Article

Journal of Interdisciplinary Studies in Education

Volume 15, Issue 8 (2026)
*Journal of Interdisciplinary Studies in Education*
ISSN: 2166-2681 Print 2690-0408 Online | https://ojed.org/jise
https://doi.org/10.32674/ed2yma36

# Mapping the Sources of Mathematical Controversy: A Source-Based Typology of Pre-service Teachers' Reasoning on Misleading Media, Procedures, and Operations

Erry Hidayanto*
*Faculty of Mathematics and Natural Sciences, Universitas Negeri Malang, Indonesia*
ORCID: 0000-0001-9412-0799

Subanji
*Faculty of Mathematics and Natural Sciences, Universitas Negeri Malang, Indonesia*
ORCID: 0000-0002-4281-1923

Rivan Gestiardi
*Independent Researcher, Wagir, Malang, Indonesia*
ORCID: 0000-0003-0742-1169

* Corresponding author: erry.hidayanto.fmipa@um.ac.id

**ABSTRACT**

*Research on counterintuitive mathematics tasks describes levels of resolution but rarely distinguishes where a contradiction originates. This qualitative study examines a provisional typology of misleading media, procedures, and operations. Fifty Indonesian pre-service mathematics teachers in a university teacher education programme completed four tasks on percentage growth, fraction representation, algebraic cancellation, and root–exponent rules; 16 joined follow-up interviews. Responses were coded through three cycles. Participants showed different item-level patterns, and combining source and level produced a two-axis diagnostic matrix. Because media and operation were each represented by one task, source cannot be separated from item content or difficulty. The matrix is therefore a heuristic for further validation, not a tested intervention. It may help teacher educators organise analysis of student reasoning.*

**Keywords:** cognitive conflict, diagnostic framework, mathematical reasoning, pre-service mathematics teachers, qualitative content analysis, student errors, teacher education

© Author(s), 2026. Published by Star Scholars Press.
This article is distributed under the Creative Commons Attribution 4.0 International License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. https://creativecommons.org/licenses/by/4.0/

## INTRODUCTION

Reasoning has long been treated as the cognitive engine of mathematics learning, with the National Council of Teachers of Mathematics (NCTM, 2000) placing reasoning and sense-making among the core processes of school mathematics. Pre-service mathematics teachers occupy a doubly demanding position in this regard: they must reason mathematically themselves and also notice, interpret, and respond to the reasoning of their future students (Ball et al., 2008; Jeannotte & Kieran, 2017). Among the situations that exercise both demands at once, mathematical paradoxes, fallacies, and cognitively conflictive tasks have received sustained attention (Movshovitz-Hadar & Hadass, 1990; Mueller & Yankelewitz, 2014; Zaslavsky, 2005). Such tasks present a procedure or representation that appears legitimate but yields a result that violates established mathematical knowledge, generating a productive cognitive conflict between expectation and outcome (Limón, 2001). The present study examines where such contradictions originate: in a representation that obscures an invariant, in a multi-step procedure that conceals an invalid move, or in an operation applied outside its valid domain. This diagnostic focus is relevant to performance assessment and to the metacognitive processes involved in mathematical problem solving (Barron, 2015; Rodil, 2024). Because these demands are cultivated in university-based teacher education, the study is situated in higher education and asks how a teacher education programme can analyse prospective teachers' reasoning before they enter classrooms.

Movshovitz-Hadar and Hadass (1990, 1991) initiated this line of inquiry in pre-service teacher education, demonstrating that paradoxes can sharpen prospective teachers' mathematical concepts and raise their pedagogical awareness of the constructive role of fallacious reasoning. Zaslavsky (2005) extended the argument by showing that uncertainty deliberately built into mathematical tasks adds value for both mathematical and pedagogical knowledge. Mueller and Yankelewitz (2014) further documented how exposure to fallacious arguments can productively unsettle procedure-based misconceptions. These traditions converge on a common claim: tasks that generate cognitive conflict are valuable for teacher learning. They differ, however, in their analytic foci. Some emphasise the resolution trajectory through which learners progress; others emphasise the pedagogical affordances of conflict; few have asked, systematically, where the conflict originates.

The question of origin is consequential for both theory and practice. Conflicts triggered by visual representations, algorithmic procedures, and operational rules can recruit different cognitive resources (Duval, 2017; Fischbein, 1987; Tall & Vinner, 1981). Recent studies likewise show that pre-service teachers' errors depend on the mathematical and task demands they encounter, including proof writing, problem-to-prove activity, modelling-task formulation, and responses to fraction errors (Akkus et al., 2026; An et al., 2026; Koç & Bingölbali, 2026; Segura et al., 2026). Existing level frameworks describe how far learners progress toward resolution but leave the proposed source dimension underdeveloped (Subanji et al., 2021).

We use the term controversial reasoning to denote the reasoning that learners deploy when confronted with mathematical situations in which an apparently valid procedure or representation yields a contradictory result. We retain this term, while clarifying its relation to paradoxes, fallacies, and cognitive conflict in the Theoretical Framework, because it foregrounds the dialogical and pedagogical features of the situation: the learner confronts a claim, weighs its components, and must produce a justification that is mathematically sound and pedagogically defensible. This emphasis on the social production of justification distinguishes controversial reasoning from purely intra-cognitive accounts of paradox-resolution and from purely formal accounts of fallacy-detection. Three research questions guide the study:

RQ1: How can the three candidate sources of controversy be operationalised in the four tasks?
RQ2: What item-level reasoning patterns appear across the four tasks?
RQ3: How can a provisional source-based typology complement an existing level-based framework of controversial mathematical reasoning?

By answering these questions, we contribute (a) a tentative typology that complements rather than replaces existing level accounts; (b) an empirical illustration of how the typology operates in a cohort of fifty pre-service mathematics teachers; and (c) a two-axis matrix combining source and level dimensions. Because the design does not test an intervention or validate an assessment instrument, the matrix is presented as a provisional heuristic.

## THEORETICAL FRAMEWORK

### Mathematical Reasoning and Standards of Practice

Mathematical reasoning is more than the production of correct answers; it is the construction and evaluation of arguments that justify why a given step or conclusion follows from accepted premises (Jeannotte & Kieran, 2017; Schoenfeld, 2020). Reasoning becomes especially visible in non-routine settings that resist familiar procedures, and reasoning quality predicts both conceptual depth and procedural flexibility (Rittle-Johnson, 2017; Rittle-Johnson et al., 2015). For pre-service teachers, the ability to reason and to interpret others' reasoning is at the heart of Mathematical Knowledge for Teaching (Ball et al., 2008).

### Paradoxes, Fallacies, Cognitive Conflict, and Controversial Reasoning: Terminological Clarification

Four overlapping but distinct constructs populate the literature on counterintuitive mathematical tasks. A mathematical paradox is a statement or argument that appears acceptable on each step but produces an unacceptable conclusion; Movshovitz-Hadar and Hadass (1990) used paradoxes systematically to develop pre-service teachers' conceptual sharpness. A fallacy is an argument whose form mimics a valid pattern but contains an unjustified inference; Mueller and Yankelewitz (2014) showed how engaging with fallacies productively destabilises procedural misconceptions. Cognitive conflict, more general still, is the psychological state that arises when new information clashes with prior knowledge (Limón, 2001), with Zaslavsky (2005) extending this to include task-induced uncertainty as a positive learning resource.

