Journal of Advanced Sciences and Mathematics Education
Volume 6, Issue 4, 502-521.
e_ISSN: 2798-1606
DOI: 10.58524/jasme.v6i4.1680

GASING Mountain-valley representations in integer addition and subtraction: A quasi-experimental study with seventh graders

Yenni Natalina Buknoni*
Universitas Pendidikan Indonesia, INDONESIA

Turmudi
Universitas Pendidikan Indonesia, INDONESIA

Effy Mulyasari
Universitas Pendidikan Indonesia, INDONESIA

Article Info

Article history:

Received: September 15, 2026
Revised: October 29, 2026
Accepted: November 19, 2026

Keywords:

GASING Method;
Integer Addition And Subtraction;
Learning Outcomes;
Mountain-Valley Representations;
Quasi-Experimental Design.

Abstract

Background: Seventh-grade students frequently encounter difficulties in integer addition and subtraction, particularly when operating with negative numbers. Although concrete representations may facilitate mathematical learning, empirical evidence regarding the integration of mountain-valley representations into progressive instructional sequences remains limited.
Aims: This study examined differences in immediate computational learning outcomes between students receiving mountain-valley-based GASING instruction and those receiving conventional direct instruction.
Method: A quasi-experimental nonequivalent control group design involved 62 seventh-grade students from two intact classes at SMP Negeri 3 Kota Kupang, Indonesia (31 students per group). The intervention comprised concrete manipulation, symbolic representation, and mental calculation within a 140-minute session. Learning outcomes were assessed using parallel 10-item pre-tests and post-tests. Data were analyzed using analysis of covariance (ANCOVA), Mann–Whitney U tests, normalized gain (N-gain), and effect-size estimates.
Results: ANCOVA revealed significantly higher adjusted post-test scores in the experimental group (86.20) than in the control group (77.67), with an adjusted difference of 8.53 points (95% CI [3.35, 13.71]; $F(1, 59) = 10.85$, $p = 0.002$, partial $\eta^2 = 0.155$). Nonparametric analyses supported these findings, with higher post-test performance ($U = 244.0$, $p < 0.001$) and N-gain (0.77 vs. 0.61; $p = 0.001$).
Conclusion: Mountain-valley-based GASING instruction was associated with higher immediate computational performance. However, nonrandomized group assignment and teacher-related differences limit causal interpretation, warranting further controlled investigations.

To cite this article: Buknoni, Y. N., Turmudi, & Mulyasari, E. (2026). GASING mountain-valley representations in integer addition and subtraction: A quasi-experimental study with seventh graders. *Journal of Advanced Sciences and Mathematics Education, 6*(4), 502-521.

INTRODUCTION

Integer addition and subtraction constitute fundamental mathematical competencies that enable students to progress from whole-number arithmetic toward more advanced concepts, including algebraic expressions, linear equations, and coordinate systems (Acosta & Soliba, 2024). Despite their foundational importance, these operations frequently present conceptual and procedural challenges, particularly when students encounter negative numbers and expressions involving different signs (Wessman-Enzinger & Mooney, 2021). Students may successfully recall arithmetic rules without adequately understanding the mathematical relationships underlying integer operations, resulting in inconsistent computational performance across different problem types (Lin & Riccomini, 2026). These difficulties become particularly apparent when subtraction involves negative integers, as the resulting calculations may contradict students' previously established expectations about whole-number operations. Research has demonstrated that students frequently transfer whole-number reasoning into integer arithmetic, creating misconceptions about

***Corresponding author:**
Yenni Natalina Buknoni, Universitas Pendidikan Indonesia, Indonesia
yennibuknoni26@upi.edu

numerical magnitude, operational direction, and the meaning of subtraction (Lin & Riccomini, 2026). Such misconceptions may persist when mathematical instruction emphasizes procedural accuracy without sufficiently developing conceptual representations that connect numerical symbols with meaningful operations. Representational approaches offer opportunities to establish these connections by making otherwise abstract mathematical relationships accessible through visual or physical experiences. However, the educational value of mathematical representations depends on how effectively learners connect concrete experiences with symbolic procedures and increasingly independent reasoning (Abrahamson et al., 2020; J. Kim, 2025). Consequently, instructional approaches that integrate conceptual representations with structured computational practice warrant further investigation in secondary mathematics education. Addressing these challenges is particularly important during the transition to algebra, when students must apply integer operations accurately across increasingly complex mathematical situations (Acosta & Soliba, 2024).

Students' difficulties with integer operations are also associated with the multiple mathematical meanings of negative signs and subtraction symbols, which require learners to distinguish numerical properties from operational processes. In particular, students may interpret subtraction exclusively as removing a quantity, making expressions involving negative numbers difficult to conceptualize (Bryant et al., 2020; Hansen & Jankvist, 2021; White Brahmia et al., 2020). Research on integer reasoning indicates that students employ different strategies depending on the numerical structure of a problem, suggesting that computational competence cannot be explained solely through memorized arithmetic rules (Wessman-Enzinger & Mooney, 2021). These challenges highlight the importance of instructional representations that explicitly demonstrate relationships between positive and negative quantities. Concrete models based on opposing quantities can illustrate additive inverses through cancellation, allowing students to visualize how positive and negative units combine to produce zero (Ding et al., 2024; Sari et al., 2025). Nevertheless, such representations may become procedurally complicated when subtraction requires the introduction of additional zero pairs. Without sufficient instructional guidance, students may reproduce manipulative procedures without developing transferable mathematical reasoning. Recent research on concrete-representational-abstract instruction has emphasized the potential of systematically connecting physical models, representational activities, and abstract mathematical expressions (Ebner et al., 2025). Similarly, investigations involving students with learning difficulties have documented the instructional value of structured representational sequences for developing mathematical competencies (Flores et al., 2024). These findings suggest that carefully sequenced instructional experiences may provide a productive foundation for supporting integer computation, although their effectiveness within regular secondary classrooms requires further examination.

Comparable instructional challenges have been documented in Indonesian mathematics education, where students continue to experience difficulties interpreting negative numbers and applying integer operations accurately (Maizora & Rosjanuardi, 2020; Anggraeni et al., 2025). Recent investigations have also identified confusion between negative signs and subtraction operations, reinforcing concerns about students' conceptual interpretation of integer expressions (Jeranah et al., 2025). Within this educational context, the GASING method, characterized by concrete manipulation, symbolic representation, and mental calculation, offers a structured instructional sequence intended to connect mathematical meaning with computational fluency (Surya, 2024). The approach employs mountain-valley representations in which mountains symbolize positive integers, valleys represent negative integers, and opposing units cancel to produce zero. Through physical manipulation, students can construct representations of integer operations before translating their experiences into mathematical symbols and mental calculations (Fitriasari et al., 2026; Purwasih et al., 2024). Previous research has examined GASING instruction in relation to students' mathematical understanding, including applications involving integer concepts at the junior secondary level (Dea

et al., 2025). Related studies have also explored structured representational instruction as a means of supporting mathematical learning across different content areas (Flores et al., 2024; Ebner et al., 2025). Nevertheless, the implementation of mountain-valley representations requires careful consideration of how concrete activities are progressively connected to symbolic procedures rather than treated as isolated instructional exercises. Preliminary informal interviews with mathematics teachers and eighth-grade students at SMP Negeri 3 Kota Kupang indicated that classroom instruction frequently emphasized sign rules, number-line demonstrations, and procedural exercises, while students experienced difficulties explaining the meaning of integer operations. Although these preliminary observations were not collected through a formal research protocol, they provided contextual justification for investigating an alternative instructional sequence within the school's regular mathematics classrooms.

