



Mathematical Critical Thinking of Madrasah Ibtidaiyah Students: Component Profile and Difficulties in Open-Ended Mathematics Tasks

Annisa Nurfitriani^{1*}, Turmudi², Verra Wulandary³, Asep Nurhuda⁴

^{1,2} Universitas Pendidikan Indonesia, Kota Bandung, Indonesia

³ Universitas Hiroshima, Hiroshima, Jepang

⁴ STKIP Purwakarta, Purwakarta, Indonesia

ARTICLE INFO

Article history:

Submitted: September 06, 2026

Final Revised: Sept. 21, 2026

Accepted: September 30, 2026

Published: September 30, 2026

Keywords:

Mathematical Critical Thinking;
Open-Ended Mathematics
Tasks; Madrasah Ibtidaiyah;
Elementary Education;
Qualitative Descriptive Study



ABSTRACT

Purpose: This study describes the mathematical critical thinking profile of fifth-grade students in a madrasah ibtidaiyah and identifies the difficulties they experience when solving open-ended mathematics tasks, a context that previous research has rarely examined. **Methods:** A descriptive qualitative approach was employed with 27 fifth-grade students of Madrasah Ibtidaiyah Hidayatul Mubtadiin, an Islamic elementary school under the Ministry of Religious Affairs of Indonesia. Critical thinking was examined using Facione's framework covering interpretation, analysis, evaluation, inference, and explanation. Data were collected through an open-ended prism problem instrument validated by two mathematics education experts, followed by semi-structured interviews with the classroom teacher and nine selected students. The interactive analysis model of Miles and Huberman was used with source and method triangulation. Because a single item represented each critical thinking component, the component percentages are reported as descriptive indicators of performance in this sample and not as definitive measures of ability. **Findings:** Critical thinking developed unevenly across the five components. Evaluation and inference appeared relatively well developed, whereas interpretation recorded the lowest class achievement of all five components, including among high-performing students, who also found it difficult to articulate the meaning of mathematical information. Since each component was represented by one item, this difference is reported as an observed pattern rather than as a definitive ordering of component ability. The pattern suggests that critical thinking does not automatically follow content mastery, but requires deliberately designed instructional support. **Research Implications:** Mathematics instruction in elementary madrasah should explicitly scaffold interpretive and analytical processes, for example by training students to restate problems, identify relevant information, and explain their reasoning before engaging in higher-order evaluation. As a single-school qualitative study, these implications are context-bound and are offered as pedagogical recommendations rather than as tested effects. **Originality:** Unlike studies that rely on public elementary schools or written-test scores alone, this study combines a validated open-ended instrument with interviews to profile critical thinking difficulties in the under-represented madrasah ibtidaiyah context.



Doi: <https://doi.org/10.61255/jupiter.v4i4.2028>

INTRODUCTION

Mathematics education aims at more than the acquisition of computational routines. Across national curricula and international policy discourse, the ability to reason, justify, evaluate, and create with mathematics is treated as an essential outcome of schooling rather than an optional extra (National Council of Teachers of Mathematics, 2000). This conviction reflects a broader understanding of education as the deliberate cultivation of higher-order cognitive capacities, among which critical thinking holds a privileged position in the set of competencies widely labeled twenty-

first-century skills (Saavedra & Opfer, 2012). Learning science research has shown that such competencies are not naturally acquired but must be intentionally taught and practiced within subject matter, which places the classroom at the center of efforts to develop them (Saavedra & Opfer, 2012). For mathematics, the implication is direct: students should leave school not only able to calculate, but able to decide what a problem means, which strategy is defensible, and whether an obtained answer is plausible.

Indonesian education policy affirms this orientation. National education policy identifies critical reasoning among the graduate competencies that all levels of schooling, including elementary education, are expected to develop (Kemendikbudristek, 2023). The same expectation reaches Islamic schools. Ministerial Regulation of Religious Affairs Number 347 of 2022, which governs the implementation of the Merdeka Curriculum in madrasah, requires learning experiences that strengthen the profile of students as critical, creative, and collaborative thinkers (Abidin et al., 2024). Indonesian elementary mathematics curricula likewise incorporate problem solving and reasoning as process standards that should accompany content standards (National Council of Teachers of Mathematics, 2000). Policy intent, however, is not classroom reality. Evidence from Indonesian classrooms repeatedly shows that routine exercises dominate instruction and that many students, although able to reproduce procedures, falter when a task demands judgment rather than recall (Fajari, 2021; Lestari et al., 2024).

The gap between aspiration and performance is visible in international assessment data, although such evidence must be interpreted with care. In the 2022 Programme for International Student Assessment (PISA), Indonesian fifteen-year-olds attained a mean mathematics score of 366, well below the OECD average of 472, and only 18% of Indonesian students reached Level 2 or above, the proficiency baseline considered necessary for applying basic mathematics in real-world contexts, compared with 69% of students across OECD countries (OECD, 2023). PISA measures mathematical literacy, the capacity to formulate, employ, and interpret mathematics in varied situations, rather than critical thinking as such (OECD, 2023). The two constructs nonetheless overlap substantially, because both concern the reflective use of mathematics rather than the reproduction of memorized procedures. National results therefore serve in this study as contextual background indicating a systemic challenge in higher-order mathematics; they are not treated as evidence about any particular classroom or age group, a distinction that the study maintains throughout.

