

The Influence of Game-Based Activities Within the GASING Method on the Mathematical Critical Thinking of Junior High School Learners

Abu Moh. Rasyid Ridho^{1*}, Al Jupri², Sumanang Muhtar Gozali³

^{1,2,3}Universitas Pendidikan Indonesia



DOI: <https://doi.org/10.46245/ijorer.v7i5.1563>

Sections Info

Article history:

Submitted: July 09, 2026

Final Revised: Sept. 16, 2026

Accepted: September 28, 2026

Published: September 30, 2026

Keywords:

Mathematical Games; GASING Method; Mathematical Critical Thinking; Integer Multiplication; 3T (Remote) Regions



ABSTRACT

Objective: This study examined whether instructional games embedded within the GASING method—an Indonesian acronym for “Easy, Fun, and Enjoyable”—influence junior high school learners’ mathematical critical thinking in integer multiplication. **Method:** A sequential explanatory mixed-methods design paired a nonequivalent control-group quasi-experiment with subsequent qualitative interpretation. Participants (24 experimental, 25 control) were purposively selected from two state junior high schools in a 3T (Frontier, Outermost, and Underdeveloped) region of Lembata Regency, East Nusa Tenggara. A four-item essay test anchored in Facione’s critical thinking dimensions was expert-validated (mean rating 3.54/4.00; Cronbach’s Alpha = 0.907). **Results:** The Mann-Whitney U test showed no statistically significant between-group difference (Sig. = 0.420, $r = -0.115$, very small effect); normalized gain was comparable (experimental 0.40, control 0.41, both moderate). Qualitatively, achievement varied by indicator: interpretation 71.88% (high), analysis and inference 53.13% (moderate), and evaluation 43.75% (low). **Novelty:** This study offers early empirical grounding for a low-barrier, game-enriched GASING approach in resource-limited 3T settings, evaluates games fused into GASING as one integrated package, and surfaces a spoken-written critical thinking disparity that motivates multimodal assessment.

INTRODUCTION

Within the educational priorities of the twenty-first century, the capacity to think critically about mathematics has come to occupy an increasingly prominent position (Wang & Abdullah, 2024). Building on the Delphi consensus report, Facione (1990) characterizes critical thinking as a deliberate, self-regulating form of judgment—one that yields interpretation, analysis, evaluation, and inference by weighing evidence, concepts, methods, criteria, and contextual factors. Translated into mathematics education, these four component skills serve as the core operational markers used to gauge a learner’s mathematical critical thinking, and instructional models deliberately targeting them have been shown to raise achievement measurably (Angraini & Wahyuni, 2021; Setiana et al., 2021). The GASING method itself, an Indonesian acronym for “Easy, Fun, and Enjoyable” and devised by Surya (2013), constitutes a structured route into mathematics that advances through three ordered stages: the concrete (working with physical manipulatives), the abstract (operating on symbolic representations), and mental computation (arithmetic carried out in the head, without any written working). This progression allows learners to assemble conceptual understanding from lived experience before they move into symbolic reasoning. Recent classroom evidence supports pairing this staged logic with play: Maulida et al. (2024) found that Teams-Games-Tournament cooperative learning built on the GASING method raised fourth graders’ conceptual understanding of two-digit multiplication more than conventional instruction did (N-gain 0.56 versus 0.26). Folding instructional games into the GASING framework therefore carries clear promise for enriching this learning process; yet the precise impact of such a

combined design on mathematical critical thinking has so far drawn only sparse empirical scrutiny.

The international picture sharpens the urgency of the question. In the mathematics domain of PISA 2022, Indonesia placed 68th among 81 participating nations with a score of 366—down from 379 in PISA 2018 (OECD, 2023). Such sustained underachievement mirrors pervasive shortfalls in learners' mathematical literacy, including their ability to interpret, analyze, and appraise authentic mathematical situations (Masfufah & Afriansyah, 2021), a pattern Indonesian teachers themselves attribute to limited exposure to reasoning-oriented tasks and uneven instructional resources (Wijaya et al., 2024). The same shortfalls surfaced locally in a preliminary study involving 24 Grade VIII learners at a state junior high school in Lembata Regency, East Nusa Tenggara, where the average critical thinking score reached only 49.74%, broken down by indicator as interpretation 55.2%, analysis 48.96%, evaluation 41.0%, and inference 54.2%. These figures point to a critically low level of critical thinking ability, especially around integer multiplication—a topic that functions as a foundational gateway to more advanced mathematical material.

Much of this weakness can be traced to the prevalence of conventional teaching that prizes procedural memorization—most conspicuously the rote recall of sign rules for integer operations—over conceptual understanding (Lawson et al., 2021). Integers are in any case a well-documented site of difficulty: Grade 7 learners still display misconceptions when operating on negative numbers (Harun et al., 2023), partly because they import whole-number intuitions into a domain where they no longer hold (Aqazade & Bofferding, 2021). Instruction of this kind leaves too little room for learners to cultivate the higher-order skills of analysis, evaluation, and inference, all of which depend on the active construction of knowledge (Facione, 2020). Conditions characteristic of Indonesia's Frontier, Outermost, and Underdeveloped (3T) regions (Presidential Regulation No. 63/2020, 2020), Lembata Regency among them, intensify these difficulties further. Restricted access to pedagogical innovation in such regions narrows learners' exposure to the very experiences that would otherwise stimulate higher-order reasoning. Nurhikmayati et al. (2024) supply direct evidence on this point, reporting that Indonesian junior high school learners in border and highland schools rank significantly below their lowland urban counterparts in mathematical critical thinking—a geographic gradient they trace to unequal access to educational resources and teaching innovation.

