasoning, representation, self-monitoring, mathematical communication, and metacognitive control (English, 2023; Purnomo et al., 2022; Rott et al., 2021; Santos-trigo, 2024; Schoenfeld, 1992). Within this paradigm, evaluating the quality of a student's solution requires moving beyond binary judgements of

correctness to examine how learners monitor and review their work and extract mathematical meaning from their problem-solving processes.

George Pólya's four-stage problem-solving framework (Polya, 1957) provides a foundational and widely recognized structure for analysing how learners approach mathematical tasks. This framework divides the problem-solving process into four sequential yet interrelated stages: (1) understanding the problem, in which solvers identify the given conditions, unknowns, and constraints; (2) devising a plan, in which heuristics and strategies are formulated; (3) carrying out the plan, in which procedures, representations, and calculations are executed; and (4) looking back, in which the resulting solution and the overall problem-solving process are reviewed and evaluated.

Despite its theoretical prominence, the looking-back stage remains one of the most neglected and least visible phases in students' written work and typical classroom practices. Empirical investigations have documented that novice and intermediate problem solvers often terminate their cognitive engagement as soon as an answer is obtained, treating the task as complete without undertaking retrospective evaluation (Rott et al., 2021; Santos-Trigo, 2024; Schoenfeld, 1985). Even when post-solution checking is explicitly encouraged, instructional practices and assessment rubrics frequently reduce "looking back" to a limited procedural routine, such as briefly repeating a calculation, rather than treating it as a meaningful metacognitive endeavour (Rott et al., 2021; Santos-Trigo, 2024; Schoenfeld, 1985).

Although foundational literature on metacognition conceptualises post-performance activities through broad constructs such as cognitive monitoring, regulation, evaluation, and self-reflection (Brown, 1987; Flavell, 1979; Schoenfeld, 1985), these constructs are often treated at a relatively broad level. They do not always explicitly distinguish between the technical validation of mathematical products and the conceptual abstraction of the problem-solving process. In conventional metacognitive frameworks, the broad construct of "evaluation" can encompass both checking computational correctness and reflecting on structural representations, potentially obscuring the distinct cognitive operations involved.

This conceptual conflation obscures important qualitative differences in students' thinking. For example, empirical studies on mathematical modelling reveal that students may substitute numerical values back into formulas to confirm computational accuracy yet fail to interpret whether their algebraic models adequately account for the constraints of the real-world problems (Verschaffel et al., 1994; Wahyu, 2022; Wahyu, 2026). Conversely, qualitative think-aloud studies suggest that problem solvers may discuss heuristic methods or

Kalamatika: Jurnal Pendidikan Matematika
Volume 11, No. 2, November 2026, pages 234-250
DOI: 10.22236/KALAMATIKA.vol11no2.2026pp234-250

alternative representations in broad conceptual terms while neglecting to test whether their obtained values satisfy all the original boundary conditions (Goos et al., 2002; Rott et al., 2021)

To address this theoretical and instructional ambiguity, this article elaborates Pólya's fourth stage by operationalizing it through two distinct yet complementary metacognitive mechanisms: verification and reflection.

a. *Verification* is a product-oriented mechanism that systematically examines the validity of a solution by addressing the following questions: Is the obtained answer mathematically correct and procedurally accurate? Is it dimensionally consistent and aligned with all the original problem constraints?
b. *Reflection* is a process-oriented mechanism that examines conceptual meaning and the potential for transfer by addressing questions such as: Why was the selected strategy effective? What do alternative representations reveal? What structural insights can be generalised and applied to future mathematical challenges?

