Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

From Recognition to Resolution: A Four-Phase Cognitive Model of Pre- service Teachers' Reasoning on Counterintuitive Mathematics Problems

Erry Hidayanto

erry.hidayanto.fmipa@um.ac.id

Universitas Negeri Malang

Subanji Subanji

Universitas Negeri Malang

Rivan Gestiardi

Universitas Negeri Malang

Research Article

Keywords: Cognitive conflict, Conceptual change, Counterintuitive reasoning, Pre-service mathematics teachers, Phenomenography, Teacher noticing

Posted Date: June 19th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9793104/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

From Recognition to Resolution: A Four-Phase Cognitive Model of Pre-service Teachers' Reasoning on Counterintuitive Mathematics Problems

Erry Hidayanto[1*], Subanji[1], Rivan Gestiardi[2]

[1] Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Negeri Malang, Malang, Indonesia

[2] Doctoral Program in Elementary Education, Postgraduate, Universitas Negeri Malang, Malang, Indonesia

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

Author Biographies

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.

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.

Rivan Gestiardi holds a doctoral degree in the Postgraduate Program of Universitas Negeri Malang, Indonesia. He has published peer-reviewed articles on character education, technology readiness in learning, and quality management in elementary schools, and is currently engaged in research on pre-service teachers' mathematical and pedagogical reasoning.

From Recognition to Resolution: A Four-Phase Cognitive Model of Pre-service Teachers' Reasoning on Counterintuitive Mathematics Problems

Abstract

Counterintuitive mathematics problems create cognitive conflict that can lead to conceptual change. This study identifies four cognitively distinct phases by which pre-service teachers move from sensing a contradiction to producing a teachable resolution, drawing together phenomenography, conceptual change theory, and mathematics teacher noticing. From fifty Indonesian pre-service teachers who responded to four problems on percentage growth, fraction addition, cross-multiplication, and the order of square root and exponent, sixteen cases were selected for in-depth analysis. Written responses were triangulated with clinical interviews and analysed phenomenographically. The four phases—Recognition, Identification, Controversy Awareness, and Resolution—each correspond to one of the conditions for conceptual change proposed by Posner et al. (1982). The transition from Phase 2 to Phase 3, which requires structural vocabulary, was the most demanding, and a backward loop between Phases 3 and 2 occurred in five episodes, a phenomenon warranting further study. By operationalising the conditions for conceptual change as observable, within-task behaviours, the model offers a process-level complement to outcome-based taxonomies and a transferable diagnostic lens for pre-service teacher education. Because the four phases are defined by cognitive function rather than by problem content, they are proposed as domain-general markers of conceptual change; their demonstration here in arithmetic and elementary algebra invites validation across further mathematical domains.

Keywords Cognitive conflict · Conceptual change · Counterintuitive reasoning · Pre-service mathematics teachers · Phenomenography · Teacher noticing

1 Introduction

Counterintuitive problems are central to mathematics learning. They confront the learner with arguments that look valid yet produce results that cannot be accepted, and they invite the cognitive work of locating the source of contradiction and reconstructing a justification (Movshovitz-Hadar & Hadass, 1990; Zaslavsky, 2005). For pre-service mathematics teachers, this work is doubly important. They must reason mathematically to resolve the contradiction, and they must reason pedagogically to anticipate why a student might have produced it (Ball et al., 2008; Jacobs et al., 2010).

The cognitive psychology of conceptual change provides a productive frame. Posner et al. (1982) proposed that conceptual change requires four conditions: dissatisfaction with the current conception, intelligibility of the

alternative, plausibility of the alternative, and fruitfulness of the alternative for further inquiry. Vosniadou and Skopeliti (2014) refined the account through the framework theory approach, in which counterintuitive science and mathematics tasks expose the gap between learners' intuitive frameworks and the formal structures of the discipline. Limón (2001) and Pacaci et al. (2024) have shown that cognitive conflict alone produces small effects, while conflict combined with structured scaffolding produces large effects. The path from conflict onset to resolution is therefore not automatic; it is mediated by metacognitive monitoring, structural vocabulary, and the learner's motivation to engage.

What this body of work has not yet produced is a fine-grained model of the within-task cognitive process. Most conceptual change research operates at the timescale of weeks or months and focuses on stable misconceptions about specific concepts. The within-task process, on the timescale of minutes within a single problem-solving episode, has received less attention in the mathematics education literature. Process accounts that exist (Pólya, 1945; Mason et al., 1982; Sfard, 1991) were designed for standard problem solving or for concept acquisition, not for the specific situation in which a contradiction is built into the task itself. The mathematics teacher noticing literature (Sherin et al., 2011; Jacobs et al., 2010) has developed the concept of professional vision but has not articulated a process model of how pre-service teachers themselves reason through the counterintuitive tasks they will later encounter as teachers.

This study proposes a four-phase cognitive model for the within-task process. The four phases (Recognition, Identification, Controversy Awareness, and Resolution) were grounded in phenomenographic analysis of sixteen cases drawn from a cohort of fifty Indonesian pre-service mathematics teachers. The contribution is threefold and addresses a gap that current accounts leave open at the intersection of conceptual change theory, mathematical problem solving, and teacher noticing. First, to our knowledge the model is the first to operationalise the four Posner et al. (1982) conditions for conceptual change as observable behaviours within a single counterintuitive task, turning a set of static prerequisites into a traceable process. Second, it locates the Phase 2 to Phase 3 transition as the empirical bottleneck where structural vocabulary mediates progress, pinpointing where instructional support is most consequential. Third, because the phases are defined by cognitive function rather than by problem content, the model equips teacher educators with a process-level diagnostic lens for pre-service teacher noticing that is intended to generalise beyond the specific tasks studied here. The model is consistent with the regional level account of Subanji et al. (2021) yet is grounded in the international literature on cognitive conflict and conceptual change rather than derived from a single antecedent.

Three research questions guide the study. RQ1: What cognitively distinct phases can be identified in pre-service teachers' reasoning as they work through counterintuitive mathematics problems? RQ2: What are the characteristic transitions between phases, and where do learners most often stall? RQ3: How does the proposed model relate to the international literature on cognitive conflict, conceptual change, and mathematical reasoning? The paper proceeds with the theoretical framework (Section 2), methods including a reflexivity statement and verbatim presentation of the four problems (Section 3), findings (Section 4), discussion (Section 5), and conclusion (Section 6).

2 Theoretical Framework and Related Literature

2.1 Cognitive Conflict and Conditions for Conceptual Change

The classical conceptual change framework of Posner et al. (1982) identifies four conditions for accommodation. The first is that "there must be dissatisfaction with existing conceptions" (Posner et al., 1982, p. 214); the alternative conception must then be intelligible, plausible, and fruitful for further inquiry. These conditions were originally formulated for science learning but have been applied to mathematics by Limón (2001) and others. Vosniadou and Skopeliti (2014) refined the account through the framework theory approach, arguing that children's naive conceptions form "a relatively coherent conceptual system" (Vosniadou & Skopeliti, 2014, p. 1427) that resists change because it is functionally adequate for everyday reasoning. On this view, counterintuitive tasks surface the gap between the framework theory and the formal structure of the discipline.

Recent work has continued to refine the account. Pacaci et al. (2024), in a meta-analysis of conceptual change interventions, found that cognitive conflict alone produces small effects, but conflict combined with structured scaffolding produces large effects. Sinatra and Pintrich (2003) earlier emphasised intentional engagement, distinguishing cold cognitive change (driven only by logical conflict) from warm cognitive change (driven by motivation and self-regulation). The present study contributes to this body of work by examining how the conditions for conceptual change are met or unmet at the within-task scale, in pre-service teachers facing counterintuitive mathematics problems. We argue that the four phases proposed here can be read as observable markers of the four Posner conditions: dissatisfaction marks the entry to Phase 1, intelligibility marks the entry to Phase 2, plausibility marks the entry to Phase 3, and fruitfulness marks the entry to Phase 4.

2.2 Counterintuitive Tasks in Mathematics Education

Mathematics education has a long tradition of using tasks that confront learners with surprising results. Movshovitz-Hadar and Hadass (1990, 1991) documented how mathematical paradoxes sharpen pre-service teachers' conceptual understanding. Mueller and Yankelewitz (2014) showed how fallacious arguments can

productively unsettle procedural misconceptions. Zaslavsky (2005) classified the kinds of uncertainty that mathematical tasks induce. More recent work by Buchbinder and McCrone (2023) on examples in proving has examined how pre-service teachers engage with counterexamples that reveal hidden assumptions. What unites this literature is the recognition that the value of counterintuitive tasks lies not in the correctness of the final answer but in the cognitive work that resolution requires. The present study contributes by characterising that cognitive work as a sequence of qualitatively distinct phases.

2.3 Mathematics Teacher Noticing

Mathematics teacher noticing (Sherin et al., 2011; Jacobs et al., 2010) refers to the trained ability of teachers to attend to, interpret, and decide how to respond to student mathematical thinking. Pre-service teachers typically struggle to develop this professional vision without explicit support (Buchbinder & McCrone, 2020). Recent intervention studies indicate that structured programmes can measurably develop pre-service teachers' noticing of student mathematical thinking (Weingarden & Buchbinder, 2023; Baş-Ader et al., 2025). The present model contributes a within-task process framework that pre-service teachers can use to analyse student reasoning. A teacher who can identify which phase a student is at can design a follow-up question that targets the next transition rather than a generic correction, which is the diagnostic stance Stein et al. (2008) identified as productive for orchestrating mathematical discussions.