The existing literature offers several encouraging leads. Qirom and Juandi (2023), synthesizing Indonesian classroom studies, concluded that traditional games woven into mathematics lessons benefit learners across the cognitive, affective, and psychomotor domains, with cognitive gains the most frequently documented. On a wider canvas, a meta-analysis of 13 empirical studies (40 effect sizes) by Anggoro et al. (2025) found that game-based learning exerts a small but significant positive effect on learners' mathematical higher-order thinking skills ($g = 0.134$), with intervention duration among the significant moderators. In a similar vein, Barz et al. (2024) meta-analyzed school-based digital game interventions and reported a medium effect on cognitive outcomes ($g = 0.67$) and a small effect on affective-motivational outcomes ($g = 0.32$), yet no significant effect on metacognitive outcomes. Pan et al. (2022), reviewing 43 K-12 mathematics studies, called for closer attention to how mathematical content is integrated into game mechanics, and Kärki et al. (2022) showed that a brief game intervention supported primary learners' conceptual rational number knowledge without producing a

significant group difference in adaptive number knowledge. Vankúš (2021), reviewing 57 studies, further reported that the great majority document positive effects on the affective domain – motivation, engagement, and enjoyment – while Hidayat et al. (2024) mapped the expanding range of online mathematics games now available to teachers. Turning specifically to critical thinking, Mao et al. (2022) synthesized 20 studies and detected a significant positive effect (Hedges' $g = 0.863$), and Kannadass et al. (2023) established that critical thinking is tightly bound up with other higher-order competencies in mathematics. Hui and Mahmud (2023), through a systematic review of 28 studies, additionally established that game-based learning shapes both the cognitive and affective domains of mathematics education. Even so, a clear research gap endures on two interlocking fronts. First, at the level of instructional design, studies pairing GASING with game formats have so far measured conceptual understanding rather than critical thinking (Maulida et al., 2024), and no study has yet specifically interrogated the effect of games embedded within the GASING method – as a single, unified instructional package rather than games and GASING treated separately – on learners' mathematical critical thinking. Second, and more specifically, this design has never been examined within Indonesia's 3T regions: because prior game-based and GASING research has been conducted almost exclusively in urban, better-resourced schools, it remains unknown whether such findings transfer to 3T classrooms, where limited access to teaching aids, intermittent teacher training, and geographic isolation may alter how instructional games function. Compounding both gaps, the bulk of prior work depends solely on quantitative analysis and forgoes any in-depth qualitative reading of learners' critical thinking profiles at the indicator level.

In light of the foregoing, this study explicitly aims to: (1) examine the gap in mathematical critical thinking achievement between learners taught through game-based GASING instruction and those taught conventionally; and (2) characterize learners' mathematical critical thinking profiles indicator by indicator, drawing on a qualitative reading of written responses and interviews. Its novelty rests on three pillars: (a) the empirical evaluation of educational games fused into the GASING method as one integrated instructional package; (b) a study context located in an Indonesian 3T region that the literature has chronically overlooked; and (c) a mixed-methods design joining inferential statistical testing to per-indicator qualitative analysis, thereby furnishing a fuller portrait of learners' critical thinking development than earlier single-method work could provide.

RESEARCH METHOD

A sequential explanatory mixed-methods design framed this study. In its opening phase, quantitative data were gathered and examined; a second, qualitative phase then followed in order to expand upon and add depth to the quantitative results (Creswell & Plano Clark, 2018). The quantitative phase rested on a quasi-experimental nonequivalent control group design, with a pretest-treatment-posttest sequence delivered to both groups in the absence of randomization (Sugiyono, 2017). The qualitative phase, in turn, comprised a close reading of learners' written work together with semi-structured interviews, undertaken to bring their indicator-level critical thinking processes into view. Randomization at the individual level was not pursued because the two participating schools – already operating under tight staffing and scheduling constraints typical of the 3T context – required that instruction proceed within intact, administratively fixed

classes; disaggregating and reassigning students would have disrupted the schools' regular teaching schedule and was therefore judged both impractical and ethically inadvisable. This nonequivalent control-group design is consequently treated as the most defensible naturalistic approximation of a controlled comparison available under field conditions, with the pretest serving as a covariate check on initial group comparability.

All Grade VII learners at two state junior high schools in Lembata Regency, East Nusa Tenggara, made up the study population. From this population a sample was drawn by purposive sampling, the selection resting on class equivalence and the practical accessibility of each site. The resulting sample comprised: (1) an experimental class ($n = 24$) given game-integrated GASING instruction across four sessions, each built around a distinct game—a hand-gesture competition game (Session 1: positive times positive multiplication), a mathematics puzzle (Session 2: positive times negative), a mathematics snakes-and-ladders game (Session 3: negative times positive and negative times negative), and a mathematics ludo game (Session 4: integration of all combinations); and (2) a control class ($n = 25$) taught the same content conventionally. Each game was played in small collaborative groups, a configuration chosen because peer interaction during game play has been linked to greater task efficiency and learning engagement (Moon & Ke, 2020) and because learners themselves perceive in-game scaffolding as central to their progress (Sun et al., 2021). In the qualitative phase, the written responses of every one of the 24 experimental learners were analyzed, and three learners were purposively chosen as interview informants to represent variation in ability (high, moderate, and low) as indicated by posttest scores. The school administration granted ethical clearance, and informed consent was obtained from all participants. The modest sample size ($n = 49$) reflects the structural realities of Lembata Regency as a 3T region: only two state junior high schools lie within a single day's safe travel of the research site given the available inter-village and inter-island transportation, and each school offered only one Grade VII class of the size reported here. Rather than treating this constraint as a limitation to be masked, the study frames it explicitly as a boundary condition of 3T educational research, where population access—not statistical convenience—sets the ceiling on feasible sample size; the results are consequently reported and interpreted as pilot-level evidence rather than as findings generalizable beyond comparable 3T settings. This constraint limits generalizability in three concrete ways. First, statistical power was low: with groups of 24 and 25 learners, the between-group comparison had approximately 40% power to detect a medium effect ($d = 0.50$, $\alpha = 0.05$, two-tailed) and reached 80% power only for effects of about $d = 0.82$ or larger, so a nonsignificant result cannot rule out a small or moderate effect. Second, because the sample comprised two purposively selected intact classes, each located in a different school, school- and class-level influences (including differences in teaching staff) cannot be disentangled from the treatment effect, and the findings cannot be assumed to hold for other schools, even within Lembata Regency. Third, the indicator-level qualitative profile rests on 24 written responses and three interview informants and is therefore indicative rather than representative. Accordingly, all inferences are confined to junior high school learners in 3T settings with comparable access, class size, and resource conditions.