Recent research has increasingly examined representational approaches to integer learning, yet important questions remain concerning their implementation and comparative effectiveness in regular secondary classrooms. Sari et al. (2020) investigated a neutralization-based empty number-line model, while Sari et al. (2025) subsequently examined a related mound-hollow representation through qualitative observations involving three sixth-grade students. Although these investigations contributed insights into students' mathematical reasoning, their findings did not establish whether a structured progression from concrete manipulation to independent computation produces higher immediate learning outcomes than conventional instruction. Similarly, existing research on GASING has examined mathematical understanding among seventh-grade students without specifically isolating computational performance across different combinations of integer addition and subtraction (Dea et al., 2025). Recent evidence supporting concrete-representational-abstract instruction provides a relevant theoretical foundation, but much of this literature concerns different mathematical topics or populations with specific learning difficulties (Flores et al., 2024; Ebner et al., 2025). Consequently, comparative evidence remains limited regarding the application of mountain-valley representations within a three-stage GASING sequence for teaching integer addition and subtraction in mainstream secondary education. An additional methodological concern involves the need to account for initial differences between intact classroom groups when estimating instructional outcomes in nonrandomized educational studies. The present investigation addresses these gaps by examining ten operation types, including subtraction involving negative integers, and comparing learning outcomes using baseline-adjusted statistical analysis and effect-size estimates. Its contribution lies not in introducing an entirely new integer representation but in evaluating the immediate computational outcomes associated with integrating mountain-valley manipulatives into a structured progression toward symbolic and mental calculation. Nevertheless, because instructional components and teacher characteristics cannot be independently isolated within the present design, the study provides comparative classroom evidence rather than definitive confirmation of the proposed instructional mechanism.

This study aimed to examine differences in immediate integer addition and subtraction learning outcomes between seventh-grade students receiving mountain-valley-based GASING instruction and those receiving direct instruction. Specifically, the investigation focused on the implementation of a three-stage instructional sequence comprising concrete manipulation, symbolic writing, and mental calculation. The first objective was to compare students' post-test computational performance after statistically adjusting for differences in their initial achievement. The second objective was to examine differences in normalized gain scores between the experimental and control groups. The third objective was to estimate the magnitude and statistical precision of the observed between-group differences using appropriate effect-size measures and confidence intervals. These objectives were formulated to provide empirical evidence regarding short-term computational outcomes rather than to establish changes in conceptual understanding or long-term

mathematical retention. Based on previous findings concerning structured representational instruction and its potential contribution to mathematical performance (Flores et al., 2024; Ebner et al., 2025), the study proposed two directional hypotheses. The first hypothesis predicted that students receiving mountain-valley-based GASING instruction would achieve higher pre-test-adjusted post-test scores than students receiving direct instruction. The second hypothesis predicted that the experimental group would demonstrate higher normalized gain scores than the control group. Given the quasi-experimental design and the use of intact classes taught by different teachers, the findings were interpreted as evidence of between-group differences associated with instructional conditions rather than as conclusive estimates of causal effects.

LITERATURE REVIEW

Integer addition and subtraction represent essential components of mathematical development because they require students to extend arithmetic reasoning beyond nonnegative numbers and establish meaningful relationships between positive and negative quantities. Understanding these operations involves more than applying computational procedures, as students must recognize the mathematical properties of integers, additive inverses, and the different functions of subtraction (Wessman-Enzinger & Mooney, 2021). Research indicates that students frequently experience conceptual difficulties when mathematical operations contradict their previous experiences with whole-number arithmetic (Lin & Riccomini, 2026). Such difficulties are particularly evident when students encounter expressions involving negative operands, subtraction of negative numbers, and operations with different signs (Lin & Riccomini, 2026). Students who rely primarily on memorized sign rules may obtain correct answers in familiar situations while experiencing difficulties when numerical structures become more complex (Sercenia et al., 2023). These challenges demonstrate the importance of distinguishing procedural competence from conceptual understanding when examining mathematical learning outcomes. Conceptual understanding involves recognizing why mathematical operations produce particular results, whereas procedural competence concerns the accurate execution of computational processes (Borji et al., 2021; Braithwaite & Sprague, 2021; Dorner et al., 2026; Hadi et al., 2025; Han et al., 2026; Qetrani et al., 2021; Schaathun, 2022). Recent investigations have emphasized that students' mathematical reasoning develops through interactions among numerical representations, prior knowledge, and instructional experiences (Wessman-Enzinger & Mooney, 2021). Consequently, effective integer instruction should provide opportunities for students to connect mathematical symbols with meaningful numerical relationships rather than treating sign rules as isolated procedures. Within this perspective, evaluating students' computational performance across different combinations of integer addition and subtraction provides a relevant basis for examining immediate learning outcomes.

Mathematical representations play an important role in supporting students' interpretation of abstract numerical concepts by connecting symbolic expressions with observable relationships and meaningful mathematical actions. Number-line representations, for example, provide a spatial framework for interpreting integer magnitude, numerical order, and directional movement during addition and subtraction (Anton & Abrahamson, 2025; Fitriasari et al., 2026; Satsangi et al., 2025; Sutherland et al., 2024; Vest & Alibali, 2026). Alternatively, neutralization-based representations illustrate relationships between positive and negative quantities through opposing units that combine to produce zero. These models are particularly relevant to the additive inverse property because they make cancellation processes visible through physical or graphical arrangements. However, representational models may introduce additional cognitive demands when students

encounter subtraction problems requiring the construction of zero pairs (Aqazade & Bofferding, 2021; Bouck et al., 2021; Daroczy et al., 2020; Erbilgin & Macur, 2022; Sari et al., 2025; Wilkie & Hopkins, 2024). The mound-hollow representation developed by Sari et al. (2025) offers a related approach by representing positive and negative quantities through complementary physical structures. Building on earlier investigations involving neutralization-based empty number-line models, this representation provides opportunities for students to explore integer operations through observable mathematical relationships (Sari et al., 2020). Nevertheless, evidence derived from a small qualitative investigation cannot establish whether the representation produces superior computational outcomes across broader classroom populations. The effectiveness of concrete representations therefore depends not only on their mathematical characteristics but also on the instructional processes through which students interpret and progressively internalize them (Torres-Peña et al., 2025; Zeng et al., 2025). Accordingly, mountain-valley representations warrant investigation as instructional tools that may support the transition from physical manipulation toward symbolic and mental computation.

The Concrete–Representational–Abstract (CRA) approach provides a theoretical foundation for understanding how mathematical instruction can progress from physical experiences toward increasingly abstract forms of reasoning. This instructional framework emphasizes the gradual development of mathematical understanding through concrete manipulation, representational activities, and symbolic procedures (Hadi et al., 2025; H. Kim, 2020; Simon, 2022). A meta-analysis conducted demonstrated that concrete manipulatives can support mathematical achievement, particularly when their implementation is accompanied by appropriate instructional guidance. More recent meta-analytic evidence has also indicated positive outcomes associated with CRA instruction, although much of the available research involves students with learning disabilities and single-case experimental designs (Ebner et al., 2025). Empirical investigations have further documented the application of structured representational sequences in teaching place value and related mathematical competencies (Flores et al., 2024). These findings suggest that concrete learning experiences may become more educationally meaningful when students are systematically guided toward symbolic representations. Nevertheless, the relationship between physical manipulation and mathematical abstraction should not be interpreted as an automatic developmental progression. Students may successfully perform manipulative activities without necessarily transferring the underlying mathematical relationships into independent computational strategies. Therefore, instructional sequences should establish explicit connections among concrete actions, mathematical symbols, and opportunities for increasingly independent calculation (Fries et al., 2021; Kokkonen & Schalk, 2021; Yakubova et al., 2020). In the context of integer operations, this perspective provides a theoretical rationale for examining whether a structured progression from mountain-valley manipulation to mental computation is associated with improved immediate computational performance.

The GASING method, derived from the Indonesian expression *Gampang, Asyik, dan Menyenangkan*, emphasizes accessible and engaging mathematics instruction through a structured progression from concrete experiences to abstract and mental calculation. Its instructional sequence consists of three interconnected stages involving concrete manipulation, symbolic mathematical writing, and mental computation. In integer learning, mountain-valley representations operationalize this sequence by assigning positive values to mountains and negative values to valleys. Combining opposing units produces a neutral arrangement, thereby illustrating the additive inverse relationship through observable physical actions. During subtraction activities, students can introduce equivalent positive-negative pairs when the initial arrangement does not contain sufficient units to remove. This procedure provides a concrete interpretation of zero-pair construction while maintaining the numerical value of the original expression. After completing physical activities,

students translate their manipulations into symbolic expressions and subsequently practice solving problems without relying on concrete materials. Previous investigations have examined the implementation of GASING instruction in elementary mathematics and reported promising outcomes concerning students' mathematical understanding (Merdja et al., 2026). More recent research has explored the application of GASING to integer learning among seventh-grade students, providing additional evidence of its relevance to secondary mathematics education (Dea et al., 2025). Nevertheless, the specific contribution of mountain-valley representations within the complete three-stage sequence remains insufficiently established because existing studies differ in their instructional content, research designs, and measured learning outcomes.