Understanding what critical thinking is, and what it is not, is prerequisite to studying it. The Delphi consensus statement, developed by an international panel of experts, defines critical thinking as purposeful, self-regulatory judgment that results in interpretation, analysis, evaluation, inference, explanation, and self-regulation (Facione, 1990). Alternative conceptualizations exist, including Ennis's account of critical thinking as reasonable reflective thinking focused on deciding what to believe or do, which emphasizes dispositions alongside skills (Ennis, 2018). The two frameworks agree on the central point that matters for research: critical thinking is analyzable into identifiable cognitive components rather than being an undifferentiated synonym for intelligence or general cleverness. The present study adopts the Facione framework because its components map cleanly onto observable behavior in mathematics problem solving, and it examines the first five components, interpretation, analysis, evaluation, inference, and explanation, which are sufficiently distinct to allow a fine-grained description of where student thinking succeeds and where it breaks down.

In school mathematics, these components become visible in concrete operations. Interpretation appears when a student determines what a problem actually states, distinguishes the given information from what is asked, and restates the problem in own words. Analysis appears when the student untangles the relations among the elements of the problem and identifies the structure that connects them. Evaluation appears when the student judges whether a selected strategy is appropriate and whether a produced answer is consistent with the conditions of the task. Inference appears when the student draws a conclusion from the available evidence, and explanation appears when the student communicates and justifies the reasoning that led to the conclusion (Facione, 1990). Open-ended tasks are particularly suited to exposing these operations, because they do not provide ready-made answer options and therefore require students to construct, justify, and defend a solution path. Such tasks transform an exercise in recall into an occasion for interpretation, analysis, evaluation, inference, and explanation, which is why they have become a standard instrument in studies of mathematical thinking (Fajari, 2021; Liu et al., 2022).

A substantial body of evidence indicates that critical thinking responds to deliberately designed instruction. Meta-analytic reviews covering several decades of intervention research conclude that teaching strategies that make thinking explicit, provide authentic problems, and support metacognitive reflection produce measurable gains in critical thinking (Abrami et al., 2015). The evidence available for elementary learners specifically remains thinner and more variable than the evidence base for older students, and the intensity and design of instruction determine the size of any effect (Liu et al., 2022). The distribution of this evidence, however, is uneven. Most intervention studies have been conducted with secondary and tertiary students, while elementary learners remain comparatively underrepresented in the literature (Abrami et al., 2015; Liu et al., 2022). The elementary years are nonetheless a formative period for

reasoning: it is during late childhood that children begin to move from recognizing mathematical facts toward understanding the relations and justifications that underlie them, and insufficient support during this transition can constrain later mathematical development (National Council of Teachers of Mathematics, 2000).

Recent Indonesian research illustrates both the concern and the possibility at the elementary level. A qualitative case study of 29 fifth-grade students reported an average critical thinking score of only 50 out of 100, with merely 10% of students meeting the school minimum standard on open description tasks (Fajari, 2021). Classroom studies in Indonesian public elementary schools have shown that structured, differentiated instruction can improve students' mathematical critical thinking (Lestari et al., 2024). What research in these settings rarely provides, however, is a component-by-component account of how students actually reason. Aggregate scores can show that thinking is weak, but they cannot show whether the weakness originates in misreading the problem, in failing to analyze relations, in accepting an unjustified answer, or in being unable to explain a correct result. Knowledge of such points of breakdown is precisely what teachers need in order to design targeted instruction.

Two features distinguish the present study from prior work. The first is its analytical focus on difficulties rather than outcomes. The study profiles how fifth-grade students perform on each critical thinking component when solving open-ended mathematics tasks about prisms, a grade-five geometry topic in which problems are embedded in text and diagrams. Open-ended prism tasks require students to interpret a situation, select and evaluate a strategy, draw a conclusion, and explain their reasoning, thereby generating evidence about the reasoning process itself. The aim is not to compare teaching methods or to measure learning gains, but to describe, in some detail, the reasoning patterns and task-solving difficulties that characterize students with relatively high, moderate, and low performance. The second feature is its context. Existing Indonesian research on mathematical critical thinking has concentrated on public elementary schools and secondary education, whereas madrasah ibtidaiyah, Islamic elementary schools operating under the Ministry of Religious Affairs and combining the national curriculum with religious formation, remain rarely studied (Abidin et al., 2024). The madrasah context is not incidental: it differs from public schools in governance, curriculum organization, and student population, and findings from public schools cannot be assumed to transfer without empirical support.

This study therefore profiles the mathematical critical thinking of fifth-grade students in a madrasah ibtidaiyah and identifies the difficulties they experience in solving open-ended mathematics tasks. Two questions guided the inquiry. First, how do students perform on each of the five critical thinking components of interpretation, analysis, evaluation, inference, and explanation when solving open-ended prism tasks? Second, what reasoning patterns and task-solving difficulties characterize students with relatively high, moderate, and low performance on those tasks?