Three components constituted the data collection instrumentation: (1) a mathematical critical thinking test of four open-ended essay items, each keyed to one of Facione's (1990) four indicators—interpretation (Item 1: chess game point-loss context), analysis (Item 2: hourly temperature change context), evaluation (Item 3: sea-level change

context), and inference (Item 4: fisherman's income context) – all framed within integer multiplication; (2) a semi-structured interview guide; and (3) observation sheets recording teacher and learner activity. Four expert judges (two mathematics education lecturers and two seasoned junior high school mathematics teachers) validated the test, returning a mean validity rating of 3.54 out of 4.00. Every item met acceptable item validity ($r\text{-calculated} > r\text{-table} = 0.456$), showed a moderate difficulty level, and displayed excellent discriminating power ($DP > 0.70$); across its entirety the test yielded a Cronbach's Alpha reliability coefficient of 0.907, indicating excellent internal consistency. The semi-structured interview guide and the observation sheets underwent the same four-judge expert-validation procedure as the essay test and were judged content-valid for eliciting indicator-level evidence of critical thinking; because both instruments served a qualitative, evidence-gathering function rather than a scored function, no numerical reliability coefficient was computed for them, and their trustworthiness was instead secured through source triangulation across test scores, written responses, and interview transcripts, as detailed in the data-analysis procedure below. Data were collected in the following order: pretest, four instructional sessions, posttest, documentation of answer sheets, and interviews. A GASING-certified teacher delivered the instruction, while the researcher took the role of a nonparticipant observer.

For the quantitative strand, analysis drew on descriptive statistics, the Shapiro-Wilk normality test, Levene's homogeneity test, and a suitable inferential comparison test. Since the control group's posttest data did not meet the normality assumption ($\text{Sig.} = 0.034 < 0.05$), the nonparametric Mann-Whitney U test was selected ($\alpha = 0.05$). Effect size was expressed through rank-biserial correlation ($r = Z/\sqrt{N}$), consistent with APA 7th edition reporting conventions (Cohen, 1988). Because the groups differed in their pretest scores, normalized gain ($N\text{-gain} = [\text{posttest score} - \text{pretest score}] / [100 - \text{pretest score}]$) was additionally calculated to permit a relative comparison of learning improvement. The qualitative strand followed the Miles and Huberman (1994) framework of data reduction, data display, and conclusion drawing. Data adequacy was secured through purposive selection ensuring representational variation in learner ability, and triangulation proceeded by cross-referencing three evidentiary sources: test scores (quantitative), written-response analysis (documentation), and semi-structured interviews with learners and the teacher.

RESULTS AND DISCUSSION

Results

Table 1 sets out the descriptive statistics for mathematical critical thinking scores in each group.

Table 1. Descriptive statistics of mathematical critical thinking scores
(ideal score = 100).

Group	N	Mean	SD	Max.	Min.
Pretest					
Experimental	24	25.64	14.57	68.75	0.00
Control	25	15.50	12.64	43.75	0.00
Posttest					
Experimental	24	55.47	20.30	93.75	25.00
Control	25	50.00	14.77	81.25	31.25

As Table 1 shows, the experimental group began with the higher pretest mean yet recorded the smaller raw gain (29.83 versus 34.50 points) and the wider posttest dispersion. Normalized gain scores were computed to place the two groups' improvement on a common scale despite their unequal starting points: the experimental group recorded $N\text{-gain} = (55.47 - 25.64) / (100 - 25.64) = 0.40$ (moderate), and the control group $N\text{-gain} = (50.00 - 15.50) / (100 - 15.50) = 0.41$ (moderate).

These figures indicate that both groups opened the study with low critical thinking ability and without baseline equivalence. The two groups' nearly identical N-gain values suggest that, once the unequal starting points are accounted for, relative learning improvement was comparable under the two instructional conditions rather than favoring either approach. The experimental group's larger posttest standard deviation further points to wider interindividual variation in outcomes under the game-based GASING approach, implying that the intervention may not have benefited all learners uniformly.

Shapiro-Wilk testing of the posttest data revealed normal distribution in the experimental group (Sig. = 0.088 > 0.05) but not in the control group (Sig. = 0.034 < 0.05). With the normality assumption breached for one group, the nonparametric Mann-Whitney U test was applied. Table 2 reports the outcome.

Table 2. Mann-Whitney U test results for mathematical critical thinking.

Statistic	Value
Mann-Whitney U	260.000
Wilcoxon W	585.000
Z	-0.807
Asymp. Sig. (2-tailed)	0.420
Rank-biserial correlation (r)	-0.115
Effect size interpretation	Very small ($ r < 0.3$)

Table 2 yields Sig. (2-tailed) = 0.420 > 0.05; H₀ is consequently retained, denoting no statistically significant difference in mathematical critical thinking achievement between the two groups. With $r = -0.115$, the rank-biserial effect size falls into Cohen's (1988) very small band, confirming that the practical separation between the groups is likewise negligible.