Empirical investigations comparing instructional approaches provide an important foundation for evaluating whether representational learning sequences are associated with differences in students' mathematical performance. Previous research has indicated that structured concrete learning experiences can support mathematical achievement, although the magnitude and consistency of these outcomes vary across instructional contexts and assessment characteristics (Ebner et al., 2025). Studies involving GASING instruction have reported favorable mathematical learning outcomes, but their findings cannot automatically be generalized to integer addition and subtraction in regular secondary classrooms (Dea et al., 2025). Similarly, research on mound-hollow representations has contributed valuable qualitative insights into students' reasoning without establishing comparative differences in computational achievement between instructional groups (Sari et al., 2025). These limitations highlight the importance of research designs that compare learning outcomes while considering students' initial mathematical performance. In nonrandomized classroom investigations, baseline achievement differences may influence post-intervention comparisons and should therefore be considered when interpreting observed group differences (Anghel et al., 2026; Dalgaard et al., 2022; Li et al., 2026; Pellas, 2024; Rezende et al., 2026). Analysis of covariance provides a statistical approach for estimating adjusted post-test differences, although statistical adjustment cannot eliminate unmeasured differences between intact classes. Complementary analyses involving normalized gain, nonparametric comparisons, and effect-size estimates can provide additional perspectives on the magnitude and consistency of observed learning outcomes (Kwangmuang et al., 2026; Putra et al., 2026). Within this methodological context, the present study examines seventh-grade students' immediate computational performance across ten types of integer addition and subtraction problems following mountain-valley-based GASING instruction and direct instruction. By integrating representational learning theory with baseline-adjusted comparative analysis, the study contributes empirical evidence concerning short-term computational outcomes while recognizing that its quasi-experimental design cannot independently establish the causal effects of the representations or instructional sequence.

METHOD

Research Design

This study used a quasi-experimental nonequivalent control group design with a pre-test and a post-test. Random assignment of students to conditions was not possible because Grade 7 students at the school are placed in fixed classes (*rombongan belajar*) at the beginning of the school year, and these classes could not be reorganized for the study. Without randomization, the two groups may have differed before the intervention in ways that also influence learning, such as prior achievement, class climate, and the teacher. To reduce this threat, students' pre-test scores were measured and used as a covariate in the analysis; other differences between the classes could not be controlled and are discussed as limitations. Both classes followed teaching modules with the same learning

objectives, content, and duration. They differed in the instructional approach and in the teacher who delivered it (see Procedures and Time Frame).

The study was conducted with a research permit issued by Universitas Pendidikan Indonesia to the school and with the permission of the school principal. The intervention took place during regular mathematics lessons, covered content prescribed by the national curriculum for Grade 7, and both conditions consisted of ordinary classroom instruction with the same objectives, content, and duration; the pre-test and post-test were administered as regular classroom tasks. Students were informed by their teachers that the lessons and test results would be used for research. Written parental consent and formal review by an institutional ethics committee were not obtained, which we acknowledge as a limitation of the study. To protect participants, students were identified only by codes in the data set, their test results were used solely for research and not for grading, and all results are reported anonymously at the group level.

Participants

Each of the two participating classes had 32 enrolled students. In each class, one student did not attend all sessions of the research project and was excluded from the analysis, so the analyzed sample comprised 62 students: 31 in the experimental group (16 boys and 15 girls) and 31 in the control group (16 boys and 15 girls). All analyzed students completed both the pre-test and the post-test, so there were no missing test data and no imputation was required.

Population and Sampling Method

The population consisted of the ten Grade 7 classes at SMP Negeri 3 Kota Kupang in the 2025/2026 school year. When the first author requested permission to conduct the study, the school assigned two of these classes and also determined which class would serve as the experimental group and which as the control group; the criteria used by the school were not disclosed to the researcher. Because the classes were assigned by the school rather than selected at random, they may not represent the other eight Grade 7 classes, and selection bias cannot be ruled out. The initial equivalence of the groups was examined by comparing their pre-test scores (see Results).

Instrument

Learning outcomes in integer addition and subtraction were measured with two parallel 10-item short-answer tests, one used as the pre-test and the other as the post-test (Table 1). In both forms, each item corresponds to one of the ten operation types taught during the session (Figure 2), in the same order, so both forms covered all sign combinations, including subtraction of a negative number. The two forms used different numbers: the pre-test used single-digit numbers, whereas the post-test used larger numbers, several of which exceeded what could be represented with the 20 units given to each pair (e.g., −112 − 110), so students could not answer by recreating the concrete arrangements. Using different items reduced the chance that post-test performance reflected memory of pre-test items. However, the two forms were not statistically equated and may differ in difficulty, so the change from pre-test to post-test, including N-gain, should be interpreted with caution; because both groups took the same forms, this difference does not bias the comparison between groups. The items are presented symbolically and ask only for numerical answers. The test therefore measures accuracy in computing integer sums and differences and does not capture students' reasoning, so the results are interpreted as learning outcomes rather than as direct evidence of conceptual understanding.

Each item was scored dichotomously (1 = correct, 0 = incorrect), giving a raw score of 0–10, which was multiplied by 10 to obtain a score on a 0–100 scale. Both groups took the pre-test immediately before instruction and the post-test immediately after it, within the same session, so learning outcomes were measured only immediately after instruction and retention was not assessed.

Table 1. Pre-test and post-test items by operation type

No	Operation type	Pre-test	Post-test
1	Positive + positive	4 + 2	16 + 8
2	Negative + negative	(−4) + (−2)	(−23) + (−14)
3	Positive + negative	4 + (−2)	8 + (−6)
4	Negative + positive	(−4) + 2	(−12) + 8
5	Positive − smaller positive	4 − 2	9 − 7
6	Negative − negative (smaller absolute value)	(−4) − (−2)	(−10) − (−5)
7	Positive − larger positive	2 − 4	3 − 5
8	Negative − negative (larger absolute value)	(−2) − (−4)	(−2) − (−8)
9	Positive − negative	2 − (−4)	15 − (−7)
10	Negative − positive	(−2) − 4	−112 − 110

Note. In both forms, students were instructed: "Compute the following integer additions and subtractions."

The test items were developed by the first author so that each item corresponds directly to one of the ten operation types taught during the session (Table 1; Figure 2). The two forms were not reviewed by independent experts and were not piloted before the study; content validity therefore rests on this one-to-one alignment between items and the taught operation types. Reliability was estimated from the study sample (N = 62) using KR-20: 0.54 for the pre-test and 0.68 for the post-test. Items without variance were excluded from these analyses: on the pre-test, items 1 and 5 were answered correctly and items 9 and 10 incorrectly by all students; on the post-test, items 1–5 were answered correctly by all students. Corrected item-total correlations for the remaining items ranged from 0.23 to 0.40 on the pre-test and from 0.13 to 0.69 on the post-test, with only post-test item 6 below 0.20. According to criteria, these values indicate moderate reliability for the pre-test and high reliability for the post-test; by the more conservative benchmark of 0.70, however, both are below the conventional threshold. The lower pre-test coefficient is plausible for a test administered before instruction, when many students lacked a consistent strategy for operations with negative numbers. Because the pre-test served as the covariate, its reliability was addressed in a sensitivity analysis (see Analysis Plan). The lack of independent expert validation and of a pilot study is acknowledged as a limitation.

Procedures and Time Frame

This study was part of a larger research project conducted at SMP Negeri 3 Kota Kupang from August 27 to September 8, 2025, in which both classes received a four-session unit on integers: (1) the concept of integers, (2) addition and subtraction, (3) multiplication and division, and (4) word problems. Each session was assessed with its own pre-test and post-test covering only the operations taught in that session, and the unit as a whole was assessed with a separate, broader test of conceptual understanding and problem solving. This article reports only the second session and its session-specific tests. Results for the multiplication and division session are reported in Buknoni et al. (2026), and results from the broader unit test are reported in the first author's master's thesis. The data analyzed here (the addition and subtraction pre-test and post-test) were presented orally at the International Conference on Education Technology Innovation without publication in the proceedings and have not been reported in any other publication.

The session was scheduled for three 40-minute lesson periods (120 minutes) but in practice lasted about 140 minutes in each class: a 10-minute pre-test, 10 minutes of opening, 100 minutes of core instruction, 10 minutes of closing, and a 10-minute post-test. Both classes worked through the same ten operation types in the same order (Figure 2), using identical worked examples and practice exercises.

