



ANALYZING ELEMENTARY STUDENTS' MISCONCEPTION PATTERNS IN KEY SCIENCE AND SOCIAL STUDIES (IPAS) TOPICS: IMPLICATIONS FOR INSTRUCTIONAL DESIGN

Yaman¹, Ayi Abdurahman², Dewanta Arya Nugraha³, Nuris Fattahillah⁴

^{1,2}Nusa Putra University, Sukabumi, Indonesia

³Sebelas Maret University, Solo, Indonesia

⁴Universitas Sultan Ageng Tirtayasa, Indonesia

*Corresponding Author:  ayi.abdurahman@nusaputra.ac.id

Article History

Received : 01/05/2025

Revised : 18/06/2025

Accepted : 25/07/2025

Keywords:

Conceptual change,
Science and Social
Studies (IPAS),
Misconceptions, Two-
tier diagnostic test.

Abstract. Persistent misconceptions in foundational science ideas often resist routine instruction, undermine knowledge transfer, and remain hidden when assessments capture answers but not reasoning. Within Indonesia's integrated Science–Social Studies curriculum (IPAS), such misconceptions can cluster around core topics and shape students' subsequent learning trajectories. This study maps elementary students' misconception patterns across five IPAS topics—Force–Motion, Changes of State, Energy, Weather/Water Cycle, and Ecosystems—and translates the findings into design principles for instruction. Using an explanatory sequential mixed-methods design, we collected data from 204 Grade 4–6 students in three public schools via a locally contextualized two-tier diagnostic test and clinical interviews supported by concept cartoons. Instrument quality was adequate (mean Aiken's $V = 0.90$; internal reliability per topic = 0.71 – 0.77 ; inter-rater reliability ≥ 0.78). Aggregate responses were distributed as 44.2% scientifically sound, 28.1% partial, and 27.7% strong misconceptions, with a grade-level trend toward improvement. The most persistent hotspots appeared in Ecosystems and Force–Motion. Tiered item sensitivity showed a consistent decrease in misconception rates from Q1 (everyday contexts) to Q4 (diagram/model representations), underscoring the role of multiple representations in bridging non-scientific intuitions. Approximately 9–12% of responses were false positives, highlighting the need to evaluate reasoning in addition to answers. Findings inform operational strategies: eliciting initial ideas with mini two-tier items; orchestrating multiple representations and simple empirical evidence within a 5E sequence with explicit refutation; and embedding metaconceptual reflection with formative assessment attuned to misconception patterns. Limitations include geographic scope and non-causal design. Future work should broaden contexts, test instrument invariance, and evaluate a 5E–refutation package quasi-experimentally.

How to Cite: Yaman, Y., Abdurahman, A., Nugraha, D. A., & Fattahillah, N. (2025). ANALYZING ELEMENTARY STUDENTS' MISCONCEPTION PATTERNS IN KEY SCIENCE AND SOCIAL STUDIES (IPAS) TOPICS: IMPLICATIONS FOR INSTRUCTIONAL DESIGN. *Prima Magistra: Jurnal Ilmiah Kependidikan*, 6(3), 459–470. <https://doi.org/10.37478/jpm.v6i3.5627>

Correspondence address:

Jl. Raya Cibolang Cisaat - Sukabumi No.21, Cibolang Kaler,
Kec. Cisaat, Kabupaten Sukabumi, Jawa Barat 43152.

 ayi.abdurahman@nusaputra.ac.id

Publisher:

Program Studi PGSD Universitas Flores. Jln. Samratulangi,
Kelurahan Paupire, Ende, Flores.

 primagistrauniflor@gmail.com

INTRODUCTION

Integrated Science and Social Studies (IPAS) at the elementary level plays a strategic role in building integrated science and social literacy (Desstya et al., 2024). Aligned with the spirit of the Merdeka Curriculum and the Pancasila Student Profile, IPAS guides students to develop evidence-based reasoning (Yansah & Zulfadewina, 2024), to understand the close interconnections between science, everyday life, and society (Tateo, 2014), and to participate actively as responsible citizens (Matsuda, 2009). This orientation fundamentally positions core scientific and social concepts as a crucial foundation for informed, contextual decision-making across real-world situations.

In early grades, science–social literacy is nurtured through concrete, hands-on experiences such as observing weather changes, planting, sorting waste, and sharing simple findings (Moore-Hart et al., 2004). As students' cognition develops in upper grades, these experiences grow into more complex capacities, including constructing simple, data-supported arguments, comparing alternative solutions to problems, and predicting social–environmental consequences of actions (Kuhn & Pease, 2008). Thus, IPAS functions not only as a subject but

also as an essential vehicle for 21st-century skills, including critical reasoning, collaboration, effective communication, and creativity (Fitriyanti et al., 2025).

However, the classroom implementation of IPAS faces significant challenges (Rahmani et al., 2025). Many key IPAS concepts—such as force and motion, changes of state, energy, the water cycle, and ecosystems—are not directly observable and tend to be abstract for elementary students. Without appropriate and concrete representational bridges, students often form intuitive explanations that are incorrect or misaligned with scientific concepts (Guzzetti, 2000). In the social domain, reasoning about complex cause–effect relations within socio-environmental systems frequently exceeds young learners’ personal experiences and initial cognitive capacities (Assaraf & Orion, 2010).