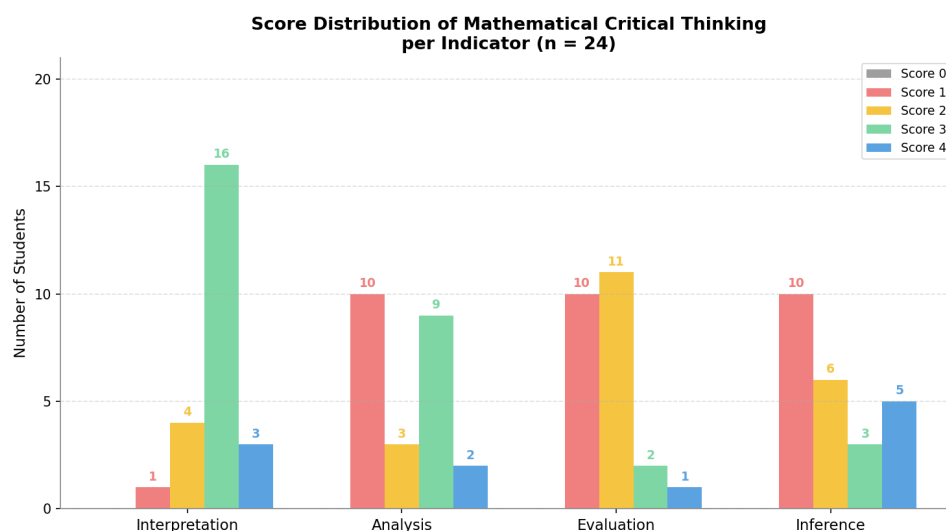


Figure 1. Score distribution of mathematical critical thinking per indicator (n = 24).

The written responses of all 24 experimental learners underwent qualitative analysis. Each test item was constructed to capture a single critical thinking indicator: Item 1 (interpretation), Item 2 (analysis), Item 3 (evaluation), and Item 4 (inference), with each item carrying a maximum of 4 points. Figure 1 displays the distribution of learner scores by indicator, and Table 3 condenses percentage achievement per indicator.

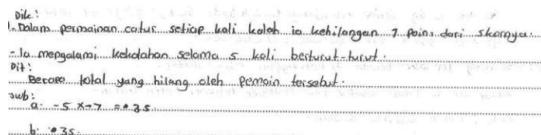
Table 3. Percentage achievement in mathematical critical thinking per indicator.

Indicator	Score (%)	Category
Interpretation	71.88	High
Analysis	53.13	Moderate
Inference	53.13	Moderate
Evaluation	43.75	Low

Interpretation Indicator

Item 1, which framed an integer multiplication problem inside an online chess game involving cumulative point losses, served to assess the interpretation indicator. Of the 24 learners, 16 scored 3, four scored 2, three scored 4, and one scored 1, producing a total achievement of 71.88% (high). Table 4 displays the item together with a representative learner response.

Table 4. Test Item 1 and representative student written response.

Item 1	Representative Written Response
In an online chess game, a player loses 7 points from their score every time they lose a match. A player loses five matches in a row. Write a mathematical expression to represent this situation and determine the total points lost. How many total points did the player lose?	

As Table 4 shows, the learner set down the expression $-7 \times 5 = -35$ and concluded that the total points lost amounted to 35, thereby reading “loss” as a negative integer and the recurring matches as a multiplication operation. The semi-structured interview excerpt below reinforces this reading:

Researcher: “Why did you write -7×5 ?”

Student: “Because loss means minus, so I wrote negative 7. Then I multiplied by 5 because they lost five times.”

Researcher: “Can you explain what the result -35 means?”

Student: “It means the total points lost are 35. The negative sign shows that the points decreased.”

The interview verifies that the learner grasped loss as a negative integer and could put the meaning of the computed result into words. High interpretation achievement most plausibly traces back to the concrete stage of the GASING method—specifically the mountain-valley physical manipulative—which trains learners to picture positive and negative integers ahead of any move into symbolic representation.

Analysis Indicator

Item 2, set within a context of hourly temperature change, addressed the analysis indicator. The score distribution was 10 learners at 1, nine at 3, three at 2, and two at 4,

giving an overall achievement of 53.13% (moderate). Table 5 presents the item and a representative response.

Table 5. Test Item 2 and representative student written response.

Item 2	Representative Written Response
The temperature in City X at 06:00 is 2 degrees Celsius. Every hour, the temperature drops by 4 degrees Celsius. This decrease continues for five consecutive hours. What information can you identify from the problem? How would you connect that information to find the temperature in City X at 11:00? What is the temperature in City X at 11:00?	dik: # suhu awal 2°C # suhu turun setiap 4 jam dit: informasi yang diperoleh bagaimana cara menghubungkan informasi untuk mencari suhu kota X pada pukul 11:00 - Berapa suhu kota X pada pukul 11:00 Jwb: a. 2°C + 4 + 4 + 4 + 4 + 4 b. suhu awal + penurunan suhu 2° + -20 = -18 c. 2 + 20 = -18

As Table 5 indicates, the learner pinpointed every relevant piece of information (initial temperature 2 degrees Celsius, a 4-degree fall per hour over five hours) and used a repeated addition strategy to break the problem down hour by hour, afterward linking the starting temperature to the cumulative decrease. The interview excerpt below further illuminates the learner's analytical work:

Researcher: "Why did you write repeated addition?"

Student: "Because I wanted to show that every hour it drops 4 degrees, so I wrote them one by one before adding them up."

Researcher: "In part (c), you wrote $2 + (-20) = -18$. Can you explain that?"

Student: "Oh, that means 2 plus negative 20, which gives minus 18."

The interview brings out sound structural understanding: the learner methodically separated the problem into its parts (initial temperature, hourly drop, number of hours) and recombined them coherently. Several learners nonetheless slipped on notation in negative number operations, a sign that conceptual consolidation remains unfinished.

Evaluation Indicator

Item 3, which asked learners to judge the truth of a claim about sea-level change while simultaneously holding two competing factors in view (a decrease and an increase), assessed the evaluation indicator. Here 11 learners scored 2 and 10 scored 1, with only two reaching 3 and one reaching 4, for an overall achievement of 43.75% (low). Table 6 sets out the item and a representative response.

Table 6. Test Item 3 and representative student written response.