Each class was taught by its regular mathematics teacher. Both teachers were certified and held equivalent qualifications. One week before the study, the first author, a certified GASING trainer,

gave the experimental-class teacher a training session of about four hours while the two of them developed the experimental teaching module together; the session covered the use of the mountain-valley media and the three-stage sequence. The first author also developed the control-class module together with the control-class teacher, following the approach that this teacher normally used. The control-class teacher received no comparable training. The two teachers therefore differed not only in the instructional approach but also in their preparation for the study, and in this design the effect of the teacher cannot be separated from the effect of the instructional approach.

To keep learning opportunities comparable, both modules specified the same learning objectives, content, and duration, and both classes worked on identical worked examples and practice exercises. Apart from the differences in teachers and their preparation described above, and the song and card game used only in the experimental class, the classes differed in how the operations were represented and taught. The first author was present in both classes throughout the session and observed the lessons informally. No observation checklist or field notes were used, so the extent to which each lesson followed its module was not formally documented. The observer's presence may have influenced the behavior of teachers and students. In addition, because the first author is a GASING trainer who prepared the experimental-class teacher, researcher allegiance is a possible source of bias. These factors are addressed in the limitations.

Experimental class. During the core instruction, students worked in pairs, and each pair received a set of 20 mountain units and 20 valley units. The 100 minutes were divided across the three stages: concrete manipulation, about 30 minutes; abstract symbolic writing and practice, about 45 minutes; and mental calculation, about 25 minutes. Teacher monitoring of pairs and presentations by student representatives at the board took place within these stages. In the opening, the experimental class sang the GASING song “Di Mana Kamu” to create a relaxed atmosphere.

Stage 1: Concrete manipulation. The teacher posed each problem, and pairs modeled it with the media. The Indonesian keywords *digabung* (combined) and *diambil* (taken) linked physical actions to addition and subtraction, respectively. In addition, joining a mountain with a valley produced flat ground, so that pairs of opposite units cancelled. In subtraction, when the arrangement did not contain enough units to be taken, the teacher introduced the zero principle through the image of digging: digging soil out of the ground to make a valley piles the removed soil into a mountain of the same size, so a mountain and a valley are always created together and the overall ground level does not change. Students then added as many mountain-valley pairs as needed and removed the required units. Figure 2 shows the action and the result for each operation type. While pairs worked, the teacher moved around the room to check their arrangements. After each operation type, a student representative came to the board, drew the arrangement, and wrote the problem and its answer for the class. Figure 3 shows students handling the media during this stage.

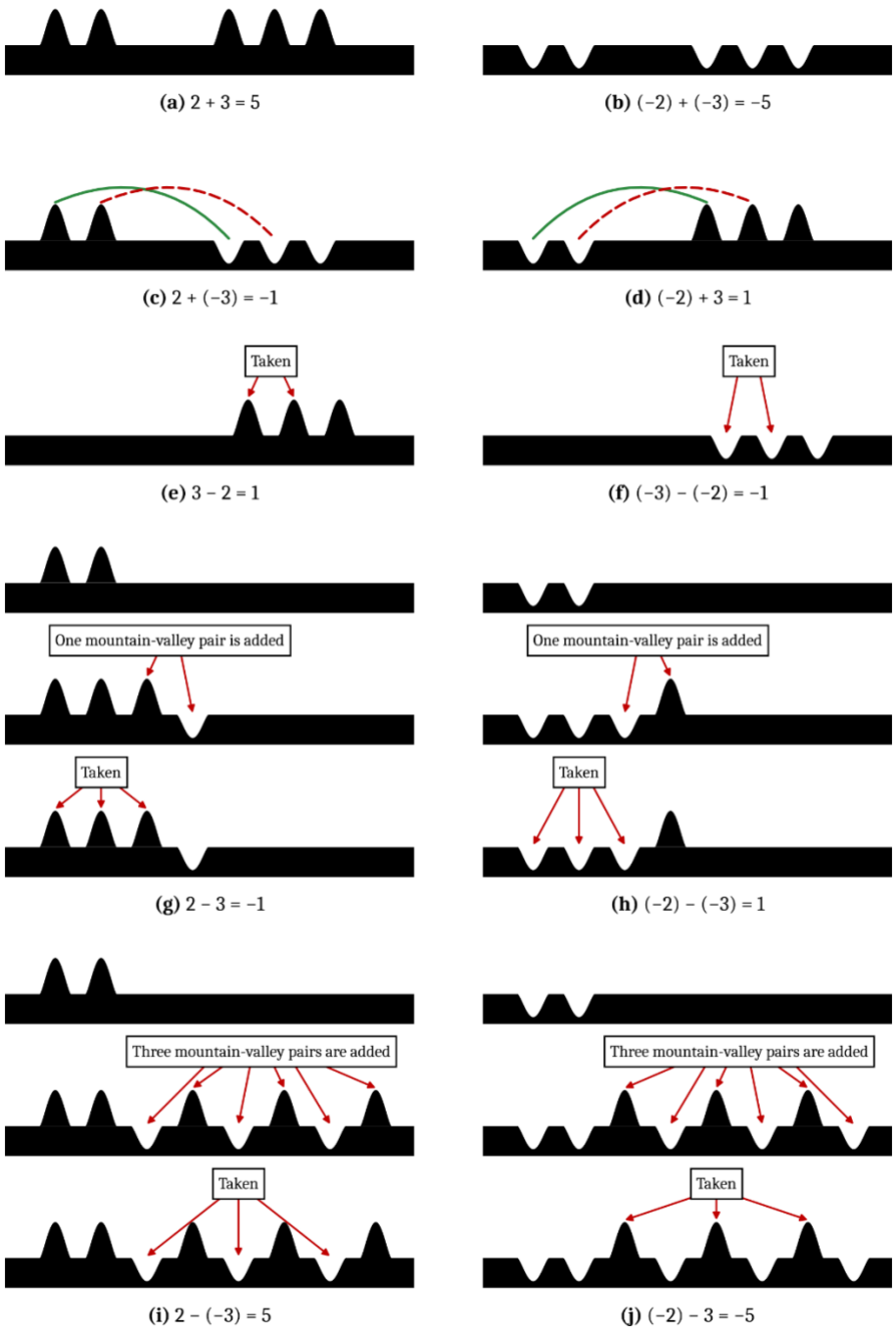


Figure 2. Mountain-valley representations of the ten operation types: (a)–(d) addition; (e)–(f) subtraction with enough units to take; (g)–(h) subtraction requiring one added mountain-valley pair; (i)–(j) subtraction of a number with the opposite sign, requiring three added pairs

Stage 2: Abstract symbolic writing. Immediately after the concrete activities, students wrote the expressions corresponding to the actions they had just carried out and then practiced on further problems. For example, after joining three valleys with five mountains, they wrote −3 + 5, and they found the result by picturing the arrangement: three valley-mountain pairs become flat ground and two mountains remain, giving +2. The progression from one stage to the next followed the planned sequence and time allocation rather than a formal mastery criterion.

Stage 3: Mental calculation (*mencongak*). The teacher read problems aloud, such as "25 + (−8)," and students answered orally without writing. Still within this stage, after the oral drills, students played a quartet card game (*kwartet*) in groups, in which they answered integer addition

and subtraction questions, and the teacher then led a discussion of the answers and acknowledged students' work. The control class had neither the song nor the game. The mental calculation stage was intended to build fluency grounded in the mountain-valley image. However, the study did not verify whether students' answers at this stage relied on that image or on memorized rules.



Figure 3. Students manipulating mountain-valley media during the concrete stage of instruction

Control class. The control class used the same 100 minutes of core instruction for the same ten operation types, with identical worked examples and practice exercises. About 40 minutes were spent on teacher explanation: for each operation type, the teacher explained the sign rule and demonstrated it on a number line on the board. The remaining 60 minutes were spent on practice: students worked on the worksheet in pairs while the teacher moved around the room to check their work, and student representatives came to the board to explain their solutions to the class. The control class thus had the same duration, content, grouping arrangement, teacher monitoring, and opportunities for student presentation as the experimental class; the two classes differed in how the operations were represented and explained.