We use controversial reasoning as a fourth, intentionally narrower term that highlights the dialogical situation: a learner is presented not only with a counterintuitive task but also with a claimant's argument and is asked to evaluate the claim, identify its problematic component, and supply a justification a teacher could use in front of students. The construct thus inherits the cognitive-conflict mechanism (Limón, 2001), the formal scrutiny demanded by paradoxes (Movshovitz-Hadar & Hadass, 1990), and the diagnostic stance demanded by fallacies (Mueller & Yankelewitz, 2014), while adding the requirement of pedagogical defensibility. The reason to retain a distinct label is that this added requirement reorganises the analytic categories one needs: we are no longer interested only in whether the learner can detect the conflict, but in whether the learner can locate the specific epistemic object—a representation, an algorithmic step, or an operational rule—that broke the chain. That is the analytic move the present typology is designed to support.

### Cognitive Conflict and Productive Uncertainty

Cognitive conflict can prompt conceptual change when the conflict is meaningful and the learner is motivated to resolve it (Limón, 2001). In mathematics, cognitively demanding tasks can disrupt procedure-based misconceptions and encourage conceptual grounding (Mueller et al., 2010; Mueller & Yankelewitz, 2014). Zaslavsky (2005) described uncertainty arising from competing claims, unknown paths or questionable conclusions, and outcomes that are not readily verifiable. Recent work with pre-service teachers also documents the cognitive and affective demands of sustaining justification in problem-to-prove activity (Akkus et al., 2026; Lee, 2026).

## Concept Image, Concept Definition, and Algorithmic Intuition

Why do certain mathematical situations generate conflict in the first place? Tall and Vinner's (1981) distinction between concept image (the total cognitive structure associated with a concept) and concept definition (the formal statement acknowledged by the mathematical community) provides a powerful explanatory lens. A learner's concept image may not be globally coherent, and a controversial task can expose precisely the gap between image and definition. When a bar illustration suggests that 1/3 + 1/3 = 2/6, for example, the figure recruits a concept image of fraction addition (combine the bars; combine the parts) that diverges from the formal concept definition (preserve the unit whole).

Fischbein's (1987) account of intuition complements this lens. Intuitive cognitions, including what Fischbein called algorithmic intuitions, are convictions of self-evidence that exceed available formal justification. A learner who applies cross-multiplication unreflectively is operating on an algorithmic intuition that the procedure preserves equality regardless of the denominators' status. The conflict produced when this intuition fails is therefore a conflict not between two formal results but between an intuitive scheme and a formal constraint. The implication for the present typology is that the source of controversy is not always at the level of representation; it can also live at the level of intuitive scheme governing operations and procedures.

## Empirical Antecedents: International and Local

Internationally, the most direct antecedents to the present study are Movshovitz-Hadar and Hadass (1990, 1991), who documented the pedagogical affordances of paradoxes for pre-service teachers in Israel; Mueller and Yankelewitz (2014), who examined fallacious arguments in informal learning settings; and Zaslavsky (2005), who proposed three forms of mathematical uncertainty with implications for task design. Recent studies of proof writing, problem-to-prove activity, modelling-task formulation, and responses to fraction errors further show that prospective teachers' reasoning varies with task demands (Akkus et al., 2026; An et al., 2026; Koç & Bingölbali, 2026; Segura et al., 2026). None of these works independently validates the three source categories proposed here. Locally, Subanji et al. (2021) documented three levels of controversial reasoning (initial, exploration, clarification) among 185 Indonesian pre-service mathematics teachers, and Hidayah et al. (2020) characterised abductive reasoning in algebraic problem solving. The Subanji et al. (2021) level framework provides a developmental trajectory that the present study uses alongside the source typology. We treat the level account as a regional antecedent rather than the sole foundation; the analytic move from "how far" to "where" requires building on the international literature on paradoxes, cognitive conflict, and concept image.

## Toward a Source-Based Typology: Three Candidate Sources

Drawing on this literature, we propose three candidate sources of controversy. A misleading medium is a representation that changes or obscures a mathematical invariant; in P2, concatenating two fraction bars quietly changes the unit whole. A misleading procedure is a multi-step sequence in which an apparently routine move conceals an invalid operation or unmet condition; in P1, additive treatment obscures multiplicative scaling, while in P3 cancellation of the factor $x - 3$ discards the admissible solution $x = 3$. A misleading operation is a rule used outside its valid domain; in P4, the principal-root identity $\sqrt{a^2} = |a|$ and restrictions on exponent laws over negative real bases are ignored. These categories are analytical proposals rather than empirically independent constructs.

The distinction between procedure and operation merits comment, because readers may object that both involve violated preconditions. The two are distinguished operationally by their grain. A procedure is a sequence of operations chained algorithmically, where the contradiction emerges from the sequence as a whole; an operation is a unitary rule, where the contradiction emerges from a single rule's scope of validity.

In Problem 3, the participant must trace a sequence of steps and recognise that cancelling $x - 3$ discards the zero-factor solution $x = 3$; in Problem 4, the participant must inspect the principal-root convention and the domain conditions governing exponent laws over negative real bases. The distinction is not always crisp at the boundary, and we return to its limits in the Limitations section. This operational distinction is provisional, and the present four-item instrument cannot establish discriminant validity between procedure and operation.

## RESEARCH METHOD

### Research Design

This study employed a qualitative exploratory–deductive design. The exploratory orientation reflects the goal of generating a tentative typology grounded in detailed analysis of pre-service teachers' reasoning. The deductive component reflects an intentional design choice: the four problems were each constructed to instantiate a candidate source of controversy identified a priori from the theoretical framework. We therefore make no claim that the three sources emerged inductively from data alone. Instead, the data were used to examine whether the preassigned labels could organise participants' responses and to characterise item-level reasoning patterns; the design does not establish that the categories are empirically independent constructs. This is, in Hsieh and Shannon's (2005) terms, a directed content analysis.

### Participants and Setting

Participants were fifty undergraduate pre-service mathematics teachers enrolled in the Mathematics Education programme of a state university in East Java, Indonesia, and taking the compulsory course Berpikir dan Bernalar Matematis (Mathematical Thinking and Reasoning). All had completed prior coursework in calculus, real analysis, abstract algebra, and mathematics pedagogy. The cohort comprised thirty-two female and eighteen male students. Their prior coursework reduced, but did not eliminate, the possibility that unresolved responses reflected limited exposure to the underlying mathematics.

### Instrument: Four Controversial Mathematics Problems

The instrument consisted of four tasks assigned a priori to the candidate source categories. For transparency and replicability, the four problems are reproduced below in English translation from the Indonesian versions administered to participants. Each problem situates the controversy in a classroom case and asks the pre-service teacher to evaluate a student's reasoning and provide their own justification.

