Journal Evaluation in Education (JEE)
Vol. 7, No. 2, April 2026, pp. 556~567
ISSN: 2716-4160, DOI: 10.37251/jee.v7i2.2810

Research Article

Effective Drill-Based Arithmetic Training for Improving Numeracy Literacy: A Quasi-Experimental Study With High N-Gain among Elementary Students

Safiru Hunas[1], Eva Puspitasari[2], Rama Wijaya Abdul Rozak[1,*], Nandang Rusmana[1]
[1] Teacher Education, Faculty of Graduate School, Universitas Pendidikan Indonesia, Jawa Barat, Indonesia
[2] Curiculum Development, Faculty of Educational Science, Universitas Pendidikan Indonesia, Jawa Barat, Indonesia

Article Info

Article history:

Received Feb 17, 2026
Revised Mar 12, 2026
Accepted Apr 14, 2026
OnlineFirst Apr 27, 2026

Keywords:

Arithmetic Operations
Elementary School
N-Gain
Numeracy Literacy
Student Training

ABSTRACT

Purpose of the study: This study aims to analyze the effectiveness of a drill-based arithmetic training program in improving elementary students' numeracy literacy.

Methodology: This quasi-experimental one-group pretest-posttest study involved 32 fourth-grade students selected through a total sampling approach. A 20-item arithmetic essay test was used as the instrument (CVR = 0.89; Cronbach's Alpha = 0.856). The intervention comprised 8 sessions (90 minutes each) integrating drill and practice, concrete manipulative media, scaffolding, peer tutoring, and corrective feedback. Data were analyzed using the Wilcoxon Signed-Rank Test ($\alpha = 0.05$) and Hake's N-Gain.

Main Findings: Pretest scores were extremely low (M = 15.31), with 0% achieving KKM (Minimum Completion Criteria). Post-intervention scores rose to M = 82.03, a gain of 66.72 points (435.71%), with 84.4% of students achieving KKM. The Wilcoxon test confirmed a highly significant improvement ($p = 0.000001$), and N-Gain yielded a mean of 0.7964 (high category). Effectiveness is attributed to the integration of drill-and-practice (behavioristic theory), concrete manipulatives (Piaget), peer tutoring within the Zone of Proximal Development (Vygotsky), and gradual scaffolding. The N-Gain exceeded prior studies. Limitations include the absence of a control group and a small sample size, restricting generalizability.

Novelty/Originality of this study: This is the first study to examine an integrated, multi-component program that addresses all four arithmetic operations simultaneously in Eastern Indonesia (Buton Regency), demonstrating that multi-component designs yield superior N-Gain outcomes and offering a replicable framework for teachers addressing low numeracy literacy.

Corresponding Author:
Rama Wijaya Abdul Rozak,
Teacher Education, Faculty of Graduate School, Universitas Pendidikan Indonesia,
Jl. Setiabudhi No. 229, Isola, Sukasari, Kota Bandung, Jawa Barat, 40154, Indonesia
Email: gede.agus.siswadi@mail.ugm.ac.id

1. INTRODUCTION

Numeracy literacy is recognized as one of the most critical 21st-century competencies that students must acquire from an early age [1], [2]. The Ministry of Education, Culture, Research, and Technology (2021) defines numeracy literacy as the ability to use numbers and mathematical symbols to solve problems across various real-life contexts. Within elementary school mathematics education, numeracy literacy is fundamentally anchored in the mastery of four basic arithmetic operations: addition, subtraction, multiplication, and division [3], [4]. These operations constitute the cognitive foundation upon which all subsequent mathematical learning is built and are essential for the application of mathematics in everyday life [5], [6]. Without solid mastery of these foundational skills, students face compounding difficulties as mathematical content becomes increasingly complex at higher educational levels.

Despite its fundamental importance, Indonesian students' numeracy literacy remains alarmingly low compared to international standards. The 2022 Programme for International Student Assessment (PISA) revealed that Indonesia scored only 366 in mathematics, significantly below the OECD average of 472—a gap of 106 points representing approximately 2.5 years of schooling difference [7]. Indonesia ranked 67th out of 81 participating countries in the mathematics category. At the national level, the 2023 National Assessment (Asesmen Nasional) results showed that 50.76% of Indonesian students remain in the category requiring special intervention in numeracy competence [8]. This persistent gap between expected standards and actual student performance constitutes a critical educational crisis that demands urgent, evidence-based intervention, particularly at the elementary school level, where foundational mathematical skills are first established [9], [57].

This national-level gap is mirrored and even more pronounced at the local level. At SD Negeri 42 Buton, Buton Regency, Southeast Sulawesi, initial observations and documentation of fourth-grade students' daily test results on arithmetic operations (September 2025) revealed a severe performance gap: not a single student out of 32 (0%) achieved the Minimum Completion Criteria (KKM) of 70, with the class average score reaching only 45.3. This represents a gap of 24.7 points below the KKM threshold, indicating that students are performing far below minimum expected standards. Interviews with the classroom teacher confirmed that students experienced pervasive difficulties across all arithmetic operations, with the most severe challenges in multiplication (inability to recall basic multiplication facts 1–10) and division (poor conceptual understanding of division as repeated subtraction and inverse of multiplication). Students also exhibited persistent procedural errors in column techniques for addition and subtraction, particularly in carrying and borrowing operations [10]. The magnitude of this gap—from the international level (PISA), through the national level (Asesmen Nasional), down to the classroom level—establishes a multi-layered performance deficit that urgently requires targeted intervention.

Analysis of the root causes underlying this performance gap reveals four interrelated factors. First, regular classroom instruction provides insufficient frequency and intensity of arithmetic practice, with students receiving limited opportunities for repeated skill-building exercises [11]. Second, the absence of concrete manipulative media in mathematics instruction prevents students—who are developmentally at Piaget's concrete operational stage—from connecting abstract numerical concepts to tangible experiences [12]. Third, constrained instructional time within the standard curriculum does not permit individualized practice tailored to each student's specific learning needs and pace [13]. Fourth, low motivation and diminished self-confidence toward mathematics, often reinforced by repeated failure experiences, create a negative affective cycle that further impedes learning [14], [15]. These compounding factors create a self-reinforcing cycle of low performance, suggesting that conventional classroom instruction alone is inadequate and that a specialized, structured intervention is urgently needed to break this cycle.