Analysis Plan

Scores are reported on a 0–100 scale. Normalized gain was calculated for each student as (post-test − pre-test) / (100 − pre-test) and averaged within each group. Hake's categories (high, $g \geq 0.70$; moderate, $0.30 \leq g < 0.70$; low, $g < 0.30$) are reported only descriptively: they were developed for conceptual physics tests and do not account for differences in baseline scores between groups, so they were not used as evidence of effectiveness.

Initial equivalence was examined by comparing pre-test scores with the Mann-Whitney U test. Because the groups were intact classes and their pre-test means differed, the primary analysis was an analysis of covariance (ANCOVA) with post-test score as the dependent variable, group as the factor, and pre-test score as the covariate. Although the raw scores within each group departed from normality (Shapiro-Wilk, $p < 0.05$), ANCOVA assumes normality of residuals rather than of raw scores. Its assumptions were checked through the normality of residuals (Shapiro-Wilk), homogeneity of regression slopes (group × pre-test interaction), and homogeneity of variance (Levene's test). Adjusted means, the adjusted mean difference with its 95% confidence interval, and partial η^2 are reported. Because measurement error in a covariate leads ANCOVA to under-adjust for baseline differences in nonrandomized designs, a sensitivity analysis was also conducted in which each student's pre-test score was corrected for unreliability by shrinking its deviation from the group mean by the pre-test KR-20 coefficient (0.54) before the ANCOVA was repeated.

Two secondary analyses were conducted to check whether the result depended on distributional assumptions. First, post-test and N-gain scores were compared with the Mann-Whitney U test, using the rank-biserial correlation as the effect size, calculated as $r_{rb} = 1 - 2U / (n_1 n_2)$, where U is the smaller of the two U statistics (Kerby, 2014). Its 95% confidence interval was obtained with 10,000 bootstrap resamples. Second, rank analysis of covariance was applied with pre-test scores as the covariate. Effect sizes were interpreted against conventional benchmarks (r: 0.10, 0.30, and 0.50; partial η^2: 0.01, 0.06, and 0.14 for small, medium, and large effects). However, because these benchmarks are general, educational significance was judged mainly from the size of the difference

on the test's score scale. Descriptive statistics and Mann-Whitney U tests were computed in SPSS version 25; ANCOVA, Quade's test, and bootstrap confidence intervals were computed in Python (SciPy and statsmodels). The significance level was set at $\alpha = 0.05$.

RESULTS AND DISCUSSION

Results

Before the intervention, the experimental group's mean pre-test score ($M = 46.45$, $SD = 11.99$) was 4.84 points higher than that of the control group ($M = 41.61$, $SD = 10.03$; 95% CI for the difference [−0.78, 10.46]). The difference was not statistically significant ($U = 361.5$, $Z = -1.757$, $p = 0.080$), but its size ($r_{rb} = 0.25$) was not negligible, so the groups cannot be regarded as fully equivalent at baseline. Pre-test scores were therefore used as a covariate in the primary analysis. Table 2 presents the descriptive statistics.

Table 2. Descriptive statistics of pre-test, post-test, and N-gain scores

Group	*n*	Pre-test *M* (*SD*) [95% CI]	Post-test *M* (*SD*) [95% CI]	N-gain *M* (*SD*) [95% CI]
Experimental	31	46.45 (11.99) [42.06, 50.85]	87.42 (10.94) [83.40, 91.43]	0.77 (0.19) [0.70, 0.84]
Control	31	41.61 (10.03) [37.93, 45.29]	76.45 (11.70) [72.16, 80.74]	0.61 (0.18) [0.54, 0.67]

Note. Scores range from 0 to 100 (ten items scored 0 or 1, multiplied by 10). Medians: pre-test 50 (experimental) and 40 (control); post-test 90 and 80; N-gain 0.80 and 0.57.

The assumptions of ANCOVA were met: the residuals were normally distributed (Shapiro-Wilk $W = 0.980$, $p = 0.405$), the regression slopes did not differ between groups (group × pre-test interaction, $p = 0.119$), and the post-test variances were homogeneous (Levene's $F = 0.26$, $p = 0.615$). After adjusting for pre-test scores, the experimental group's adjusted mean post-test score (86.20) was 8.53 points higher than that of the control group (77.67; 95% CI [3.35, 13.71]), $F(1, 59) = 10.85$, $p = 0.002$, partial $\eta^2 = 0.155$ (Table 3). The adjusted difference is smaller than the unadjusted difference of 10.97 points (95% CI [5.21, 16.73]), indicating that part of the raw difference reflected the experimental group's higher starting point. When pre-test scores were corrected for their reliability, the adjusted difference decreased further to 6.45 points (95% CI [0.98, 11.92]), $F(1, 59) = 5.57$, $p = 0.022$. The difference therefore remained statistically significant, but it was smaller and less precisely estimated.

Table 3. ANCOVA of post-test scores with pre-test scores as covariate

Source	*SS*	*df*	*F*	*p*	Partial η^2
Pre-test (covariate)	1861.41	1	18.80	< 0.001	0.242
Group	1074.39	1	10.85	0.002	0.155
Residual	5841.82	59			

The nonparametric analyses led to the same conclusion (Table 4). The experimental group had higher post-test scores (mean ranks 39.13 vs. 23.87; $U = 244.0$, $Z = -3.417$, $p < 0.001$, $r_{rb} = 0.49$, 95% CI [0.24, 0.72]) and higher N-gain scores (mean ranks 38.94 vs. 24.06; $U = 250.0$, $Z = -3.271$, $p = 0.001$, $r_{rb} = 0.48$, 95% CI [0.22, 0.72]). Quade's rank analysis of covariance, which adjusts for pre-test scores without assuming normality, also showed a significant difference, $F(1, 60) = 9.84$, $p = 0.003$.

Table 4. Mann-Whitney U tests on pre-test, post-test, and N-gain scores

Measure	Mean rank (E)	Mean rank (C)	U	Z	p	r_{rb} [95% CI]
Pre-test	35.34	27.66	361.5	−1.757	0.080	0.25 [−0.03, 0.51]
Post-test	39.13	23.87	244.0	−3.417	< 0.001	0.49 [0.24, 0.72]
N-gain	38.94	24.06	250.0	−3.271	0.001	0.48 [0.22, 0.72]

Note. E = experimental (n = 31); C = control (n = 31). r_{rb} = rank-biserial correlation, 1 − 2U / (n_1n_2); 95% CI from 10,000 bootstrap resamples.

In practical terms, the adjusted difference of 8.53 points corresponds to almost one additional correct item out of ten. Nine of the 31 experimental students (29%) and three of the 31 control students (10%) answered all post-test items correctly; this ceiling may have limited the observable difference. Descriptively, the experimental group's mean N-gain fell in Hake's high category and the control group's in the moderate category, but these categories should be read alongside the between-group difference reported above rather than as independent evidence. The pattern of results is consistent with the hypothesis that students taught with mountain-valley representations would obtain higher pre-test-adjusted post-test scores and higher N-gain scores than students receiving direct instruction. Given the nonrandomized design and the differences between teachers, however, the results do not establish that the representations caused the difference.

Discussion

The findings of this study indicate that seventh-grade students receiving mountain-valley-based GASING instruction demonstrated higher immediate computational performance in integer addition and subtraction than students receiving direct instruction. After adjusting for initial achievement, the experimental group obtained a mean post-test score of 86.20, compared with 77.67 in the control group, resulting in a statistically significant adjusted difference of 8.53 points (95% CI [3.35, 13.71]; $F(1, 59) = 10.85$, $p = 0.002$, partial $\eta^2 = 0.155$). This difference corresponds to approximately one additional correct response on the ten-item assessment, indicating an observable advantage in immediate computational accuracy. The consistency of the findings across post-test comparisons ($U = 244.0$, $p < 0.001$, rrb = 0.49) and normalized gain analyses ($U = 250.0$, $p = 0.001$, rrb = 0.48) further supports the observed direction of the between-group differences. These results are consistent with the findings of Ebner et al. (2025), who examined the educational outcomes associated with concrete-representational-abstract instruction and highlighted the potential benefits of systematically connecting mathematical representations with symbolic procedures. Similarly, Flores et al. (2024) investigated structured representational instruction in mathematics learning, providing relevant evidence regarding the educational value of progressively organized instructional experiences. However, differences in mathematical content, participant characteristics, instructional duration, and assessment instruments limit direct comparisons between these studies and the present investigation. Unlike previous research examining broader mathematical competencies, the present study specifically evaluated computational accuracy across ten types of integer addition and subtraction problems immediately following a single instructional session. The findings therefore suggest that the GASING instructional sequence may be associated with improved short-term computational performance, although they do not establish whether the observed advantage would persist over time. Overall, the results contribute to the growing literature on representational mathematics instruction by providing comparative classroom evidence concerning the immediate learning outcomes associated with mountain-valley-based GASING activities.