2.4 Phenomenography as Methodological Warrant

Phenomenography studies variation in the ways people experience a phenomenon (Marton & Booth, 1997). Categories of description that emerge from phenomenographic analysis represent qualitatively different ways of experiencing the same phenomenon. When these categories are ordered and linked by transition conditions, they constitute a process model (Åkerlind, 2005). Phenomenography is well suited to the present study because it foregrounds variation across cases and treats the cognitive structure of experience as the primary object of analysis. Åkerlind (2005) argues that ten to twenty cases are typically sufficient when variation has been adequately mapped, which justifies the present sample of sixteen interview cases.

2.5 Concept Image, Concept Definition, and Dual Process

The conflict in counterintuitive problems often originates in a gap between a learner's concept image—"the total cognitive structure that is associated with the concept" (Tall & Vinner, 1981, p. 152)—and the formal concept definition. Phase 2 and Phase 3 of the proposed model can be interpreted as the learner progressively narrowing this gap. Fischbein's (1987) account of algorithmic intuition (schemes that feel self-evidently valid but lack formal justification) explains why the gap is not visible at the start, and why Phase 1 is often accompanied by confusion

rather than diagnosis. Recent work in dual-process theory (Leron & Hazzan, 2009) suggests that algorithmic intuition operates at the System 1 level (fast, automatic), while structural articulation requires System 2 effort (slow, deliberate). The Phase 2 to Phase 3 transition can be read as the activation of System 2 reasoning under cognitive conflict.

2.6 Positioning the Proposed Model

The four-phase model occupies the intersection of three research traditions. From conceptual change research (Posner et al., 1982; Vosniadou & Skopeliti, 2014; Pacaci et al., 2024), it inherits the focus on conditions for cognitive accommodation. From the mathematical paradox tradition (Movshovitz-Hadar & Hadass, 1990; Zaslavsky, 2005; Mueller & Yankelewitz, 2014), it inherits the task genre of counterintuitive problems. From the teacher noticing literature (Sherin et al., 2011; Jacobs et al., 2010), it inherits the pedagogical orientation toward diagnostic engagement with student reasoning. The model is not an extension of Pólya (1945), Mason et al. (1982), or Sfard (1991); those accounts were designed for different task genres or different timescales. The level taxonomy of Subanji et al. (2021) provides a regional outcome framework with which the present model is consistent, but the empirical and theoretical grounding lies in the international cognitive conflict literature.

3 Methods

3.1 Research Design

The study used a phenomenographic research design with purposive case selection. The full dataset comprised written responses from fifty pre-service mathematics teachers to four counterintuitive problems, supplemented by clinical interviews with sixteen selected participants. Phenomenographic analysis focused on the sixteen interview cases. Following Åkerlind (2005), variation in the ways participants experienced the counterintuitive task was the primary object of analysis. We treated the reasoning episode (the full sequence of cognitive moves a participant made on a given problem) as a unit of observation rather than as an independent case; the cases for phenomenographic comparison are the sixteen participants, and the four episodes per participant provide breadth of variation within each case. This addresses the concern that phenomenography typically operates on inter-individual variation: each of the sixteen individuals contributes the primary case-level variation, while the four problems provide within-case variation that supports the identification of source-specific patterns. We adopted the developmental variant of phenomenography (Åkerlind, 2005), in which the categories of description are treated as a hierarchically related outcome space rather than as a set of unrelated types; this variant is well suited to building a process model because it foregrounds the transitions between qualitatively different ways of experiencing the task. Sampling and analysis continued until no new category of description or

inter-category transition emerged from additional cases, which we took as evidence that the outcome space had stabilised.

3.2 Reflexivity and Positionality

Phenomenographic studies require explicit reflexivity (Lincoln & Guba, 1985; Åkerlind, 2005). The first author is a mathematics education lecturer at the institution where data were collected. He has previously taught the course Berpikir dan Bernalar Matematis but did not teach the cohort during the data collection period; his relationship with the participants was that of a post-course evaluator, not their direct instructor. The other authors have no institutional relationship with the participants. The first author's familiarity with the curriculum and with Indonesian mathematics teacher education is a methodological asset, supporting interpretation of student reasoning, but it also carries a risk of confirmatory interpretation. Three steps mitigated this risk. First, all coding was conducted by both researchers independently before joint discussion. Second, six participants were invited to member-check the phase classifications applied to their interviews. Third, two external mathematics education researchers (not authors) reviewed a subsample of coded transcripts and noted that the phase distinctions were defensible from outside the institutional context.

3.3 Participants and Setting

Participants were fifty pre-service mathematics teachers enrolled in the compulsory course Berpikir dan Bernalar Matematis (Mathematical Thinking and Reasoning) at a state university in East Java, Indonesia. All had completed prior coursework in calculus, real analysis, abstract algebra, and mathematics pedagogy. The cohort comprised thirty-two female and eighteen male students. The examination was administered in December 2025 under standard closed-book conditions. Each participant gave written informed consent and was assured of anonymity through pseudonyms (S1 to S50). Ethical approval was obtained from the institutional research ethics committee.

3.4 The Four Counterintuitive Problems (Verbatim Instrument)

The instrument consisted of four problems designed to instantiate qualitatively different sources of cognitive conflict. The problems are reproduced verbatim below, translated from Indonesian. The original Indonesian text was used in the examination, and the complete instrument in the original Indonesian, together with the full phase-coding rubric, is provided in Online Resource 1.

Table 1 *The four counterintuitive problems used in the study (translated verbatim from the original Indonesian instrument)*

P1	Pak Andi has a rectangular field with length p and width l. The price of corn is rising, so he plans to enlarge the length and width by 30% each. His two children give different answers about the

	result. Puput says: perimeter increases by 60%, area increases by 9%. Sari says: perimeter increases by 30%, area increases by 69%. Whose answer is correct, and why?
P2	A teacher asks her student to compute 1/3 + 1/3. The student answers 2/6, supporting the answer with this picture: [two bars each shaded in 1 of 3 segments are physically merged into a single bar with 6 segments, 2 of which are shaded]. (a) Does the student's reasoning make sense? Explain. (b) Do you agree with the answer? Give your reasoning.
P3	A student writes: $(x - 3)/(x + 1) = (x - 3)/(x + 2)$. Cross-multiplying: $(x + 2)(x - 3) = (x + 1)(x - 3)$. Expanding: $x + 2 = x + 1$, therefore $2 = 1$. Does the student's reasoning make sense? Where, if anywhere, does the argument fail?
P4	A student writes: $2 = \sqrt{4} = \sqrt{((-2)^2)} = ((-2)^2)\text{^}(1/2) = (-2)^1 = -2$, therefore $2 = -2$. Does the student's reasoning make sense? If not, where does it break down, and how would you explain this to the student?

The four problems were reviewed by three expert reviewers (two senior mathematics educators and one professor of mathematics) and revised through two rounds before administration. Reviewers judged whether the controversy in each problem was clearly traceable to the intended source, whether the mathematics was appropriate for the cohort, and whether the problem was open enough to allow multiple solution paths. P1 and P3 were designed as misleading-procedure tasks; P2 as misleading-media; P4 as misleading-operation.

3.5 Case Selection for Phenomenographic Analysis

Sixteen cases were selected from the cohort using stratified purposive logic. Three criteria guided the selection. Phase diversity required at least two cases that appeared, from written responses alone, to stall at each provisional phase. Source diversity required at least three cases per source type to allow source-specific patterns to be probed. Transition cases required four cases whose written responses showed evidence of incomplete or interrupted transitions, providing the richest data for boundary description. The final sample comprised nine females and seven males. Six were initially classified at the clarification level, six at exploration, and four at initial on their primary analysis problem.

3.6 Data Collection

Each of the sixteen selected participants completed a 25- to 40-minute clinical interview within two weeks of the examination. Interviews followed a semi-structured protocol with three components. First, a think-aloud reconstruction of the participant's reasoning on their primary analysis problem (Ericsson & Simon, 1993). Second, probing questions targeting phase transitions. Third, a transfer probe with a structurally analogous problem not seen in the examination. All interviews were audio-recorded and transcribed verbatim. Translations of interview excerpts in this paper are by the first author and were independently back-translated by a bilingual mathematics educator to check fidelity.

3.7 Data Analysis

Analysis proceeded in four stages following Åkerlind (2005). Stage 1 mapped each reasoning episode onto a provisional timeline of cognitive moves using in-vivo codes. Stage 2 compared episodes across cases to identify qualitatively distinct types of cognitive moves. Stage 3 located phase boundaries by identifying the cognitive trigger that marked the transition between phases. Stage 4 examined cases that did not fit the provisional model. Two researchers coded a 30% subsample independently, with inter-coder agreement of 91% (Cohen's $\kappa = 0.87$). Discrepancies were resolved through joint review. The operational definitions and observable indicators used to assign each phase are summarised in Appendix A.

3.8 Trustworthiness

The four criteria of Lincoln and Guba (1985) were addressed. Credibility was supported by triangulation among written response, interview transcript, and transfer probe, and by member checking with six participants. Dependability was supported by a full audit trail of coding decisions. Confirmability was supported by reflexive analytic memos. Transferability was supported by thick description of the sixteen cases, the institutional context, and the four problems.