To address this multi-level performance gap, this study proposes a structured arithmetic operation training program as a targeted problem-solving intervention. Arithmetic operation training is a learning activity specifically and intensively designed to improve students' skills in performing basic mathematical operations [16]. Unlike conventional classroom instruction, this training program integrates five key pedagogical components grounded in established learning theories: (1) drill and practice methods to build procedural fluency through repeated, systematic exercises, supported by behavioristic learning theory which posits that repetitive practice strengthens stimulus-response bonds [17]-[19]; (2) concrete manipulative media (marbles, matchsticks, number cards) to facilitate conceptual understanding among students at the concrete operational stage, aligned with Piaget's cognitive development theory [20], [21]; (3) gradual scaffolding from simple to complex operations to manage cognitive load, consistent with Sweller's cognitive load theory; (4) peer tutoring to create collaborative learning opportunities within Vygotsky's Zone of Proximal Development (ZPD); and (5) immediate corrective feedback and positive reinforcement to accelerate error correction and sustain motivation. This multi-component design directly addresses each identified root cause of the performance gap, providing a comprehensive and theoretically grounded intervention framework.

Several previous studies have demonstrated the potential effectiveness of intensive training approaches for improving elementary students' arithmetic abilities, yet significant research gaps remain. Susilowati and Retnawati (2021) found that drill-method training improved the speed and accuracy of elementary students' arithmetic operations, but achieved only moderate effectiveness (N-Gain = 0.68) and focused primarily on single operations [22]. Handayani, reported a 72% improvement in fourth-grade students' understanding of multiplication and division through manipulative media, but did not examine addition and subtraction simultaneously [23]. Wijaya et al., demonstrated that drill-and-practice-based remedial programs were effective for addition and subtraction operations (effect size = 0.85), but excluded multiplication and division [24]. Three critical gaps emerge from this literature review: (1) no existing study has comprehensively examined the effectiveness of an integrated training program addressing all four arithmetic operations simultaneously within a single intervention; (2) research on numeracy interventions in Eastern Indonesia—a region characterized by limited educational resources, geographic isolation, and below-average academic outcomes—remains severely underrepresented in the literature [25], [26]; and (3) none of the previous studies combined all five pedagogical components (drill and practice, concrete media, gradual scaffolding, peer tutoring, and direct feedback) into a unified training framework. These gaps underscore both the scientific justification and practical urgency for the present study.

Based on the identified performance gaps and the existing research gaps described above, this study aims to analyze the effectiveness of a structured, multi-component arithmetic operation training program on improving the numeracy literacy skills of fourth-grade elementary students. Specifically, the research objectives are: (1) to describe the initial arithmetic operation abilities of fourth-grade students before the training intervention; (2) to examine the implementation process of the arithmetic operation training program using drill and practice methods, concrete media, gradual scaffolding, peer tutoring, and direct feedback across 8 intensive sessions; and (3) to determine whether there is a statistically significant improvement in students' arithmetic operation abilities following the training program, as measured by N-Gain effectiveness analysis.

Corresponding to these objectives, this study addresses the following research questions: (RQ1) What is the level of arithmetic operation abilities of fourth-grade students before participating in the training program? (RQ2) How is the arithmetic operation training program implemented across 8 sessions using an integrated multi-component approach? (RQ3) Is there a statistically significant difference in students' arithmetic operation abilities between pre-training and post-training assessments? (RQ4) What is the level of effectiveness of the arithmetic operation training program as measured by the Normalized Gain (N-Gain) index?

The results of this study are expected to provide both practical and theoretical contributions. Practically, this research offers an evidence-based, replicable training framework that teachers and schools can adopt to address low numeracy literacy among elementary students, particularly in regions with limited educational resources. Theoretically, this study contributes to the body of knowledge on the integration of multiple evidence-based pedagogical strategies—drill and practice, concrete manipulatives, scaffolded instruction, peer tutoring, and corrective feedback—as a unified intervention model for mathematics education at the elementary level [27], [28].

2. RESEARCH METHOD

This study employed a quantitative research approach with a quasi-experimental design, specifically the one-group pretest-posttest design. In this design, a single group of participants is measured before (pretest) and after (posttest) receiving an intervention, and the difference between the two measurements is analyzed to determine the effect of the treatment [29]. This design was selected because it is appropriate for evaluating the effectiveness of a specific intervention (arithmetic operation training) on student learning outcomes within a naturally occurring classroom setting where random assignment to control and experimental groups was not feasible. The study was conducted at SD Negeri 42 Buton, Buton Regency, Southeast Sulawesi Province, during November–December 2025. The research design can be represented as: $O1 \rightarrow X \rightarrow O2$, where O1 = pretest, X = arithmetic operation training intervention (8 sessions), and O2 = posttest.

The research population consisted of all fourth-grade students at SD Negeri 42 Buton for the 2025/2026 academic year, totaling 32 students. The sampling technique used was total sampling (also referred to as saturated sampling or census sampling), in which the entire population was selected as the research sample. This technique was chosen because the population size was relatively small (N = 32) and fell below the minimum threshold of 100 commonly recommended for probability sampling techniques, making total sampling the most appropriate approach to ensure adequate statistical power [30]. Three inclusion criteria were applied: (1) actively registered as a fourth-grade student at SD Negeri 42 Buton in the 2025/2026 academic year; (2) present during both the pretest and posttest administrations; and (3) attended at least 80% of the total training sessions (minimum 6 out of 8 sessions). All 32 students met these criteria, resulting in a final sample of 32 students comprising 18 males (56.25%) and 14 females (43.75%), with ages ranging from 9 to 10 years. No students were excluded from the study.

The primary instrument used in this study was an arithmetic operation ability test in essay (constructed-response) form. The instrument was adapted from the standardized arithmetic competency assessment framework developed by Van de Walle, Karp, and Bay-Williams (2019) in *Elementary and Middle School Mathematics: Teaching Developmentally* [3], and further aligned with the competency standards outlined in the Indonesian Independent Curriculum (Kurikulum Merdeka) for fourth-grade mathematics [31]. The adaptation process involved: (1) selecting item types and difficulty levels appropriate for the Indonesian elementary school context; (2) adjusting numerical ranges and contextual word problems to reflect local cultural and daily-life situations relevant to students in Buton Regency; and (3) translating and reformulating items into Bahasa Indonesia while maintaining the mathematical rigor of the original framework.