METHOD

This study employed a descriptive qualitative design to construct a profile of mathematical critical thinking among fifth-grade students and to identify the difficulties they experience when solving open-ended mathematics tasks. The design was selected because the research objective was not to test causal relationships or to measure the effect of an intervention, but to describe reasoning processes in their natural setting through written responses and interview accounts (Creswell & Poth, 2016). Within this qualitative frame, the numerical scores derived from the written test played a limited and transparent role. They served as a descriptive indicator of relative performance across the five critical thinking components and as the basis for purposively selecting cases for in-depth analysis, while all substantive interpretations were anchored in the students' actual reasoning rather than in the scores alone. Consistently with that limited role, the performance groups are labelled relative - relatively high, moderate, and relatively low - and they serve only to structure the qualitative comparison. The rubric developed for this study assigned each response a score from 0 to 2 per item, distinguishing no understanding, partial understanding, and complete and correct reasoning, so that partial reasoning could be acknowledged. The score bands reported in the Results were not derived from a published instrument; they were set by the researchers to partition the observed distribution of total scores into three groups of comparable size for case selection, and they are therefore descriptive rather than psychometric.

The study was conducted at Madrasah Ibtidaiyah Hidayatul Mubtadiin in Sukatani District, Purwakarta, Indonesia, an Islamic elementary school under the Ministry of Religious Affairs located in a semi-rural area near the Jatiluhur Reservoir, where the livelihood of most families is connected to freshwater fish farming. The madrasah context was chosen deliberately, because previous studies of mathematical critical thinking in Indonesian elementary education have concentrated on public schools, leaving madrasah ibtidaiyah settings largely unexamined. The participants were 27 fifth-grade students, comprising 11 boys (40.7%) and 16 girls (59.3%) aged 10 to 11 years. Purposive sampling was applied to obtain information-rich cases; the sample intentionally spanned different levels of test performance so that reasoning patterns could be compared across the spectrum.

The primary instrument was a mathematics critical thinking test consisting of five open-ended items on the topic of prisms, each constructed to correspond to one component of critical thinking as defined by [Facione \(1990\)](#), namely interpretation, analysis, evaluation, inference, and explanation. The open-ended format was chosen deliberately, because this format does not present ready-made answer options and therefore requires students to externalize their reasoning rather than merely select among alternatives ([Facione, 1990](#)). A semi-structured interview guide was developed as a supporting instrument to probe the students' problem-solving strategies, reasoning processes, and perceived difficulties. Both instruments were validated for content by two senior mathematics education experts from Universitas Pendidikan Indonesia and were revised twice in response to their feedback on clarity, content adequacy, and alignment with the research objectives. The same two experts also reviewed the scoring rubric against the five critical thinking indicators.

Data collection proceeded sequentially. The written test was first administered to all 27 students in their regular classroom setting. Each response was examined using the five critical thinking indicators, and students were assigned to relative performance groups on the basis of the total score obtained from the rubric together with the overall quality, consistency, and completeness of their responses across the items. These groups were used only to select cases for the qualitative comparison and are not presented as a validated classification of ability. The interviews were conducted individually in a quiet room at the school, lasted approximately twenty minutes each, were audio-recorded with consent, and were transcribed verbatim before analysis. Semi-structured interviews were then conducted with the mathematics teacher and with nine students purposively selected to represent the three performance groups, exploring how the students interpreted the problems, the steps they took in solving them, and the difficulties they perceived while working on the open-ended items. Ethical procedures were observed throughout the study. Approval was obtained from the relevant institutional authority, written consent was secured from the students' parents or guardians, and verbal assent was obtained from the students themselves. Participation was voluntary, anonymity was guaranteed, and the data were used exclusively for research purposes.

Data analysis followed the interactive model of qualitative data analysis, which comprises data reduction, data display, and conclusion drawing with verification ([Miles et al., 2014](#)). During data reduction, the written responses and interview transcripts were coded systematically according to the five critical thinking components. The coded data were then organized into displays that made patterns and contrasts across performance levels visible. Conclusions were drawn by examining recurring reasoning difficulties, and these conclusions were continually checked against the data. To sharpen the description of reasoning patterns, two students from each performance level were selected for detailed case analysis, using the total score obtained from the rubric together with the consistency and representativeness of the responses, yielding six case profiles that illustrate how interpretation, analysis, evaluation, inference, and explanation operate in practice. The trustworthiness of the study was addressed through the four criteria proposed by Lincoln and [Guba \(1985\)](#). Credibility was pursued through triangulation between test results and interview findings, complemented by expert validation of the instruments. Transferability was supported by the detailed description of the school context, the participants, and the data collection procedures ([Lincoln & Guba, 1985](#)). Dependability was addressed through transparent and systematic analytic procedures, including documentation of coding decisions, and confirmability was sought by grounding all interpretations in direct evidence from the students' written responses and interview accounts. Transferability was approached through the thick description of the setting reported above, and its limits are made explicit in the discussion of limitations.

RESULTS

The test consisted of five open-ended items, each aligned with one of the five components of critical thinking ([Facione, 1990](#)). Every written response was scored against a rubric developed for this study, and each item carried the same maximum score, so the maximum attainable class score was identical for every component. Percentages were calculated as the class total score divided by the class maximum score, which was obtained by multiplying the number of students by the maximum score per item. Table 1 presents the class achievement for each component.