4 Findings

4.1 Overview of the Four-Phase Model

Phenomenographic analysis of the sixteen interview cases yielded four cognitively distinct phases of reasoning on counterintuitive mathematics problems. The phases are summarised in Table 2, and the model is presented graphically in Fig. 1. Each phase corresponds to a qualitatively distinct way of experiencing the counterintuitive task and to one of the conditions for conceptual change in Posner et al. (1982). Recognition corresponds to dissatisfaction; Identification corresponds to intelligibility (the contradiction has become a locatable object); Controversy Awareness corresponds to plausibility (the structural alternative has been articulated); Resolution corresponds to fruitfulness (the alternative has been deployed in a justification suitable for teaching).

Table 2 *The four cognitive phases, with characteristic actions, Posner condition met, and Subanji et al. (2021) level equivalent*

Phase	Name	Cognitive Action	Posner Condition	Level Equivalent
1	**Recognition**	Detects that something is wrong, but cannot diagnose the source.	Dissatisfaction	Initial
2	**Identification**	Locates the problematic element but cannot name why it fails.	Intelligibility	Exploration (early)

Phase	Name	Cognitive Action	Posner Condition	Level Equivalent
3	**Controversy Awareness**	Names the structural rule that the element violates.	Plausibility	Exploration (late)
4	**Resolution**	Reconstructs a complete justification suitable for teaching.	Fruitfulness	Clarification

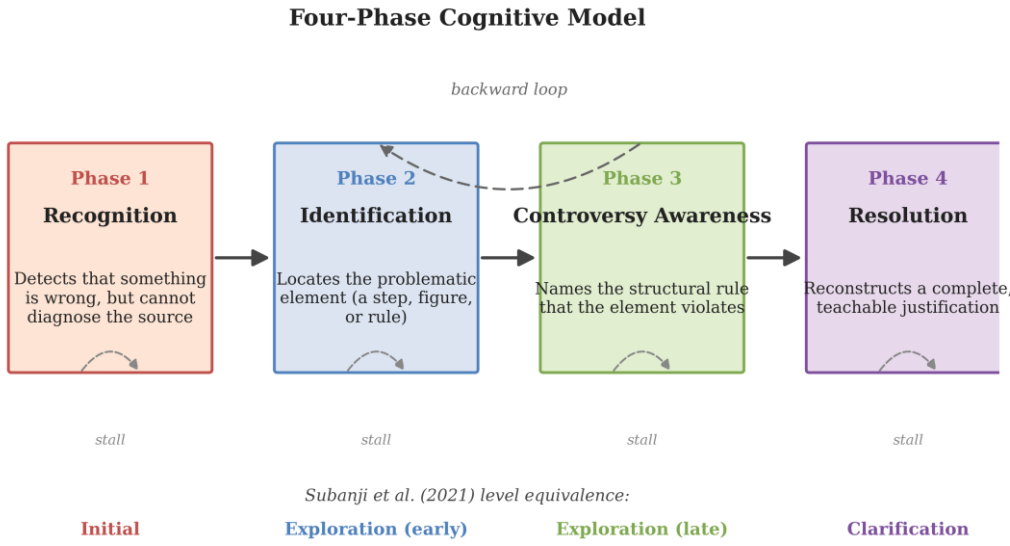


Fig. 1 The four-phase cognitive model. Forward transitions are shown by solid arrows; stalling within a phase is shown by dashed self-arrows; the documented Phase 3 to Phase 2 backward loop is shown by the dashed curve

Of the sixty-four episodes analysed, forty-five followed the linear sequence Phase 1 to Phase 2 to Phase 3 to Phase 4 with no regression. Fourteen episodes ended at an earlier phase (the participant did not progress further within the available time). Five episodes showed a Phase 3 to Phase 2 backward loop. We report the backward loop as a documented phenomenon worthy of further study, not as a structural feature of the model. The following sections describe each phase in detail and provide evidence from the dataset.

4.2 Phase 1, Recognition: Dissatisfaction without Diagnosis

The cognitive signature of Phase 1 is the gap between sensing the contradiction and diagnosing it. The learner notices that the offered answer cannot be right but lacks the structural vocabulary to locate the fault. In the framework of Posner et al. (1982), the dissatisfaction condition is met but the alternative is not yet intelligible. Subject S38's response to Problem 1, reproduced in Fig. 2, is paradigmatic. S38 computes K = 2 × (30% + 30%) = 120% for the perimeter and L = 30% × 30% = 900% for the area, and concludes that "both answers are wrong because neither 120% nor 900% is correct." The participant's own computation is itself incorrect, but more importantly, S38 cannot say which step in the original arguments produced the contradiction. In the interview, S38 said: "I knew something was wrong with the percentages, but I was not sure which rule they were breaking."

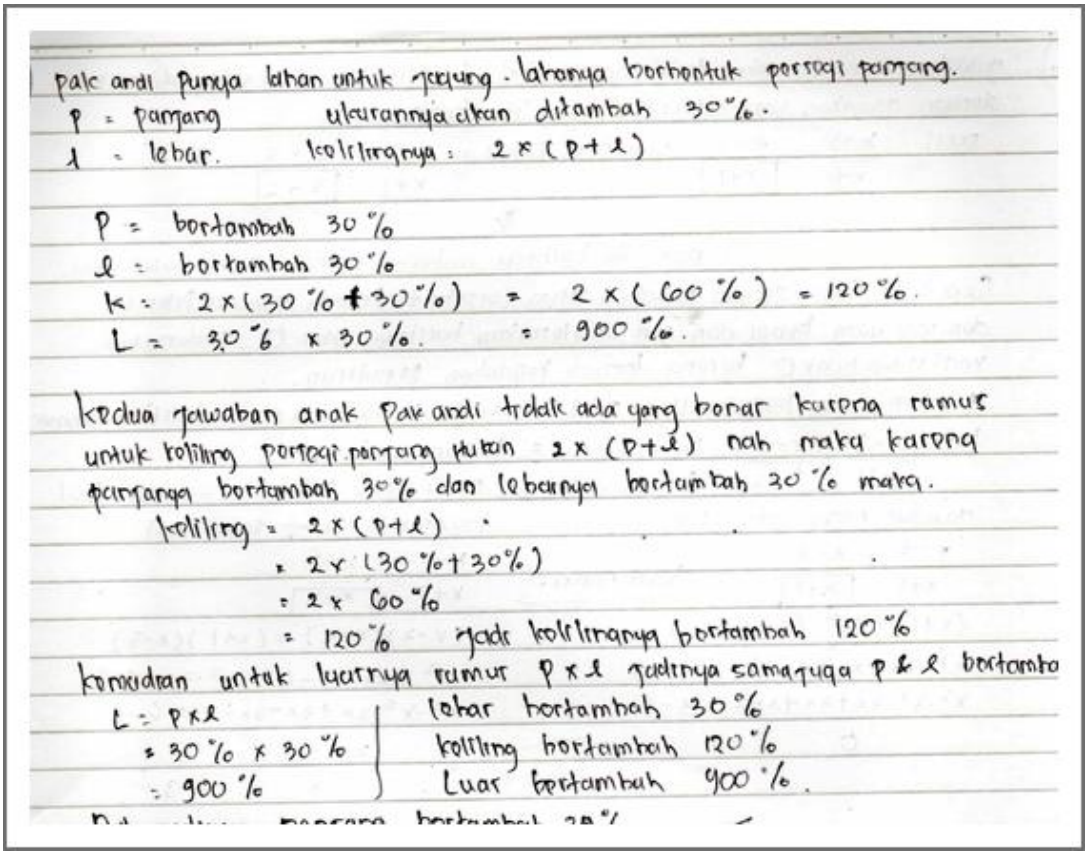


Fig. 2 Subject S38, Problem 1, Phase 1 (Recognition). The participant produces alternative computations (K = 120%, L = 900%) but cannot diagnose which step in the original argument is responsible

Phase 1 took a different form on Problem 2. Subject S32, shown in Fig. 3, accepted the bar illustration as a valid representation, writing that 2/6 is plausible "because both numerator and denominator are added." In the interview, S32 acknowledged that the answer felt strange but did not engage with the figure as the source of the strangeness. On procedure problems, learners often arrived at a partial result through their own (sometimes incorrect) computation but could not diagnose the offered argument; on media problems, learners often accepted the offered argument because the visual illustration looked persuasive. In both forms, the dissatisfaction condition is met (the learner is uneasy with the result) but the intelligibility condition (locating the fault) is not.