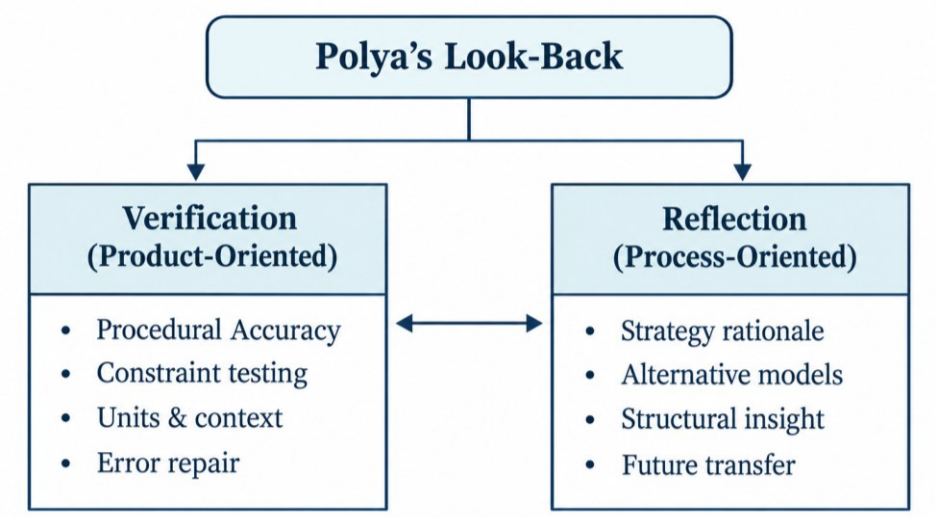


Figure 1. Elaboration of Polya's Look-Back Stage into Verification and Reflection

As illustrated in Figure 1, the looking-back stage functions as an interactive and recursive metacognitive space rather than a terminal, linear checkpoint. The primary aim of this article is not to replace Pólya's classical model but to refine its fourth stage into an actionable and fine-grained analytical framework that can inform research instruments, classroom prompts, and diagnostic assessment rubrics. Accordingly, the study addresses the following research question: How can verification and reflection be operationalized as complementary metacognitive mechanisms within Polya-based mathematical problem solving?

Kalamatika: Jurnal Pendidikan Matematika
Volume 11, No. 2, November 2026, pages 234-250
DOI: 10.22236/KALAMATIKA.vol11no2.2026pp234-250

To establish the theoretical foundation and justify this operationalization, Table 1 summarizes the positioning of the proposed framework in relation to existing strands of literature and highlights the specific research gaps addressed in this article.

Table 1. Literature Positioning and Research Gaps for The Verification-Reflection Elaboration

Research Strand	Representative Emphasis	Theoretical Implication	Research Gap & How This Paper Addresses It
Problem-Solving Processes	Problem solving involves recursive cycles of understanding, planning, monitoring, and validation rather than a linear answer-producing routine (Polya, 1957; Rott et al., 2021; Santos-Trigo, 2024).	The look-back stage must be analyzed as an active, multi-action metacognitive phase rather than an omitted or trivial step.	*Unresolved Gap*: Looking back is frequently skipped or treated as a passive, terminal step in problem solving. *Paper Contribution*: Unpacks the phase into distinct observable indicators that make student post-solution actions visible and measurable.
Metacognitive Architecture	Metacognition encompasses knowledge, monitoring, regulation, and evaluation of cognitive systems (Brown, 1987; Flavell, 1979; Schoenfeld, 1985).	Verification and reflection operate as distinct metacognitive functions in mathematical reasoning.	*Unresolved Gap*: Broad constructs of "evaluation" fail to distinguish procedural constraint-checking from conceptual strategy evaluation. *Paper Contribution*: Analytically decouples product validity (*verification*) from process meaning and transfer (*reflection*).
Mathematical Representation	Reasoning is mediated through symbolic, tabular, visual, and contextual representations (English, 2023; Kholid et al., 2024).	Post-solution evaluation must inspect coherence across diverse representational forms.	*Unresolved Gap*: Checking is predominantly confined to symbolic recalculations, neglecting visual or tabular validation. *Paper Contribution*: Integrates multi-representational comparison and constraint testing into operational indicators.
Instructional Assessment	Classroom assessment rubrics routinely collapse checking, explanation, and generalization into a single undifferentiated score.	Assessment tools require diagnostic granularity to support targeted instructional feedback.	*Unresolved Gap*: Single-score assessment rubrics fail to identify students who achieve correct answers without strategic understanding. *Paper Contribution*: Establishes a dual-axis diagnostic rubric (Table 5) and tiered classroom prompts (Table 6).

METHODS

Research Design and Methodological Framework

This article adopts a conceptual framework development design, following the systematic conceptual synthesis procedures outlined by Jabareen (2009) and Torraco (2005).