#### *Problem 1: Perimeter and Area*

Pak Andi is a farmer who owns a relatively large rice field. A portion of his land is planted with corn. The corn-planted area is rectangular, with length p and width l.

Because corn prices have risen, Pak Andi wishes to expand his corn cultivation by increasing both the length and the width by 30%. To determine the amount of seed and fertiliser required, he must know the new perimeter and area, and asks his two children, Puput and Sari, to compute the percentage increase in perimeter and area of the corn field. The two children give different answers.

*Puput's answer:*

Length: increased by 30%; width: increased by 30%; perimeter: increased by 60%; area: increased by 9%.

*Sari's answer:*

Length: increased by 30%; width: increased by 30%; perimeter: increased by 30%; area: increased by 69%.

In your view, are either of the children's answers correct? If one is correct, provide an explanation. If neither is correct, provide what you believe to be the correct answer with a complete justification.

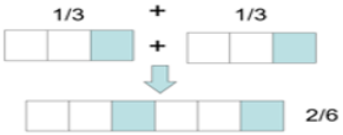
***Problem 2: Fraction Operation***

In a classroom case, a teacher discovered a student's error in answering a fraction-addition problem: the student wrote 1/3 + 1/3 = 2/6. The teacher traced the student's reasoning through the following interview.

Teacher: Tell me how you arrived at 1/3 + 1/3 = 2/6.
Student: Once I was given the problem 3 + 3 = 6 with this illustration:




Teacher: How does that lead to 1/3 + 1/3 = 2/6?
Student: It can be drawn like this, sir:




Student: So, the answer is 2/6, sir.

Based on this classroom case:
(1) Does the student's answer make sense? Explain.
(2) Do you agree with the student's answer? Justify your reasoning.

***Problem 3: Algebraic Expressions and Equations***

A student previously encountered a problem on simplifying an algebraic expression and solved it successfully:

$$\frac{x^2-4}{x-2}=\frac{(x-2)(x+2)}{x-2}=x+2$$

On another occasion, the same student also successfully solved a rational algebraic equation:

$$\frac{x-3}{x+3}=\frac{x-1}{x+2}$$
$$(x+2)(x-3)=(x-1)(x+3)$$
$$x^2-x-6=x^2+2x-3$$
$$-2x-x=6-3$$
$$x=-1$$

When the teacher subsequently posed the problem "find the solution of the equation $\frac{x-3}{x+1}=\frac{x-3}{x+2}$", the student carried out the following steps:

$$\frac{x-3}{x+1}=\frac{x-3}{x+2} \implies (x+2)(x-3)=(x+1)(x-3) \implies x+2=x+1 \implies 2=1$$

Based on this classroom case:
(1) Does the student's answer make sense? Explain.

(2) Do you agree with the student's answer? Justify your reasoning.

***Problem 4: Square Root and Exponentiation***

A teacher has explained the exponential identities $\sqrt{a^2} = a$ and $(a^m)^n = a^{mn}$. Based on these identities, the teacher asked students to evaluate the following argument:

$$2 = \sqrt{4} = \sqrt{(-2)^2} = ((-2)^2)^{1/2} = (-2)^1 = -2$$

(1) What is your view of this argument?
(2) Justify your answer.

A brief summary of how each problem maps onto the candidate sources of controversy is given in the next paragraph; the detailed expert-validation procedure follows.

P1 was assigned to misleading procedure, P2 to misleading media, P3 to misleading procedure, and P4 to misleading operation. For P3, the given equation has the valid solution x = 3 (with $x \neq -1, -2$). The student's cancellation of x − 3 produces 2 = 1 only because it divides by a factor that is zero at the solution, thereby losing that solution. For P4, the contradiction arises because the identity $\sqrt{a^2} = a$ presented in the task holds only for $a \geq 0$; the correct identity is $\sqrt{a^2} = |a|$, and exponent laws involving negative real bases require domain restrictions.

Three expert reviewers (two senior mathematics educators and one professor of mathematics) judged whether the controversy in each problem was plausibly traceable to the intended dominant source, whether the mathematics was appropriate for the cohort, and whether the problem was sufficiently open to allow multiple solution paths. After two rounds of revision, all four problems were judged satisfactory. We acknowledge that designing problems for a target source carries an implicit risk of confirmation: the typology cannot, by itself, demonstrate that no other sources exist. The Discussion and the Limitations section return to this limitation. This expert review supports mathematical appropriateness and content clarity, but it should not be interpreted as independent construct validation because the source labels were specified a priori and remained tied to the particular tasks.

**Data Collection**

Data were collected during a 100-minute closed-book end-of-semester examination administered in December 2025. Participants worked individually without electronic devices. The instructions asked participants, for each problem, to (a) judge whether the offered solution made sense, (b) explain the reasoning behind the judgement, and (c) propose, where appropriate, the response they would give as a teacher. After the analysis of written responses, sixteen participants were selected for follow-up clinical interviews (25–40 minutes each, audio-recorded with consent).

**Stratified Purposive Sampling for Interview**

Selection of interviewees followed three explicit criteria. First, breadth across levels: at least four interviewees were selected whose written work suggested initial-level responses, four with exploration-level responses, and four with clarification-level responses, so that the interview sample spanned the full level range of Subanji et al. (2021). Second, breadth across sources: at least three interviewees per source category (media, procedure, operation) were chosen, ensuring within-source variation could be probed. Third, ambiguity-driven selection: four interviewees were selected because their written work was difficult to classify, so that interviews could clarify borderline cases. The combination yielded sixteen interviewees: nine female and seven male; six initially classified at clarification, six at exploration, and four at initial. Interview questions probed how the participant detected the controversy, how they localised the offending step, and how they justified their resolution. Because media and operation were each represented by one

task, this sampling increased the diversity of accounts but could not separate source effects from item effects.

### Data Analysis

Analysis proceeded through three coding cycles. The first cycle was open coding: each written response and interview transcript was read line by line and assigned in-vivo codes. The second cycle was axial coding: the open codes were grouped into emerging categories that described both the candidate source and the resolution level. The third cycle was selective coding, in which the categories were related to one another to generate the source-based typology and the two-axis matrix. Two researchers coded a 30% subsample independently. Inter-coder agreement was 96% for application of the preassigned source labels (Cohen's $\kappa = .92$) and 88% for level assignment (Cohen's $\kappa = .81$). Agreement on source labels indicates coding consistency within this instrument, not construct validation of the typology. Discrepancies were resolved through joint review of the script and, where necessary, the interview transcript. The analysis was qualitative throughout; we therefore report frequencies for descriptive purposes only and do not undertake inferential testing.