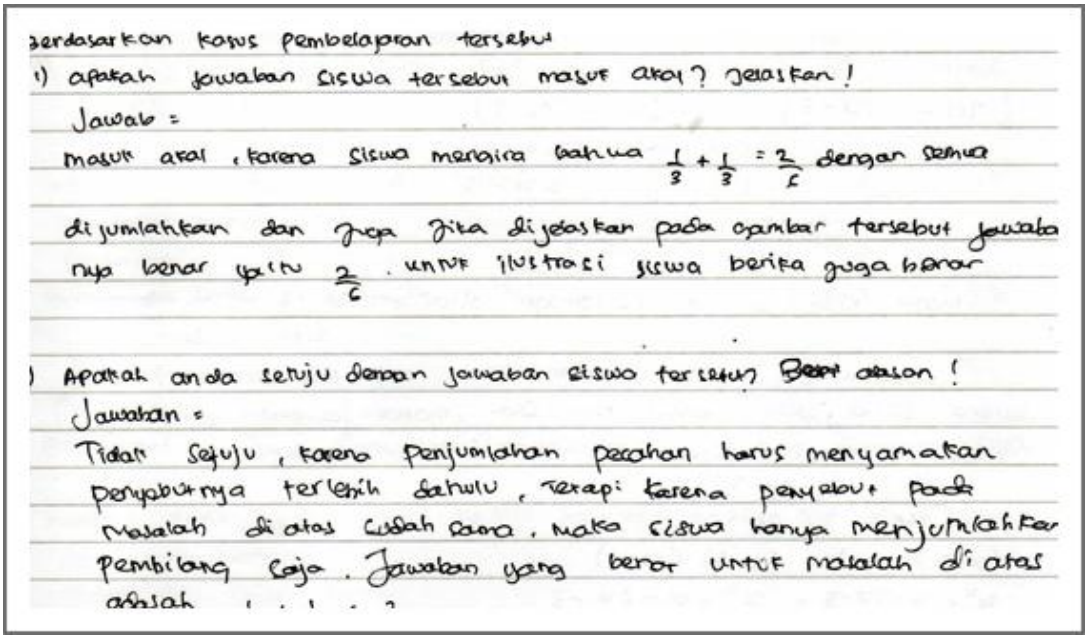


Fig. 3 Subject S32, Problem 2, Phase 1 (Recognition). S32 accepts the answer 2/6 as plausible because numerator and denominator are added, without engaging the bar illustration as the source of the misleading reasoning

Stalling at Phase 1 was observed in four cases on at least one problem. In each case, interview probes such as "Can you point to the exact line where things went wrong?" produced verbal acknowledgement but no structural diagnosis. This confirms Phase 1 as a stable stalling point rather than a momentary pause. What unites the Phase 1 episodes across cases is a metacognitive feature, not a mathematical knowledge gap. Most participants who stalled at Phase 1 had access to the underlying mathematics: they knew the formulas for perimeter and area, they knew that 1/3 + 1/3 should not equal 2/6, they knew that 2 cannot equal 1 or −2. What they lacked was the analytic stance that treats the offered argument as an object to be inspected step by step. This finding is consistent with Schoenfeld's (2014) account of monitoring as a key but underdeveloped competence in mathematical problem solving.

4.3 Phase 2, Identification: Localisation without Articulation

The cognitive signature of Phase 2 is localisation without articulation. The learner identifies a specific step, feature, or operation as the source of the contradiction but cannot yet name the structural rule that the element violates. The Posner condition that is newly met is intelligibility: the contradiction has become a locatable object. The transition from Phase 1 was triggered, in eleven of sixteen cases, by a shift from evaluating the answer to tracing the argument step by step, prompted either by a re-reading of the problem or by an interview probe.

Subject S5's response to Problem 1 (Fig. 4) lists both students' answers in parallel columns and writes that "neither answer is fully correct, only Sari is correct on the perimeter." S5 has localised the issue to the perimeter computation but does not explain why 30% + 30% does not produce 60% perimeter growth in this context. In the

interview, S5 said: “I knew the area answer was wrong, but I was not sure how to fix the formula.” Stalling at Phase 2 was the most common pattern in the dataset; eight of sixteen cases had at least two episodes that ended at Phase 2. The Phase 2 to Phase 3 transition is the most demanding cognitive transition in the model, because it requires not only attention to the relevant element but also access to the structural vocabulary needed to articulate the violation.

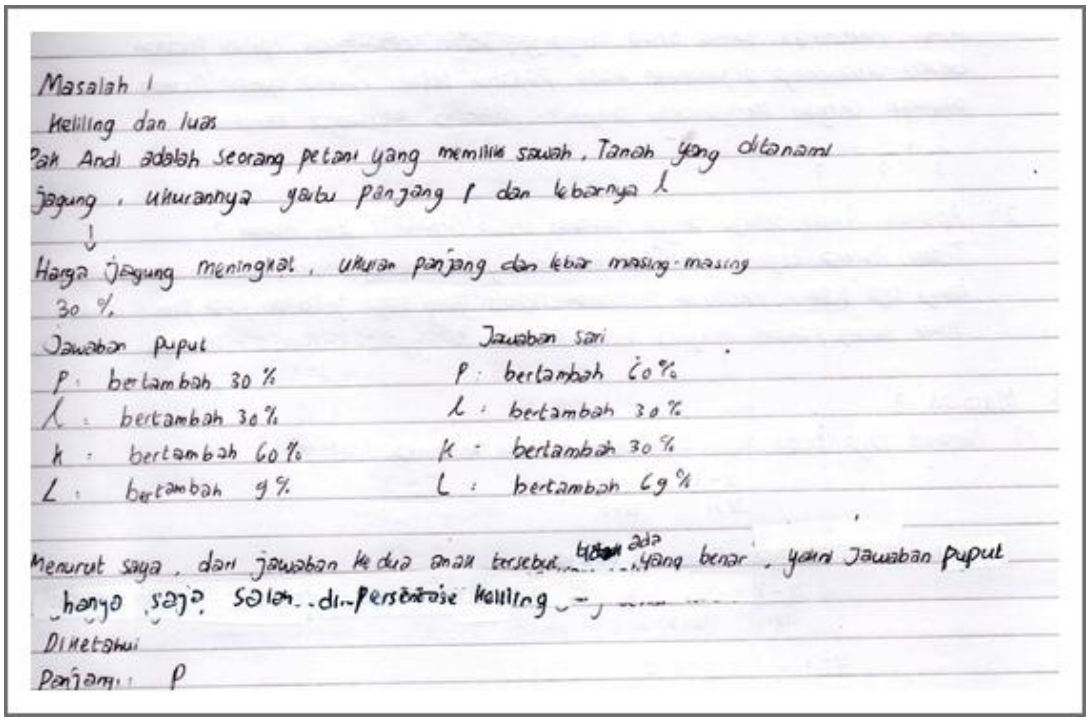


Fig. 4 Subject S5, Problem 1, Phase 2 (Identification). S5 lists the two answers in parallel and identifies the perimeter as a key issue, but does not yet name the rule that distinguishes additive from multiplicative percentage change

Phase 2 took a similar form across other procedure problems. On Problem 3, three participants identified the cancellation by $(x - 3)$ as the suspect step but could not articulate the precondition (the cancellation is invalid when both sides reduce to the same expression, which makes the equation an identity). On Problem 4, four participants identified the swap of root and exponent as suspicious but could not invoke the principal-root convention. The shared signature is the same: the offending element has been located, the offered argument has been rejected, but the structural language that makes the rejection complete is not yet available. This signature distinguishes Phase 2 sharply from Phase 1 (where no localisation has occurred) and from Phase 3 (where the rule has been named).

4.4 Phase 3, Controversy Awareness: Naming the Structural Rule

The cognitive signature of Phase 3 is the explicit naming of the structural rule that the problematic element violates. This is distinct from Phase 2 because naming the rule makes the contradiction fully intelligible: the

learner can say not only where the argument fails but why it must fail. In Posner's framework, the alternative has become plausible. Subject S25's response to Problem 1 (Fig. 5) sits at the Phase 2 to Phase 3 boundary. S25 writes that the perimeter formula $2 \times 30\%(p+l)$ gives $60\%(p+l)$, then notes that this represents "30 times the original." The participant identifies that the result of the procedure exceeds what is reasonable but does not yet articulate the multiplicative scaling rule. In the interview, S25 began to formulate the rule: "If each side is multiplied by 1.3, the perimeter becomes 1.3 times the original, so the increase is 30%, not 60%." This articulation marked the entry into Phase 3.

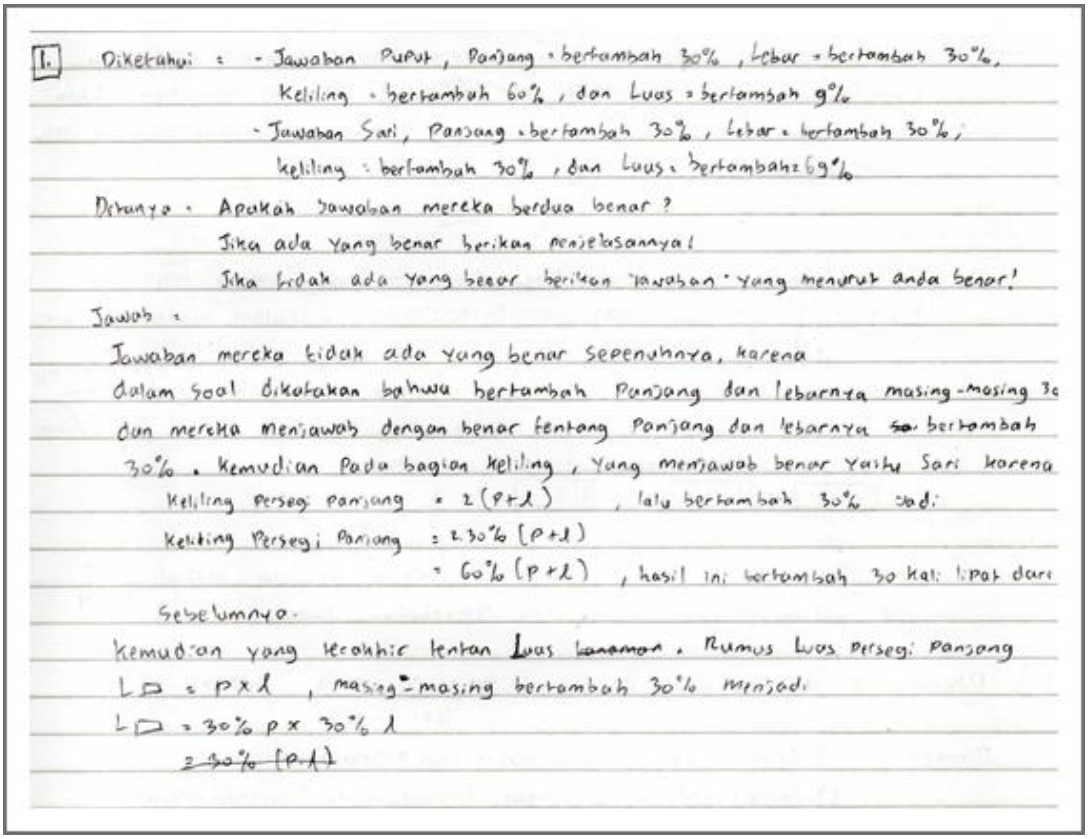


Fig. 5 Subject S25, Problem 1, Phase 2 to Phase 3 transition. S25 begins to articulate the multiplicative scaling rule but does not complete the structural account in the written response