Item 3	Representative Written Response
At a harbor, an observer records that the sea level falls by 2 meters every hour. However, another factor causes the sea level to rise by 1 meter every 2 hours. If this receding continues for six consecutive hours, then the total decrease in sea level reaches 50 meters. Is the observer's claim correct or incorrect? Explain your answer. What is the actual change in sea level after six hours?	dik: # air laut surut 2 meter setiap jam # air laut surut setiap 6 jam dit: bagaimana faktor yang benar atau salah besar dan kecil - Penurunan air laut setiap 6 jam Jwb: a. Air laut surut 2 meter setiap jam x air surut 6 jam berturut-turut b. Jadi Penurunan air laut setiap 6 jam adalah 12 m

As Table 6 reveals, the learner logged only the decreasing factor and left the rising factor out of the working entirely. The learner produced $2 \times 6 = 12$ meters as the final answer,

disregarding the 1-meter rise occurring every two hours. The interview exchange that follows, however, exposes a latent evaluative capacity:

Researcher: "Is all the information in the problem included in your solution?"

Student: "Hmm, there is also a 1-meter rise every 2 hours, but I did not include it in my calculation."

Researcher: "Can you try recalculating with that factor included?"

Student: "The rise is 1 meter every 2 hours. Over 6 hours that is 3 rises, so 3 meters. The actual decrease is therefore 12 minus 3, which is 9 meters."

The interview lays bare a telling divide between oral and written evaluative performance: once prompted, the learner could name the missing information and recompute the answer correctly (9 meters net decrease).

Inference Indicator

Item 4, built around a fisherman's income after spoiled fish are deducted across several baskets, assessed the inference indicator. The distribution showed 10 learners at 1, six at 2, five at 4, and three at 3, yielding an overall achievement of 53.13% (moderate). Table 7 presents the item and a representative response.

Table 7. Test Item 4 and representative student written response.

Item 4	Representative Written Response
A fisherman catches 8 baskets of fish. Each basket has a total weight of 50 kg. Upon inspection, 3 kg per basket is found to be spoiled and unsellable. The fisherman plans to sell the good fish at IDR 50,000 per kilogram. Calculate how many kilograms of fish are still sellable. Determine the fisherman's total income from selling the good fish. A classmate concludes that the fisherman's income is IDR 20,000,000. Determine whether this conclusion is correct or incorrect, and explain your reasoning.	Dik : - Seorang nelayan menangkap ikan sebanyak 8 keranjang. - Setiap keranjang memiliki berat 50 kg. - 3 kg ikan busuk. - Ikan yang masih baik dijual dengan harga Rp. 50.000 per kg. Dit : Hitunglah berapa kg ikan yang masih layak dijual oleh nelayan tersebut. Jawab : a. 47 kg. (50 - 3 = 47) b. 17.800.000 (47 x 8 = 356 x 50.000 = 17.800.000) (50 kg - 3 kg = 47 kg) c. Salah, karena ia masih hitung dengan ikan yang busuk, maka hasilnya Rp. 20.000.000 (50 x 8 = 400 x 50.000 = 20.000.000). Jika harus mengurangi terlebih dahulu : ada 8 keranjang beratnya 50 kg setiap keranjang 50 dan 3 kg setiap keranjang busuk jadi 50 kg - 3 kg = 47 lalu 47 x 8 = 356 x 50.000 = 17.800.000.

As Table 7 shows, the learner adopted a per-basket approach ($50 - 3 = 47$ kg per basket, then $47 \times 8 = 356$ kg). Although an arithmetic error crept in (the correct figure being 376 kg), the inferential reasoning chain held together conceptually. The learner also rightly judged the classmate's claim to be incorrect on the grounds that it neglected to subtract the spoiled fish. The interview excerpt below confirms this inferential capacity:

Researcher: "Can you recalculate? Is 47×8 really 356?"

Student: "Hmm, 40×8 is 320, and 7×8 is 56. So 320 plus 56 is 376. I made an arithmetic error earlier."

Researcher: "In part (c), you wrote the claim is incorrect. Can you explain why?"

Student: "Because the claim did not subtract the spoiled fish. If you take $50 \times 8 = 400$ kg and multiply by IDR 50,000 you get 20 million. But you must subtract the spoiled fish first before calculating the income."

The interview makes plain that the learner sustained a conceptually sound inferential chain and could recognize and self-correct arithmetic slips when guided. The error in 47×8 suggests that computational fluency had not yet matured within four sessions.

Teacher Interview

To round out the per-indicator analysis, the following passage is drawn from the semi-structured interview with the classroom teacher:

Researcher: "How did you observe students' critical thinking developing over the course of this instruction?"

Teacher: "Looking at how they approached the problems, there was noticeable progress. Before, they often guessed without writing out their work. By the third and fourth sessions, many students had started writing down what was given and what was asked before calculating. Some were even able to check their own answers, although not all students could do that yet. In my view, the hand-gesture competition and puzzle activities at the beginning helped them think more carefully, because an incorrect answer meant they immediately lost."

The teacher's account corroborates the pattern surfacing through the quantitative and written analysis: interpretation-level skills (identifying and articulating given information) developed most visibly, while evaluation-level skills (verifying answers) stayed limited.

Discussion

The absence of a statistically significant between-group difference (Sig. = 0.420, $r = -0.115$) should be read as inconclusive rather than as evidence that game-integrated GASING instruction is ineffective, and its meaning depends on how three competing explanations are weighed. The first is methodological: with 24 and 25 learners, the comparison could reliably detect only large effects, so the study can reasonably exclude a large benefit but not a small or moderate one. The second concerns baseline non-equivalence. The experimental group began 10.14 points higher, which left less room for growth and helps explain why the control group posted the larger raw gain (34.50 versus 29.83 points) even though normalized gains were virtually identical (0.40 versus 0.41). Because the posttest comparison did not adjust for this head start, the unadjusted test was, if anything, tilted in favor of the intervention, which makes the absence of a large effect more credible rather than less. The third explanation is substantive: four sessions may fall below the exposure that critical thinking skills require, a reading consistent with Anggoro et al. (2025), who identified intervention duration as a significant moderator of game effects on mathematical higher-order thinking, and with Facione's (2020) view that critical thinking matures through sustained, repeated practice. Since the present design cannot separate these explanations, the quantitative result is best treated as a baseline for a larger and longer replication rather than as a verdict on the approach.