Heterogeneity in cognitive readiness and limited scientific vocabulary further heighten the risk of conceptual misunderstanding (Suprpto, 2020). Common everyday expressions such as “clouds break” or “energy runs out” can inadvertently reinforce misconceptions if not addressed through careful instructional design, including precise language, appropriate visualizations, and manipulatives that bridge understanding. In addition, limited teaching aids, loosely structured student worksheets, and constrained access to information and communication technology often restrict opportunities for deep inquiry (Putriyani et al., 2024). Teaching practices still emphasizing broad content coverage and summative assessment tend to promote lecturing and rote practice. Such patterns make it difficult for teachers to surface students’ initial ideas, induce productive cognitive conflict, or conduct formative assessments sensitive to specific misconception types (Haug & Ødegaard, 2015). As a result, conceptual misunderstandings often go undetected early on, becoming more resistant and harder to change later.

Misconceptions are defined as erroneous understandings that students believe to be true and that are relatively stable—not merely ignorance or conceptual gaps. They can arise from multiple sources, including non-scientific intuitive experiences, ambiguous language, analogies applied without clear boundaries, and overly abstract presentations (Coley & Tanner, 2012; Suprpto, 2020). Common examples among elementary students include beliefs that objects require continuous force to keep moving (Sari et al., 2019), that mass “disappears” during evaporation (Stavy, 1991), that heat is identical to temperature (Sullivan & Edmondson, 2008), that rain occurs because “clouds break” (Brewer, 1952), and that food chains are linear without recognizing the central role of decomposers (Mougi, 2020).

The consequences of misconceptions are layered and detrimental. For students, they hinder the development of coherent, structured conceptual schemas (Kuczmann, 2017), reduce the transfer of knowledge across contexts (Argote & Ingram, 2000), and trigger repeated failures in advanced topics that depend on basic conceptual understanding (Lima et al., 2019). For teachers and schools, undetected misconceptions misdirect instructional strategies, render summative assessments less representative of true understanding, and ultimately weaken a classroom culture of inquiry (Lakin & Wallace, 2015). Early identification and targeted intervention are therefore essential prerequisites for improving IPAS instructional quality.

Reliable detection of misconceptions requires instruments that assess not only answers but also the underlying reasoning (Hirsch & O'Donnell, 2001). Two-tier diagnostic tests (Tsui & Treagust, 2010) assess concept choice in Tier 1 and reasoning in Tier 2, enabling answer–reason patterns to reveal specific misconceptions. Three-tier variants even add confidence ratings (Ardiansah et al., 2018). These approaches can be complemented by clinical interviews, concept cartoons, concept mapping, and misconception-based distractors in multiple-choice items. Item analysis and, where feasible, Rasch modeling can further strengthen validity and reliability evidence.

Internationally, studies have mapped consistent misconception patterns on topics such as force and motion (e.g., Métioui & Trudel (2021)), matter and changes of state (e.g., Crespo & Pozo (2004)), energy (e.g., Dib (2013)), weather/water cycle (e.g., Halidullin et al. (2018)), and ecosystems (e.g., Fernandez et al. (2016)), including impetus-like reasoning (Madsen et al., 2015). In Indonesia, similar findings have been reported from elementary to lower secondary



levels, highlighting the roles of everyday language, imprecise textbook illustrations, and limited practical work as reinforcing factors (Chiu, 2007).

Nevertheless, important gaps remain in the Indonesian context. Comprehensive mapping of misconception patterns across key IPAS topics that integrate science and social dimensions within the Merdeka Curriculum remains limited; many studies focus on a single science topic (e.g., Desi et al. (2025) and Waseso et al. (2024)). The need for Indonesian-language, locally contextualized two-tier diagnostics (e.g., dialect, culture, local examples) is also not fully met. Moreover, comparative data across grade levels (e.g., Grades 4–6), school types, and student characteristics are scarce for constructing locally grounded learning progressions.

This study addresses these gaps with two main contributions. First, it systematically maps Indonesian elementary students' misconception patterns across five key IPAS topics—force and motion, changes of state, energy, weather/water cycle, and ecosystems—using a two-tier diagnostic test developed and validated in context. Second, it translates these patterns into operational pedagogical recommendations aligned with the Merdeka Curriculum, including a 5E learning sequence emphasizing elicitation of prior ideas, cognitive-conflict prompts, explicit refutation, multiple representations, and mini two-tier formative checks to monitor shifts in misconceptions over time.

The novelty of this study lies in its integrated cross-topic focus spanning science and social domains, the use of linguistically and culturally contextualized two-tier instruments, and a direct bridge from misconception maps to classroom-applicable instructional design. The urgency is driven by the need to strengthen students' science–social literacy and to ensure that the Merdeka Curriculum is implemented on a robust conceptual foundation from the earliest grades.

Research objectives are: (1) to map students' misconception pattern profiles across five key IPAS topics and identify the most prevalent specific misconceptions; (2) to translate these findings into specific, contextualized, and operational instructional strategies and formative evaluations to measurably reduce misconception proportions; and (3) to generate actionable empirical evidence for teachers to improve IPAS instructional practice in elementary classrooms, strengthen students' conceptual foundations, and ultimately enhance science–social literacy sustainably.

RESEARCH METHODS

This study employed an explanatory sequential mixed-methods design (Acquah et al., 2021). The quantitative phase mapped the prevalence and patterns of misconceptions across IPAS topics, while the qualitative phase (clinical interviews supported by concept cartoons) probed the reasoning structures underlying response patterns, validated interpretations of the quantitative findings, and enriched the pedagogical recommendations. Integration occurred at the interpretation stage through joint displays and meta-inferences to bridge misconception maps with instructional design implications (Guetterman et al., 2015). The study was conducted in three public elementary schools in Sukabumi Regency (pseudonymized as SDN-A, SDN-B, SDN-C) with participants from Grades 4, 5, and 6. Quantitative sampling was purposive (Memon et al., 2024) via principal and teacher networks, with inclusion criteria: (a) active enrollment in Grades 4–6, (b) willingness to complete the diagnostic test, and (c) parental/guardian consent. The qualitative sub-sample was selected using purposive maximum variation based on quantitative response patterns (strong misconception, partial, scientifically sound) and expanded via snowballing as needed to achieve adequate variation.