Five episodes (across three participants) showed a Phase 3 to Phase 2 backward loop. In each instance, the participant correctly named a rule, applied it incorrectly, produced a confused result, then returned to re-examining the specific step before re-entering Phase 3 with a corrected articulation. We report this pattern as a documented phenomenon, not as a structural feature of the model. The frequency (five out of sixty-four episodes) is too low to support a structural claim, but the pattern is theoretically interesting because it is invisible to outcome-based level taxonomies (the participant remains within the exploration level throughout) and because it suggests that Phase 3 entry is not always permanent. Further work with larger samples is needed to determine the prevalence and conditions of backward loops.

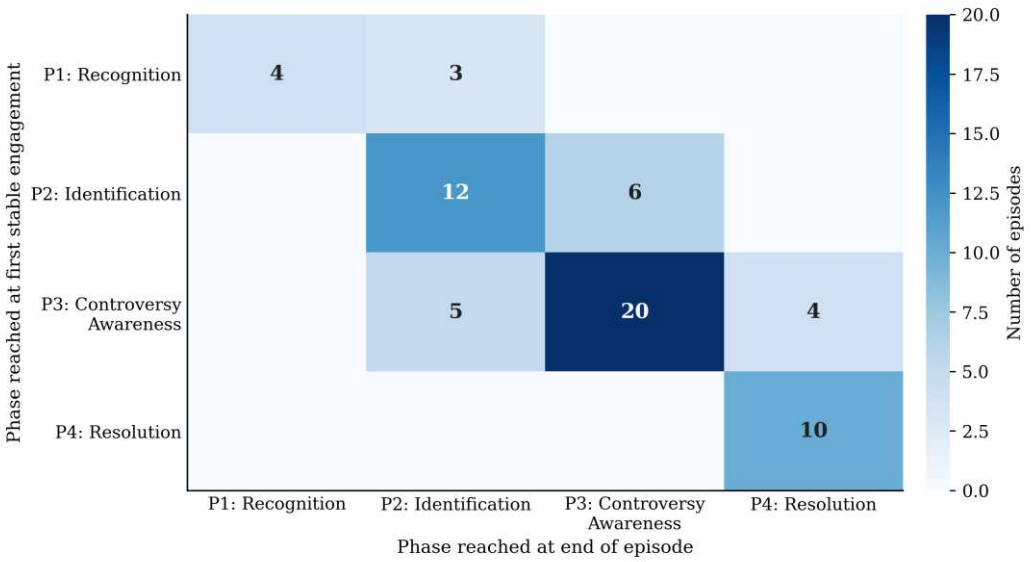


Fig. 6 Phase outcome matrix across the sixty-four episodes. Cells show the number of episodes that started at each phase (rows) and ended at each phase (columns). The off-diagonal cell at (P3 start, P2 end) records the five backward-loop episodes

4.5 Phase 4, Resolution: Constructing a Teachable Justification

The cognitive signature of Phase 4 is the production of a complete, mathematically sound reconstruction that includes both an explanation of why the offered argument fails and a correct alternative, framed in language sufficiently explicit to serve as a pedagogical justification. The Posner condition that is newly met is fruitfulness: the alternative has been deployed in a justification that supports further inquiry, in this case the inquiry of teaching. The marker that distinguishes Phase 4 from a strong Phase 3 articulation is the framing for pedagogical use.

Subject S22's Phase 4 work on Problem 1 (Fig. 7) used a numerical strategy: $p = 20$ m, $l = 10$ m, perimeter 60 m → 78 m (30% growth), area 200 m² → 338 m² (69% growth). In the interview, S22 framed the choice pedagogically: "I thought a specific example would be clearer for a student than symbols." Subject S10's Phase 4 work on the same problem (Fig. 8) used a symbolic strategy, expanding $(p + 0.3p)(l + 0.3l) = 1.69pl$. The contrast between numerical and symbolic Phase 4 strategies is significant. Both produce complete and teachable resolutions but recruit different competences and would speak to different student audiences. Outcome-based taxonomies classify both as clarification, while the present model captures the variation as different routes through the same Phase 4.

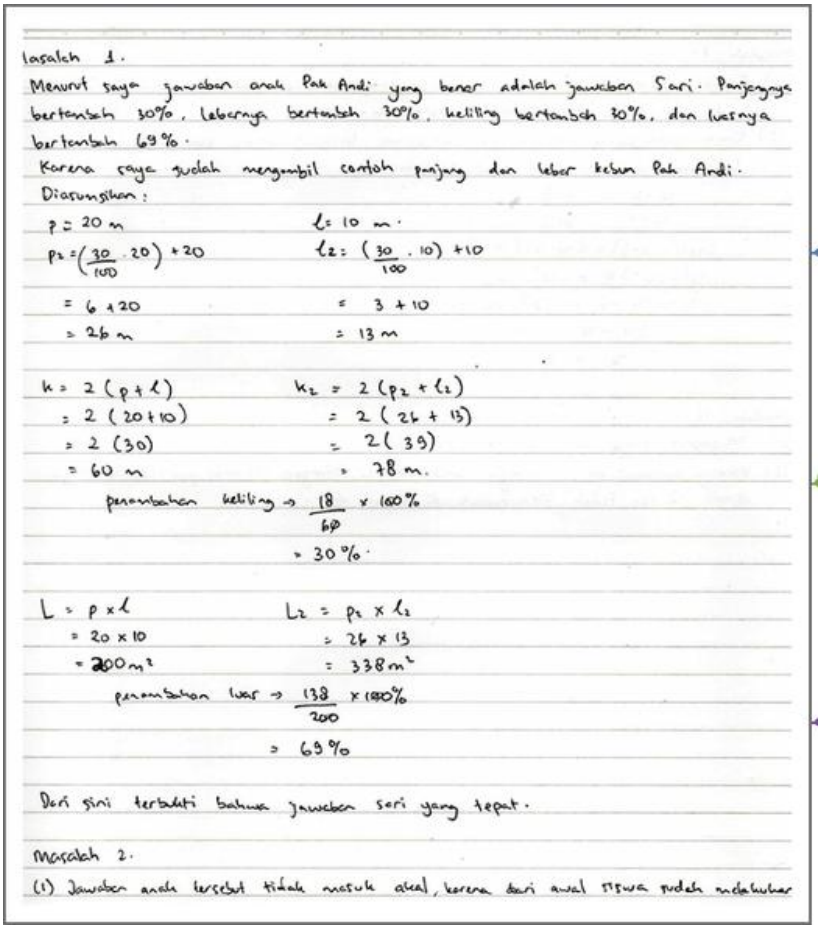


Fig. 7 Subject S22, Problem 1, Phase 4 (Resolution, numerical route). The participant uses a concrete numerical case ($p = 20, l = 10$) to compute both the perimeter change (30%) and the area change (69%), framed for pedagogical use



English transcription:

P1 (Recognition)

"In my opinion both answers contain something correct, namely Sari's answer."

P2 (Identification)

Length = p, Width = l
Perimeter K = 2(p + l), Area L = p × l
After the length and width are enlarged by 30%:
enlarged length = (p × 30/100) + p
enlarged width = (l × 30/100) + l

P3 (Controversy Awareness)

New perimeter $K_1 = 2[(p\cdot30/100) + p + (l\cdot30/100) + l]$
$= 2[p + l + 30p/100 + 30l/100]$
$= 2(p + l) + (30p + 30l)/50$
So the original 2(p + l) becomes 2(p + l) + (30p+30l)/50, i.e. perimeter increases by 30%.

P4 (Resolution)

New area $L_1 = (p\cdot30/100 + p) \times (l\cdot30/100 + l)$
$= pl + 30pl/100 + 30pl/100 + 9pl/100$
$= pl + 69pl/100$
Therefore area increases by 69%.
Conclusion: Sari is correct (perimeter +30%, area +69%).

Fig. 8 Subject S10, Problem 1, Phase 4 (Resolution, algebraic route). The symbolic expansion of $(p + 0.3p)(l + 0.3l)$ produces $1.69pl$, making the multiplicative scaling rule explicit

Phase 4 was reached by eleven of sixteen cases on at least one problem. In every Phase 4 episode, the participant included a teacher-facing element: an explicit explanation framed for a student, a redrawn illustration, a numerical example chosen for clarity, or a teacher-facing question. This pedagogical framing distinguishes Phase 4 from a strong Phase 3 articulation. In phenomenographic terms, Phase 4 represents experiencing the same problem as a teaching opportunity rather than a personal puzzle. The transition from Phase 3 to Phase 4 therefore involves a perspective shift, not only a knowledge increment.