The observed differences may be interpreted through the mathematical characteristics of mountain-valley representations, particularly their emphasis on the relationships between positive quantities, negative quantities, and additive inverses. Within this instructional model, mountains symbolize positive integers, valleys represent negative integers, and opposing units cancel to

produce a neutral value. This structure provides a physical interpretation of integer operations that may help students relate symbolic expressions to observable mathematical relationships. Sari et al. (2020) investigated a neutralization-based empty number-line model and identified the importance of instructional design in supporting students' interpretation of integer addition and subtraction. Their findings also indicated that students could emphasize procedural use of the representation rather than employ it as a model for mathematical reasoning, suggesting that the presence of a representation alone does not guarantee conceptual understanding. Extending this research, Sari et al. (2025) examined the mound-hollow model and reported that three sixth-grade students used the representation to solve integer problems through cancellation and the construction of opposing pairs. These findings provide a relevant conceptual comparison because the mountain-valley approach similarly illustrates zero-pair relationships through complementary positive and negative quantities. Nevertheless, the present investigation differs from Sari et al. (2025) by examining comparative computational outcomes among 62 seventh-grade students rather than analyzing the reasoning processes of a small group of learners. The higher adjusted post-test performance observed in the experimental class is compatible with the possibility that physical representations supported students' interpretation of integer operations, but the assessment did not directly measure representational understanding. Consequently, the findings extend previous research by documenting an association between representational instruction and computational accuracy while leaving the underlying cognitive processes open to further investigation.

The instructional sequence employed in this study also provides a theoretical basis for interpreting the observed differences because GASING integrates concrete manipulation, symbolic writing, and mental calculation within a progressively structured learning experience. During the concrete stage, students manipulated mountain-valley units to represent addition and subtraction, including operations requiring the introduction of zero pairs. These activities were followed by symbolic exercises in which students translated physical arrangements into mathematical expressions before completing mental calculations without manipulatives. Such progression is consistent with the principles of concrete-representational-abstract instruction, which emphasizes systematic connections between physical experiences and abstract mathematical procedures (Ebner et al., 2025). Flores et al. (2024) similarly examined structured representational instruction, providing relevant evidence that explicitly organized mathematical learning experiences can support students' acquisition of computational competencies. Within integer education, Sari et al. (2020) emphasized the importance of designing representational activities that encourage students to interpret mathematical relationships rather than merely reproduce operational procedures. The present study extends these perspectives by incorporating a planned transition from mountain-valley manipulation to symbolic and mental computation within regular seventh-grade mathematics instruction. Nevertheless, the intervention included several interconnected components, such as teacher-guided manipulation, paired practice, oral calculation, a GASING song, and a quartet card game, which were not independently evaluated. Furthermore, students progressed through the instructional stages according to predetermined time allocations rather than demonstrated mastery, and the study did not directly examine whether mental calculations relied on internalized mountain-valley representations. Therefore, although the findings are consistent with the educational rationale underlying structured representational instruction, they cannot establish that the gradual withdrawal of concrete materials independently explains the observed differences in computational performance.

The statistical findings require careful interpretation because the study employed a quasi-experimental design involving two intact classes without random assignment. Although the initial difference in pre-test performance was not statistically significant ($p = 0.080$), the experimental group obtained a higher mean score than the control group (46.45 versus 41.61), indicating a

potentially meaningful baseline imbalance. Adjusting for pre-test achievement through ANCOVA reduced the estimated between-group difference from 10.97 unadjusted points to 8.53 adjusted points, demonstrating the relevance of initial performance to the observed post-test comparison. The adjusted difference remained statistically significant ($p = 0.002$), and Quade's rank analysis of covariance produced a consistent result ($F(1, 60) = 9.84$, $p = 0.003$). Nevertheless, statistical adjustment cannot fully eliminate selection bias or account for unmeasured classroom characteristics in nonrandomized educational investigations. The sensitivity analysis addressing pre-test measurement unreliability yielded a smaller adjusted difference of 6.45 points (95% CI [0.98, 11.92]; $p = 0.022$), suggesting that the estimated magnitude was sensitive to assumptions concerning baseline measurement. This consideration is particularly important because the pre-test and post-test reliability coefficients were 0.54 and 0.68, respectively, and the two test forms were not statistically equated. Moreover, the experimental and control classes were taught by different teachers, and only the experimental teacher received specialized GASING preparation, making it impossible to separate instructional effects from teacher-related influences. Because each condition consisted of a single classroom, student-level statistical analyses could not adequately distinguish instructional differences from classroom-level variation, thereby limiting the strength of causal interpretation. Accordingly, the statistically significant findings should be understood as evidence of differences between the participating instructional conditions rather than definitive proof that mountain-valley representations independently caused superior mathematical achievement.

The findings offer several implications for mathematics education, particularly regarding the integration of concrete representations into instruction on integer addition and subtraction. The mountain-valley approach provides a potentially accessible means of illustrating positive-negative relationships, additive inverses, and zero-pair construction through physical activities before students encounter symbolic computation. This instructional rationale aligns with the findings of Sari et al. (2025), who demonstrated how mound-hollow representations could support students in explaining and solving integer operations involving negative quantities. Similarly, the representational progression examined by Flores et al. (2024) and the broader evidence synthesized by Ebner et al. (2025) reinforce the relevance of structured transitions between concrete activities and abstract mathematical procedures. Nevertheless, the present results should not be interpreted as evidence that GASING is universally superior to direct instruction because the control class also received teacher explanations, number-line demonstrations, paired exercises, and opportunities to present mathematical solutions. The relatively high post-test means in both groups indicate that students demonstrated substantial immediate computational performance under both instructional conditions, although the experimental group achieved higher adjusted scores. Additionally, the occurrence of perfect post-test scores among nine experimental students and three control students suggests a possible ceiling effect that may have restricted the assessment's sensitivity to differences among higher-performing learners. Future research should employ independently validated and appropriately equated assessment instruments that measure not only computational accuracy but also students' explanations, representational strategies, and conceptual interpretation of integer operations. More rigorous investigations involving multiple classrooms, balanced teacher preparation, systematic instructional fidelity assessments, and delayed post-tests are needed to determine the reproducibility, sustainability, and potential mechanisms of the observed differences. In conclusion, this study provides preliminary empirical evidence that mountain-valley-based GASING instruction is associated with higher immediate computational accuracy in integer addition and subtraction, while emphasizing that stronger experimental designs and more comprehensive assessments are necessary before drawing broader conclusions about its instructional effectiveness.

Implications

The study indicates that mountain-valley representations can be used within a regular Indonesian Grade 7 lesson: the materials are simple to produce, and the session required only a modest extension of the scheduled lesson time. Whether the approach is particularly suited to under-resourced classrooms was not examined and cannot be inferred from a single school. For teacher preparation, the study offers one concrete observation: the experimental-class teacher received about four hours of training on the media and the sequence before teaching. Professional development modeled on this study should therefore cover at least the use of the media, the justification for zero pairs, and the planned withdrawal of the media, although the amount of training required for effective implementation remains an open question.

Limitations and Suggestions for Future Research

The findings are subject to several limitations. First, the two classes were taught by different teachers, and only the experimental-class teacher received training, so teacher effects are fully confounded with the intervention and the difference between the classes may reflect the teacher as much as the approach. In addition, only the experimental class sang the GASING song and played the quartet card game, which may have raised engagement independently of the representations. Second, the classes were intact and assigned by the school, the experimental class started with somewhat higher scores, and only one class per condition was studied, so selection effects cannot be ruled out even after statistical adjustment. Third, the first author, a GASING trainer, prepared the experimental-class teacher and observed both lessons without a fidelity checklist, so researcher allegiance and observer effects may have influenced implementation. Fourth, the tests were short computation tests that had not been validated by independent experts or piloted, their internal consistency was moderate for the pre-test (KR-20 = 0.54) and the post-test (KR-20 = 0.68), and a sensitivity analysis correcting for pre-test unreliability reduced the adjusted difference from 8.53 to 6.45 points; the pre-test and post-test forms were not equated, and nine experimental students reached the maximum post-test score; the results therefore concern computational performance rather than conceptual understanding, and the ceiling may have limited the observed difference. Fifth, the intervention consisted of a single session followed immediately by the post-test, so the findings establish neither retention nor sustained conceptual development. Finally, written parental consent and institutional ethics review were not obtained.