The adapted test consisted of 20 items covering four types of arithmetic operations with the following distribution: (1) addition (items 1–5); (2) subtraction (items 6–10); (3) multiplication (items 11–15); and (4) division (items 16–20). Each item was scored on a scale of 0–5 points based on a scoring rubric assessing both procedural accuracy and conceptual understanding, yielding a total maximum score of 100. Items within each operation type were arranged in increasing difficulty: items 1–2 (basic operations without carrying/borrowing), items 3–4 (operations with carrying/borrowing or multi-digit), and item 5 (contextual word problem requiring operation selection and execution).

The instrument underwent a rigorous two-stage validation process. In the first stage (content validation), three experts consisting of two mathematics education lecturers from Universitas Pendidikan Indonesia and one experienced elementary mathematics teacher with over 15 years of teaching experience evaluated each item for content relevance, construct alignment, and language clarity. The expert validation yielded an average Content Validity Ratio (CVR) of 0.89, indicating strong content validity according to the Lawshe (1975) criterion [32]. In the second stage (empirical validation), the instrument was pilot-tested on 30 fourth-grade students at a neighboring elementary school. Item validity was assessed using Pearson product-moment correlation, with the criterion r-count > r-table (0.361 at $\alpha = 0.05$, df = 28). Results showed that 18 out of 20 items met the validity threshold; the two invalid items were revised based on item analysis before inclusion in the final instrument. Instrument reliability was assessed using Cronbach's Alpha, yielding a coefficient of $\alpha = 0.856$, which is categorized as very high reliability ($\alpha > 0.80$) according to the classification by George and Mallery (2019) [32]. This reliability coefficient indicates excellent internal consistency among the test items, confirming that the instrument consistently measures arithmetic operation ability across all four operation types.

Data collection was conducted through a systematic three-stage procedure. The first stage (preparation, October 2025) included: (1) conducting initial observations of student conditions and identifying arithmetic operation difficulties through teacher interviews and documentation analysis; (2) developing and adapting the research instrument; (3) performing expert validation and pilot testing of the instrument; (4) developing structured training modules for 8 sessions; and (5) obtaining research permission from school administration and informed consent from parents/guardians. The second stage (implementation, November 2025) began with the administration of the pretest on November 7, 2025 (duration: 90 minutes), conducted under standardized testing conditions. The pretest served as the baseline measurement (O1) of students' arithmetic operation abilities. Subsequently, the arithmetic operation training intervention (X) was conducted over 8 sessions, each lasting 90 minutes, with the following design:

Session 1 (November 8, 2025): Introduction to training and diagnostic assessment. Students were introduced to training objectives, given pre-assessment to identify specific difficulties of each student, and reintroduced to basic arithmetic operation concepts using concrete media (marbles, matchsticks). Sessions 2-3 (November 10-14, 2025): Training in addition and subtraction operations. Students practiced addition and subtraction without column techniques (1-2 digit numbers), then continued with column techniques (carrying and borrowing) for 3-digit numbers. Each student completed at least 25 graded practice problems with direct feedback from the teacher. Sessions 4-5 (November 15-17, 2025): Training in multiplication operations. Students reviewed multiplication tables 1-10 with multiplication song methods and card games. Students practiced 1 digit × 1 digit multiplication, followed by column multiplication for 2 digit × 1 digit. Contextual word problems were given. Sessions 6-7 (November 21-22, 2025): Training in division operations. Students were reminded of the division concept as repeated subtraction and the inverse of multiplication using concrete media. Practiced simple division (divisible), division with remainder, and short column division technique. Daily life word problems were given. Session 8 (November 24, 2025): Intensive mixed practice and review. Students completed mixed problems of all four operations randomly, discussed common errors, reinforced weak concepts, and test simulation for posttest preparation.

Each training session employed drill and practice methods with three variations to maintain engagement and address diverse learning needs: (a) individual practice for building personal fluency, (b) paired practice through peer tutoring to leverage collaborative learning within Vygotsky's ZPD, and (c) group competitions to enhance motivation through positive social dynamics. Teachers provided immediate corrective feedback on each exercise and positive reinforcement (verbal praise, star stickers, progress charts) for students demonstrating improvement [33], [34]. The posttest (O2) was administered on November 28, 2025 using the identical instrument as the pretest, under the same standardized testing conditions (90 minutes, supervised administration). The third stage (data processing, December 2025) involved scoring, data entry, and statistical analysis.

Data analysis was conducted in three sequential stages using Python software (version 3.11) with the SciPy (version 1.11) and Pandas (version 2.1) libraries. In the first stage, descriptive statistical analysis was performed to characterize the distribution of pretest and posttest scores, including measures of central tendency (mean, median, mode) and variability (standard deviation, variance, range, minimum, and maximum values). Completion rates were calculated based on the KKM threshold of 70.

In the second stage, inferential statistical analysis was conducted to test the research hypothesis. Prior to hypothesis testing, a normality assumption test was performed using the Shapiro-Wilk test, selected for its superior statistical power with small sample sizes ($N < 50$). The normality criterion was: data are considered normally distributed if p-value > 0.05. Based on the normality test results, the appropriate hypothesis test was selected: (a) Paired Sample t-test if both pretest and posttest data are normally distributed, or (b) Wilcoxon Signed-Rank Test (non-parametric alternative) if one or both datasets violate the normality assumption [35]. The research hypotheses were formulated as: H_0: There is no significant difference between students' arithmetic operation abilities before and after the training program ($\mu pre = \mu post$); H_1: There is a significant difference between students' arithmetic operation abilities before and after the training program ($\mu pre \neq \mu post$). The decision criterion was: reject H_0 if p-value < 0.05 at a significance level of $\alpha = 0.05$. In the third stage, the effectiveness of the training program was measured using Hake's (1999) Normalized Gain (N-Gain) formula: $g = (S\_post - S\_pre) / (S\_max - S\_pre)$, where S_post = posttest score, S_pre = pretest score, and S_max = maximum possible score (100). N-Gain values were categorized according to Hake's classification: high ($g > 0.7$), medium ($0.3 \leq g \leq 0.7$), and low ($g < 0.3$) [36]. Both individual and class-average N-Gain values were calculated and classified to provide a comprehensive picture of training effectiveness across all students.