4.6 Source-Specific Patterns

Fig. 9 presents the phase reached by each of the sixteen cases on each problem. Three source-specific patterns emerge. First, on misleading-media problems (P2), thirteen of sixteen cases reached Phase 3, the highest of any source. The Phase 2 to Phase 3 transition was typically triggered by visual reconstruction (Fig. 10). Second, on misleading-procedure problems (P1 and P3), Phase 2 stalling was most common; Subject S20's Phase 4 work on Problem 3 (Fig. 11) shows what successful resolution required, namely the explicit distinction between an algebraic expression and an algebraic equation. Third, on misleading-operation problems (P4), Phase 1 was rarely a stable stalling point because the contradiction ($2 = -2$) is starkly visible; Subject S45's response (Fig. 12) invokes the principal-root convention to justify that $((-2)^2)^{(1/2)}$ equals 2, not −2.

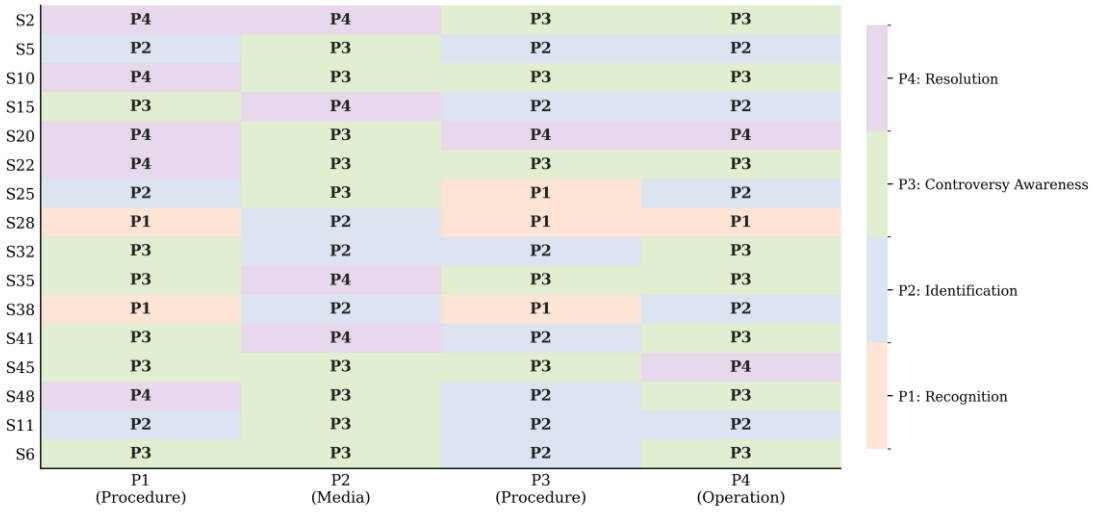


Fig. 9 Phase reached by each of the sixteen selected cases on the four problems

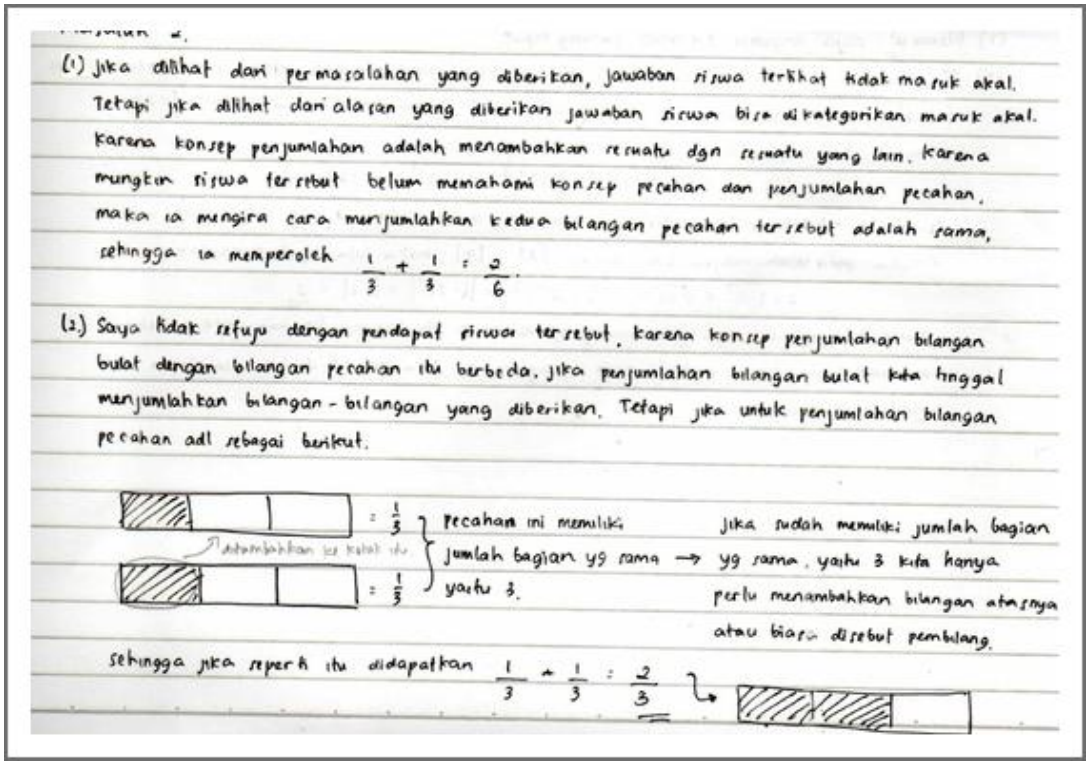


Fig. 10 Subject S2, Problem 2, Phase 4 (Resolution). S2 redraws the bar illustration with the unshaded portions retained and articulates the unit-whole rule

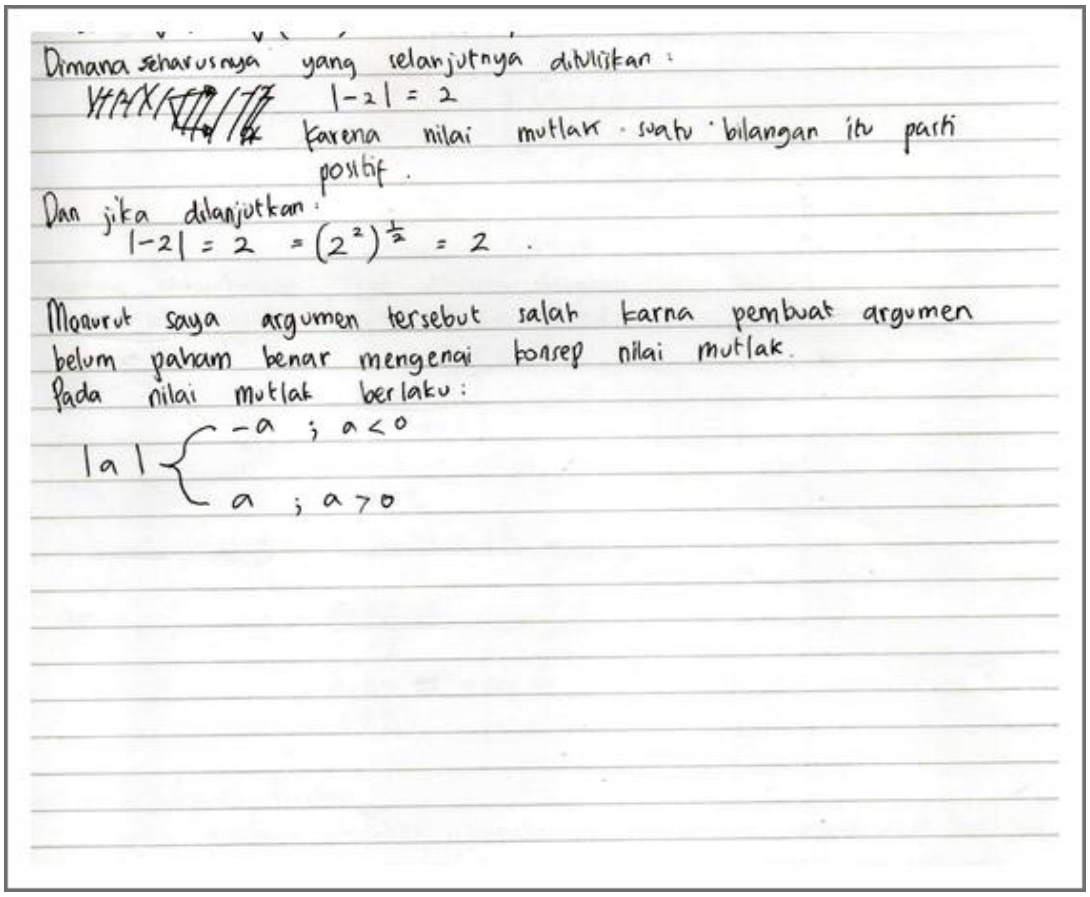


Fig. 11 Subject S20, Problem 3, Phase 4 (Resolution). S20 distinguishes an algebraic expression from an algebraic equation, traces the cross-multiplication, and concludes that $0x = -1$ has no solution