Table 1. Class Achievement across the Five Critical Thinking Components (N = 27)

Component	Item	Class Total Score	Maximum Score per Item	Class Maximum Score	Percentage
Interpretation	1	22	2	54	41%
Analysis	2	31	2	54	57%
Evaluation	3	46	2	54	85%
Inference	4	42	2	54	78%
Explanation	5	36.5	2	54	68%

Note. Percentage = class total score divided by the class maximum score (27 students x 2 points per item). Responses were scored in whole and half points so that partially correct reasoning could be acknowledged.

As shown in Table 1, student performance was uneven across the five components. Of the five components, interpretation recorded the lowest class achievement (41%), followed by analysis (57%), which suggests that many students struggled to make sense of the problem situation and to identify relevant mathematical relationships. By contrast, evaluation reached the highest percentage (85%), followed by inference (78%), which suggests that once the structure of a problem was recognized, most students were able to assess statements and draw logical conclusions. Explanation stood at 68%, reflecting a continuing challenge in articulating mathematical reasoning clearly. Since every item contributed the same maximum score to the total, the percentages permit direct comparison across components. This pattern points to an association between weaker performance on the initial interpretive and analytical components and lower engagement with higher-order reasoning, rather than a causal sequence, because the data only capture performance at a single point in time.

On the basis of their total scores, students were assigned to three relative performance groups using researcher-defined bands: relatively high (71 to 100), moderate (61 to 70), and relatively low (20 to 60). Because the instrument contained a single item per component, the total score mainly served to distinguish levels of overall task performance for the purpose of case selection, not to measure component ability. Table 2 shows the distribution of students across these groups.

Table 2. Distribution of Students across Relative Performance Groups (N = 27)

Performance Group	Score Range	Number of Students	Percentage
High	71-100	6	22.2%
Moderate	61-70	9	33.3%
Low	20-60	12	44.4%

Note. The bands are researcher-defined thresholds used only as a descriptive device for case selection in this qualitative study, as explained in the Method section.

As presented in Table 2, six students (22.2%) were classified in the relatively high group, nine (33.3%) in the moderate group, and twelve (44.4%) in the relatively low group. The distribution indicates that the majority of participants faced considerable challenges in solving open-ended mathematics tasks. To examine the reasoning behind these percentages, two students from each group were selected for in-depth qualitative analysis following the operational criteria described in the Method section. Table 3 lists the six selected cases.

Table 3. Selected Cases for Qualitative Analysis

Nu.	Student Code	Performance Group
1	S-04, S-14	High
2	S-20, S-26	Moderate
3	S-03, S-22	Low

The cases were S-04 and S-14 from the high group, S-20 and S-26 from the moderate group, and S-03 and S-22 from the low group. Their written responses are examined below with reference to the corresponding figures, which reproduce the students' original work. Interview excerpts are presented in English translation; the interviews themselves were conducted in Indonesian.

The two higher performing students illustrate that procedural strength was not always accompanied by explicit interpretive reasoning. In the response reproduced in Figure 1, S-04 compared two three-dimensional shapes mainly by counting the number of edges and left out other determining characteristics such as the faces and the bases, so the response received only partial credit. Although Questions 2 to 5 were answered correctly, this contrast suggests that successful task completion was not always accompanied by systematic interpretation of the problem itself. This reading was supported in the interview, in which S-04 explained, "When I saw the two shapes, I counted their edges right away. That is what my teacher usually asks us to do, so I did not think about the bases."

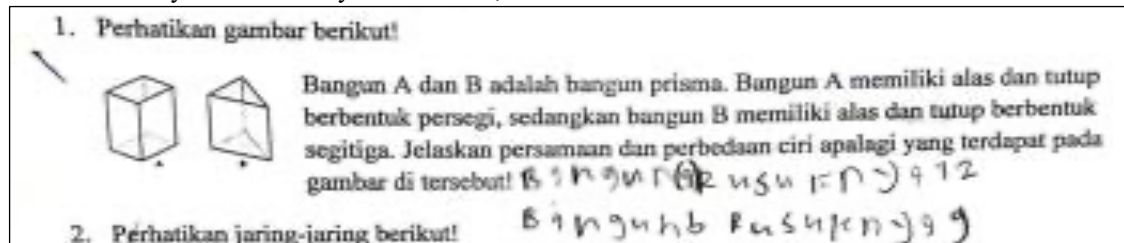


Figure 1. Written response of S-04 to Question 1

A similar pattern appeared in the work of S-14, shown in Figure 2. S-14 identified the difference at the vertices and noted similarities in the sides and the bases, but the explanation provided for these observations was incomplete. In the interview, however, the same student articulated the similarities and differences correctly: "If you ask me directly, I can say that both shapes have two bases and the same number of sides. But when I have to write it, I do not know where to start." This discrepancy points to a gap between mathematical understanding and written mathematical communication rather than to an absence of understanding, since the study did not measure conceptual mastery separately.