The primary instrument was an Indonesian two-tier diagnostic test locally contextualized for five key IPAS topics: force and motion, changes of state, energy, weather/water cycle, and ecosystems (Ningroom et al., 2025). Each topic comprised four tiered items (Q1–Q4): Q1 elicited common intuitive ideas that often trigger misconceptions, followed by Q2–Q4 varying contexts/representations to assess the stability of students' explanatory schemas (Herrmann-Abell & DeBoer, 2011). Tier 1 (conceptual) consisted of four-option multiple-choice items with distractors derived directly from misconceptions documented in the literature (Shin et al., 2019).



Tier 2 (reasoning) offered targeted rationale options mapping misconception types, plus an “other reason, specify” option for brief open-ended responses (Haslam & Treagust, 1987; Kamcharean & Wattanakasiwich, 2014). A confidence rating was provided as an optional Tier 3 for exploratory purposes but was not a focus of the main analyses (Yousefi et al., 2024). The qualitative instrument was a clinical interview protocol using concept cartoons that juxtaposed scientific and intuitive explanations in local contexts to stimulate reasoning, clarify terminology, and explore conceptual boundaries (Kruit & Bredeweg, 2020).

Data collection proceeded in two phases: a classroom-based two-tier diagnostic test (approximately 35–45 minutes) emphasizing Tier-2 reasoning, followed by individual clinical interviews (approximately 10–15 minutes) with a selected sub-sample based on response patterns. Two-tier test scoring classified Tier 1 as correct/incorrect and Tier 2 as scientific/misconception/partial, then combined categories into “scientifically sound,” “partial,” or “strong misconception” (including false positives), guided by an operational rubric. Instrument quality was established via content validity (Aiken’s $V \geq .80$), internal reliability (Cronbach’s alpha per topic), and inter-rater reliability (Cohen’s kappa/Krippendorff’s alpha $\geq .70$) for Tier-2 coding and combined categorization, alongside item analysis (difficulty, discrimination, distractor functioning).

Quantitative data were analyzed for category proportions by topic/grade (with 95% CIs), trend tests, and exploratory multilevel logistic models for clustering effects; qualitative data were analyzed using directed thematic analysis with triangulation and double-coding. Quantitative and qualitative findings were integrated through joint displays to formulate meta-inferences and operational pedagogical recommendations. All procedures adhered to research ethics (permissions, informed consent, data anonymization). Given purposive sampling and limited geographic scope, generalizability is constrained; findings are positioned as contextual mapping. Detailed instruments, rubrics, and the codebook are provided in the appendix for audit and replication.

RESULTS AND DISCUSSION

Overview

This study collected data from 204 Grade 4–6 students across three public elementary schools (pseudonyms: SDN-A, SDN-B, SDN-C). Grade-level composition was balanced and gender was relatively even, ensuring that emerging patterns were not dominated by any single group. All students completed 20 two-tier items across five IPAS topics, yielding 4,080 responses for quantitative analysis, alongside 36 clinical interviews to probe reasoning behind answers. Tables 1 and Table 2 confirm the balanced sample profile as context for interpreting the findings.

Table 1. Sample distribution by school and grade (n = 204)

School	Grade 4	Grade 5	Grade 6	Total
SDN-A	24	23	22	69
SDN-B	23	22	22	67
SDN-C	23	23	22	68
Total	70	68	66	204

Table 2. Gender composition by grade

Grade	Male	Female	Total
Grade 4	34 (48.6%)	36 (51.4%)	70
Grade 5	33 (48.5%)	35 (51.5%)	68
Grade 6	33 (50.0%)	33 (50.0%)	66
Total	100 (49.0%)	104 (51.0%)	204

Instrument quality

Before interpreting outcomes, instrument credibility was established. Content validity was high, internal reliability per topic was adequate, and Tier-2 reasoning codes were consistent across raters. A summary appears in Table 3.



Table 3. Instrument quality summary

Aspect	Summary result
Content validity (Aiken's V, 20 items)	Mean 0.90 (range 0.86–0.94)
Internal reliability (alpha)	Force–Motion 0.74; Changes of State 0.77; Energy 0.71; Weather/Water Cycle 0.73; Ecosystems 0.76
Inter-rater reliability (open Tier-2)	Cohen's kappa 0.78; Krippendorff's alpha 0.80
Item functioning	Moderate difficulty, good discrimination; misconception- based distractors functioning
Rasch (exploratory)	Most item infit–outfit 0.7–1.3

Category distribution: overall picture

Aggregated responses fell into three categories: scientifically sound 44.2%, partial 28.1%, and strong misconception 27.7%. Thus, over one-quarter of responses were still supported by relatively stable, incorrect explanatory schemas. Grade-level trends showed progressive improvement (Table 4).

Table 4. Category distribution by grade

Category	Grade 4	Grade 5	Grade 6
Scientifically sound	39.1%	43.4%	49.0%
Partial	28.4%	29.6%	27.9%
Strong misconception	32.5%	27.0%	22.8%

Narratively, Table 4 shows two things: first, a consistent increase in “scientifically sound” from Grade 4 to Grade 6; second, “strong misconception” decreases by about ten percentage points between Grade 4 and Grade 6, indicating that cumulative instructional exposure and learning experiences help consolidate conceptual schemas.