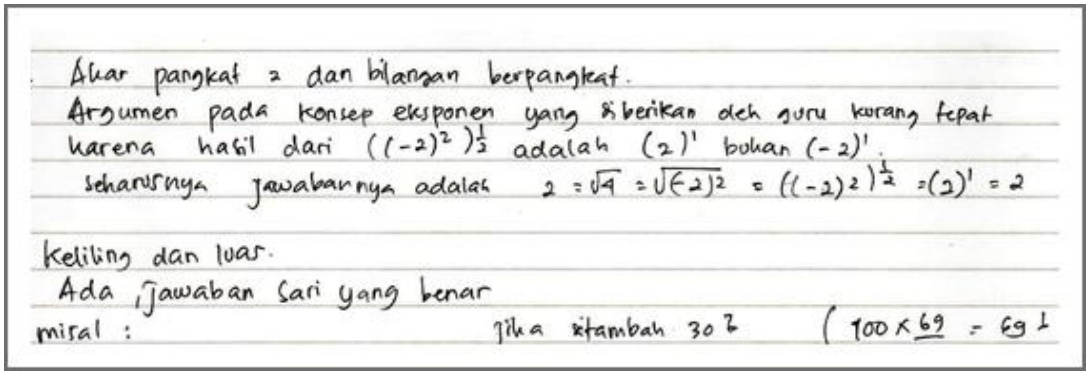


Fig. 12 Subject S45, Problem 4, Phase 4 (Resolution). S45 invokes the principal-root convention to justify why ((−2)²)^(1/2) equals (2)¹, not (−2)¹

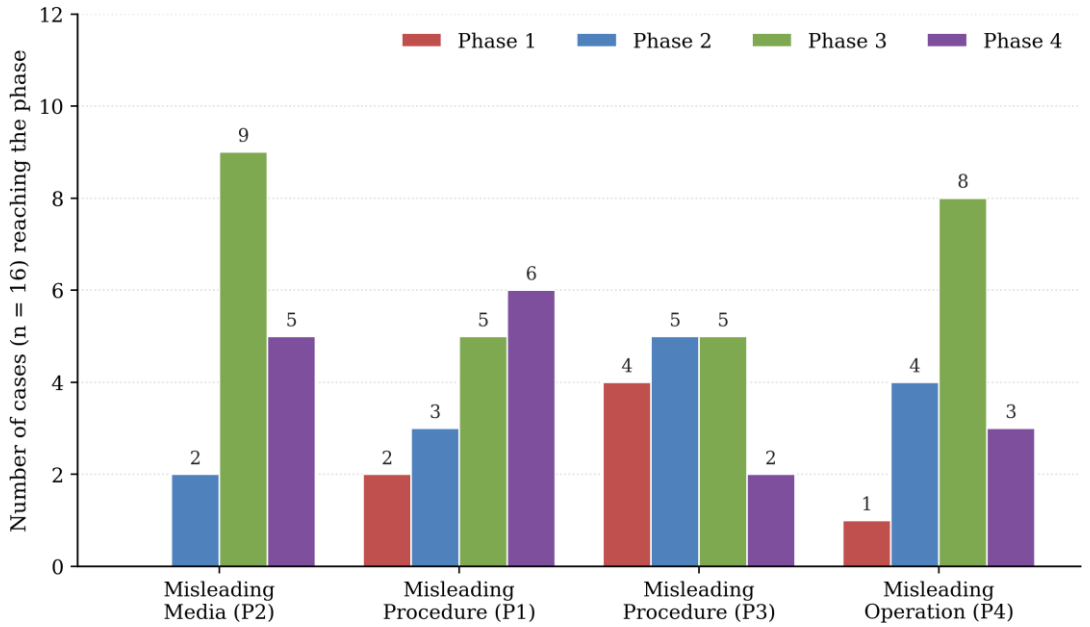


Fig. 13 Number of cases (n = 16) reaching each phase by source type. Misleading-media problems show the highest Phase 3 reach; misleading-procedure problems show the highest Phase 2 stalling; misleading-operation problems show an exploration-heavy distribution

5 Discussion

5.1 The Within-Task Process Behind the Posner Conditions

The four-phase model contributes to the international conceptual change literature by operationalising the Posner et al. (1982) conditions as observable behaviours within a single counterintuitive task. The conditions for conceptual change have typically been treated as static prerequisites: the learner either is or is not dissatisfied, finds the alternative either intelligible or not, and so on. The present findings suggest that the conditions are met sequentially within the cognitive work of resolution. Phase 1 marks the entry of dissatisfaction; Phase 2 the entry of intelligibility; Phase 3 the entry of plausibility; Phase 4 the deployment of fruitfulness. This finer-grained reading of the Posner conditions is consistent with Vosniadou and Skopeliti's (2014) account of conceptual change as a slow restructuring of framework theories, and with Pacaci et al.'s (2024) finding that scaffolded conflict

produces larger effects than unscaffolded conflict. Where the present model adds value is in identifying which condition is being addressed at each moment of the within-task process, supporting more targeted instructional support.

5.2 Why Counterintuitive Tasks Demand a Different Cognitive Model

Existing process accounts (Pólya, 1945; Mason et al., 1982; Sfard, 1991) were designed for task genres in which the goal is given and the path must be found. In a counterintuitive task, the path is given (in the form of the offered argument) and the goal must be constructed: what is wrong with this argument, and how can it be repaired? This inversion is not cosmetic. It changes the cognitive resources that resolution requires. Standard problem solving recruits heuristic search through a problem space (Schoenfeld, 2014). Counterintuitive resolution recruits a diagnostic stance toward an existing argument: the learner must adopt the analytic posture of an evaluator, not the productive posture of a solver.

This shift in stance has cognitive consequences that the standard problem-solving models cannot capture. Pólya's four phases all assume the learner is moving toward an unknown goal; if the goal is given, "understand the problem" and "devise a plan" collapse into a single act of evaluation. Mason et al.'s entry-attack-review structure assumes that the learner has not yet attempted a solution; in counterintuitive tasks, an attempt is what the problem itself supplies. Sfard's operational-to-structural transition describes how a learner moves from process to object; counterintuitive tasks ask the learner to move in the opposite direction, from a polished object (the offered argument) back to its constituent processes, to find which one fails. The four-phase model is therefore not a refinement of these accounts but a cognitive complement, designed for the specific situation of within-task contradiction. The diagnostic stance is precisely what the mathematics teacher noticing literature (Sherin et al., 2011; Jacobs et al., 2010) has shown pre-service teachers to lack without explicit support, and the four phases can be read as a developmental sequence of the diagnostic stance: from sensing (Phase 1) to localising (Phase 2) to articulating (Phase 3) to deploying (Phase 4).

5.3 The Phase 2 to Phase 3 Bottleneck

The most demanding transition in the model is from Phase 2 to Phase 3. Eight of sixteen cases had at least two episodes that stalled at Phase 2, and the Phase 2 stall was most common on procedure problems. The cognitive resource that the transition requires is structural vocabulary: the precise language for the rule the problematic element violates. This is consistent with Booth and Newton's (2012) account of procedural fragility and with Rittle-Johnson and Schneider's (2015) account of how procedures detached from conceptual preconditions become brittle.

From a dual-process perspective (Leron & Hazzan, 2009), Phase 2 corresponds to the stage at which System 1 algorithmic intuition has been challenged but System 2 structural reasoning has not yet engaged. The Phase 2 to Phase 3 transition therefore requires the deliberate effort of System 2: the learner must move from sensing that the procedure failed to articulating why it failed at the level of mathematical structure. This finding has direct implications for instructional support. Pacaci et al.'s (2024) meta-analysis showed that scaffolded conflict produces large effects, while unscaffolded conflict produces only small effects. The present model identifies the specific point where scaffolding is most needed: the moment between Phase 2 and Phase 3, when an instructional prompt that supplies or elicits the relevant structural vocabulary is likely to be more productive than a prompt that simply asks the learner to think harder.

5.4 Source-Specific Cognitive Resources

The phase distribution across sources reveals that different kinds of cognitive conflict recruit different cognitive resources for resolution. Misleading-media problems benefited from visual reconstruction as a Phase 2 to Phase 3 bridge, consistent with Duval's (2017) account of register conversion. Misleading-procedure problems required the articulation of preconditions that classroom instruction typically leaves implicit. Misleading-operation problems required structural vocabulary from formal algebra. Subject S2's redrawn illustration on Problem 2 (Fig. 10) is not just a visual aid; it is a cognitive externalisation that supports the Phase 2 to Phase 3 transition by making the unit-whole invariant explicit.

5.5 Implications for Pre-service Teacher Education

Three implications follow, each tied directly to the empirical findings rather than offered as speculation. First, instructional attention should concentrate on the Phase 2 to Phase 3 transition, because it is the most common stalling point in the dataset and because it requires structural vocabulary that classroom instruction often leaves implicit. A teacher educator who recognises this bottleneck can design specific scaffolds, for example a routine in which pre-service teachers practise articulating the structural rule that an offered student argument violates, before being asked to construct an alternative justification. Analysing the opportunities for reasoning embedded in such tasks offers one concrete route for designing these scaffolds (Weingarden & Buchbinder, 2026).

Second, source-specific scaffolds should be paired with their characteristic cognitive resource. The S2 case (Fig. 10) suggests that visual reconstruction prompts can support media-induced controversies; the S20 case (Fig. 11) suggests that precondition-articulation prompts can support procedure-induced controversies; the S45 case (Fig. 12) suggests that structural-rule invocation prompts can support operation-induced controversies. The pairing is not arbitrary. Each scaffold draws on the cognitive resource that learners successfully used to bridge

from Phase 2 to Phase 3 in the dataset. A teacher educator can use this pairing to design diagnostic activities in which pre-service teachers identify the source of a student's controversy and select the matching scaffold.