3. RESULTS AND DISCUSSION

3.1. Pre-Training Arithmetic Operation Abilities

Descriptive statistical analysis was performed on the pretest scores of 32 students. The results revealed that students' initial arithmetic operation abilities were in the very low category. The mean pretest score was 15.31 out of a maximum of 100 (SD = 15.96), indicating extremely poor mastery of basic arithmetic operations. The distribution was heavily right-skewed, with a median of 10.00 (meaning half of all students scored below 10) and a mode of 0.00 (the most frequent score was zero). The range extended from a minimum of 0 to a maximum of 55. Critically, not a single student (0 out of 32, 0%) achieved the Minimum Completion Criteria (KKM) of 70, meaning all 32 students (100%) performed below the expected minimum standard. This finding directly answers RQ1: the pre-training arithmetic operation ability level was extremely low, with a 100% failure rate relative to the KKM benchmark. Descriptive statistics of both pretest and posttest results are presented in Table 1.

Table 1. Descriptive Statistics of Pretest and Posttest Results

Statistics	Pretest	Posttest
Number of Students (N)	32	32
Mean	15.31	82.03
Median	10.00	85.00
Mode	0.00	85.00
Standard Deviation	15.96	12.11
Variance	254.74	146.55
Minimum	0.00	60.00
Maximum	55.00	100.00
Range	55.00	40.00
Students Passing (≥70)	0 (0%)	27 (84.4%)
Students Not Passing	32 (100%)	5 (15.6%)

Further analysis of pretest scores by operation type revealed a hierarchy of difficulty. Division yielded the lowest sub-scores, followed by multiplication, subtraction, and addition (in ascending order of performance). Specifically, 28 students (87.5%) scored zero on division items, and 25 students (78.1%) scored zero on multiplication items, while addition items showed relatively higher (though still very low) scores with a sub-mean of 6.25 out of 25. This pattern of differential difficulty is consistent with developmental mathematics research indicating that division requires the most complex procedural and conceptual integration [3]. The pretest data distribution, confirmed by the Shapiro-Wilk normality test ($p = 0.0004 < 0.05$), was not normally distributed, which informed the subsequent selection of non-parametric inferential tests as described in the methodology section (2.5).

3.2. Post-Training Arithmetic Operation Abilities and Training Implementation

The training intervention was implemented across 8 sessions (90 minutes each) following the data collection procedure outlined in section 2.4. All 32 students attended 100% of the sessions, meeting the minimum 80% attendance criterion. The posttest, administered under standardized conditions identical to the pretest, revealed a substantial improvement. The mean posttest score was 82.03 (SD = 12.11), representing an increase of 66.72 points (435.71%) from the pretest mean of 15.31. The reduced standard deviation (from 15.96

to 12.11) indicates that post-training abilities became more homogeneous across students. The posttest distribution was approximately normal (Shapiro-Wilk p = 0.1264 > 0.05), with a median of 85.00 and mode of 85.00, showing that the majority of students clustered around high scores. The KKM completion rate increased dramatically from 0% (0/32) to 84.4% (27/32). The five students who did not achieve KKM scored between 60–65, placing them close to the threshold and representing a marked improvement from their pretest scores of 0–15.

Individual-level analysis confirmed that all 32 students (100%) experienced score improvements, with individual gains ranging from 40 to 100 points. Of these, 24 students (75%) achieved gains exceeding 60 points. This universal improvement across all students—regardless of their initial ability level—provides preliminary evidence that the multi-component training design (drill and practice, concrete media, gradual scaffolding, peer tutoring, and direct feedback) was effective for heterogeneous learners. These descriptive results are further examined through inferential statistical testing in sections 3.3 and 3.4 to determine statistical significance and effect magnitude.

Table 2. Summary of Training Session Implementation and Post-Training Results by Operation Type

Session	Date	Focus / Activity	Method / Media	Post-Test Sub-Score (Mean/25)	Students Reaching KKM per Operation (%)
1	Nov 8, 2025	Introduction & diagnostic; concept review with concrete media	Marbles, matchsticks; direct instruction	–	–
2–3	Nov 10–14, 2025	Addition & subtraction (with/without carrying and borrowing); column technique; 25+ graded problems	Drill & practice; peer tutoring; corrective feedback	20.87 / 25	90.6%
4–5	Nov 15–17, 2025	Multiplication: tables 1–10; 1-digit × 1-digit; 2-digit × 1-digit (column); word problems	Multiplication songs; card games; drill & scaffolding	19.53 / 25	81.3%
6–7	Nov 21–22, 2025	Division: concept as repeated subtraction & inverse of multiplication; simple & remainder division; column technique; word problems	Concrete media; drill & scaffolding; peer tutoring	19.50 / 25	78.1%
8	Nov 24, 2025	Intensive mixed practice (all 4 operations); error review; test simulation	Mixed drill; group competition; positive reinforcement	22.13 / 25	84.4%
Overall	Nov 8–24, 2025	8 sessions × 90 min = 720 total minutes; 100% attendance (32/32 students)	Drill, concrete media, scaffolding, peer tutoring, feedback	82.03 / 100	84.4% (27/32)

3.3. Statistical Significance of Pre-Post Differences

Inferential statistical analysis was conducted following the analytical framework described in section 2.5. As a prerequisite, the Shapiro-Wilk normality test was applied to both datasets. Results showed that the pretest data violated the normality assumption (W = 0.864, p = 0.0004 < 0.05), while the posttest data was normally distributed (W = 0.952, p = 0.1264 > 0.05). Because at least one dataset was not normally distributed, the non-parametric Wilcoxon Signed-Rank Test was selected as the appropriate hypothesis test, in accordance with the decision criteria established in the methodology. Normality test results are presented in Table 3.

Table 3. Normality Test Results (Shapiro-Wilk)

Data	p-value	Conclusion
Pretest	0.0004	Not Normal
Posttest	0.1264	Normal

The Wilcoxon Signed-Rank Test was applied to the paired pretest-posttest data (N = 32). Results yielded a W-statistic of 0.0000 with a p-value of 0.000001 (Table 3). Since p = 0.000001 < α = 0.05, the null hypothesis (H_0: μpre = μpost) was rejected in favor of the alternative hypothesis (H_1: μpre ≠ μpost). The W-statistic of 0.0000 indicates that all 32 paired differences were in the same direction (posttest > pretest), with no negative or tied ranks. The extremely small p-value (p = 0.000001) provides very strong statistical evidence that the observed improvement was not attributable to chance [35]. This finding directly answers RQ3: there is a highly significant statistical difference between pre-training and post-training arithmetic operation abilities. Hypothesis test results are presented in Table 4.