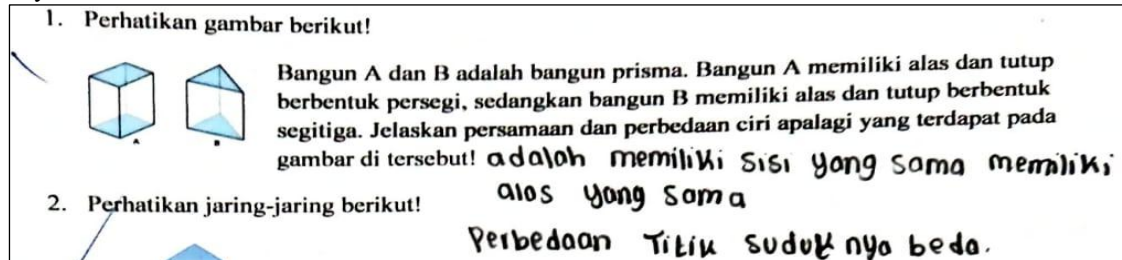


Figure 2. Written response of S-14 to Question 1

In the moderate group, the responses of S-20, reproduced in Figure 3, show difficulty in moving beyond the surface of the problem. S-20 tended to restate information from the problem statement and to identify shapes by name without analyzing their properties or relationships. The response received partial credit for Question 1 and was incorrect for Question 2, while Questions 3 to 5 were answered correctly. The interview indicated that the student understood the context of the problem but found it difficult to compare the shapes analytically: "I knew the question was about prisms, but I only wrote what I remembered about each shape. I was not sure which parts I had to compare."

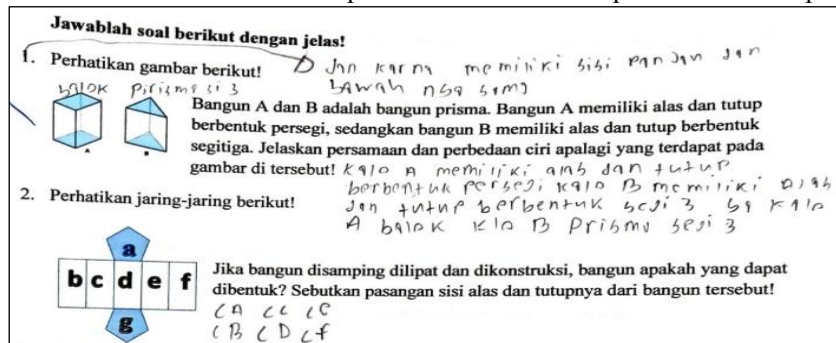


Figure 3. Written response of S-20 to Questions 1 and 2

The work of S-26, reproduced in Figure 4, showed a comparable profile. The written responses to Questions 1 to 3, which correspond to interpretation, analysis, and evaluation, were incomplete or incorrect, whereas the items aligned with inference and explanation were answered relatively well. Table 4 summarizes the quality of S-26's written responses across the five indicators.

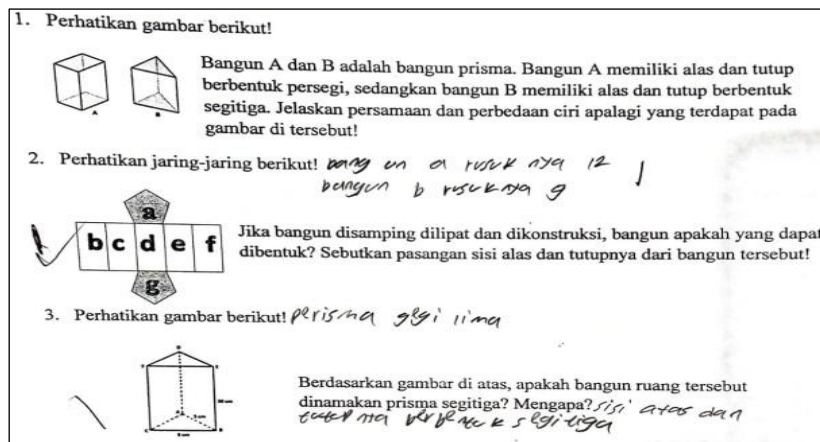


Figure 4. Written response of S-26 to Questions 1, 2, and 3

Table 4. Written Response Quality of Case S-26 across the Five Indicators

Indicator	Item	Response Quality
Interpretation	1	Incomplete or incorrect
Analysis	2	Incomplete or incorrect
Evaluation	3	Incomplete or incorrect
Inference	4	Relatively strong
Explanation	5	Relatively strong

Note. Ratings were derived from S-26's written responses reproduced in Figure 4 and confirmed through the follow-up interview.

The interview with S-26 suggested that the student understood what the questions required but struggled to present reasoning in writing: "I understood what the question meant, but when I had to write the reason, I could not arrange the words. It is easier when the answer is only a number." It should be emphasized that these patterns describe the two cases analyzed here; they cannot be assumed to represent every student in the moderate group. Figure 5 shows the written work of S-03 from the low group. S-03 answered Questions 1 and 2 incorrectly and demonstrated limited ability to evaluate statements in Question 3. The interview indicated that misunderstanding of the problem statement contributed to the incorrect answers: "I did not understand what the question asked. I saw the picture and I guessed, because I was not sure which shape was being talked about." Overall, the response showed limited engagement with the components assessed by the rubric.