Topic hotspots

By topic, the two most challenging areas were Ecosystems and Force–Motion. These topics had the highest proportions of “strong misconception” and the lowest “scientifically sound” compared with other topics. Changes of State showed relatively stronger understanding. See Table 5.

Table 5. Category distribution by topic

Topic	Scientifically sound	Partial	Strong misconception
Force–Motion	42.1%	27.1%	30.8%
Changes of State	47.9%	27.3%	24.8%
Energy	44.6%	28.5%	26.9%
Weather/Water Cycle	43.6%	27.7%	28.7%
Ecosystems	39.2%	27.9%	32.9%

This pattern aligns with the conceptual nature of items: unobservable processes (e.g., particles, energy flow) and system structure (e.g., food webs, decomposers) tend to trigger persistent intuitive explanations.

Item sensitivity across Q1–Q4

The tiered item-writing strategy (Q1–Q4) functioned as designed. Q1—constructed with everyday contexts to elicit intuitive ideas—showed the highest share of “strong misconception.” As contexts and representations varied in Q2–Q4, that proportion declined consistently. Table 6 summarizes the downward trend in “strong misconception” (percent) by topic. Note: the table below reports Q1–Q3 where data were available.

Table 6. Strong misconception rates (%) by item tier

Topic	Q1	Q2	Q3
Force–Motion	55	41	33
Changes of State	47	35	29
Energy	49	34	25
Weather/Water Cycle	52	38	30
Ecosystems	58	43	36

The sharpest drops occurred when representation shifted from verbal narratives to diagrams and models (e.g., force and friction sketches for Force–Motion; particle models for



Changes of State; condensation–coalescence sequences for Weather/Water Cycle; and explicit matter–energy flow diagrams positioning decomposers for Ecosystems). This underscores the centrality of representational support in bridging everyday intuition and scientific concepts.

Quality of reasoning and false positives

Combined Tier-1 and Tier-2 analyses identified approximately 9–12% “false positives”: students selected the correct Tier-1 answer but provided Tier-2 reasoning that contained a core misconception. This occurred most often in Energy and Weather/Water Cycle. In classroom terms, a correct answer does not necessarily signal a robust conceptual schema; teachers should check reasoning to ensure genuine conceptual change.

Qualitative insights: how students’ reason

Clinical interviews clarified the reasoning structures behind the numbers. In Force–Motion, students often analogized force as the “fuel” of motion; when pushing stops, “speed runs out.” Visualizing friction helped correct parts of this reasoning but did not automatically break the assumption that a continuous push is required. In Changes of State, “vapor disappears” was used to explain the lack of visible gas; a closed-system demonstration helped students state “the mass stays the same,” though confusion between heat and temperature persisted. In Energy, flow to the environment was treated as “lost” rather than transformed; energy-flow diagrams shifted narratives toward “converted to heat,” yet understanding of efficiency and “waste” energy remained weak. In Weather/Water Cycle, the “clouds break” idiom was deeply ingrained; visual sequences of condensation helped, but metaphorical phrasing still surfaced. In Ecosystems, many perceived food chains linearly and placed decomposers outside the system; concept cartoons featuring organic remains reinstated decomposers’ roles, but generalizing to food webs required further facilitation.

This study maps elementary students’ misconception patterns across five key IPAS topics and reveals three consistent patterns. First, the proportion of strong misconceptions remains in the range of roughly one-quarter to one-third of responses, with a clear improvement trend across grade levels. Second, tiered item sensitivity (Q1–Q4) shows that intuitive ideas most readily surface in everyday contexts (Q1) and can be attenuated through variation in representations and contexts (Q2–Q4). Third, approximately one in ten responses are false positives—correct answers supported by misconception-based reasoning—underscoring the importance of assessing reasoning rather than end answers alone. Substantively, the most persistent hotspots are Ecosystems and Force–Motion, followed by Weather/Water Cycle; Changes of State is relatively stronger, though non-observable aspects (gas phase, conservation of mass) still pose challenges.

Overall, these findings align with classic international misconception maps: impetus-like theories in force–motion, errors about conservation of matter during phase change, energy conceived as something that is “used up,” the “clouds break” metaphor, and linear food-chain modeling that sidelines decomposers (Mougi, 2020; Zou et al., 2016). In Indonesia, prior studies at elementary to lower secondary levels have reported similar patterns, reinforced by linguistic idioms, textbook illustrations, and limited practical/representational experience (Ert et al., 2014). The novelty of this study lies in three contributions: cross-topic mapping across five domains within a single, consistent design; integrated quantitative–qualitative data that trace reasoning behind answers (not just prevalence); and a systematic test of representational sensitivity (Q1–Q4) showing how varied forms of representation progressively reduce misconception intensity. In addition, the false-positive finding corroborates longstanding cautions in conceptual assessment research that correct answers are not synonymous with robust conceptual understanding (e.g., Haslam & Treagust (1987) dan Hirsch & O’Donnell (2001)), strengthening the case for two-tier classroom assessment.

Cognitively, both Ecosystems and Force–Motion demand systemic relations and intermediate-level reasoning that are not yet fully developed for many elementary students. In Force–Motion, the idea of “force as the fuel of motion” is an adaptive intuitive schema grounded in daily experience (push–move; stop pushing–slow/stop) and reinforced by the everyday neglect of friction (Ma & Rodriguez, 2018). Without explicit representations of



friction, inertia, and the distinction between force and velocity, this schema can persist even when answers become correct in particular contexts. In Ecosystems, students struggle to integrate two threads: energy flow (unidirectional and degrading) and matter cycling (circulating through producers–consumers–decomposers). Textbooks and classroom practices that emphasize simple “chains” often fail to foreground food-web structure and decomposers’ role in closing matter cycles (Samanta et al., 2019). This study shows that when webs and decomposer roles are visualized with local contexts (organic remains/soil), some students revise their ideas, but generalization remains weak—signaling the need for richer, repeated experiences.