### Trustworthiness

Trustworthiness was established through the four criteria proposed by Lincoln and Guba (1985). Credibility was supported by triangulation between written work and interview, by member checking with eight participants, and by peer debriefing with two external mathematics education researchers. Dependability was supported by an audit trail. Confirmability was supported by reflexive memos. Transferability was supported by thick description of participants and the institutional context. Institutional permission to conduct the study was obtained from Universitas Negeri Malang. Participants gave written informed consent and were assured of anonymity through pseudonyms (S1–S50).

## FINDINGS

### Overview

The four tasks were assigned by design to the three candidate categories: P2 to media, P1 and P3 to procedure, and P4 to operation. Table 1 records those assignments and the mathematically correct resolution of each contradiction. The analysis can describe response patterns for these tasks and illustrate how the candidate labels organise them; it cannot establish independent source effects.

**Table 1**

*Four Tasks, Assigned Candidate Sources, and Mathematically Correct Resolutions*

| Task | Mathematical content | Assigned source | Correct resolution / diagnostic issue |
|---|---|---|---|
| P1 | Perimeter and area when both rectangle dimensions increase by 30% | Procedure | Perimeter scales by 1.30 (30% increase); area scales by 1.69 (69% increase). |
| P2 | 1/3 + 1/3 represented by concatenated fraction bars | Media | The unit whole must remain a three-part bar; the sum is 2/3, not 2/6. |
| P3 | $\frac{x-3}{x+1} = \frac{x-3}{x+2}$, $x \neq -1, -2$ | Procedure | The solution is x = 3. Cancelling x − 3 divides by zero at the solution and loses it. |
| P4 | $2 = \sqrt{4} = \sqrt{(-2)^2} = ((-2)^2)^{1/2} = (-2)^1 = -2$ | Operation | $\sqrt{a^2} = \lvert a \rvert$; exponent laws require domain restrictions for negative real bases. |

*Note.* Source labels were assigned a priori and are not independent empirical classifications.

**Figure 1**

*Provisional Source-Based Typology Used to Organise the Four Tasks*

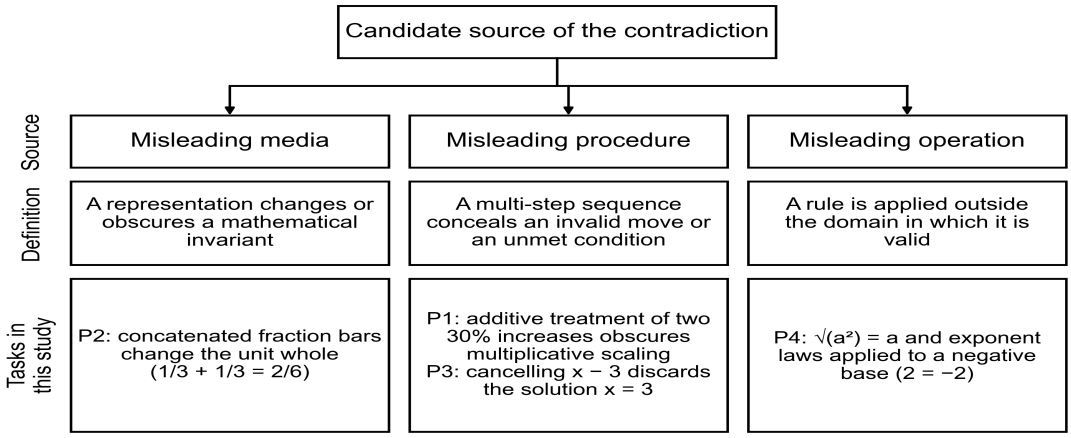



**Figure 2**

*Distribution of Reasoning Levels Across the Four Tasks (N = 50)*

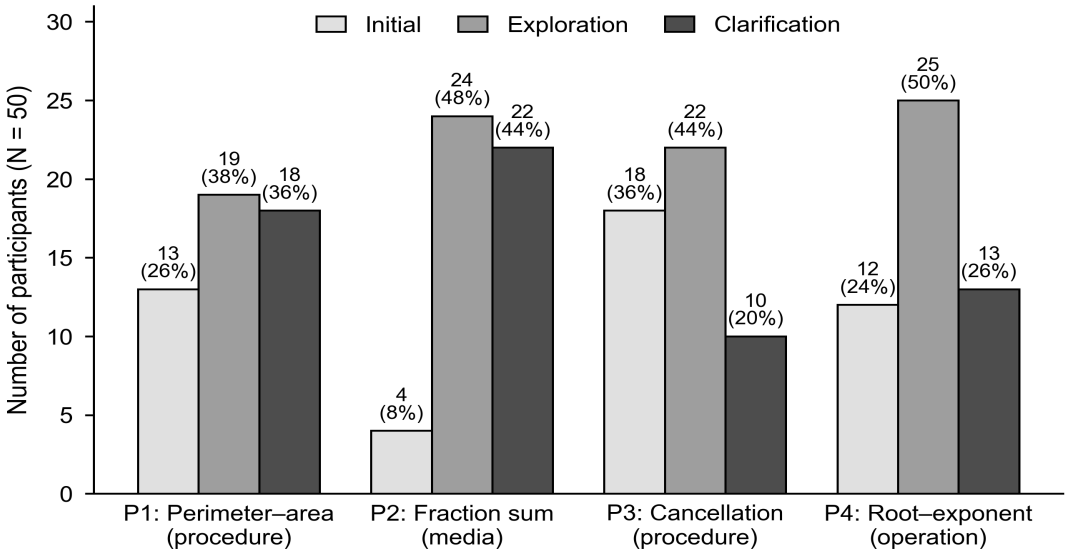



Figure 2 shows item-level distributions. On P2, 46 participants reached exploration or clarification, whereas P3 had the lowest clarification count (10). P4 had 25 exploration responses. These are descriptive differences among four tasks. Because P2 is the only media item and P4 the only operation item, the differences cannot be attributed to source independently of mathematical content, wording, representation, or difficulty.

## Why Results Are Reported by Task Rather Than Aggregated by Source

Responses are not aggregated by assigned source as evidence of source effects, because the denominators are unequal: procedure contains two tasks (P1 and P3; 100 task responses), whereas media and operation each contain one task (50 responses). Pooling would also conceal the substantial difference between P1 and P3. Results are therefore reported by task, and source labels are used only as provisional organising descriptors.

**Misleading-Media Controversy: The Bar Illustration in Problem 2**

Forty-six of fifty participants (92%) reached at least the exploration level on P2. Initial responses (n = 4) recognised that 2/6 was wrong but did not engage the figure. Exploration responses (n = 24) localised the illustration as the source of trouble but did not fully articulate a general principle. Clarification responses (n = 22) made the visual frame an explicit object of analysis. S1 (Figure 3A) redrew the illustration with the unshaded portions retained and concluded 2/3; the reconstruction made the invariant visible because the original bars and the resulting bar all preserve a three-segment whole.

S1's reasoning illustrates the relation between concept image and concept definition (Tall & Vinner, 1981): the original illustration fragmented the unit whole, whereas the reconstruction restored it. S41 reached the same conclusion verbally, distinguishing "one part of three" from "two parts of three" and explaining that the whole still contains three equal parts and that empty spaces should not be counted as additional thirds. The visual and verbal routes converge on preservation of the relevant invariant, consistent with Goldin's (2020) account of representational competence.