The indicator profile is more informative than the group comparison. Interpretation (71.88%) clearly outpaced evaluation (43.75%), a gradient that follows the ordering of Bloom's (1956) taxonomy, in which comprehension precedes analysis and evaluation sits near the top (Marzuki et al., 2021). This gradient, however, reflects not only task difficulty but also what the games rewarded. The four games scored the speed and accuracy of computation, a reward structure that sustains engagement and fluency (Bhagat et al., 2021; Yeh et al., 2019) but never required learners to judge whether a claim was adequate. Item 3 shows the consequence: the representative learner computed the decreasing factor correctly yet ignored the competing rise, an omission that a computation-focused game would never penalize. Two syntheses reinforce this interpretation. Barz et al. (2024)

found reliable cognitive but no significant metacognitive effects of digital game interventions, and Mao et al. (2022) reported that game effects on critical thinking skills were considerably smaller than those on critical thinking dispositions and largest for role-playing games. Games that do cultivate higher-order reasoning oblige learners to construct, test, and revise solutions (Ke, 2019; Ke & Clark, 2020) or embed conceptually demanding content in the core mechanic (Kärki et al., 2022). Evaluation and inference, in other words, must be designed into the game rules – for example, by awarding points for detecting and justifying an opponent's error – rather than expected to emerge from computational play.

The interviews complicate the written-test picture in an important way. The representative learner who omitted the competing factor in writing named and corrected the omission as soon as the researcher asked whether all information had been used, which suggests that part of the low evaluation score reflects difficulty in externalizing reasoning on paper rather than a complete absence of evaluative capacity. In Vygotskian terms (Vygotsky, 1978), the interviewer's prompt supplied the scaffolding that the games did not. Collaborative play appears to have supported learners within their zone of proximal development at the interpretation level, where GASING's concrete-abstract-mental computation sequence parallels Bruner's (1966) enactive-iconic-symbolic modes and where staged GASING instruction has previously supported understanding (Nuari et al., 2019), yet no comparable support was built in for evaluation. The missing ingredient was therefore more likely explicit instruction than additional play: gains in evaluative competence have been reported when critical thinking is taught as an explicit object within mathematics lessons (Arisoy & Aybek, 2021) or when cooperative structures require learners to interrogate one another's reasoning (Silva et al., 2023). Learners' epistemological beliefs may form a further barrier, since viewing mathematics as fixed procedures to be memorized reduces readiness for evaluative thinking (Rott, 2021); a game culture that rewards the fastest correct answer risks reinforcing precisely that belief.

Two further observations qualify these conclusions. First, the experimental group's larger posttest dispersion ($SD = 20.30$ versus 14.77) indicates that game-integrated GASING instruction widened rather than narrowed individual differences, consistent with Hussein et al.'s (2022) finding that game-based learning effects vary with learners' cognitive readiness. For 3T classrooms with heterogeneous prior knowledge, this implies that games should be accompanied by differentiated support for lower-performing learners rather than applied uniformly. Second, the notational slips with negative numbers show that game practice did not dislodge the habit of recalling sign rules without their conceptual rationale (Lawson et al., 2021; Zhang et al., 2017). At the same time, the teacher's report that the immediate-penalty mechanism made learners more careful, together with the apparent contribution of the Session 2 puzzle to systematic analysis (Melero & Hernandez-Leo, 2014), indicates that the games' feedback loops did foster attentiveness and self-monitoring (Gentile et al., 2019; Hui & Mahmud, 2023). This self-monitoring, however, seems to have operated on computational accuracy rather than on the soundness of reasoning, which again locates the limitation of the intervention in what its feedback targeted rather than in game-based learning as such.

CONCLUSION

Fundamental Finding: Game-integrated GASING instruction did not produce a statistically significant difference in mathematical critical thinking on integer

multiplication compared with conventional instruction (Mann-Whitney U, Sig. = 0.420, $r = -0.115$), and both groups achieved comparable moderate normalized gains (0.40 and 0.41). At the indicator level, interpretation was highest (71.88%, high), analysis and inference were moderate (53.13%), and evaluation was lowest (43.75%, low), indicating that the approach more readily fosters lower- and mid-order critical thinking than higher-order evaluative competence. **Implication:** Game-integrated GASING instruction is a workable and accessible pedagogical alternative for Indonesia's 3T regions; however, because interviews revealed reasoning that written responses did not capture, test-based measurement alone may understate learners' critical thinking, which calls for multimodal assessment and GASING teacher training tailored to 3T contexts. **Limitation:** The intervention was brief (four sessions); the games targeted conceptual reinforcement more than evaluation and inference; the non-randomized design with unequal baseline scores limited causal inference; the small purposive sample ($n = 49$) drawn from two intact classes afforded low statistical power and restricts generalization to comparable 3T settings; the scope covered one topic and one grade level; and a single teacher carried both instructional and facilitation roles. **Future Research:** Future studies should extend the intervention to 8–12 sessions, design games that explicitly require answer verification and reasoning about competing factors, add structured post-game reflection, replicate the study across topics and grade levels, and adopt multimodal assessment combining written tests, oral assessment, and portfolios.

ACKNOWLEDGEMENTS

The authors thank the principals, teachers, and students of the two-state junior high schools in Lembata Regency, East Nusa Tenggara, for their participation and support throughout the data collection process. The authors also thank the expert judges who validated the research instrument.