Future research should address these weaknesses in order of priority. First, teacher effects should be separated from the intervention, for example by having the same trained teacher teach both conditions or by randomly assigning several classes and teachers to each condition and analyzing the data with multilevel models. Second, outcomes should include measures of reasoning, such as items that ask students to explain answers like $3 - (-2)$, or interviews, so that changes in understanding can be distinguished from computational accuracy. Third, implementation fidelity should be documented with observation checklists, and the media and the three-stage sequence could be tested in separate conditions to identify their respective contributions. Fourth, delayed post-tests should be included to examine retention. These steps should precede the study of additional outcomes.

CONCLUSION

In one Grade 7 setting, students taught integer addition and subtraction with mountain-valley representations through the GASING sequence obtained higher post-test scores than students taught through direct instruction, after adjustment for pre-test scores (adjusted difference of 8.53 points on a 0–100 scale, about one additional correct item out of ten). The difference appeared on a test of

computational accuracy administered immediately after a single session. Because the classes were intact, came from a single school, and were taught by different teachers with unequal preparation, the result shows an association between the approach and short-term performance rather than a causal effect, and it does not demonstrate improved conceptual understanding. The proposed mechanism, in which a concrete image of cancelling units supports symbolic and mental computation, was not examined directly. Priority directions for further research are designs that separate teacher effects from the intervention, measures of students' reasoning, and delayed tests of retention.

AUTHOR CONTRIBUTIONS STATEMENT

Yenni Natalina Buknoni contributed to the conceptualization, research design, methodology, instructional material development, data collection, data analysis, and preparation of the original manuscript. Turmudi contributed to the conceptualization, methodological guidance, research supervision, interpretation of findings, and critical review of the manuscript. Effy Mulyasari contributed to methodological refinement, research supervision, interpretation of results, and critical revision of the manuscript. All authors reviewed and approved the final version of the manuscript and agreed to be accountable for its content.

REFERENCES

Abrahamson, D., Nathan, M. J., Williams-Pierce, C., Walkington, C., Ottmar, E. R., Soto, H., & Alibali, M. W. (2020). The future of embodied design for mathematics teaching and learning. *Frontiers in Education*, *5*, Article 147. https://doi.org/10.3389/feduc.2020.00147

Acosta, C. P., & Soliba, B. L. (2024). Pedagogical approach of grade 7 teachers in teaching the learning competency of integers. *Educational Technology Quarterly*, *2024*(1), 38-55. https://doi.org/10.55056/etq.652

Anggraeni, F., Andayani, S., & Rahmawati, D. (2025). Analysis of students' mastery levels in using integer operations in elementary school mathematics learning. *International Journal of Multidisciplinary Research of Higher Education*, *8*(1), 1-13.

Anghel, T., Hogea, L., Costea, I., Marinca, A., Dumache, R., Nussbaum, L., & Trăilă, I.-A. (2026). Enhancing emotional intelligence in adolescents through creativity-based experiential group training: A controlled pretest-posttest study. *Adolescents*, *6*(5), Article 67. https://doi.org/10.3390/adolescents6050067

Anton, J., & Abrahamson, D. (2025). Walking the number line: Towards an enactive understanding of integer arithmetic. *Instructional Science*, *53*(4), 657-677. https://doi.org/10.1007/s11251-025-09707-w

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

Borji, V., Radmehr, F., & Font, V. (2021). The impact of procedural and conceptual teaching on students' mathematical performance over time. *International Journal of Mathematical Education in Science and Technology*, *52*(3), 404-426. https://doi.org/10.1080/0020739X.2019.1688404

Bouck, E. C., Long, H., & Jakubow, L. (2021). Teaching students to solve subtraction problems online via the virtual-representational-abstract instructional sequence. *Investigations in Mathematics Learning*, *13*(3), 197-213. https://doi.org/10.1080/19477503.2021.1963157

Braithwaite, D. W., & Sprague, L. (2021). Conceptual knowledge, procedural knowledge, and metacognition in routine and nonroutine problem solving. *Cognitive Science*, *45*(10), Article e13048. https://doi.org/10.1111/cogs.13048

Bryant, D. P., Bryant, B. R., Dougherty, B., Roberts, G., Pfannenstiel, K. H., & Lee, J. (2020). Mathematics performance on integers of students with mathematics difficulties. *The Journal of Mathematical Behavior*, *58*, Article 100776. https://doi.org/10.1016/j.jmathb.2020.100776

Buknoni, Y., Wahyudin, W., & Iriawan, S. B. (2026). Efektivitas metode GASING untuk meningkatkan pemahaman konsep bilangan bulat di sekolah. *Scholaria: Jurnal Pendidikan dan Kebudayaan*, *16*(2), 194-203. https://doi.org/10.24246/j.js.2026.v16.i2.p194-203

Dalgaard, N. T., Bondebjerg, A., Viinholt, B. C. A., & Filges, T. (2022). The effects of inclusion on academic achievement, socioemotional development and wellbeing of children with special educational needs. *Campbell Systematic Reviews*, *18*(4), Article e1291. https://doi.org/10.1002/cl2.1291

Daroczy, G., Meurers, D., Heller, J., Wolska, M., & Nürk, H.-C. (2020). The interaction of linguistic and arithmetic factors affects adult performance on arithmetic word problems. *Cognitive Processing*, *21*(1), 105-125. https://doi.org/10.1007/s10339-019-00948-5

Dea, M. F., Bantas, M. G. D., & Mbagho, H. M. (2025). Efektivitas metode GASING dalam meningkatkan pemahaman matematis siswa kelas VII SMP Negeri 2 Ende. *Math Didactic: Jurnal Pendidikan Matematika*, *11*(2), 199-207. https://doi.org/10.33654/math.v11i2.181

Ding, M., Mochaourab, R., & Spiro, B. (2024). Promoting teaching improvement in elementary mathematics: A case with inverse relations. *Asian Journal for Mathematics Education*, *3*(2), 143-173. https://doi.org/10.1177/27527263241256678

Dorner, C., Ableitinger, C., & Krammer, G. (2026). Revealing the nature of mathematical procedural knowledge by analysing students' deficiencies and errors. *International Journal of Mathematical Education in Science and Technology*, *57*(3), 413-434. https://doi.org/10.1080/0020739X.2024.2445666

Ebner, S., MacDonald, M. K., Grekov, P., & Aspiranti, K. B. (2025). A meta-analytic review of the concrete-representational-abstract math approach. *Learning Disabilities Research & Practice*, *40*(1), 31-42. https://doi.org/10.1177/09388982241292299

Erbilgin, E., & Macur, G. M. A. (2022). A subtraction game to scaffold primary students' word problem solving skills. *Journal on Mathematics Education*, *13*(2), 307-322.

Fitriasari, P., Suryadi, D., Nurlaelah, E., Suhendra, S., Hermanto, H., & Ningsih, Y. L. (2026). A didactical transposition study on secondary students' number sense of integers. *Frontiers in Education*, *11*, Article 1739073. https://doi.org/10.3389/feduc.2026.1739073

Flores, M. M., Hinton, V. M., Shadoan, L., & Monroe, C. (2024). Teaching place value concepts and their application using the concrete-representational-abstract integrated sequence. *Learning Disabilities Research & Practice*, *39*(3), 132-145. https://doi.org/10.1177/09388982241245464

Fries, L., Son, J. Y., Givvin, K. B., & Stigler, J. W. (2021). Practicing connections: A framework to guide instructional design for developing understanding in complex domains. *Educational Psychology Review*, *33*(2), 739-762. https://doi.org/10.1007/s10648-020-09561-x

Hadi, S., Dolk, M., Kamaliyah, & Hidayanto, T. (2025). Mathematical reasoning: How students learn mathematics? Journal on Mathematics Education, 16(3), 937-954.