Interview accounts added process detail. Several participants described first asking what the picture was supposed to represent; one asked, "Are these the same whole? If they are, why are there six segments at the end?" This metarepresentational stance is consistent with Duval's (2017) account of register conversion. It documents strategies used on P2, the sole media item, and should not be interpreted as evidence that media-based controversies are generally easier or that the cohort has a stable source-specific competence.

**Misleading-Procedure Controversy: P1 and P3**

On P1, initial responses (n = 13) reproduced the linear-transfer assumption. S38 (Figure 3B) computed $K = 2 \times (30\% + 30\%) = 120\%$ and $L = 30\% \times 30\% = 900\%$, applying percentage expressions without first converting each dimension to a multiplicative scale factor. The work is useful as evidence of an algorithmic routine applied without inspection of its mathematical base (Fischbein, 1987).

Clarification responses (n = 18) made the multiplicative structure explicit. S20 (Figure 3C) derived $K_1 = 2(1.3p + 1.3l) = 1.3K_0$ and $L_1 = (1.3p)(1.3l) = 1.69L_0$, giving increases of 30% and 69%. The decisive move was the factorisation $1.30 \times 1.30 = 1.69$ rather than additive transfer of the two percentage increases.

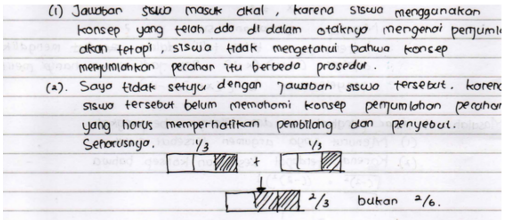
S22 and S48 reached the same result through numerical exemplification rather than symbolic derivation. S22 used p = 20 and l = 10 to obtain perimeters of 60 and 78 and areas of 200 and 338; S48 used p = 40 and l = 20 and verified the same proportional changes. The convergence of symbolic and numerical routes shows that clarification need not take one epistemic form, provided that the multiplicative structure is made explicit.

P3 produced the most resistant distribution: 18 participants (36%) remained at the initial level, 22 (44%) reached exploration, and 10 (20%) were coded at clarification. The correct mathematics is that $\frac{x-3}{x+1} = \frac{x-3}{x+2}$, with $x \neq -1, -2$, has the solution x = 3. Cross-multiplication gives $(x-3)(x+2) = (x-3)(x+1)$, hence $(x-3)[(x+2) - (x+1)] = 0$. Cancelling $x - 3$ before considering the zero-factor case divides by zero at the solution and produces the false statement 2 = 1.

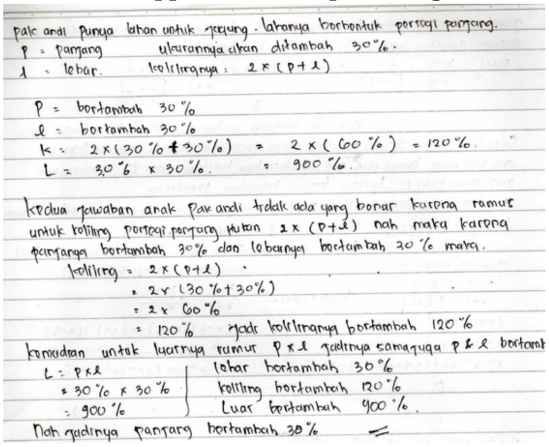
**Figure 3**

*Selected Written Responses Used in the Item-Level Analysis*

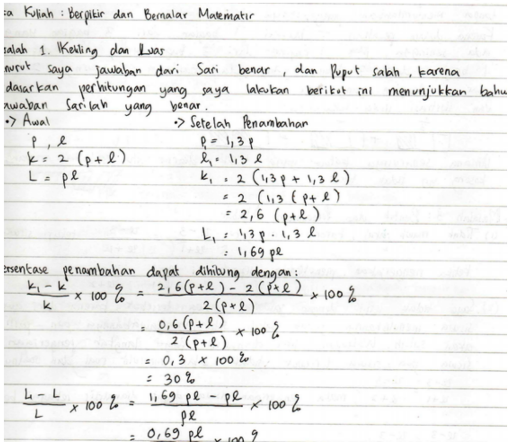
**(A) P2: S1 reconstructs the shared whole**



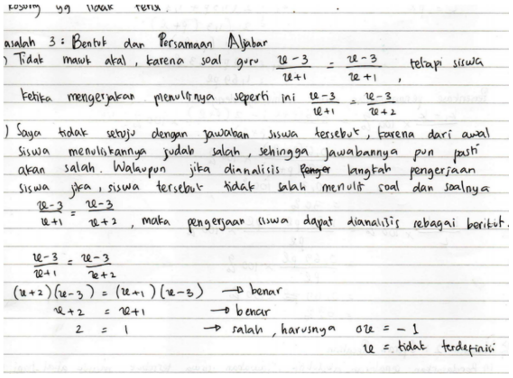

**(B) P1: S38 applies additive percentages**



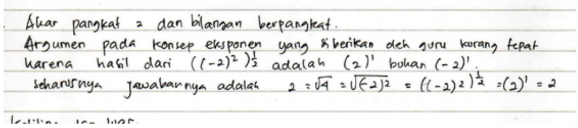

**(C) P1: S20 exposes multiplicative scaling**




**(D) P3: invalid cancellation retained as misconception evidence**




**(E) P4: S45 preserves the principal root**




*Note.* (A) S1 reconstructs the fraction bar; (B) S38 applies additive percentage reasoning; (C) S20 makes multiplicative scaling explicit; (D) S20 reproduces the invalid cancellation on P3 and concludes that no solution exists, retained as misconception evidence, not as a valid solution; (E) S45 preserves the principal-root convention.

Figure 3D shows the P3 response of S20: it follows the invalid cancellation and then treats the resulting contradiction, $0x = -1$, as evidence that the equation has no solution. It is included as evidence of the procedural difficulty, not as a mathematically valid clarification. The P3 level counts reported above come from the coding cycles described in the Data Analysis subsection, in which a response that located the cancellation step, rejected the student's conclusion, and supplied a mathematical reason was accepted

at the clarification level even when that reason ended in the claim that the equation has no solution; the response in Figure 3D was coded at that level. Because the claim is itself incorrect, the P3 clarification count should be read as the number of responses that localised and rejected the invalid step with a stated reason, not as the number that recovered x = 3, and under the stricter criterion adopted for the diagnostic markers in Table 2, which requires preserving or recovering x = 3 as the zero-factor case, the count would be lower. The central issue is therefore not an inconsistent cross-product or an absent solution, but a cancellation that changes the solution set when the cancelled factor may be zero.