The downward pattern in strong misconceptions from Q1 to Q4 underscores the importance of layered representational design. Everyday narratives activate intuitive schemas; force diagrams condition the differentiation of force versus velocity; particle models support reasoning about phases and conservation; process sequences help map condensation–coalescence; and matter–energy flow schematics place decomposers explicitly. These results are consistent with literature on representational competence and dual coding, which shows that shifting modes of representation can facilitate the reorganization of conceptual schemas (Amin, 2017). However, representational variation alone is insufficient; affectively safe cognitive conflict and explicit refutation that ties evidence to scientific concepts are necessary. Clinical interviews indicated that students were willing to reconsider when visual/experiential evidence was presented in sequence and teachers guided the reasoning rather than merely providing answers.

Qualitative findings reinforce the role of everyday language and local idioms as misconception amplifiers, notably “clouds break” and “energy runs out.” While idioms aid communication, they bind inappropriate causal schemas. In the elementary IPAS context, guided translanguaging—starting from students’ language and bridging to scientific terminology via analogies with explicit constraints—can reduce distortion (Clement, 1993). This requires task designs that focus students on mechanism (how and why) rather than surface labels.

These findings align with the Merdeka Curriculum’s emphasis on meaningful understanding, formative assessment, and differentiation. Practically, three implications stand out. First, eliciting prior ideas should be an instructional routine, for example through mini two-tier prompts at the start of a unit. Second, core activities should orchestrate multiple representations and simple empirical evidence (micro-experiments, simulations/visualizations) within a 5E sequence with refutation. Third, lesson closure should include metaconceptual reflection—why initial answers seemed plausible, what evidence shifted thinking, and how to generalize to other contexts. The presence of false positives indicates that formative rubrics should score reasoning, not just answers, and use feedback keyed to specific misconception patterns.

From a conceptual change perspective, the Q1→Q4 pattern can be read as a form of fragile convergence: reduced misconceptions in certain contexts without stable schema reorganization (Gouvea & Simon, 2018). From a knowledge-in-pieces perspective, student responses reflect the activation of context-sensitive p-prims/heuristics; certain representations deactivate/mute problematic p-prims and activate more appropriate reasoning (Fotou & Abrahams, 2016). Integrating these perspectives helps design interventions: rather than merely “correcting” facts, instruction should reconfigure activation conditions and provide experiences that strengthen cross-context conceptual coherence.

The false-positive rate (approximately 9–12%) underscores that single-tier tests risk overestimating understanding. In classroom practice, mini two-tier formative assessments can be routinized: simple items with targeted reasons and an “other reason” option to capture diverse reasoning. Confidence ratings, used judiciously, can also help map “fragile zones” where correct answers co-occur with low confidence or inconsistent reasons (e.g., Ardiansah et al. (2018) dan Tsui & Treagust (2010)). At the school level, item banks keyed to specific



misconceptions—standardized at least for content validity and internal reliability—can improve the quality of formative assessment across classes.

Although the analysis did not find consistent gender differences after controlling for grade and school, between-class variation indicated that discussion practices using images/diagrams correlated with reduced misconceptions on representation-heavy items. This suggests a need for differentiation based on representational access: students with lower text literacy may benefit from visualizations and manipulatives; students with limited scientific vocabulary require explicit language bridges. Within the Pancasila Student Profile framework, collaborative practices and evidence-based argumentation can modulate readiness differences without diluting conceptual targets (Lidinillah et al., 2024; Ndari et al., 2024).

These interpretations should be read in light of design limitations. First, schools were limited to one regency with purposive sampling; generalization warrants caution. Second, although reliability and content validity were adequate, the instrument has not been tested for broader invariance (e.g., Rasch DIF across schools/grades) (Józsa et al., 2023). Third, the design does not test instructional causality; Q1→Q4 findings are diagnostic rather than evaluative of instructional impact. Future work should expand regional coverage; test multilevel models with larger samples; develop and trial a 5E–refutation package grounded in these findings via quasi-experimental designs; and examine translanguaging and multimodal literacy in sustainably reducing misconceptions.

Building on the identified misconception map, the following recommendations can be implemented directly in elementary IPAS classrooms. First, use a 3–5 minute two-tier starter at unit launch to elicit prior ideas and map false-positive risk. Second, design activity sequences that orchestrate topic-appropriate multiple representations: force sketches and friction experiments (Force–Motion); particle models and closed systems (Changes of State); energy-flow diagrams and transformation-tracking tasks (Energy); visual sequences of condensation–coalescence (Weather/Water Cycle); and explicit food webs positioning decomposers, paired with soil/compost sampling (Ecosystems). Third, conduct structured metaconceptual reflection at the end of the cycle and use mini two-tier formative checks to monitor conceptual shifts.

Overall, this study contributes at two levels. At the practice level, it provides a cross-topic misconception map readily translatable into instruction and formative assessment that are sensitive to students' reasoning. At the school/curriculum level, it underscores the urgency of supporting contextualized Indonesian two-tier instrument development and strengthening teacher capacity in multiple-representation design and facilitation of evidence-based argumentation. In this way, Merdeka Curriculum implementation for IPAS can proceed from empirical evidence that maps students' conceptual schemas, rather than assumptions alone (Desi et al., 2025; Sri Hartuti et al., 2024).