Han, X., Lin, X., & Song, X. (2026). The impact of mathematical-content complexity on students' mathematics-writing performance. *Reading and Writing*, *39*(4), 1187-1210. https://doi.org/10.1007/s11145-025-10647-7

Hansen, M. E., & Jankvist, U. T. (2021). Designing tasks for a dynamic online environment: Applying research into students' difficulties with linear equations. *Mathematics*, *9*(5), Article 557. https://doi.org/10.3390/math9050557

Jeranah, Nisraeni, Novianti, R., Huda, N., & Ahmad, A. (2025). The use of the Rasch model to analyze misconceptions on the topic of integer numbers. *Journal for Lesson and Learning Studies*, *8*(2), 385-395.

Kerby, D. S. (2014). The simple difference formula: An approach to teaching nonparametric correlation. *Comprehensive Psychology*, *3*, Article 1. https://doi.org/10.2466/11.IT.3.1

Kim, H. (2020). Concreteness fading strategy: A promising and sustainable instructional model in mathematics classrooms. *Sustainability*, *12*(6), Article 2211. https://doi.org/10.3390/su12062211

Kim, J. (2025). How representations support young students' conceptual understanding of multiplication? A study on exploring third-grade students' representational reasoning. *Asian*

Journal for Mathematics Education, *4*(3), 311-331. https://doi.org/10.1177/27527263251335686

Kokkonen, T., & Schalk, L. (2021). One instructional sequence fits all? A conceptual analysis of the applicability of concreteness fading in mathematics, physics, chemistry, and biology education. *Educational Psychology Review*, *33*(3), 797-821. https://doi.org/10.1007/s10648-020-09581-7

Kwangmuang, P., Najui, S., Duangngern, P., & Gessala, N. (2026). Design and evaluation of an AR-integrated STEM-BCG learning kit to foster creative problem solving in secondary chemistry education. Humanities and Social Sciences Communications. https://doi.org/10.1057/s41599-026-09001-1

Li, Y., Zhang, X., & Heo, G. (2026). The effects of educational drama on students' key competencies: A systematic review and meta-analysis. *Frontiers in Psychology*, *17*, Article 1870089. https://doi.org/10.3389/fpsyg.2026.1870089

Lin, T.-H., & Riccomini, P. J. (2026). Understanding error patterns in integer operations of students with and without mathematics difficulty: A descriptive analysis. *Learning Disabilities Research & Practice*, *41*(1), 34-47. https://doi.org/10.1177/09388982251366454

Maizora, S., & Rosjanuardi, R. (2020). Konsepsi siswa kelas tiga sekolah dasar tentang bilangan bulat. *Pythagoras: Jurnal Pendidikan Matematika*, *15*(1), 50-61. https://doi.org/10.21831/pg.v15i1.31526

Merdja, J., Mulyatiningsih, E., Rosnawati, R., Apriyansyah, C., Rahmawati, L., Saputra, G., & Disnawati, H. (2026). CATPO as an innovative evaluation framework for GASING-based mathematics learning: Bridging educational evaluation and instructional innovation. F1000Research. https://doi.org/10.12688/f1000research.179375.1

Pellas, N. (2024). Effects of Kahoot! on K-12 students' mathematics achievement and multi-screen addiction. *Multimodal Technologies and Interaction*, *8*(9), Article 81. https://doi.org/10.3390/mti8090081

Purwasih, R., Turmudi, T., Dahlan, J. A., Irawan, E., & Minasyan, S. (2024). A semiotic perspective of mathematical activity: The case of integer. *Infinity Journal*, *13*(1), 271-284. https://doi.org/10.22460/infinity.v13i1.p271-284

Putra, D. E., Rakhmad, H., Andini, D. P., Firgiyanto, R., Hudori, H. A., Antika, E., Umami, M. R., Mukhlisoh, N. A., Setyohadi, D. P. S., & Rachmanita, R. E. (2026). Topic-level effectiveness of the entrepreneur development class in Indonesia's Wirausaha Merdeka program. *Journal of Economics, Entrepreneurship, Management Business and Accounting*, *4*(4), 371-392. https://doi.org/10.61255/jeemba.v4i4.1421

Qetrani, S., Ouailal, S., & Achtaich, N. (2021). Enhancing students' conceptual and procedural knowledge using a new teaching approach of linear equations based on the equivalence concept. *Eurasia Journal of Mathematics, Science and Technology Education*, *17*(7), Article em1978. https://doi.org/10.29333/ejmste/10938

Rezende, M. B., Nacife, J. M., Barbosa, D. P., Guimarães, Í. J. B., Ouro-Salim, O., & Freitas, T. M. (2026). Effects of leadership and entrepreneurship training on family farmers' competencies: A quasi-experimental study in Brazil. *Sustainability*, *18*(18), Article 9394. https://doi.org/10.3390/su18189394

Sari, P., Dindyal, J., & Shutler, P. M. E. (2025). The "mound-hollow" model for solving integer addition and subtraction problems. *Journal on Mathematics Education*, *16*(1), 365-382. https://doi.org/10.22342/jme.v16i1.pp365-382

Sari, P., Hajizah, M. N., & Purwanto, S. (2020). The neutralization on an empty number line model for integer additions and subtractions: Is it helpful? Journal on Mathematics Education, 11(1), 1-16. https://doi.org/10.22342/jme.11.1.9781.1-16

Satsangi, R., Park, S., & Raines, A. R. (2025). Teaching fraction computations to children with mathematics difficulties using a virtual number line intervention. *Learning Disabilities Research & Practice*, *40*(4), 219-231. https://doi.org/10.1177/09388982251358399

Schaathun, H. G. (2022). On understanding in mathematics. *Teaching Mathematics and Its Applications: An International Journal of the IMA*, *41*(4), 318-328. https://doi.org/10.1093/teamat/hrac016

Sercenia, J. C., Ibañez, E. D., & Pentang, J. T. (2023). Thinking beyond thinking: Junior high school students' metacognitive awareness and conceptual understanding of integers. *Mathematics Teaching Research Journal*, *15*(1), 4-24.

Simon, M. A. (2022). Contributions of the learning through activity theoretical framework to understanding and using manipulatives in the learning and teaching of mathematical concepts. *The Journal of Mathematical Behavior*, *66*, Article 100945. https://doi.org/10.1016/j.jmathb.2022.100945

Surya, Y. (2024). *Modul pelatihan GASING matematika dasar: Bilangan bulat, pecahan, dan desimal.* GA Press.

Sutherland, M., Furjanic, D., Hermida, J., & Clarke, B. (2024). Using the number line to develop understanding of whole number magnitude and operations. *Intervention in School and Clinic*, *59*(3), 158-164. https://doi.org/10.1177/10534512231156869

Torres-Peña, R. C., Peña-González, D., Lara-Orozco, J. L., Ariza, E. A., & Vergara, D. (2025). Enhancing numerical thinking through problem solving: A teaching experience for third-grade mathematics. *Education Sciences*, *15*(6), Article 667. https://doi.org/10.3390/educsci15060667

Vest, N. A., & Alibali, M. W. (2026). Flexibility in abstract thought: Context-dependent representations of integer magnitude. *Psychonomic Bulletin & Review*, *33*(7), Article 243. https://doi.org/10.3758/s13423-026-02977-6

Wessman-Enzinger, N. M., & Mooney, E. S. (2021). Conceptual models for integer addition and subtraction. *International Journal of Mathematical Education in Science and Technology*, *52*(3), 349-376. https://doi.org/10.1080/0020739X.2019.1685136

White Brahmia, S., Olsho, A., Smith, T. I., & Boudreaux, A. (2020). Framework for the natures of negativity in introductory physics. *Physical Review Physics Education Research*, *16*(1), Article 010120. https://doi.org/10.1103/PhysRevPhysEducRes.16.010120

Wilkie, K. J., & Hopkins, S. (2024). Primary students' relational thinking and computation strategies with concrete-to-symbolic representations of subtraction as difference. *The Journal of Mathematical Behavior*, *73*, Article 101121. https://doi.org/10.1016/j.jmathb.2023.101121

Yakubova, G., Hughes, E. M., & Baer, B. L. (2020). Supporting students with ASD in mathematics learning using video-based concrete-representational-abstract sequencing instruction. *Preventing School Failure: Alternative Education for Children and Youth*, *64*(1), 12-18. https://doi.org/10.1080/1045988X.2019.1627999

Zeng, J., Zhang, P., Zhou, J., Shang, J., & Black, J. B. (2025). The impact of embodied scaffolding sequences on STEM conceptual learning. *Educational Technology Research and Development*, *73*(2), 767-792. https://doi.org/10.1007/s11423-024-10438-x