**Misleading-Operation Controversy: Problem 4**

On P4, 12 participants (24%) remained at the initial level, 25 (50%) reached exploration, and 13 (26%) reached clarification. Initial responses recognised that $2 = -2$ was untenable but could not localise the responsible step. Exploration responses identified the root–exponent transition as suspicious without stating a governing rule. Clarification responses invoked the principal-root convention or $\sqrt{a^2} = |a|$ and explained why exponent laws cannot be transferred without restriction to negative real bases. S20 wrote that $\sqrt{(-2)^2} = |-2| = 2$ because absolute value is non-negative; S45 (Figure 3E) preserved the non-negative root through the chain of equalities.

The P4 accounts show a shift from recognising an unacceptable conclusion to naming a formal invariant. In interviews, exploration-level participants commonly described the offending step as strange or wrong but could not specify the violated condition; clarification-level participants appealed directly to a convention, identity, or domain restriction. This is an item-level pattern from the sole operation task, not evidence of a general operation-source effect.

**Synthesising the Two Dimensions: Source × Level**

Table 2 combines the candidate source labels with the initial, exploration, and clarification levels. The cells retain the recurring progression visible in the written and interview evidence: exploration localises a problematic component, whereas clarification names and applies the relevant invariant or precondition. The specific rule differs by task—the unit whole for P2, multiplicative scaling or the zero-factor condition for P1 and P3, and the principal-root convention for P4. The matrix remains a provisional coding and discussion heuristic; its reliability beyond this data set and its discriminant validity remain to be tested.

**Table 2**

*Diagnostic Markers Used in the Provisional Source × Level Matrix*

| Level | Media | Procedure | Operation |
|---|---|---|---|
| Initial | Rejects the result without analysing the representation. | Repeats a routine sequence without checking its conditions. | Rejects the conclusion without localising the invalid rule use. |
| Exploration | Questions whether the representation preserves the same whole. | Locates a suspect scaling or cancellation step but gives no complete repair. | Locates the root–exponent step but does not state the governing invariant. |
| Clarification | Reconstructs the representation and states the preserved invariant. | Makes the hidden condition explicit; for P3, preserves x = 3 as the zero-factor case. | Invokes $\sqrt{a^2} = \|a\|$ and restrictions on exponent laws. |

*Note.* These markers describe the present tasks and require validation with broader, balanced item sets.

The matrix should guide questions rather than generate scores. Reading across a row can help a teacher educator compare what localisation or justification looks like in different tasks, while reading down a column can clarify the transition from recognition to structural repair. Apparent asymmetries across tasks

may motivate follow-up questions, but the present design does not establish stable learner profiles or source-specific abilities. No instructional intervention was implemented, and no learning or teaching outcomes were measured.

## DISCUSSION

### Position Relative to Existing Frameworks

The source-based typology is positioned alongside, rather than in opposition to, accounts of paradoxes (Movshovitz-Hadar & Hadass, 1990, 1991), fallacies (Mueller & Yankelewitz, 2014), cognitive conflict (Limón, 2001; Zaslavsky, 2005), and levels of controversial reasoning (Subanji et al., 2021). The paradox tradition asks what a counterintuitive task can do for a learner, the fallacy tradition examines how invalid argumentation can unsettle misconceptions, and the level framework describes how far a learner progresses toward resolution. The proposed source labels instead ask where the contradiction appears within a task's epistemic objects. The present evidence shows that source and level labels can be combined descriptively; it does not establish that they are psychometrically orthogonal dimensions.

Three positioning points follow. First, the typology complements the level framework by recording a task feature alongside the observed reasoning level; it does not establish a stable diagnostic profile for an individual learner. Second, it complements the paradox literature by distinguishing whether the relevant task feature is representational, sequential, or operational. Third, it complements cognitive-conflict accounts by identifying the object around which conflict is organised, not by claiming that a distinct psychological mechanism has been demonstrated. The typology is therefore a provisional task-analytic vocabulary rather than a validated assessment.

### Concept Image–Definition and Algorithmic Intuition as Lenses

Tall and Vinner's (1981) concept image–definition distinction illuminates P2: the bar illustration recruits a concept image that fragments the unit whole, while clarification reconstructs alignment between image and definition through visual or verbal reasoning. Fischbein's (1987) account of algorithmic intuition illuminates P1 and P3, where percentage scaling and cancellation are deployed as self-evident routines until their preconditions are examined, a pattern consistent with the interdependence of procedural and conceptual knowledge (Rittle-Johnson & Schneider, 2015). P4 concerns a unitary formal constraint: the principal square root is non-negative, $\sqrt{a^2} = |a|$, and exponent laws over negative real bases require domain restrictions. These interpretations connect the detailed item-level evidence to existing theory without treating four tasks as a definitive partition of mathematical error.

### Task-Level Differences and Design Constraints

The four tasks produced different descriptive distributions. On P2, 92% of participants reached exploration or clarification; P3 had the lowest clarification count (10 of 50); and P4 had an exploration-heavy pattern (25 of 50). These task-level observations should not be converted into a source-level difficulty ranking. P2 and P4 each instantiate a category with one item, and P1 and P3 differ substantially despite sharing the procedure label. Item content, wording, familiarity, representation, and inherent difficulty are therefore confounded with source. The distributions motivate hypotheses for a balanced multi-item study; they do not demonstrate that media controversies are easier or procedure controversies harder.

## Implications for Pre-service Teacher Education

Three cautious implications follow for higher education institutions that prepare mathematics teachers. First, methods courses can invite pre-service teachers to ask whether an error is connected to a representation, procedure, or operation, while treating these as prompts rather than fixed diagnoses. Second, analysis should make mathematical preconditions explicit, including preservation of the unit whole, multiplicative scaling, zero-factor cases, and principal-root conventions. Third, teacher educators can compare multiple tasks within each proposed category before drawing conclusions about learner profiles. These practices align with work on performance assessment, powerful classroom reasoning, problem-solving processes, and case-based analysis of student errors (Barron, 2015; Koç & Bingölbali, 2026; Rodil, 2024; Schoenfeld, 2014).

One possible instructional sequence would begin with reconstruction of a misleading representation, continue with step-by-step inspection of the preconditions licensing an algorithmic procedure, and end with articulation of the domain governing an operation. Such a sequence would connect register conversion (Duval, 2017), concept image and definition (Tall & Vinner, 1981), and algorithmic intuition (Fischbein, 1987). It is an implication generated from the analysis, not an intervention implemented or tested in this study. Its effectiveness would require comparison with alternative instructional designs.

The framework could also be piloted as a structured discussion activity using authentic student errors. Working in pairs, pre-service teachers might identify the suspected task feature, state the violated mathematical invariant or precondition, and design a response targeted to that feature. The two-axis matrix could organise discussion of both the proposed source and the quality of justification. In the present state, however, it is a discussion scaffold rather than a validated marking rubric or assessable course component. Research should establish scoring reliability, validity, and effects on teacher learning before consequential use.

At programme level, a proportionate next step would be a small pilot within existing methods or assessment courses. Programmes could compare student analyses of representation-, procedure-, and operation-focused tasks and examine whether the vocabulary improves diagnostic discussion without adding a mandated curriculum requirement. A source-aware error analysis might also be trialled in a teaching-practicum portfolio. The present findings do not justify certification standards, required curricular changes, or claims of effectiveness.