Rather than reporting data from a field-based empirical intervention, this study focuses on theoretical refinement and operationalisation. Conceptual framework development is particularly suited to resolving construct ambiguities, bridging fragmented theoretical perspectives, and producing practical tools—such as indicators, diagnostic rubrics, and instructional prompts—that can guide subsequent empirical investigations and classroom interventions.

Corpus Selection and Literature Synthesis Procedure

The framework was developed through a four-phase analytical process based on a systematically screened corpus of 38 foundational and contemporary peer-reviewed studies published between 1957 and 2025 and identified through searches of major international databases (Scopus, Web of Science, and ERIC). The literature searches combined Boolean operators targeting the core themes: ("Polya" OR "problem solving") AND ("metacognition" OR "monitoring" OR "evaluation") AND ("verification" OR "reflection" OR "looking back"). The four progressive analytical phases were conducted as follows:

1. *Literature Screening and Identification.* Foundational models of mathematical problem solving (e.g., Polya, 1957; Rott et al., 2021; Schoenfeld, 1985) and contemporary empirical studies on mathematical metacognition (e.g., Aljura et al., 2025; Fleur et al., 2021; Hidayati et al., 2025; Kholid et al., 2025; Ramlah et al., 2024) were reviewed to synthesise how post-solution behaviours are currently conceptualised and operationalised.
2. *Functional De-construction and Boundary Setting.* Post-solution actions were functionally differentiated according to their cognitive orientation. Verification was defined in terms of product-oriented actions aimed at determining validity, mathematical consistency, and satisfaction of the original problem constraints. Reflection was defined in terms of process-oriented actions aimed at explaining strategic decisions, comparing alternative representations, and abstracting transferable insights. Affective commentary without mathematical substance (e.g., stating that "the problem was easy") was explicitly excluded.
3. *Operationalization and Task Grounding.* These functional distinctions were translated into fine-grained, observable indicators that could be identified in written student solutions, think-aloud protocols, and classroom discussions. To examine the internal logic and applicability of these indicators, the framework was conceptually

contextualized using a representative multi-condition algebraic word problem, referred to as the Rectangle Dimensions Task.

4. *Tool Synthesis and Assessment Translation.* The operational indicators were synthesised into a functional distinction matrix (Table 2), a set of operational indicators (Table 3), an analytical task illustration (Table 4), a two-dimensional diagnostic assessment rubric (Table 5), and tiered classroom instructional prompts (Table 6).

Analytical Task Anchor

To demonstrate how the indicators can be applied, a representative algebraic word problem—the Rectangle Dimensions Task—is used as an analytical anchor (see Figure 2). The task requires students to determine the dimensions of a rectangular garden from its perimeter and a given linear relationship between its length and width, and subsequently to calculate its total area.

> *"A rectangular garden has a perimeter of 44 meters. Its length is 4 meters more than twice its width. Determine the dimensions of the garden and find its total area."*

Figure 2. Representative Task Used to Illustrate Verification and Reflection Indicators

This specific task was selected for its analytical value because it incorporates a multi-layered constraint structure: (1) a perimeter constraint [$2(l + w) = 44$]; (2) a relational constraint [$l = 2w + 4$]; (3) a secondary goal requirement [Area = $1 \times w$]; and (4) contextual plausibility, requiring positive physical dimensions and the use of appropriate square units. As shown in Figure 2, solving this task illustrates why verification must extend beyond simply repeating the primary equation. For example, a solver might calculate dimensions that satisfy the perimeter constraint but violate the length-to-width relationship. Alternatively, a solver might determine the correct dimensions but fail to complete the secondary area calculation or include the appropriate units. The task also provides a clear context for reflection, allowing learners to compare algebraic substitution with alternative approaches, such as tabular modelling or visual area diagrams.

Functional Matrix of Verification and Reflection

To establish a clear conceptual separation between the two mechanisms, Table 2 compares verification and reflection across key functional dimensions.