REFERENCES

- Anggoro, B. S., Dewantara, A. H., Suherman, S., Muhammad, R. R., & Saraswati, S. (2025). Effect of game-based learning on students' mathematics high order thinking skills: A meta-analysis. *Revista de Psicodidáctica (English ed.)*, 30(1), 500158. <https://doi.org/10.1016/j.psicoe.2024.500158>
- Angraini, L. M., & Wahyuni, A. (2021). The effect of concept attainment model on mathematical critical thinking ability. *International Journal of Instruction*, 14(1), 727–742. <https://doi.org/10.29333/iji.2021.14144a>
- Aqazade, M., & Bofferding, L. (2021). Second and fifth graders' use of knowledge-pieces and knowledge-structures when solving integer addition problems. *Journal of Numerical Cognition*, 7(2), 82–103. <https://doi.org/10.5964/jnc.6563>
- Arisoy, B., & Aybek, B. (2021). The effects of subject-based critical thinking education in mathematics on students' critical thinking skills and virtues. *Eurasian Journal of Educational Research*, 21(92), 99–120. <https://doi.org/10.14689/ejer.2021.92.6>
- Barz, N., Benick, M., Dörrenbächer-Ulrich, L., & Perels, F. (2024). The effect of digital game-based learning interventions on cognitive, metacognitive, and affective-motivational learning outcomes in school: A meta-analysis. *Review of Educational Research*, 94(2), 193–227. <https://doi.org/10.3102/00346543231167795>
- Bhagat, K. K., Yang, F.-Y., Cheng, C.-H., Zhang, Y., & Liou, W.-K. (2021). Tracking the process and motivation of math learning with augmented reality. *Educational*

- Technology Research and Development*, 69(6), 3153-3178.
<https://doi.org/10.1007/s11423-021-10066-9>
- Bloom, B. S. (1956). *Taxonomy of educational objectives: Handbook I: Cognitive domain*. Longmans, Green and Co.
- Bruner, J. S. (1966). *Toward a theory of instruction* (Nachdr.). Belknap Press of Harvard University Press.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
- Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). SAGE.
- Facione, P. A. (1990). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction (the delphi report)*. ERIC.
- Facione, P. A. (2020). *Critical thinking: What it is and why it counts*. Measured Reasons LLC.
- Gentile, M., Città, G., Perna, S., Signa, A., Dal Grande, V., Ottaviano, S., La Guardia, D., & Allegra, M. (2019). The role of disposition to critical thinking in digital game-based learning. *International Journal of Serious Games*, 6(3), 51-63.
<https://doi.org/10.17083/ijsg.v6i3.316>
- Harun, N. J., Cuevas, K. A., Asakil, O. I., Alviar, J. V., & Solon, L. V. (2023). Assessing students' mastery and misconceptions in the fundamental operations on integers. *International Journal of Science, Technology, Engineering and Mathematics*, 3(3), 36-55.
<https://doi.org/10.53378/353000>
- Hidayat, R., Qi, T. Y., Ariffin, P. N. B. T., Hadzri, M. H. B. M., Chin, L. M., Ning, J. L. X., & Nasir, N. (2024). Online game-based learning in mathematics education among generation z: A systematic review. *International Electronic Journal of Mathematics Education*, 19(1), em0763. <https://doi.org/10.29333/iejme/14024>
- Hui, H. B., & Mahmud, M. S. (2023). Influence of game-based learning in mathematics education on the students' cognitive and affective domain: A systematic review. *Frontiers in Psychology*, 14, 1105806. <https://doi.org/10.3389/fpsyg.2023.1105806>
- Hussein, M. H., Ow, S. H., Elaish, M. M., & Jensen, E. O. (2022). Digital game-based learning in K-12 mathematics education: A systematic literature review. *Education and Information Technologies*, 27(2), 2859-2891. <https://doi.org/10.1007/s10639-021-10721-x>
- Kannadass, P., Hidayat, R., Siregar, P. S., & Husain, A. P. (2023). Relationship between computational and critical thinking towards modelling competency among pre-service mathematics teachers. *TEM Journal*, 12(3), 1370-1382.
<https://doi.org/10.18421/TEM123-17>
- Kärki, T., McMullen, J., & Lehtinen, E. (2022). Improving rational number knowledge using the NanoRoboMath digital game. *Educational Studies in Mathematics*, 110(1), 101-123. <https://doi.org/10.1007/s10649-021-10120-6>
- Ke, F. (2019). Mathematical problem solving and learning in an architecture-themed epistemic game. *Educational Technology Research and Development*, 67(5), 1085-1104.
<https://doi.org/10.1007/s11423-018-09643-2>
- Ke, F., & M. Clark, K. (2020). Game-based multimodal representations and mathematical problem solving. *International Journal of Science and Mathematics Education*, 18(1), 103-122. <https://doi.org/10.1007/s10763-018-9938-3>