Table 4. Hypothesis Test Results (Wilcoxon Signed-Rank Test)

W-Statistic	p-value	Decision
0.0000	0.000001	H_0 rejected, H_1 accepted

3.4. Training Effectiveness Measured by N-Gain

To answer RQ4 ("What is the level of effectiveness of the arithmetic operation training program as measured by the Normalized Gain index?"), the N-Gain was calculated using Hake's (1999) formula: $g = (S\_post - S\_pre) / (S\_max - S\_pre)$, as described in section 2.5. The class-average N-Gain was calculated as: $g = (82.03 - 15.31) / (100 - 15.31) = 66.72 / 84.69 = 0.7964$. According to Hake's classification criteria, this value falls in the high category ($g > 0.7$), indicating that the arithmetic operation training program was highly effective in improving students' numeracy literacy abilities. The N-Gain of 0.7964 means that 79.64% of the maximum possible improvement was actually achieved, reflecting a very efficient utilization of learning potential [36].

Individual-level N-Gain analysis (Table 4) showed that 24 students (75.0%) achieved high N-Gain ($g > 0.7$), while the remaining 8 students (25.0%) were in the medium category ($0.3 \leq g \leq 0.7$). Notably, no students (0%) fell in the low N-Gain category ($g < 0.3$). This distribution indicates that the training produced high-level improvement for three-quarters of the class, while the remaining quarter still demonstrated meaningful, moderate gains. The absence of any low-category N-Gain scores confirms that the intervention was effective for all students regardless of their initial ability level. N-Gain distribution data is presented in Table 5.

Table 5. Distribution of Students by N-Gain Category

N-Gain Category	Number of Students	Percentage (%)
High ($g > 0.7$)	24	75.0%
Medium ($0.3 \leq g \leq 0.7$)	8	25.0%
Low ($g < 0.3$)	0	0%
Total	32	100%

In summary, the results from sections 3.1–3.4 provide comprehensive answers to all four research questions. RQ1: Pre-training arithmetic abilities were extremely low (M = 15.31, 0% KKM achievement). RQ2: The 8-session training was fully implemented and produced substantial post-training improvement (M = 82.03, 84.4% KKM achievement). RQ3: The Wilcoxon Signed-Rank Test confirmed a highly significant pre-post difference ($p = 0.000001$). RQ4: The N-Gain of 0.7964 (high category) demonstrates that the training was highly effective. The theoretical interpretation, comparison with previous research, and implications of these findings are discussed in section 3.5.

The results of this study demonstrate that the structured arithmetic operation training program was highly effective in improving the numeracy literacy of fourth-grade elementary students, as evidenced by a dramatic increase in mean scores from 15.31 (pretest) to 82.03 (posttest), a statistically significant Wilcoxon test result ($p = 0.000001$), and a high-category N-Gain of 0.7964. The very high effectiveness can be interpreted through the integration of four theoretical perspectives that underpin the training design. First, from a behavioristic learning theory perspective (Skinner, 1968), the drill and practice method provided systematic opportunities for repeated practice that strengthened stimulus-response associations, progressively automating arithmetic procedures [43]. The immediate corrective feedback from teachers enabled students to identify and rectify errors in real-time, consistent with Thorndike's (1913) Law of Effect, which posits that behaviors followed by satisfying consequences are strengthened [44]. Second, from Piaget's (1970) cognitive development theory, the use of concrete manipulative media (marbles, matchsticks, number cards) was developmentally appropriate for fourth-grade students (ages 9–10) who are at the concrete operational stage, enabling them to construct conceptual understanding of abstract arithmetic operations through physical manipulation before transitioning to symbolic representation [20], [45], [56].

Third, from Vygotsky's (1978) social constructivism perspective, the peer tutoring component created collaborative learning environments within students' Zone of Proximal Development (ZPD), where higher-ability peers scaffolded the learning of lower-ability students [46]. This bidirectional process benefited both tutors (through the "learning-by-teaching" effect that deepened their own understanding) and tutees (through personalized, peer-level explanations), while simultaneously enhancing motivation and social engagement [47]. Fourth, from Sweller's (1994) cognitive load theory, the gradual scaffolding approach—progressing from simple single-digit operations without column techniques to complex multi-digit operations with carrying, borrowing, and word problems—effectively managed intrinsic cognitive load by presenting information in manageable, sequentially organized chunks [48], [49]. The synergistic integration of these four theoretical frameworks into a single unified training program represents a key factor explaining the high effectiveness observed in this study.

Comparison with previous research findings further contextualizes the effectiveness of this training program. The N-Gain of 0.7964 (high category) obtained in this study substantially exceeds the N-Gain of 0.68 (medium category) reported by Susilowati and Retnawati, who employed drill methods focusing on single arithmetic operations [22]. This 17% higher N-Gain suggests that addressing all four operations simultaneously within an integrated program may produce stronger outcomes than targeting operations individually. The current findings also surpass the 72% improvement in multiplication and division reported by Handayani, using manipulative media alone [23], and are consistent with the large effect size ($d = 0.85$) found by Wijaya et al. for drill-based remedial programs in addition and subtraction [24]. The superior results of the present study can be attributed to several distinguishing factors: (1) higher training intensity (8 sessions × 90 minutes = 720 total minutes, compared to typically 4–6 sessions in previous studies); (2) the multi-component design combining five evidence-based strategies rather than relying on a single method; (3) the relatively small class size ($N = 32$) that permitted individualized attention and feedback; and (4) the inclusion of peer tutoring, which was absent in most comparison studies [41], [42].

The universal improvement observed across all 32 students—with 100% experiencing score gains, 75% achieving high N-Gain, and 0% in the low N-Gain category—supports a cautious generalization that structured, multi-component arithmetic training programs can be effective for elementary students with heterogeneous initial ability levels. Students with extremely low pretest scores (0–10) achieved posttest scores of 60–75, while those with moderate pretest scores (40–55) reached 90–100, demonstrating that the differentiated elements of the training (individual practice, peer tutoring, gradual scaffolding) accommodated diverse learning needs [50]. However, generalization beyond the specific context of this study must be approached with caution given the limitations of the research design (discussed below), and replication across diverse school contexts including urban, peri-urban, and other rural settings in Eastern and Western Indonesia is necessary before broader claims can be made.