Table 2. Distinguishing Verification and Reflection in The Look-Back Stage

Dimension	Verification	Reflection
Primary orientation	Product validity, satisfaction of problem constraints, and procedural	Process meaning, strategic reasoning, and conceptual transfer.

	correctness.	
Core question	Is the obtained answer mathematically correct, procedurally accurate, and consistent with all the original problem constraints?	Why is this strategy suitable? How do alternative representations compare, and what principles can be transferred to future tasks?
Typical actions	Back-substituting values, recalculating using inverse operations, checking units, and evaluating contextual plausibility.	Explaining the rationale behind strategy choices, evaluating efficiency, comparing alternative representations, and abstracting general principles.
Evidence in written work	Explicitly checking each problem constraint through supporting calculations or diagrams.	Written explanations justifying method choices, identifying difficulties, or articulating generalisable heuristics.
Risks when absent	Undetected procedural or arithmetic errors, or violations of contextual constraints.	Correct but superficial problem solving, with limited conceptual development or strategic adaptability.

As detailed in Table 2, the two constructs serve fundamentally different instructional and cognitive purposes. Verification acts as a safeguard against mathematical errors, whereas reflection transforms a completed problem-solving task into an opportunity for deeper and more lasting learning.

Methodological Trustworthiness and Empirical Validation Pathway

The trustworthiness of this conceptual study is established through conceptual transparency. This is achieved by clearly presenting definitions, operational indicators, analytical task illustrations, and diagnostic scoring descriptors, enabling researchers and practitioners to examine, replicate, and adapt the proposed framework. As this article presents a theoretical and operational framework that has not yet been empirically validated, future validation is envisaged through a two-stage methodological process:

1. Expert Content Validity Index (CVI): The proposed operational indicators (see Table 3) and scoring rubric (see Table 5) will be evaluated by a panel of experts in mathematics education and metacognition research to establish item-level content validity (I-CVI) and scale-level content validity (S-CVI).
2. Empirical Field Testing: The operationalised framework will be implemented in authentic secondary and tertiary mathematics classrooms using concurrent think-aloud protocols, video-recorded peer discussions, and students' written work. Inter-rater reliability will be assessed statistically using Cohen's kappa (κ) and Fleiss' kappa to evaluate the consistency of coding across independent raters.

Kalamatika: Jurnal Pendidikan Matematika
Volume 11, No. 2, November 2026, pages 234-250
DOI: 10.22236/KALAMATIKA.vol11no2.2026pp234-250

Table 3. Operational indicators for verification and reflection

Construct	Code	Operational Indicator	Observable Evidence in Student Work
Verification	V1	Checks procedural accuracy	Recalculates solution steps or performs inverse mathematical operations to confirm computational accuracy.
Verification	V2	Checks the original problem conditions	Substitutes the obtained results into the relevant expressions to test whether all stated constraints are satisfied independently.
Verification	V3	Checks representations and units	Inspects labels, visual diagrams, units of measurement, and contextual plausibility.
Verification	V4	Identifies and corrects errors	Locates mathematical inconsistencies, diagnoses their causes, and corrects the solution.
Reflection	R1	Explains why the strategy worked	Articulates the mathematical rationale linking the selected method to the structure of the task.
Reflection	R2	Compares alternative strategies	Compares algebraic, tabular, or visual representations and evaluates their relative efficiency.
Reflection	R3	Articulates transferable insights	Formulates generalisable rules, heuristics, or lessons that can be applied to related tasks.
Reflection	R4	Evaluates efficiency and difficulty	Analyses cognitive bottlenecks, sources of confusion, and potentially more efficient solution approaches.

RESULT AND DISCUSSION

Theoretical Extension, Challenge, and Synthesis

The theoretical analysis indicates that verification and reflection serve distinct yet complementary cognitive functions within Pólya's looking-back stage. Verification ensures mathematical validity and satisfaction of the original problem constraints, whereas reflection enables learners to interpret the strategic significance of their problem-solving processes. When these two mechanisms operate together, retrospective evaluation is transformed from a passive checking routine into an active metacognitive practice that supports both conceptual understanding and computational accuracy.