- Lawson, A. P., Ayala, B., & Son, J. Y. (2021). Priming students to calculate inhibits sense-making. *Journal of Cognitive Science*, 22(1), 41–69. <https://doi.org/10.17791/jcs.2021.22.1.41>
- Mao, W., Cui, Y., Chiu, M. M., & Lei, H. (2022). Effects of game-based learning on students' critical thinking: A meta-analysis. *Journal of Educational Computing Research*, 59(8), 1682–1708. <https://doi.org/10.1177/07356331211007098>
- Marzuki, Wahyudin, Cahya, E., & Juandi, D. (2021). Students' critical thinking skills in solving mathematical problems; a systematic procedure of grounded theory study. *International Journal of Instruction*, 14(4), 529–548. <https://doi.org/10.29333/iji.2021.14431a>
- Masfufah, R., & Afriansyah, E. A. (2021). Analysis of students' mathematical literacy ability through PISA problems. *Mosharafa: Jurnal Pendidikan Matematika*, 10(2), 291–300. <https://doi.org/10.31980/mosharafa.v10i2.662>
- Maulida, S. A., Isrokatun, I., & Julia, J. (2024). Pengaruh pembelajaran kooperatif TGT dengan metode GASING terhadap pemahaman konsep matematis perkalian dua angka di SD [The effect of TGT cooperative learning with the GASING method on elementary students' conceptual understanding of two-digit multiplication]. *SJME (Supremum Journal of Mathematics Education)*, 8(1), 101–113. <https://doi.org/10.35706/sjme.v8i1.10789>
- Melero, J., & Hernandez-Leo, D. (2014). A model for the design of puzzle-based games including virtual and physical objects. *Educational Technology and Society*, 17(3), 192–207. Scopus.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Sage Publications.
- Moon, J., & Ke, F. (2020). Exploring the relationships among middle school students' peer interactions, task efficiency, and learning engagement in game-based learning. *Simulation & Gaming*, 51(3), 310–335. <https://doi.org/10.1177/1046878120907940>
- Nuari, L. F., Prahmana, R. C. I., & Fatmawati, I. (2019). Learning of division operation for mental retardations' student through math gasing. *Journal on Mathematics Education*, 10(1), 127–142. <https://doi.org/10.22342/jme.10.1.6913.127-142>
- Nurhikmayati, I., Priatna, N., Dahlan, J. A., & Minasyan, S. (2024). An ex post facto study of critical thinking skills in mathematics learning based on school geography. *Al-Jabar : Jurnal Pendidikan Matematika*, 15(1), 15–31. <https://doi.org/10.24042/ajpm.v15i1.19330>
- OECD. (2023). *PISA 2022 results (Volume I): The state of learning and equity in education*. OECD. <https://doi.org/10.1787/53f23881-en>
- Pan, Y., Ke, F., & Xu, X. (2022). A systematic review of the role of learning games in fostering mathematics education in K-12 settings. *Educational Research Review*, 36, 100448. <https://doi.org/10.1016/j.edurev.2022.100448>
- Presidential Regulation of the Republic of Indonesia Number 63 of 2020 on the Designation of Disadvantaged, Frontier, and Outermost Regions 2020-2024. (2020). <https://peraturan.bpk.go.id/Details/136563/perpres-no-63-tahun-2020>
- Qirom, M. S., & Juandi, D. (2023). Learning from Indonesia: A systematic literature review on the implementation of traditional Indonesian games for mathematics education. *Beta: Jurnal Tadris Matematika*, 16(2), 136–150. <https://doi.org/10.20414/betajtm.v16i2.570>

- Rott, B. (2021). Inductive and deductive justification of knowledge: Epistemological beliefs and critical thinking at the beginning of studying mathematics. *Educational Studies in Mathematics*, 106(1), 117–132. <https://doi.org/10.1007/s10649-020-10004-1>
- Setiana, D. S., Purwoko, R. Y., & Sugiman, S. (2021). The application of mathematics learning model to stimulate mathematical critical thinking skills of senior high school students. *European Journal of Educational Research*, 10(1), 509–523. <https://doi.org/10.12973/eu-jer.10.1.509>
- Silva, H., Lopes, J., Morais, E., & Dominguez, C. (2023). Fostering critical and creative thinking through the cooperative learning jigsaw and group investigation. *International Journal of Instruction*, 16(3), 261–282. <https://doi.org/10.29333/iji.2023.16315a>
- Sugiyono. (2017). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D* [Quantitative, qualitative, and R&D research methods]. CV. ALFABETA.
- Sun, L., Ruokamo, H., Siklander, P., Li, B., & Devlin, K. (2021). Primary school students' perceptions of scaffolding in digital game-based learning in mathematics. *Learning, Culture and Social Interaction*, 28, 100457. <https://doi.org/10.1016/j.lcsi.2020.100457>
- Surya, Y. (2013). *Modul Pelatihan Matematika GASING SD* [GASING mathematics training module for primary school]. PT. Kandel.
- Vankúš, P. (2021). Influence of game-based learning in mathematics education on students' affective domain: A systematic review. *Mathematics*, 9(9), 986. <https://doi.org/10.3390/math9090986>
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- Wang, Q., & Abdullah, A. H. (2024). Enhancing students' critical thinking through mathematics in higher education: a systemic review. *Sage Open*, 14(3), 1–18. <https://doi.org/10.1177/21582440241275651>
- Wijaya, T. T., Hidayat, W., Hermita, N., Alim, J. A., & Talib, C. A. (2024). Exploring contributing factors to PISA 2022 mathematics achievement: Insights from Indonesian teachers. *Infinity Journal*, 13(1), 139–156. <https://doi.org/10.22460/infinity.v13i1.p139-156>
- Yeh, C. Y. C., Cheng, H. N. H., Chen, Z.-H., Liao, C. C. Y., & Chan, T.-W. (2019). Enhancing achievement and interest in mathematics learning through Math-Island. *Research and Practice in Technology Enhanced Learning*, 14(1), 5. <https://doi.org/10.1186/s41039-019-0100-9>
- Zhang, D., Ding, Y., Lee, S., & Chen, J. (2017). Strategic development of multiplication problem solving: Patterns of students' strategy choices. *Journal of Educational Research*, 110(2), 159–170. Scopus. <https://doi.org/10.1080/00220671.2015.1060928>



***Abu Moh. Rasyid Ridho (Corresponding Author)**

Universitas Pendidikan Indonesia
Jl. Dr. Setiabudi No. 229, Sukasari, Bandung City, West Java, Indonesia
Email: abumohrasyidridhos@upi.edu

Al Jupri (Corresponding Author)

Universitas Pendidikan Indonesia
Jl. Dr. Setiabudi No. 229, Sukasari, Bandung City, West Java, Indonesia
Email: aljupri@upi.edu

Sumanang Muhtar Gozali

Universitas Pendidikan Indonesia
Jl. Dr. Setiabudi No. 229, Sukasari, Bandung City, West Java, Indonesia
Email: sumanang.muhtar.gozali@upi.edu