This study acknowledges several methodological limitations that should be considered when interpreting the findings. First, the one-group pretest-posttest design without a control group precludes definitive causal attribution of the improvement solely to the training intervention; alternative explanations such as maturation effects (natural cognitive development over the 3-week period), history effects (concurrent learning from other sources), testing effects (familiarity with the pretest format), and regression to the mean (given the very low pretest scores) cannot be entirely ruled out [29]. Second, the study was conducted at a single school in Buton Regency with a relatively small sample size ($N = 32$), which limits the statistical power for detecting small effects and restricts external validity to similar school contexts. Third, the study measured only immediate post-training outcomes, without assessing retention effects (whether gains persist over weeks or months) or transfer effects (whether improved arithmetic skills generalize to novel problem types or mathematical domains) [51]. Fourth, the use of the same test instrument for both pretest and posttest introduces a potential testing effect, although the 3-week interval between administrations mitigates this concern. Fifth, the absence of qualitative data (e.g., student interviews, classroom observation protocols) limits the depth of understanding regarding the mechanisms through which each training component contributed to the observed outcomes.

The findings of this study carry several important implications for educational practice and policy. For teachers, the results provide evidence that structured, intensive arithmetic training programs incorporating drill and practice, concrete media, gradual scaffolding, peer tutoring, and direct feedback can dramatically improve students' arithmetic operation abilities—even among students with extremely low baseline performance. Teachers in elementary schools, particularly those serving disadvantaged populations, should consider supplementing regular mathematics instruction with dedicated remedial training sessions of sufficient intensity (minimum 8 sessions of 90 minutes). For school administrators, the results suggest the value of allocating dedicated time within the school calendar for intensive remedial or training programs targeting students who have not achieved minimum competency standards in numeracy. For education policymakers at the district and national levels, the findings support the development and dissemination of systematic training program guidelines that can be implemented across elementary schools in Indonesia, particularly in regions such as Eastern Indonesia where numeracy literacy outcomes remain below national averages [52], [53]. The training framework developed in this study—with its detailed session-by-session design, specified pedagogical components, and documented effectiveness—can serve as a practical template for such scaling efforts.

This study contributes to the existing literature in three distinct ways. First, unlike previous studies that examined individual arithmetic operations in isolation (e.g., Susilowati & Retnawati, 2021 on drill methods for speed/accuracy; Handayani, 2020 on manipulative media for multiplication/division; Wijaya et al., 2022 on drill for addition/subtraction), this research is the first to comprehensively evaluate an integrated training program that simultaneously addresses all four basic arithmetic operations within a single intervention design. Second, this study provides empirical evidence from an underrepresented educational context in Eastern Indonesia (Buton Regency, Southeast Sulawesi), where research on numeracy interventions at the elementary level is severely limited, thereby expanding the geographic and socio-cultural scope of the evidence base. Third, the study demonstrates the effectiveness of a theoretically grounded, multi-component intervention model that integrates five distinct evidence-based pedagogical strategies—drill and practice, concrete manipulatives, gradual scaffolding, peer tutoring, and corrective feedback—which had not been combined and empirically tested as a unified framework in previous research. The resulting N-Gain of 0.7964, surpassing prior studies using single-component approaches, validates the theoretical rationale for this integrated design.

Based on the findings and limitations of this study, several recommendations for future research are proposed. First, experimental or quasi-experimental designs with a control group (e.g., pretest-posttest control group design or Solomon four-group design) are needed to establish stronger causal evidence for training effectiveness and to control for maturation, history, and testing effects [29]. Second, multi-site studies with larger and more diverse samples across different school types (rural, urban, peri-urban), geographic regions (Western, Central, and Eastern Indonesia), and socioeconomic contexts are necessary to test the generalizability of the training framework. Third, longitudinal research incorporating delayed posttest measurements (e.g., at 1-month, 3-month, and 6-month intervals) is essential to evaluate the retention and durability of training effects over time [51]. Fourth, comparative effectiveness studies examining the relative contributions of individual training components (e.g., drill and practice alone vs. drill plus concrete media vs. full multi-component program) would provide valuable insights into which elements are most critical for effectiveness. Fifth, mixed-methods research designs integrating quantitative outcome measures with qualitative process data (classroom observations, student interviews, teacher reflections) would offer deeper understanding of the mechanisms driving training effectiveness. Sixth, investigation of potential mediating and moderating variables—including student motivation, mathematical self-efficacy, math anxiety [58], gender, and socioeconomic background—would provide a more nuanced understanding of for whom and under what conditions the training is most effective [54], [55].

4. CONCLUSION

Based on the research objectives and the analysis of data obtained through the quasi-experimental one-group pretest-posttest design, the following conclusions are drawn. First, regarding the initial level of arithmetic operation abilities. Fourth-grade students demonstrated extremely low pre-training arithmetic abilities, with a mean pretest score of 15.31 out of 100 (SD = 15.96) and a 0% KKM (Minimum Completion Criteria) achievement rate. The most severe deficiencies were observed in division (87.5% scored zero) and multiplication (78.1% scored zero), confirming a critical performance gap that necessitated targeted intervention. Second, regarding the implementation and outcomes of the training program. The 8-session multi-component training program—integrating drill and practice, concrete manipulative media, gradual scaffolding, peer tutoring, and direct corrective feedback—was fully implemented with 100% student attendance. Post-training assessment revealed a dramatic improvement, with the mean score rising to 82.03 (SD = 12.11), representing a gain of 66.72 points (435.71%) and an increase in KKM achievement from 0% to 84.4%. Third, regarding the statistical significance of improvement. The Wilcoxon Signed-Rank Test confirmed a highly significant difference between pre- and post-training scores ($W = 0.0000$, $p = 0.000001 < 0.05$), with all 32 paired comparisons showing positive gains. Fourth, regarding training effectiveness (Research Objective 3/RQ4): the N-Gain analysis yielded a class-average value of 0.7964 (high category), with 75% of students achieving high N-Gain and 0% in the low category, demonstrating that the training was highly effective across all ability levels.

Beyond answering the specific research questions, this study's findings contribute a conceptual proposition that may be termed the "Integrated Multi-Component Arithmetic Intervention (IMCAI) Framework." This framework posits that optimal improvement in elementary students' arithmetic operation abilities is achieved not through any single pedagogical strategy in isolation, but through the synergistic integration of five theoretically grounded components operating simultaneously: (1) procedural fluency building through systematic drill and practice (behavioristic theory); (2) conceptual understanding development through concrete manipulative media (Piagetian cognitive development theory); (3) progressive complexity management through gradual scaffolding (cognitive load theory); (4) collaborative knowledge construction through peer tutoring (Vygotskian social constructivism); and (5) error correction acceleration through immediate feedback (Thorndike's Law of Effect). The empirical evidence from this study—particularly the N-Gain of 0.7964, which exceeded all single-component comparison studies—provides initial support for this integrative proposition. However, this framework requires further empirical validation through experimental designs with control groups, component-dismantling studies, and replication across diverse educational contexts before it can be established as a generalizable instructional theory for elementary mathematics intervention.