This framework extends Polya's (1957) fourth stage by transforming what has often been treated as a static, linear checking process into an interactive and recursive metacognitive loop. As discussed in recent literature (Aljura et al., 2025; English, 2023; Rott et al., 2021; Santos-Trigo, 2024), looking back is unlikely to be productive when treated merely as a terminal routine. By distinguishing product-oriented verification, which focuses on mathematical correctness and satisfaction of the original constraints, from process-oriented

reflection, which focuses on extracting strategic insights and identifying opportunities for transfer, the proposed framework provides researchers and educators with an explicit vocabulary for describing the recursive nature of retrospective evaluation.

Furthermore, this framework challenges conventional mathematics assessment models that assign a single, undifferentiated score to the act of checking one's work. In traditional rubrics, a student who obtains a correct answer and performs a superficial recalculation may receive full credit for looking back, despite demonstrating little evidence of metacognitive awareness of the rationale behind the chosen strategy or of alternative representations. Conversely, a student who encounters an initial error, traces its source, and reorganises their mathematical model may receive a low score because of initial procedural inaccuracies. By distinguishing verification and reflection as two analytically distinct dimensions, the proposed framework highlights that high procedural accuracy does not necessarily indicate well-developed metacognitive awareness, just as an initial error does not necessarily indicate limited problem-solving capability.

This conceptual framework complements Schoenfeld's (1985) metacognitive framework, particularly its emphasis on managerial control and the allocation of cognitive resources, and aligns with contemporary research on metacognitive awareness in mathematical problem solving (Fleur et al., 2021; Hidayati et al., 2025; Kholid et al., 2025; Ramlah et al., 2024; Subba et al., 2025). While Schoenfeld established the importance of executive monitoring in problem solving, the proposed framework specifies how such monitoring can be examined during retrospective evaluation. It provides observable indicators (see Table 3) that help bridge theoretical metacognitive constructs and evidence of students' classroom performance.

The Recursive Interplay in Problem Solving: Analysing the Anchor Task

The interplay between verification and reflection does not follow a fixed chronological sequence; rather, it constitutes a dynamic and self-correcting process. Verification can trigger reflection: when an independent check of a problem constraint fails, the solver is prompted to reconsider where their mathematical model diverged from the problem structure. Conversely, reflection can strengthen verification: reflecting on the relationship between length and width may lead a solver to recognise that checking only the perimeter equation is insufficient, prompting an additional verification of the relational constraint.

Kalamatika: Jurnal Pendidikan Matematika
Volume 11, No. 2, November 2026, pages 234-250
DOI: 10.22236/KALAMATIKA.vol11no2.2026pp234-250

To illustrate how these theoretical indicators can be applied in practice, Table 4 analyses representative student responses to the Rectangle Dimensions Task alongside targeted teacher feedback.

Table 4. Analytic illustration using the rectangle dimensions task

Look-Back Evidence in Students Work	Verification Interpretation	Reflection Interpretation	Diagnostic Teacher Response
Substituted the derived length (16 m) and width (6 m) into the perimeter equation: 2(16 + 6) = 44.	Strong V2: Independently verifies that the perimeter constraint is satisfied.	Limited evidence: No explicit reflection on why the algebraic model was appropriate for the problem.	*"Your perimeter check is correct. Which other constraint in the problem must also be checked?"*
Checked only that the length and width sum to 22, overlooking the requirement that the length be 4 m greater than twice the width.	Weak V2: The constraint check is incomplete because the relational condition has been omitted.	No explicit evidence of reflection.	*"If the width is 6 m, is the length 4 m greater than twice the width? How can you verify this?"*
Constructed a tabular model showing values for the width, the corresponding expression (2w + 4), and the resulting perimeter.	Supports V2 and V3: Uses an alternative representation to check the consistency of the problem constraints.	Strong R2: Compares algebraic substitution with tabular exploration.	*"How does the table support your algebraic solution? Which method would you choose if the numbers were much larger?"*
Wrote: *"I must always check all the given relationships, not just the perimeter formula, before calculating the area."* *before computing area*."	Provides evidence of awareness that all relevant constraints must be verified.	Strong R3: Formulates a transferable heuristic for solving multi-condition problems. .	*"How could you apply this checking principle to a problem involving the volume of a three-dimensional object?"*

Diagnostic Assessment Rubric and Classroom Scaffolding

While the analytical illustration in Table 4 contextualises students' thinking through a single algebraic task, classroom assessment across different mathematical topics requires these specific qualitative observations to be translated into a more generalisable assessment instrument. Moving from task-specific behaviours to a multi-tiered scoring matrix enables teachers and researchers to systematically assess and categorise varying levels of metacognitive competence across diverse mathematical contexts. To establish this broader assessment framework Table 5 presents an analytical rubric that defines performance levels for verification and reflection along a four-point scoring continuum, ranging from 0 to 3.