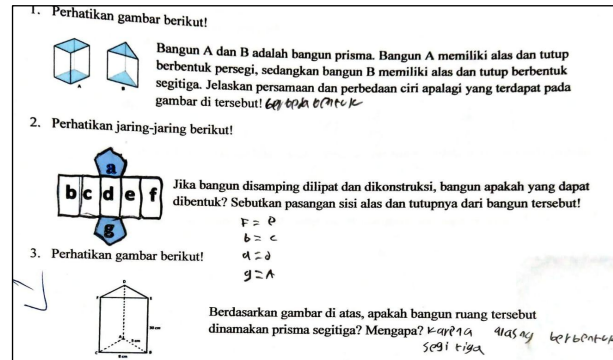


Figure 5. Written response of S-03 to Questions 1, 2, and 3

A similar profile appeared in the work of S-22, shown in Figure 6. In answering Question 5, S-22 misinterpreted the requirements of the problem, misidentified geometric properties, and applied an irrelevant formula. The interview confirmed limited understanding across the test items: "I do not really remember the formula for a prism, so I used the formula for a rectangle. I thought the shapes were the same kind."

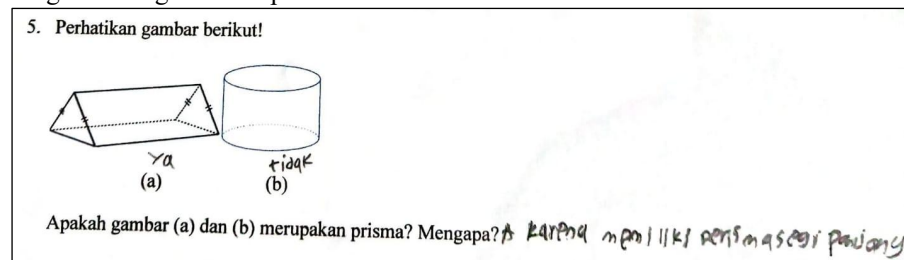


Figure 6. Written response of S-22 to Question 5

Across the six cases, two difficulties recurred. Students in every group found it hard to express their mathematical reasoning clearly in writing, and several struggled to distinguish the base from the top surface of three-dimensional shapes. These observations characterize the cases analyzed in this study and provide a descriptive basis for the discussion that follows, without implying that the same pattern holds for all students in each group.

DISCUSSION

The main contribution of this study lies in its component-level profile of mathematical critical thinking among fifth graders in a madrasah ibtidaiyah, a context that remains underrepresented in the literature. The results show that evaluation (85%) and inference (78%) recorded the highest class achievement, whereas interpretation (41%) and analysis (57%) recorded the lowest. Because each component was represented by a single item, this ordering is reported

as an observed pattern in this sample rather than as a definitive ranking of component ability. This pattern suggests that many students could assess statements and draw conclusions once the structure of a problem was recognized, but far fewer could actively construct meaning from the problem situation in the first place (Facione, 1990). Theoretically, the weakness in interpretation is hardly surprising. Interpretation is the entry point of critical thinking; it demands that a student decode a mathematical situation presented in multiple semiotic registers, namely written text, geometric figures, and mathematical symbols, and then translate those codes into a coherent personal understanding. In the prism tasks, this process required students to identify which attributes were determining features of a shape. Several students instead fixated on surface attributes such as the number of edges, suggesting that their difficulty was located in the transition from recognizing a shape to reasoning about the relations that define it. The linguistic load of open-ended items, the demand to verbalize geometric relationships, the difficulty of reading the diagrammatic information, and limited prior experience in comparing and justifying mathematical structures are all plausible sources of this difficulty. These sources were not measured in the present study, however, so they are offered as hypotheses for further research and must not be read as explanations established by these data (Hardianto et al., 2024).

The uneven profile documented here resonates with earlier studies reporting that elementary students often operate at a procedural level without engaging in sense making (Hiebert & Grouws, 2007; Schoenfeld, 2016). Students who could complete routine calculations were not necessarily able to interpret, analyze, or explain the underlying ideas, a pattern that has also been observed in Indonesian elementary classrooms (Fajari, 2021; Hardianto et al., 2024). The present findings add a nuance to this body of work by showing that the bottleneck is not located in the higher-order components alone. Evaluation and inference were performed relatively well once a problem had been structured, whereas the earlier components of interpretation and analysis constrained what students could do with the problem. This configuration implies that the components of critical thinking develop unevenly and that a deficit at the foundational layer can shape, but does not causally determine, performance in the reasoning components. Because the data are cross-sectional and descriptive, the claim here is limited to an observed association rather than a developmental mechanism (Abrami et al., 2015).

The interviews also caution against treating written responses as a transparent record of critical thinking ability. S-14, for instance, could articulate the similarities and differences between the shapes correctly in conversation but produced an incomplete written explanation. Cases of this kind indicate that limited written output may reflect difficulties in mathematical communication rather than the absence of reasoning. This distinction matters both conceptually and practically. Conceptually, it warns that paper-based profiles of critical thinking can underestimate students' actual competence when writing fluency is low; practically, it suggests that teachers should provide multiple opportunities for students to express reasoning orally as well as in writing. The finding reinforces the position that communication is not an accessory to mathematical thinking but an integral component of it (National Council of Teachers of Mathematics, 2000).