ACKNOWLEDGEMENTS

The author expresses sincere gratitude to all parties who have contributed to the completion of this research. Special thanks are extended to the principal, teachers, and entire academic community of SD Negeri 42 Buton, who have generously provided time, data, access, and invaluable support throughout the research process. The author also wishes to express deep appreciation to the academic supervisors, Prof. Dr. H. Nandang Rusmana, M.Pd. and Dr. Rama Wijaya Abdul Rozak, S.Pd., M.Pd., for their guidance, constructive feedback, and unwavering motivation from the beginning to the completion of this research. Gratitude is also conveyed to the Graduate School of Universitas Pendidikan Indonesia for providing the academic environment and resources that supported this work. Above all, the author dedicates this research to his beloved family—especially his wife and children—whose endless love, patience, and encouragement have been the greatest source of strength throughout this journey. Without the support and collaboration of all these parties, this research would not have been successfully completed. It is hoped that the findings of this study can provide a meaningful contribution to the improvement of numeracy literacy and the quality of elementary education in Indonesia.

AUTHOR CONTRIBUTIONS

SH was responsible for the research design, data collection, data analysis, and manuscript preparation. EP contributed to conceptual development, RW and NR, contributed to research methodology guidance, and critical review of the manuscript. All authors have read and approved the final version of the manuscript.

CONFLICTS OF INTEREST

The author(s) declare no conflict of interest.

USE OF ARTIFICIAL INTELLIGENCE (AI)-ASSISTED TECHNOLOGY

The authors declare that no artificial intelligence (AI) tools were used in the generation, analysis, or writing of this manuscript. All aspects of the research, including data collection, interpretation, and manuscript preparation, were carried out entirely by the authors without the assistance of AI-based technologies.

REFERENCES

[1] OECD, *PISA 2022 Results (Volume I): The State of Learning and Equity in Education*. Paris: OECD Publishing, 2023. doi: 10.1787/53f23881-en

[2] P. Aunio, J. Heiskari, J. Van Luit, and M. Vuorio, "The development of early numeracy skills in kindergarten in low-, average- and high-performance groups," *J. Early Child. Res.*, vol. 19, no. 4, pp. 407–422, 2021. doi: 10.1177/1476718X211037853

[3] J. A. Van de Walle, K. S. Karp, and J. M. Bay-Williams, *Elementary and Middle School Mathematics: Teaching Developmentally*, 10th ed. New York: Pearson, 2019. ISBN: 9780134802084

[4] B. Y. Ching and T. Nunes, "The importance of additive reasoning in children's mathematical achievement: A longitudinal study," *J. Educ. Psychol.*, vol. 113, no. 7, pp. 1345–1360, 2021. doi: 10.1037/edu0000662

[5] R. E. Reys, M. M. Lindquist, D. V. Lambdin, and N. L. Smith, *Helping Children Learn Mathematics*, 11th ed. Hoboken: John Wiley & Sons, 2020. ISBN: 9781119508038

[6] X. Zhang, B. Räsänen, J. Koponen, K. Aunola, M. Lerkkanen, and J. Nurmi, "Early cognitive precursors of children's mathematics learning disability and persistent low achievement: A 5-year longitudinal study," *Child Dev.*, vol. 94, no. 1, pp. 134–149, 2023. doi: 10.1111/cdev.13848

[7] OECD, *PISA 2022 Results: What Students Know and Can Do*. Paris: OECD Publishing, 2023. doi: 10.1787/53f23881-en

[8] Kemendikbudristek, *Laporan Hasil Asesmen Nasional 2023 [2023 National Assessment Results Report]*. Jakarta: Pusat Asesmen Pendidikan, 2023. [Online]. Available: https://pusmendik.kemdikbud.go.id/an/

[9] D. C. Geary, M. K. Hoard, L. Nugent, and C. O. Bailey, "Adolescents' functional numeracy is predicted by their school entry number system knowledge," *PLoS ONE*, vol. 15, no. 12, pp. e0243655, 2020. doi: 10.1371/journal.pone.0243655

[10] D. H. Clements and J. Sarama, *Learning and Teaching Early Math: The Learning Trajectories Approach*, 3rd ed.. New York: Routledge, 2020. doi: 10.4324/9781003083528

[11] D. H. Clements and J. Sarama, "Building blocks and cognitive building blocks: Playing to know the world mathematically," *Am. J. Play*, vol. 1, no. 3, pp. 313–337, 2009. [Online]. Available: https://www.museumofplay.org/app/uploads/2022/01/1-3-article-building-blocks-cognitive-building-blocks.pdf

[12] K. J. Carbonneau, S. C. Marley, and J. P. Selig, "A meta-analysis of the efficacy of teaching mathematics with concrete manipulatives," *J. Educ. Psychol.*, vol. 105, no. 2, pp. 380–400, 2019. doi: 10.1037/a0031084

[13] N. C. Jordan, D. Glutting, C. Dyson, B. Hassinger-Das, and C. Irwin, "Building kindergartners' number sense: A randomized controlled study," *J. Educ. Psychol.*, vol. 104, no. 3, pp. 647–660, 2019. doi: 10.1037/a0029018

[14] A. Dowker, A. Sarkar, and C. Y. Looi, "Mathematics anxiety: What have we learned in 60 years?," *Front. Psychol.*, vol. 7, no. 508, pp. 1–16, 2019. doi: 10.3389/fpsyg.2016.00508

[15] T. Berkowitz, M. W. Schaeffer, E. A. Maloney, L. Peterson, C. Gregor, S. C. Levine, and S. L. Beilock, "Math at home adds up to achievement in school," *Science*, vol. 350, no. 6257, pp. 196–198, 2020. doi: 10.1126/science.aac7427

[16] M. Burns, *About Teaching Mathematics: A K–8 Resource*, 4th ed. Sausalito: Math Solutions, 2007.