CONCLUSIONS AND SUGGESTIONS

This study confirms that misconceptions in key IPAS topics at the elementary level remain relatively high and persistent, particularly in Ecosystems and Force–Motion. Although there is improvement across grade levels, roughly one quarter of responses are still supported by flawed reasoning, and the proportion of false positives indicates that correct answers are not synonymous with robust understanding. Tiered evidence from Q1 to Q4 shows that varying representations and contexts systematically reduces misconception intensity, underscoring the importance of instructional designs that orchestrate multiple representations, introduce cognitively safe conflict, and employ explicit refutation.

Practically, teachers are advised to integrate elicitation of prior ideas using mini two-tier items; facilitate evidence-based activities with appropriate visualizations (for example, particle models, energy-flow diagrams, and food webs that explicitly position decomposers); and close lessons with metaconceptual reflection that evaluates reasoning rather than answers alone. At the school level, priorities include developing a contextualized Indonesian-language two-tier item bank and providing teacher training in representational literacy and evidence-based argumentation. Limitations include a narrow geographic scope, purposive sampling, and the



absence of causal tests of instructional interventions. Future studies should broaden contexts, test instrument invariance, and evaluate a 5E-refutation intervention package using quasi-experimental designs to ensure sustained reductions in misconceptions.

REFERENCES

- Acquah, I. S. K., Naude, M. J., & Soni, S. (2021). How the dimensions of culture influence supply chain collaboration: An explanatory sequential mixed-methods investigation. *Revista de Gestão*, 28(3), 241–262. <https://doi.org/10.1108/REGE-11-2020-0105>
- Amin, T. G. (2017). Representation, concepts, and concept learning. In *Converging perspectives on conceptual change* (pp. 129–149). Routledge. <https://doi.org/10.4324/9781315467139-17>
- Ardiansah, M., Masykuri, M., & Rahardjo, S. B. (2018). Student certainty answering misconception question: Study of three-tier multiple-choice diagnostic test in acid-base and solubility equilibrium. *Journal of Physics: Conference Series*, 1006, 012018. <https://doi.org/10.1088/1742-6596/1006/1/012018>
- Argote, L., & Ingram, P. (2000). Knowledge transfer: A basis for competitive advantage in firms. *Organizational Behavior and Human Decision Processes*, 82(1), 150–169. <https://doi.org/10.1006/obhd.2000.2893>
- Assaraf, O. B., & Orion, N. (2010). System thinking skills at the elementary school level. *Journal of Research in Science Teaching*, 47(5), 540–563. <https://doi.org/10.1002/tea.20351>
- Brewer, A. W. (1952). Why does it rain? *Weather*, 7(7), 195–198. <https://doi.org/10.1002/j.1477-8696.1952.tb01494.x>
- Chiu, M. (2007). A national survey of students' conceptions of chemistry in Taiwan. *International Journal of Science Education*, 29(4), 421–452. <https://doi.org/10.1080/09500690601072964>
- Clement, J. (1993). Using bridging analogies and anchoring intuitions to deal with students' preconceptions in physics. *Journal of Research in Science Teaching*, 30(10), 1241–1257. <https://doi.org/10.1002/tea.3660301007>
- Coley, J. D., & Tanner, K. D. (2012). Common origins of diverse misconceptions: Cognitive principles and the development of biology thinking. *CBE—Life Sciences Education*, 11(3), 209–215. <https://doi.org/10.1187/cbe.12-06-0074>
- Crespo, M. Á. G., & Pozo, J. I. (2004). Relationships between everyday knowledge and scientific knowledge: Understanding how matter changes. *International Journal of Science Education*, 26(11), 1325–1343. <https://doi.org/10.1080/0950069042000205350>
- Desi, D. I., Meilina, F., Khairiyah, K. Y., & Sanusi, R. (2025). Implementation of the Independent Curriculum in science learning in elementary schools. *TOFEDU: The Future of Education Journal*, 4(6), 2447–2451. <https://doi.org/10.61445/tofedu.v4i6.834>
- Desstya, A., Sayekti, I. C., Rahmawati, F. P., Hidayati, Y. M., Kaltsum, H. U., Mahya, Y. M. Z., Kamila, S. J. P., Khotimah, N. A., Fidyastuti, N., & Diaz, A. M. (2024). Development of a natural and social sciences learning model based on socio scientific issues to strengthen scientific literacy of Islamic elementary school students. *Jurnal Penelitian Pendidikan IPA*, 10(11), 8703–8711. <https://doi.org/10.29303/jppipa.v10i11.9094>
- Dib, C. (2013). Mass as a form of energy in a simple example. *The Physics Teacher*, 51(9), 546–548. <https://doi.org/10.1119/1.4830069>
- Ert, E., Raz, O., & Heiman, A. (2014). (Poor) Seeing is believing: When direct experience impairs product promotion. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2724084>
- Fernandez, E., Yoshioka, N., Washizaki, H., & Syed, M. (2016). Modeling and security in cloud ecosystems. *Future Internet*, 8(2), 13. <https://doi.org/10.3390/fi8020013>
- Fitriyanti, D., Mudiono, A., & Suciptaningsih, O. A. (2025). Analysis of critical thinking skills of primary students in IPAS learning. *Journal of Research in Instructional*, 5(1), 52–61. <https://doi.org/10.30862/jri.v5i1.539>