Table 5. Proposed Analytical Rubric for Assessing Verification and Reflection

Score	Verification Descriptor (Product Validity)	Reflection Descriptor (Process Meaning & Transfer)
0 (None)	No checking is attempted, or the checking performed is entirely irrelevant to the problem requirements.	No explanation of the strategy, evaluation of the problem-solving process, comparison of representations, or transferable insight is provided.

Kalamatika: Jurnal Pendidikan Matematika
Volume 11, No. 2, November 2026, pages 234-250
DOI: 10.22236/KALAMATIKA.vol11no2.2026pp234-250

1 (Superficial)	Superficial checking: Merely repeats the original calculation without using an independent operation or checks only an isolated numerical result.	Superficial reflection: Provides brief affective comments (e.g., "the problem was easy") without offering any mathematical rationale.
2 (Partial)	Partial checking: Verifies the answer against at least one primary condition but leaves other constraints or units unchecked.	Partial reflection: Explains the mechanics of the chosen strategy or identifies difficulties but does not compare alternative methods or formulate generalisable principles.
3 (Proficient)	Thorough checking: Independently verifies all original problem constraints, units, and contextual plausibility using appropriate mathematical operations or alternative representations.	In-depth reflection: Clearly explains the rationale for the chosen strategy, compares alternative representations, and articulates transferable principles or heuristics for solving future problems.

The descriptive criteria established in the assessment rubric provide a foundation for structuring formative interventions during problem-solving lessons. Rather than relying on generic and uninformative directives such as "check your work," teachers can use the rubric dimensions to formulate targeted, dialogic questions that guide students towards proficient verification and reflection at different stages of the problem-solving process.

To support teachers in incorporating these practices into authentic classroom interactions, Table 6 presents specific instructional prompts tailored to key phases of problem solving.

Table 6. Targeted classroom questioning prompts derived from the framework

Phase	(Teacher Questioning)	(Teacher Questioning)
Immediate post-solution	*"Which specific condition in the problem does your numerical answer satisfy, and which conditions have you not yet checked?"*	*"Why did you choose an algebraic equation rather than a diagram or a table to solve this problem?"*
Pre-Submission Review	*"Can you demonstrate that your answer is correct using an inverse operation or an alternative representation?"*	*"If you were explaining this problem to a classmate, which key mathematical step would you emphasise?"*
Peer Collaborative Dialogue	*"Show your partner exactly how your calculated values satisfy all the given conditions in the word problem."*	*"Which of your two solution approaches was more efficient, and what makes it easier to understand?"*
Post-Error Correction	*"At which specific constraint did your initial calculation fail, and how does your corrected answer resolve the discrepancy?"*	*"What did this error reveal about your initial assumptions, and what checking strategy will you use in future problems?"*
Summative Assessment	*"Provide two independent checks to demonstrate the mathematical validity of your final answer."*	*"State one transferable mathematical principle you learned from solving this task that could be applied to future problems."*

Implications for Teacher Education, Instrument Design, and Limitations

For mathematics teacher education, the verification–reflection framework provides a diagnostic lens for analysing students' mathematical thinking. Pre-service and in-service

teachers should be trained not to infer deep mathematical understanding solely from a correct final answer or the repetition of a calculation. When teachers observe students comparing representations, diagnosing inconsistencies between their solutions and the original problem constraints, and generalising problem structures, they gain valuable diagnostic evidence of reflective metacognitive activity. This evidence can inform more responsive instructional interventions.