Third, the existence of two distinct Phase 4 routes (numerical and symbolic, exemplified by S22 and S10) suggests that pre-service teachers should be exposed to both routes. A future teacher who can recognise both routes is better equipped to value student responses that take different paths to the same understanding (Stein et al., 2008).

5.6 Relation to Existing Frameworks

The proposed model is consistent with the level taxonomy of Subanji et al. (2021) but is grounded in international literature. Phase 1 corresponds to the initial level. Phase 4 corresponds to the clarification level. Phases 2 and 3 split the exploration level into two cognitively distinct sub-states; this is the most consequential refinement, because it identifies the Phase 2 to Phase 3 transition as the empirical bottleneck. The model is not a mere relabelling of an existing taxonomy. It is a process account anchored in cognitive conflict and conceptual change theory, with empirical correspondence to a regional outcome account.

The dual grounding (international cognitive conflict literature plus regional outcome correspondence) is one of the strengths of the model. It supports cross-cultural applicability while remaining sensitive to local educational contexts. The institutional context shapes the specific structural vocabulary available to learners, but it does not determine the phase structure itself. Further cross-cultural studies are needed.

6 Conclusion, Limitations, and Future Research

6.1 Summary

The three research questions are addressed as follows. Four cognitively distinct phases (Recognition, Identification, Controversy Awareness, Resolution) were identified, each corresponding to one of the Posner et al. (1982) conditions for conceptual change (RQ1). The Phase 2 to Phase 3 transition was the most demanding, and a Phase 3 to Phase 2 backward loop in five episodes is reported as a documented phenomenon (RQ2). The model relates to the international literature by operationalising the Posner conditions within a single counterintuitive task and by providing a process-level diagnostic framework for pre-service teacher noticing (RQ3).

6.2 Theoretical and Practical Implications

Theoretically, the model offers a within-task process reading of conceptual change conditions, complementing the longer-timescale accounts of Vosniadou and Skopeliti (2014) and Pacaci et al. (2024). Practically, the model supports the design of source-specific scaffolds and process-level diagnostic activities in

pre-service teacher education; each of these implications—the Phase 2 to Phase 3 bottleneck, the source-specific cognitive resources, and the dual Phase 4 routes—is grounded directly in the dataset.

6.3 Limitations

Five limitations should be acknowledged. First, the study was conducted at a single university in Indonesia; replication is needed before claims of transferability are warranted. Second, the empirical demonstration draws on four problems in arithmetic and elementary algebra; although the phases are defined by cognitive function rather than by content and are therefore expected to generalise, their applicability to probability, geometry, and calculus tasks remains to be tested, and such tasks may surface additional phases or different stalling patterns. Third, the phenomenographic sample (sixteen cases) supports depth but not breadth; quantitative validation with a larger cohort would strengthen the claims. Fourth, the backward-loop finding is based on five out of sixty-four episodes and should be treated as a hypothesis-generating observation, not as a structural component of the model. Fifth, all participants were pre-service teachers in their middle undergraduate years; the model has not been tested with in-service teachers or with school students.

6.4 Future Research

Three directions are promising. First, intervention studies should test whether explicitly teaching the four phases helps pre-service teachers move past the Phase 2 stall, drawing on the scaffolded-conflict findings of Pacaci et al. (2024). Second, larger-sample studies should examine the prevalence and conditions of backward loops, which the present study can only document. Third, the model should be tested with counterintuitive tasks in domains beyond arithmetic and elementary algebra, to determine whether the four phases are domain-general or domain-specific.

Appendix A. Operational Definitions and Observable Indicators of the Four Phases

Phase assignment required that both the written response and the interview satisfy the indicators for that phase. The full rubric with anchor examples for each source type is provided in Online Resource 1.

Phase	Operational definition	Observable indicators (written and verbal)
1 Recognition	The learner registers that the offered result is unacceptable but cannot locate the faulty step.	Expresses unease or rejects the answer; may produce an alternative computation; makes no reference to a specific step or rule.
2 Identification	The learner localises the specific step, feature, or operation at fault.	Points to a particular line or operation and rejects it, but cannot state the structural rule that it violates.
3 Controversy Awareness	The learner names the structural rule that the faulty element violates.	States the rule explicitly (e.g., multiplicative versus additive scaling; the principal-root

Phase	Operational definition	Observable indicators (written and verbal)
		convention) and explains why the step must fail.
4 Resolution	The learner reconstructs a complete, mathematically sound, and teachable justification.	Gives a correct alternative plus an explanation framed for a student: a redrawn figure, a chosen numerical example, or a teacher-facing question.

Statements and Declarations

Funding This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The authors gratefully acknowledge the institutional support provided by the Department of Mathematics, Faculty of Mathematics and Natural Sciences, and the Postgraduate Program of Universitas Negeri Malang, which made the data collection and analysis possible.

Competing Interests The authors declare that they have no known competing financial interests, personal relationships, or institutional affiliations that could have appeared to influence the work reported in this paper. The research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author Contributions Erry Hidayanto: conceptualisation, methodology, study design, supervision, writing (review and editing), and final approval of the manuscript. Subanji: conceptualisation, theoretical framework, validation of the four-phase cognitive model, writing (review and editing), and final approval of the manuscript. Rivan Gestiardi: data curation, formal analysis, phenomenographic coding of 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.

Ethics Approval Ethical approval for the study was obtained from the institutional research ethics committee of Universitas Negeri Malang. The study was conducted in accordance with the principles of the Declaration of Helsinki.

Consent to Participate Each participant gave written informed consent to take part in the study with a clear understanding of what participation entailed. Anonymity was assured through the use of pseudonyms (S1 to S50).

Data Availability The data that support the findings of this study (anonymised written responses and interview transcripts) are available from the corresponding author upon reasonable request, subject to ethical restrictions protecting participant confidentiality.

References

Åkerlind, G. S. (2005). Variation and commonality in phenomenographic research methods. *Higher Education Research & Development, 24*(4), 321–334. https://doi.org/10.1080/07294360500284672

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

Baş-Ader, S., Ader, E., & Taylan, R. D. (2025). Supporting prospective mathematics teachers in noticing students' proportional reasoning. *Journal of Mathematics Teacher Education, 28*(3), 681–711. https://doi.org/10.1007/s10857-024-09621-x

Booth, J. L., & Newton, K. J. (2012). Fractions: Could they really be the gatekeeper's doorman? *Contemporary Educational Psychology, 37*(4), 247–253. https://doi.org/10.1016/j.cedpsych.2012.07.001

Buchbinder, O., & McCrone, S. (2020). Preservice teachers learning to teach proof through classroom implementation: Successes and challenges. *The Journal of Mathematical Behavior, 58*, 100779. https://doi.org/10.1016/j.jmathb.2020.100779

Buchbinder, O., & McCrone, S. (2023). Preparing prospective secondary teachers to teach mathematical reasoning and proof: The case of the role of examples in proving. *ZDM Mathematics Education, 55*(4), 779–792. https://doi.org/10.1007/s11858-023-01493-4

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

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (Rev. ed.). MIT Press. https://doi.org/10.7551/mitpress/5657.001.0001

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

Jacobs, V. R., Lamb, L. L. C., & Philipp, R. A. (2010). Professional noticing of children's mathematical thinking. *Journal for Research in Mathematics Education, 41*(2), 169–202. https://doi.org/10.5951/jresematheduc.41.2.0169

Leron, U., & Hazzan, O. (2009). Intuitive vs analytical thinking: Four perspectives. *Educational Studies in Mathematics, 71*(3), 263–278. https://doi.org/10.1007/s10649-008-9175-8

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.

Marton, F., & Booth, S. (1997). *Learning and awareness*. Lawrence Erlbaum.

Mason, J., Burton, L., & Stacey, K. (1982). *Thinking mathematically*. Addison-Wesley.

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

Pacaci, C., Ustun, U., & Ozdemir, O. F. (2024). Effectiveness of conceptual change strategies in science education: A meta-analysis. *Journal of Research in Science Teaching, 61*(6), 1263–1325. https://doi.org/10.1002/tea.21887

Pólya, G. (1945). *How to solve it*. Princeton University Press.

Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education, 66*(2), 211–227. https://doi.org/10.1002/sce.3730660207

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

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

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

Sherin, M. G., Jacobs, V. R., & Philipp, R. A. (Eds.). (2011). *Mathematics teacher noticing: Seeing through teachers' eyes*. Routledge. https://doi.org/10.4324/9780203832714

Sinatra, G. M., & Pintrich, P. R. (Eds.). (2003). *Intentional conceptual change*. Lawrence Erlbaum.

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

Vosniadou, S., & Skopeliti, I. (2014). Conceptual change from the framework theory side of the fence. *Science & Education, 23*(7), 1427–1445. https://doi.org/10.1007/s11191-013-9640-3

Weingarden, M., & Buchbinder, O. (2023). Teacher learning to teach mathematics via reasoning and proving: A discursive analysis of lesson plans modifications. *Frontiers in Education, 8*, Article 1154531. https://doi.org/10.3389/feduc.2023.1154531

Weingarden, M., & Buchbinder, O. (2026). Exploring the opportunities for reasoning and proving in mathematical tasks designed by secondary prospective teachers. *Educational Studies in Mathematics, 121*(2), 309–332. https://doi.org/10.1007/s10649-025-10440-x

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