Two sets of pedagogical implications follow, and they should be distinguished by their source. The first set derives directly from the difficulties observed in the six cases. Students at all performance levels struggled to identify determining features of geometric shapes and to distinguish the base from the top surface, so instruction should deliberately train students to name the givens of a task, to compare relevant attributes, and to justify why an attribute matters before any computation begins. In addition, because the gap between oral and written reasoning appeared even among high performers, assessment tasks should separate the quality of reasoning from the fluency of written expression, for example by allowing students to explain their solutions verbally or with the aid of structured prompts before they are expected to produce formal written justifications. The second set of implications rests on the wider literature rather than on the data of this study. Reviews and intervention research consistently indicate that critical thinking does not develop spontaneously and that explicit, well-designed instruction is required to cultivate it (Abrami et al., 2015; Ennis, 2018; Saavedra & Opfer, 2012). Structured mathematical discourse, visual representations, concrete manipulatives, and metacognitive prompts are among the instructional strategies recommended in that literature (National Council of Teachers of Mathematics, 2000).

Contextual teaching and learning and realistic mathematics education are often proposed as particularly suitable vehicles for such instruction because they embed mathematical ideas in meaningful situations (Sitorus et al., 2024). The present study did not test either approach, so they should be regarded as potentially relevant pedagogical options rather than as approaches validated by the current data. Intervention studies in the Indonesian elementary context have reported encouraging effects of contextual learning designs on students' critical thinking and mathematical reasoning (Sitorus et al., 2024), and differentiated forms of instruction have likewise been associated with growth in elementary students' mathematical critical and creative thinking (Lestari et al., 2024). Experimental and design-based studies are therefore needed to determine whether approaches of this kind can lift the particular bottleneck identified here, namely the interpretive and analytic foundation of critical thinking.

Finally, several methodological limitations qualify the interpretation of these findings. The instrument contained a single item for each critical thinking component, which limits the breadth of evidence for each component; the scoring of open-ended responses necessarily involves judgment, even when guided by a rubric; content validation was conducted by two subject experts; and the cross-group interpretation was built on six purposefully selected cases from one school and one mathematics topic. These features are consistent with the study's descriptive qualitative design, but they mean that the findings should be read as context-bound insights rather than as generalizable estimates (Lincoln & Guba, 1985). Future research should involve multiple independent raters with reported agreement indices, a larger item pool per component, several schools, and longitudinal or intervention designs that can trace how the interpretive and analytic components of critical thinking develop and how they respond to targeted instruction (Abrami et al., 2015; Schoenfeld, 2016).

CONCLUSION

In response to the first research question, the profile of mathematical critical thinking among fifth graders at the madrasah ibtidaiyah was uneven and generally limited. Of the five components, interpretation recorded the lowest class achievement, reaching only 41% of the class maximum, followed by analysis at 57% and explanation at 68%, whereas inference reached 78% and evaluation 85%. Only six of the twenty-seven students (22.2%) were classified in the relatively high performance group, nine students (33.3%) in the moderate group, and twelve students (44.4%) in the low group. These findings indicate that many students could evaluate statements and draw conclusions once the problem structure was clear, but substantially fewer could interpret a mathematical situation and analyze its determining features in the first place.

In response to the second research question, the analysis of the six cases revealed distinct reasoning patterns across the performance groups. Higher performing students generally completed the items correctly, yet even they sometimes relied on remembered attributes, such as counting edges, and some could explain the similarities and differences between shapes orally but not in writing. Students in the moderate group understood the context of the tasks but had difficulty analyzing relationships between shapes and articulating their reasoning in written form, while lower performing students often misinterpreted the problem statement, misidentified geometric properties, and applied irrelevant formulas. Across all groups, two difficulties recurred: expressing mathematical reasoning clearly in writing and distinguishing the base from the top surface of three-dimensional shapes. Because these patterns rest on written responses that may underestimate students' actual reasoning, they should be interpreted as difficulties in demonstrating mathematical thinking rather than as definitive evidence of absent thinking.

Taken together, the findings provide evidence from an underrepresented madrasah ibtidaiyah context that interpretation and analysis emerged as the most fragile components of mathematical critical thinking among the students studied here, and they should not be interpreted as a developmental hierarchy because the design was cross-sectional. The pedagogical implication, which follows directly from the observed difficulties, is that instruction and assessment should explicitly support meaning making, geometric representation, and mathematical communication, for example through tasks that require students to identify relevant attributes, justify their comparisons, and express reasoning orally as well as in writing. These implications are drawn from the present descriptive evidence and are not claims about developmental mechanisms or the effectiveness of particular interventions, which would require longitudinal and experimental research.

REFERENCE

- Abidin, Z., Wiarsih, N., Faishol, R., & Baharun, H. (2024). Enhancing Students' Critical Thinking Skills in the Implementation of the Merdeka Curriculum in Madrasah. *Al-Mudarris (Jurnal Ilmiah Pendidikan Islam)*, 7(2), 177–196.
- Abrami, P. C., Bernard, R. M., Borokhovski, E., Waddington, D. I., Wade, C. A., & Persson, T. (2015). Strategies for Teaching Students to Think Critically. *Review of Educational Research*, 85(2), 275–314. <https://doi.org/10.3102/0034654314551063>
- Creswell, J. W., & Poth, C. N. (2016). *Qualitative inquiry and research design: Choosing among five approaches*. Sage publications.
- Ennis, R. H. (2018). Critical Thinking Across the Curriculum: A Vision. *Topoi*, 37(1), 165–184. <https://doi.org/10.1007/s11245-016-9401-4>
- Facione, P. A. (1990). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction (The Delphi Report)*. <https://eric.ed.gov/?id=ED315423>
-