For educational researchers, an important implication is the potential development of validated assessment instruments and coding rubrics. Future studies could build on the indicators presented in Table 3 and rubric descriptors in Table 5 to develop standardised video-coding protocols and written assessment instruments. Psychometric validation, including the assessment of content validity using the Content Validity Index (CVI) and inter-rater reliability analysis using Cohen's kappa or Fleiss' kappa, as appropriate, would help establish the empirical robustness and applicability of these instruments across diverse student populations.

A key limitation of this article is its conceptual nature: it does not report empirical distributions of students' performance, intervention effect sizes, or learner developmental trajectories. Furthermore, the proposed operational indicators are illustrated using an algebraic word problem. In other mathematical domains, such as geometric proof, statistical modelling, and calculus, the specific manifestations of verification (e.g., checking geometric invariants, testing asymptotic behaviour, or examining the appropriateness of statistical assumptions) and reflection (e.g., evaluating the elegance of a proof, interpreting statistical variability, or identifying transferable mathematical structures) will require domain-specific adaptation. Nevertheless, the proposed distinction between product-oriented validity checking and process-oriented conceptual reflection provides a potentially useful foundation for examining retrospective evaluation across different areas of mathematics.

CONCLUSION

This conceptual article refines George Pólya's looking-back stage by proposing an operational framework comprising two integrated metacognitive mechanisms: verification and reflection. Verification focuses on the validity of the mathematical product by examining computational accuracy, satisfaction of problem constraints, dimensional consistency, and contextual plausibility. Reflection focuses on the meaning of the problem-solving process by examining the rationale behind strategy choices, comparing alternative representations, identifying sources of difficulty, and formulating transferable principles for future tasks. By

distinguishing these two mechanisms, the framework clarifies why retrospective evaluation should not be reduced to an undifferentiated checking routine.

The proposed framework contributes to mathematics education at both theoretical and pedagogical levels. Theoretically, it articulates a recursive structure for metacognitive evaluation after solution generation, connecting broader metacognitive perspectives with observable mathematical practices. Pedagogically, it provides educators and researchers with a functional distinction matrix, operational indicators, an analytical assessment rubric, and targeted classroom questioning prompts. These tools are intended to help teachers foster deeper mathematical reasoning rather than merely procedural compliance. Future empirical research should examine the framework in authentic classroom settings to evaluate its diagnostic usefulness and investigate the potential longitudinal effects of explicit instruction in verification and reflection on students' mathematical problem-solving development.

ACKNOWLEDGMENTS

The authors would like to express their sincere appreciation to everyone who contributed to the completion of this study. This article is the result of a collaborative effort among researchers from different institutions, whose commitment, expertise, and constructive discussions enriched the research process and the development of the manuscript. The authors are grateful for the academic support and professional environment provided by the Department of Mathematics Education, Wisnuwardhana University, Indonesia; the Mechanical Engineering Technology Study Program, Surabaya Merchant Marine Polytechnic, Indonesia; Sultan Idris Education University, Malaysia; and the Department of Mathematics Education, University of Muhammadiyah Prof. Dr. Hamka, Indonesia. Their support and encouragement were invaluable throughout the research process and the preparation of this article.

REFERENCES

Aljura, A. N., Retnawati, H., Dewanti, S. R., Kassymova, G. K., Sotlikova, R., & Septiana, A. R. (2025). Mathematical reasoning and communication word problems with a mathematical problem-solving orientation: A relation between the skills. Journal on Mathematics Education, *16*(2), 529-558. https://doi.org/10.22342/jme.v16i2.pp529-558

Brown, A. L. (1987). Metacognition, executive control, self-regulation, and other more mysterious mechanisms. In F. E. Weinert & R. H. Kluwe (Eds.), Metacognition, motivation, and understanding (pp. 65-116). Lawrence Erlbaum Associates.

English, L. D. (2023). Ways of thinking in STEM-based problem solving. ZDM - Mathematics Education, *55*, 1219-1230. https://doi.org/10.1007/s11858-023-01474-7.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. American Psychologist, *34*(10), 906-911. https://doi.org/10.1037/0003-066X.34.10.906

Fleur, D. S., Bredeweg, B., & Van Den Bos, W. (2021). Metacognition: Ideas and insights from neuro- and educational sciences. NPI Science of Learning, *6*, Article 13. https://doi.org/10.1038/s41539-021-00089-5.