- Fotou, N., & Abrahams, I. (2016). Students' analogical reasoning in novel situations: Theory-like misconceptions or p-prims? *Physics Education*, 51(4), 044003. <https://doi.org/10.1088/0031-9120/51/4/044003>
- Gouvea, J. S., & Simon, M. R. (2018). Challenging cognitive construals: A dynamic alternative to stable misconceptions. *CBE—Life Sciences Education*, 17(2), ar34. <https://doi.org/10.1187/cbe.17-10-0214>
- Guetterman, T. C., Feters, M. D., & Creswell, J. W. (2015). Integrating quantitative and qualitative results in health science mixed methods research through joint displays. *The Annals of Family Medicine*, 13(6), 554–561. <https://doi.org/10.1370/afm.1865>
- Guzzetti, B. J. (2000). Learning counter-intuitive science concepts: What have we learned from over a decade of research? *Reading & Writing Quarterly*, 16(2), 89–98. <https://doi.org/10.1080/105735600277971>
- Halidullin, O., Nurushev, M. Z., & Duskaev, K. (2018). The man and the cycle of water in nature. *Journal of Marine Biology and Aquaculture*, 4(1), 27–29. <https://doi.org/10.15436/2381-0750.18.1780>
- Haslam, F., & Treagust, D. F. (1987). Diagnosing secondary students' misconceptions of photosynthesis and respiration in plants using a two-tier multiple choice instrument. *Journal of Biological Education*, 21(3), 203–211. <https://doi.org/10.1080/00219266.1987.9654897>
- Haug, B. S., & Ødegaard, M. (2015). Formative assessment and teachers' sensitivity to student responses. *International Journal of Science Education*, 37(4), 629–654. <https://doi.org/10.1080/09500693.2014.1003262>
- Herrmann-Abell, C. F., & DeBoer, G. E. (2011). Using distractor-driven standards-based multiple-choice assessments and Rasch modeling to investigate hierarchies of chemistry misconceptions and detect structural problems with individual items. *Chemistry Education Research and Practice*, 12(2), 184–192. <https://doi.org/10.1039/C1RP90023D>
- Hirsch, L. S., & O'Donnell, A. M. (2001). Representativeness in statistical reasoning: Identifying and assessing misconceptions. *Journal of Statistics Education*, 9(2). <https://doi.org/10.1080/10691898.2001.11910655>
- Józsa, K., Oo, T. Z., Borbélyová, D., & Zentai, G. (2023). Exploring the accuracy and consistency of a school readiness assessment tool for preschoolers: Reliability, validity and measurement invariance analysis. *Journal of Intelligence*, 11(10), 189. <https://doi.org/10.3390/jintelligence11100189>
- Kamcharean, C., & Wattanakasiwich, P. (2014, March 26). A two-tier multiple choice questions to diagnose thermodynamic misconception of Thai and Laos students. In *Proceedings of the 12th Asia Pacific Physics Conference (APPC12)*. <https://doi.org/10.7566/JPSCP.1.017008>
- Kruit, P., & Bredeweg, B. (2020). Interactive concept cartoons: Exploring an instrument for developing scientific literacy (pp. 404–409). https://doi.org/10.1007/978-3-030-57717-9_35
- Kuczmann, I. (2017). The structure of knowledge and students' misconceptions in physics. *AIP Conference Proceedings*, 1916, 050001. <https://doi.org/10.1063/1.5017454>
- Kuhn, D., & Pease, M. (2008). What needs to develop in the development of inquiry skills? *Cognition and Instruction*, 26(4), 512–559. <https://doi.org/10.1080/07370000802391745>
- Lakin, J. M., & Wallace, C. S. (2015). Assessing dimensions of inquiry practice by middle school science teachers engaged in a professional development program. *Journal of Science Teacher Education*, 26(2), 139–162. <https://doi.org/10.1007/s10972-014-9412-1>
- Lidinillah, U., Hajidah, G. B. N., Azmi, S. N. Al, & Fisher, D. (2024). The implementation of the Pancasila student profile in mathematics learning for grade XI vocational high school students. *Jurnal Prinsip Pendidikan Matematika*, 7(2), 182–194. <https://doi.org/10.33578/prinsip.v7i2.228>