## Comparison with International Studies

The findings echo international work showing that paradoxes and counterintuitive tasks can generate productive uncertainty (Movshovitz-Hadar & Hadass, 1990, 1991; Zaslavsky, 2005) and that pre-service teachers' errors vary with proof-writing, modelling, and other task demands (An et al., 2026; Segura et al., 2026). Zaslavsky's categories of uncertainty also map loosely onto the present concern with where a difficulty is located, although the correspondence is not exact. The narrower contribution of this study is a provisional vocabulary and coding scheme for locating contradictions within these four tasks; cross-cultural robustness and source-level effects remain untested.

## Sources Beyond the Three Proposed

The three categories are not closed. Probability controversies may involve misleading definitions, such as a formal account of independence that conflicts with an everyday concept image. Geometry tasks may involve misleading analogies in which a familiar relation is projected onto a structurally different figure, while calculus tasks may involve intuitive limits that diverge from formal definition. Hybrid cases

are also plausible. The typology should therefore be treated as a working framework that can absorb additional sources only after their epistemic distinctness is demonstrated in broader inductive studies.

## CONCLUSION

### Summary

Three findings address the research questions. First, four tasks designed a priori for the candidate labels illustrated how media, procedure, and operation can be operationalised, without showing that the categories emerge independently. Second, the four tasks produced different descriptive reasoning patterns; those differences remain confounded with item content and difficulty and cannot be attributed to source. Third, source and level labels can be combined in a two-axis matrix that offers a provisional way to organise analysis. The evidence does not establish source effects, a validated instrument, or an effective educational intervention.

### Theoretical and Practical Implications

Theoretically, the typology shifts attention from how far a learner has progressed to the task feature around which the contradiction is organised. As a practical heuristic, it can prompt teacher educators in university programmes to ask what representation, procedural step, or operational condition requires attention. It complements rather than replaces existing level-based accounts, and its use should remain exploratory until validated with balanced item sets and independent samples.

The practical implication concerns the diagnostic stance that teacher educators may wish to cultivate. A future teacher who interprets an incorrect fraction sum as connected to a misleading bar illustration may respond differently from one who attributes it to an algorithmic shortcut. Source-aware questions can therefore enrich anticipation and discussion of student thinking (Stein et al., 2008) and case-based work on fraction errors (Koç & Bingölbali, 2026). The present study did not test whether use of the typology changes teaching decisions or student outcomes.

### Limitations

Several limitations qualify the contribution. First, source and specific problem are confounded: P2 is the sole media item, P4 the sole operation item, and only procedure is represented by two tasks. Differences in content, wording, representation, familiarity, or difficulty cannot therefore be separated from the assigned source. Second, the tasks and source labels were designed a priori; expert review supported mathematical and content appropriateness but did not provide independent construct validation. Third, source-coding agreement reflects consistent application of known task labels rather than evidence that the typology has discriminant validity. Fourth, the boundary between procedure and operation remains porous, and the four tasks cannot establish that the categories are exhaustive. Fifth, the sample came from one university, so percentages describe this cohort only. Sixth, the study did not implement or evaluate an educational intervention, and the matrix should not be presented as a demonstrated instructional effect or validated assessment tool. Seventh, the P3 level counts were produced under a coding criterion that accepted the incorrect conclusion that the equation has no solution; relative to the stricter criterion stated in the Misleading-Procedure Controversy subsection of the Findings and Table 2, the P3 clarification count is an upper bound. Finally, gender differences are not analysed because the subgroup sizes, the absence of baseline knowledge measures, and the qualitative design do not support an inferential claim.

The evidence base comprises written responses from 50 pre-service teachers at one Indonesian state university and follow-up interviews with 16 participants. It supports detailed interpretation of reasoning within these four tasks and shows how the proposed labels can be applied to this material. It does not

establish how frequently the patterns occur among pre-service teachers generally, nor does it support causal claims, population estimates, source-difficulty rankings, or claims that the categories will transfer unchanged across programmes, countries, and mathematical domains.

**Future Research**

Future research should use several matched items per proposed source, counterbalance order and surface features, and include tasks not designed for the typology. Independent judges should classify tasks without access to intended labels, and confirmatory work should examine reliability, discriminant validity, and alternative category structures. Comparative studies across cultural and curricular contexts could then examine whether task-level patterns are robust. Only after these steps should intervention studies compare source-aware analysis with alternative instructional or level-aware approaches.

Longitudinal research could track the same pre-service teachers across courses to determine whether progress on one task type transfers to others. Classroom-based studies should also observe whether the diagnostic stance appears when pre-service teachers interact with school students, rather than only when they analyse written prompts. Together, multi-item validation, longitudinal follow-up, and classroom observation would clarify whether the matrix can move from a promising heuristic toward an evidence-supported educational resource.

**Author Contributions:** The authors confirm contribution to the paper as follows. Erry Hidayanto: conceptualisation, methodology, study design, supervision, writing – review and editing, and final approval of the manuscript. Subanji: conceptualisation, theoretical framework, validation of the source-based typology, writing – review and editing, and final approval of the manuscript. Rivan Gestiardi: data curation, formal analysis, coding of written responses and interview transcripts, visualisation, and writing – original draft preparation. All authors discussed the results, contributed to the interpretation of the findings, and approved the final version of the manuscript.

**Funding:** This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. The authors acknowledge institutional support from Universitas Negeri Malang for data collection and analysis.

**Institutional Review Board Statement:** Institutional permission to conduct the study was obtained from Universitas Negeri Malang. All participants provided written informed consent before data collection.

**Acknowledgments:** The authors thank the participants and the expert reviewers who examined the mathematical tasks.

**Declaration of Generative AI and AI-Assisted Technologies in the Writing Process:** During final revision, the authors used OpenAI Codex as a verification aid for reference details, DOI and URL validity, mathematical consistency, and compliance with the journal's formatting and reviewer checklist. It was not used to collect or analyse the research data. The authors reviewed and approved all revisions and take full responsibility for the study, its interpretation, and the content of the publication.

**Conflicts of Interest:** The authors declare no competing financial interests, personal relationships, or institutional affiliations that could have appeared to influence the work reported in this paper.

**Disclaimer/Publisher's Note:** The statements, opinions, and data in this article are solely those of the authors. The publisher and editors disclaim responsibility for any injury to people or property resulting from ideas, methods, instructions, or products referred to in the content.