Goos, M., Galbraith, P., & Renshaw, P. (2002). Socially mediated metacognition: Creating collaborative zones of proximal development in small group problem solving. Educational Studies in Mathematics, *49*(2), 193–223. https://doi.org/10.1023/A:1016209010120

Hidayati, N., Sukestiyarno, Y. L., Pujiastuti, E., Sugiman, S., & Waluya, S.B. (2025). Students' mathematical literacy in terms of metacognitive awareness. Infinity Journal, *14*(4), 839-860. https://doi.org/10.22460/infinity.v14i4.p839-860

Jabareen, Y. (2009). Building a conceptual framework: Philosophy, definitions, and procedure. International Journal of Qualitative Methods, *8*(4), 49-62. https://doi.org/10.1177/160940690900800406

Kholid, M. N., Fadhilah, H., Loc, N. P., & Magdas, I. C. (2024). Classifying analogical thinking for mathematical problem-solving. Journal on Mathematics Education, *15*(3), 793-814. https://doi.org/10.22342/jme.v15i3.pp793-814

Kholid, M. N., Widyasari, A., Santosa, Y. T., Hidayati, Y. M., Fenyvesi, K., Maharani, S., Ikram, M., Wijaya, A. P., Prabowo, A., & Hendriyanto, A. (2025). Student metacognition levels for solving PISA-like problems: A hierarchy. Infinity Journal, *14*(3), 817-836. https://doi.org/10.22460/infinity.v14i3.p817-836

Polya, G. (1957). How to solve it: A new aspect of mathematical method (2nd ed.). Princeton University Press.

Purnomo, Y. W., Shahrill, M., Pandansari, O., Susanti, R., & Winarni, W. (2022). Cognitive demands on geometrical tasks in Indonesian elementary school mathematics textbook. *Jurnal Elemen*, *8*(2), 466–479. https://doi.org/10.29408/jel.v8i2.5235

Ramlah, R., Siswono, T. Y. E., & Lukito, A. (2024). Revealing the uniqueness of variations in prospective teachers' metacognitive activities in solving mathematical problems based on gender. Infinity Journal, *13*(2), 477-500. https://doi.org/10.22460/infinity.v13i2.p477-500

Rott, B., Specht, B., & Knipping, C. (2021). A descriptive phase model of problem-solving processes. ZDM - Mathematics Education, *53*, 737-752. https://doi.org/10.1007/s11858-021-01244-3

Santos-Trigo, M. (2024). Problem solving in mathematics education: Tracing its foundations and current research-practice trends. ZDM - Mathematics Education, *56*, 211-222. https://doi.org/10.1007/s11858-024-01578-8

Schoenfeld, A. H. (1985). Mathematical problem solving. Academic Press.

Schoenfeld, A. H. (1992). Learning To Think Mathematically: Sense-Making in Mathematics. In *In D. Grouws (Ed.), Handbook for research on mathematics teaching and learning. New York: MacMillan.*

Subba, B. H., Chanunan, S., & Poonpaiboonpipat, W. (2025). A proposed constructivist instructional model to enhance metacognition and mathematical problem-solving skills among Bhutanese grade 9 students. Journal on Mathematics Education, *16*(1), 51-72. https://doi.org/10.22342/jme.v16i1.pp51-72

Torraco, R. J. (2005). Writing integrative literature reviews: Guidelines and examples. Human Resource Development Review, *4*(3), 356-367. https://doi.org/10.1177/1534484305278283.

Verschaffel, L., De Corte, E., & Lasure, S. (1994). Realistic modelling with problem solving in realistic situations. Educational Studies in Mathematics, *27*(3), 273–294. https://doi.org/10.1007/BF01273663

Wahyu, R. (2022). Degenerative thinking disposition of mathematics education students in solving quadrilateral classification problems [Doctoral dissertation, Universitas Negeri Malang].

Wahyu, R. (2026). Degenerative thinking theory: The tendency to overgeneralize in mathematical problem-solving. Journal on Mathematics Education, *16*(2), 529-558. https://doi.org/10.22342/jme.v16i2.pp529-558