- Lima, P. da S. N., Silva, L. das A., Felix, I. M., & Brandao, L. de O. (2019). Difficulties in basic concepts of mathematics in higher education: A systematic review. In 2019 IEEE Frontiers in Education Conference (FIE) (pp. 1–7). <https://doi.org/10.1109/FIE43999.2019.9028658>
- Ma, D., & Rodriguez, A. (2018). Friction variability in planar pushing data: Anisotropic friction and data-collection bias. *IEEE Robotics and Automation Letters*, 3(4), 3232–3239. <https://doi.org/10.1109/LRA.2018.2851026>
- Madsen, S., Nielsen, L., Hautopp, H., & Jensen, I. (2015). An empirical investigation of the practices and challenges specific to international user studies (pp. 132–148). https://doi.org/10.1007/978-3-319-22698-9_10
- Matsuda, N. (2009). Citizens in the era of governance: Citizen participation and active citizenship. *Interdisciplinary Information Sciences*, 15(2), 231–242. <https://doi.org/10.4036/iis.2009.231>
- Memon, M. A., Thurasamy, R., Ting, H., & Cheah, J.-H. (2024). Purposive sampling: A review and guidelines for quantitative research. *Journal of Applied Structural Equation Modeling*, 9(1), 1–23. [https://doi.org/10.47263/JASEM.9\(1\)01](https://doi.org/10.47263/JASEM.9(1)01)
- Métoui, A., & Trudel, L. (2021). Children and preservice teachers' misconceptions and scientifically acceptable conceptions about movement, force, and gravity. In *Modern perspectives in language, literature and education* (Vol. 5, pp. 26–42). Book Publisher International. <https://doi.org/10.9734/bpi/implle/v5/9032D>
- Moore-Hart, M. A., Liggit, P., & Daisey, P. (2004). Making the science literacy connection: After-school science clubs. *Childhood Education*, 80(4), 180–186. <https://doi.org/10.1080/00094056.2004.10522228>
- Mougi, A. (2020). Coupling of green and brown food webs and ecosystem stability. *Ecology and Evolution*, 10(17), 9192–9199. <https://doi.org/10.1002/ece3.6586>
- Ndari, M. S., Sukarno, S., & Karsono, K. (2024). Analysis of collaboration profiles in the strengthening project of Pancasila student profile in primary schools. *Social, Humanities, and Educational Studies (SHES): Conference Series*, 7(1), 184. <https://doi.org/10.20961/shes.v7i1.84308>
- Ningroom, R. A. A., Yamtinah, S., & Riyadi, R. (2025). A two-tier multiple-choice diagnostic test to find student misconceptions about the change of matter. *Journal of Education and Learning (EduLearn)*, 19(2), 1144–1156. <https://doi.org/10.11591/edulearn.v19i2.21478>
- Putriyani, P., Masriani, M., & Lestari, I. (2024). Development of guided inquiry-based student worksheets on element, compound, and mixture materials. *Jurnal Pendidikan Glasser*, 8(1), 78–86. <https://doi.org/10.32529/glasser.v8i1.996>
- Rahmani, C. D., Adrias, A., & Suciana, F. (2025). Penggunaan media pembelajaran berbasis teknologi dalam pembelajaran IPAS di sekolah dasar. *Sinar Dunia: Jurnal Riset Sosial Humaniora dan Ilmu Pendidikan*, 4(1), 268–278. <https://doi.org/10.58192/sidu.v4i1.3193>
- Samanta, O. da S., Cibebe, P. T., & Solange, X. dos S. (2019). Decomposição e ciclagem de nutrientes: Uma análise da abordagem do livro didático e da prática docente no ensino médio. *Tecné Episteme y Didaxis: TED*, 45, 57–70. <https://doi.org/10.17227/ted.num45-9833>
- Sari, D. R., Ramdhani, D., & Surtikanti, H. K. (2019). Analysis of elementary school students' misconception on force and movement concept. *Journal of Physics: Conference Series*, 1157, 022053. <https://doi.org/10.1088/1742-6596/1157/2/022053>
- Shin, J., Guo, Q., & Gierl, M. J. (2019). Multiple-choice item distractor development using topic modeling approaches. *Frontiers in Psychology*, 10, 825. <https://doi.org/10.3389/fpsyg.2019.00825>
- Sri Hartuti, Wahira, & Amir Pada. (2024). Analisis implementasi Kurikulum Merdeka terhadap prestasi belajar siswa kelas IV pada pembelajaran IPAS di SDN 25 Mattirowalie. *Jurnal Pustaka Cendekia Pendidikan*, 2(3), 410–416. <https://doi.org/10.70292/jpcp.v2i3.107>
- Stavy, R. (1991). Using analogy to overcome misconceptions about conservation of matter. *Journal of Research in Science Teaching*, 28(4), 305–313. <https://doi.org/10.1002/tea.3660280404>



- Sullivan, G., & Edmondson, C. (2008). Heat and temperature. *Continuing Education in Anaesthesia Critical Care & Pain*, 8(3), 104–107. <https://doi.org/10.1093/bjaceaccp/mkn014>
- Suprpto, N. (2020). Do we experience misconceptions? An ontological review of misconceptions in science. *Studies in Philosophy of Science and Education*, 1(2), 50–55. <https://doi.org/10.46627/sipose.v1i2.24>
- Tateo, L. (2014). Science at the supermarket: Multiplication, personalization and consumption of science in everyday life. *Integrative Psychological and Behavioral Science*, 48(2), 161–175. <https://doi.org/10.1007/s12124-014-9263-5>
- Tsui, C., & Treagust, D. (2010). Evaluating secondary students' scientific reasoning in genetics using a two-tier diagnostic instrument. *International Journal of Science Education*, 32(8), 1073–1098. <https://doi.org/10.1080/09500690902951429>
- Waseso, H. P., Sekarinasih, A., & Prasetyo, S. (2024). Implementasi pembelajaran sains dalam Kurikulum Merdeka: Membangun kemandirian berpikir siswa sekolah dasar. *Nusantara: Jurnal Pendidikan Indonesia*, 4(4), 1001–1016. <https://doi.org/10.14421/njpi.2024.v4i4-8>
- Yansah, E. A., & Zulfadewina, Z. (2024). Analysis of the implementation of the independent curriculum for IPAS subject in Class IV-A at SDN Tengah 03. *Journal of Social Science and Education Research*, 1(1), 45–59. <https://doi.org/10.59613/jeywys97>
- Yousefi, M., Lindsay, M. D., & Kreuzer, O. (2024). Mitigating uncertainties in mineral exploration targeting: Majority voting and confidence index approaches in the context of an exploration information system (EIS). *Ore Geology Reviews*, 165, 105930. <https://doi.org/10.1016/j.oregeorev.2024.105930>
- Zou, K., Thébault, E., Lacroix, G., & Barot, S. (2016). Interactions between the green and brown food web determine ecosystem functioning. *Functional Ecology*, 30(8), 1454–1465. <https://doi.org/10.1111/1365-2435.12626>