## REFERENCES


Akkus, R., Yılmaz-Akkurt, Y., & Yüksel, M. E. (2026). Exploring self-reported experiences of pre-service mathematics teachers within a problem-to-prove task through a cognitive-affective lens. *International Journal of Science and Mathematics Education*, *24*(5), Article 52. https://doi.org/10.1007/s10763-026-10678-8

An, T., Clark, D., & Zimerle, I. (2026). Analyzing errors in preservice middle-grade mathematics teachers' proof writing. *School Science and Mathematics*. Advance online publication. https://doi.org/10.1111/ssm.70017

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

Barron, L. (2015). Preparing pre-service teachers for performance assessments. *Journal of Interdisciplinary Studies in Education*, *3*(2), 68–75. https://ojed.org/jise/article/view/1475

Duval, R. (2017). *Understanding the mathematical way of thinking — The registers of semiotic representations*. Springer. https://doi.org/10.1007/978-3-319-56910-9

Fischbein, E. (1987). *Intuition in science and mathematics: An educational approach*. D. Reidel.

Goldin, G. A. (2020). Mathematical representations. In S. Lerman (Ed.), *Encyclopedia of mathematics education* (2nd ed., pp. 566–572). Springer. https://doi.org/10.1007/978-3-030-15789-0_103

Hidayah, I. N., Sa'dijah, C., Subanji, & Sudirman. (2020). Characteristics of students' abductive reasoning in solving algebra problems. *Journal on Mathematics Education*, *11*(3), 347–362. https://doi.org/10.22342/jme.11.3.11869.347-362

Hsieh, H.-F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, *15*(9), 1277–1288. https://doi.org/10.1177/1049732305276687

Jeannotte, D., & Kieran, C. (2017). A conceptual model of mathematical reasoning for school mathematics. *Educational Studies in Mathematics*, *96*(1), 1–16. https://doi.org/10.1007/s10649-017-9761-8

Koç, Y., & Bingölbali, E. (2026). Exploring pre-service teachers' handling of student errors in fractions: A case-based study. *Croatian Journal of Education*, *28*(3), 937–963. https://doi.org/10.15516/cje.v28i3.41363

Lee, G. C.-Y. (2026). Conditions for change in pre-service teachers' beliefs about proof and proof teaching. *Mathematics Education Research Journal*, *38*(1), 187–218. https://doi.org/10.1007/s13394-025-00525-4

Limón, M. (2001). On the cognitive conflict as an instructional strategy for conceptual change: A critical appraisal. *Learning and Instruction*, *11*(4–5), 357–380. https://doi.org/10.1016/S0959-4752(00)00037-2

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Sage.

Movshovitz-Hadar, N., & Hadass, R. (1990). Preservice education of math teachers using paradoxes. *Educational Studies in Mathematics*, *21*(3), 265–287. https://doi.org/10.1007/BF00305093

Movshovitz-Hadar, N., & Hadass, R. (1991). More about mathematical paradoxes in preservice teacher education. *Teaching and Teacher Education*, *7*(1), 79–92. https://doi.org/10.1016/0742-051X(91)90059-X

Mueller, M., & Yankelewitz, D. (2014). Fallacious argumentation in student reasoning: Are there benefits? *European Journal of Science and Mathematics Education*, *2*(1), 27–38. https://doi.org/10.30935/scimath/9398

Mueller, M. F., Yankelewitz, D., & Maher, C. (2010). Promoting student reasoning through careful task design: A comparison of three studies. *International Journal for Studies in Mathematics Education*, *3*(1), 87–107.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. NCTM.

Rittle-Johnson, B. (2017). Developing mathematics knowledge. *Child Development Perspectives*, *11*(3), 184–190. https://doi.org/10.1111/cdep.12229

Rittle-Johnson, B., & Schneider, M. (2015). Developing conceptual and procedural knowledge of mathematics. In R. Cohen Kadosh & A. Dowker (Eds.), *The Oxford handbook of numerical cognition* (pp. 1118–1134). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199642342.013.014

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

Rodil, M. S. (2024). Gearing towards efficient problem solving: Exploring the determinants that influence success. *Journal of Interdisciplinary Studies in Education*, *13*(2), 120–151. https://doi.org/10.32674/thv32d10

Schoenfeld, A. H. (2014). What makes for powerful classrooms, and how can we support teachers in creating them? A story of research and practice, productively intertwined. *Educational Researcher*, *43*(8), 404–412. https://doi.org/10.3102/0013189X14554450

Schoenfeld, A. H. (2020). Mathematical practices, in theory and practice. *ZDM Mathematics Education*, *52*(6), 1163–1175. https://doi.org/10.1007/s11858-020-01162-w

Segura, C., Ferrando, I., Gallart, C., & Barquero, B. (2026). Pre-service teachers posing modelling problems: Analysis of strategies and errors. *Canadian Journal of Science, Mathematics and Technology Education*, *26*(3), Article 57. https://doi.org/10.1007/s42330-026-00487-0

Stein, M. K., Engle, R. A., Smith, M. S., & Hughes, E. K. (2008). Orchestrating productive mathematical discussions: Five practices for helping teachers move beyond show and tell. *Mathematical Thinking and Learning*, *10*(4), 313–340. https://doi.org/10.1080/10986060802229675

Subanji, Rosyadi, A. A. P., & Emanuel, E. P. L. (2021). Levels of controversial reasoning of the pre-service teachers to solve mathematical problems. *Journal of Southwest Jiaotong University*, *56*(4), 645–658. https://doi.org/10.35741/issn.0258-2724.56.4.55

Tall, D., & Vinner, S. (1981). Concept image and concept definition in mathematics with particular reference to limits and continuity. *Educational Studies in Mathematics*, *12*(2), 151–169. https://doi.org/10.1007/BF00305619

Zaslavsky, O. (2005). Seizing the opportunity to create uncertainty in learning mathematics. *Educational Studies in Mathematics*, *60*(3), 297–321. https://doi.org/10.1007/s10649-005-0606-5

---

**ERRY HIDAYANTO** is a faculty member of the Department of Mathematics, Faculty of Mathematics and Natural Sciences (FMIPA), Universitas Negeri Malang, Indonesia. His research interests include mathematical thinking and reasoning, mathematics teacher education, and the design of cognitively challenging tasks for pre-service mathematics teachers. He has long-standing experience teaching the Mathematical Thinking and Reasoning course (Berpikir dan Bernalar Matematis) and supervises undergraduate and graduate research in mathematics education. Email: erry.hidayanto.fmipa@um.ac.id

**SUBANJI** is a senior faculty member of the Department of Mathematics, Faculty of Mathematics and Natural Sciences (FMIPA), Universitas Negeri Malang, Indonesia. His research focuses on mathematical reasoning, problem-solving processes, covariational reasoning, controversial reasoning of pre-service mathematics teachers, and the development of mathematics teacher education in Indonesia. He has authored numerous publications in national and international journals on mathematics learning processes and serves as a doctoral supervisor in the Postgraduate Program of Universitas Negeri Malang. Email: subanji.fmipa@um.ac.id

**RIVAN GESTIARDI** is an independent researcher based in Wagir, Malang, Indonesia. He holds a doctoral degree from the Postgraduate Program of Universitas Negeri Malang and has published peer-reviewed articles on character education, technology readiness in learning, and quality management in elementary schools. His current research concerns pre-service teachers' mathematical and pedagogical reasoning. Email: gestiardirr@gmail.com