- Fajari, L. E. W. (2021). Critical Thinking Skills and Their Impacts on Elementary School Students. *Malaysian Journal of Learning and Instruction*, 18(2). <https://doi.org/10.32890/mjli2021.18.2.6>
- Hardianto, H., Sabar, N., & Safitri, R. D. (2024). Deskripsi Proses Berpikir Kritis dalam Pemecahan Masalah Matematika berdasarkan Kemampuan Awal Siswa Sekolah Dasar. *Jurnal Studi Guru Dan Pembelajaran*, 7(1), 465–477. <https://doi.org/10.30605/jsgp.7.1.2024.4507>
- Hiebert, J., & Grouws, D. A. (2007). The effects of classroom mathematics teaching on students' learning. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 371–404). Information Age Publishing.
- Kemendikbudristek. (2023). *Rapor Pendidikan Indonesia 2023*. Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi.
- Lestari, F., Alim, J. A., & Noviyanti, M. (2024). Implementation of Differentiated Learning to Enhance Elementary School Students' Mathematical Critical and Creative Thinking Skills. *International Journal of Elementary Education*, 8(1), 178–187. <https://doi.org/10.23887/ijee.v8i1.64295>
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic Inquiry*. Sage Publications.
- Liu, Q., Wang, F., Lin, F., Yao, C., & Wang, Y. (2022). Research on Differentiated Training Program of Primary School Mathematics Education Based on Artificial Intelligence. *Int. Conf. Inf. Educ. Technol., ICIET*, 268–273. Scopus. <https://doi.org/10.1109/ICIET55102.2022.9779050>
- Miles, M. B., Huberman, A. M., & Saldana, J. (2014). *Qualitative Data Analysis: A Methods Sourcebook* (3rd ed.). SAGE Publications.
- National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. National Council of Teachers of Mathematics.
- OECD. (2023). *PISA 2022 results: The state of learning and equity in education (Volume I)*. OECD Publishing. <https://doi.org/10.1787/53f23881-en>
- Saavedra, A. R., & Opfer, V. D. (2012). Learning 21st-Century Skills Requires 21st-Century Teaching. *Phi Delta Kappan*, 94(2), 8–13. <https://doi.org/10.1177/003172171209400203>
- Schoenfeld, A. H. (2016). Learning to Think Mathematically: Problem Solving, Metacognition, and Sense Making in Mathematics (Reprint). *Journal of Education*, 196(2), 1–38. <https://doi.org/10.1177/002205741619600202>
- Sitorus, F. D., Hasratuddin, H., & Azhar, E. (2024). Pengembangan LKPD Berbasis HOTS melalui CTL untuk Meningkatkan Kemampuan Berpikir Kritis dan Penalaran Matematis Peserta Didik. *Jurnal Ilmiah Ilmu Pendidikan (JIIP)*, 7(4), 3695–3701. <https://doi.org/10.54371/jiip.v7i4.4297>

ACKNOWLEDGEMENT

The authors express their sincere gratitude to the principal, teachers, and students of Madrasah Ibtidaiyah Hidayatul Muftadiin, Sukatani, Purwakarta, whose cooperation and participation made this study possible. Special appreciation is extended to the mathematics teacher for the assistance provided during the data collection process. The authors also thank their academic supervisors and colleagues for valuable guidance and constructive feedback throughout the research, as well as their families for their unwavering support and encouragement.

AUTHOR CONTRIBUTION STATEMENT

AN: conceptualization, methodology, investigation, formal analysis, and writing of the original draft. TU: conceptualization, supervision, validation, and writing, review and editing. VW: validation and writing, review and editing. AN: methodology, data curation, and writing, review and editing. All authors have read and agreed to the published version of the manuscript.

AI DISCLOSURE STATEMENT

The authors used a generative AI tool during the preparation of this work for drafting assistance, reference verification, translation from Indonesian into English, and language refinement. After using the tool, the authors thoroughly reviewed and edited the content as needed and take full responsibility for the content of the publication.

***Annisa Nurfitriani (Corresponding Author)**

Department of Elementary Education Study Program Faculty of Education,
Universitas Pendidikan Indonesia,
Jl. Dr. Setiabudhi No. 229, Bandung, West Java, 40154, Indonesia
Email: annisa10@upi.edu

Turmudi

Department of Elementary Education Study Program Faculty of Education,
Universitas Pendidikan Indonesia,
Jl. Dr. Setiabudhi No. 229, Bandung, West Java, 40154, Indonesia
Email: turmudi@upi.edu

Verra Wulandary

International Education Development Program,
Universitas Hiroshima,
1 Chome-3-2 Kagamiyama, Higashihiroshima, Hiroshima 739-0046, Jepang
Email: verra.wulandary@gmail.com

Asep Nurhuda

Department of Elementary Education Study Program Faculty of Education,
STKIP Purwakarta,
Jl. Laks. Laut RE. Martadinata No.43, Nagri Tengah, Kec. Purwakarta, Kabupaten Purwakarta, Jawa Barat 41114
Email: asepnurhuda@upi.edu