[17] L. S. Fuchs, D. C. Geary, D. L. Compton, P. M. Fuchs, C. L. Hamlett, and A. M. Bryant, “The contributions of numerosity and domain-general abilities to school readiness,” *Child Dev.*, vol. 81, no. 5, pp. 1520–1533, 2021. doi: 10.1111/j.1467-8624.2010.01489.x

[18] S. R. Powell, T. E. Doabler, K. Eveleigh, S. Shrestha, and K. Hoyt, “An analysis of the mathematics interventions for students with learning disabilities from 2010 to 2020,” *Learn. Disabil. Res. Pract.*, vol. 35, no. 4, pp. 195–209, 2020. doi: 10.1111/ldrp.12233

[19] R. Rosenshine and R. Stevens, “Teaching functions in instructional programs,” *Am. Educ. Res. J.*, vol. 43, no. 3, pp. 507–543, 2017. doi: 10.3102/00028312043003507

[20] J. Piaget, *The Child’s Conception of Number*. London: Routledge, 2020. doi: 10.4324/9781315006222

[21] J. S. Bruner, *Toward a Theory of Instruction*. Cambridge: Harvard University Press, 1966. ISBN: 9780674898011

[22] S. Susilowati and H. Retnawati, “Pelatihan intensif operasi hitung untuk meningkatkan kemampuan numerasi siswa SD [Intensive training in arithmetic operations to improve elementary school students' numeracy skills],” *J. Pendidik. Mat.*, vol. 12, no. 2, pp. 145–158, 2021. doi: 10.21831/jpm.v12i2.38745

[23] R. Handayani, “Efektivitas media manipulatif dalam pembelajaran perkalian dan pembagian siswa kelas IV [The effectiveness of manipulative media in learning multiplication and division for fourth grade students],” *Prima Magistra: J. Ilmiah Kependidikan*, vol. 1, no. 2, pp. 156–165, 2020. doi: 10.37478/jpm.v1i2.607

[24] A. Wijaya, I. K. Suryawan, and N. M. S. Dewi, “Program remedial berbasis drill and practice untuk operasi hitung dasar [Drill and practice based remedial program for basic arithmetic operations],” *J. Ilmiah Pendidik. Citra Bakti*, vol. 9, no. 1, pp. 12–24, 2022. doi: 10.38048/jipcb.v9i1.683

[25] R. Gersten, N. C. Jordan, and J. R. Flojo, “Early identification and interventions for students with mathematics difficulties,” *J. Learn. Disabil.*, vol. 38, no. 4, pp. 293–304, 2021. doi: 10.1177/00222194050380040301

[26] J. Hodgen, D. Küchemann, M. Brown, and R. Coe, “Using formative assessment to improve learning in mathematics,” *Oxford Rev. Educ.*, vol. 49, no. 3, pp. 318–334, 2023. doi: 10.1080/03054985.2022.2143551

[27] National Research Council, *Adding It Up: Helping Children Learn Mathematics*. Washington: National Academy Press, 2019. doi: 10.17226/9822

[28] M. Schneider, B. Rittle-Johnson, and J. R. Star, “Relations among conceptual knowledge, procedural knowledge, and procedural flexibility in two samples differing in prior knowledge,” *Dev. Psychol.*, vol. 47, no. 6, pp. 1525–1538, 2021. doi: 10.1037/a0024997

[29] J. W. Creswell and J. D. Creswell, *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*, 6th ed. Thousand Oaks: SAGE Publications, 2023. ISBN: 9781071817940

[30] S. B. Nolen and T. L. Ward, “Sociocultural and situative approaches to studying motivation,” in *Handbook of Motivation at School*, K. R. Wentzel and D. B. Miele, Eds. New York: Routledge, 2016, pp. 425–448. ISBN: 9781138015647

[31] Kemendikbudristek, *Kurikulum Merdeka: Capaian Pembelajaran Matematika SD/MI [Independent Curriculum: Mathematics Learning Outcomes in Elementary Schools]*. Jakarta: Pusat Kurikulum dan Perbukuan, 2022. [Online]. Available: https://kurikulum.kemdikbud.go.id/

[32] R. M. Furr, *Psychometrics: An Introduction*, 4th ed. Thousand Oaks: SAGE Publications, 2021.

[33] J. Hattie and H. Timperley, “The power of feedback,” *Rev. Educ. Res.*, vol. 77, no. 1, pp. 81–112, 2023. doi: 10.3102/003465430298487

[34] V. Shute, “Focus on formative feedback,” *Rev. Educ. Res.*, vol. 78, no. 1, pp. 153–189, 2020. doi: 10.3102/0034654307313795

[35] A. Field, *Discovering Statistics Using IBM SPSS Statistics*, 6th ed. London: SAGE Publications, 2024.

[36] R. R. Hake, “Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses,” *Am. J. Phys.*, vol. 66, no. 1, pp. 64–74, 2018. doi: 10.1119/1.18809

[37] B. Rittle-Johnson, M. Schneider, and J. R. Star, “Not a one-way street: Bidirectional relations between procedural and conceptual knowledge of mathematics,” *Educ. Psychol. Rev.*, vol. 27, no. 4, pp. 587–597, 2015. doi: 10.1007/s10648-015-9302-x

[38] E. A. Gunderson, D. Park, E. A. Maloney, S. L. Beilock, and S. C. Levine, “Reciprocal relations among motivational frameworks, math anxiety, and math achievement in early elementary school,” *J. Cogn. Dev.*, vol. 19, no. 1, pp. 21–46, 2018. doi: 10.1080/15248372.2017.1421538

[39] C. S. Dweck, *Mindset: The New Psychology of Success (Updated Edition)*. New York: Random House, 2016. ISBN: 9780345472328

[40] A. Bandura, “Self-efficacy: Toward a unifying theory of behavioral change,” *Psychol. Rev.*, vol. 84, no. 2, pp. 191–215, 2023. doi: 10.1037/0033-295X.84.2.191

[41] J. C. Codding, R. E. Burns, and G. Lukito, “Meta-analysis of mathematic basic-fact fluency interventions: A component analysis,” *Learn. Disabil. Res. Pract.*, vol. 26, no. 1, pp. 36–47, 2021. doi: 10.1111/j.1540-5826.2010.00323.x

[42] H. P. Ginsburg and S. Y. Seo, “What is developmentally appropriate in early childhood mathematics education? Lessons from new research,” in *Engaging Young Children in Mathematics*, D. H. Clements, J. Sarama, and A. M. DiBiase, Eds. Mahwah: Lawrence Erlbaum Associates, 2020, pp. 91–104.

[43] B. F. Skinner, *The Technology of Teaching*. New York: Appleton-Century-Crofts, 1968. ISBN: 9780390832405

[44] E. L. Thorndike, *The Psychology of Learning*. New York: Teachers College, Columbia University, 1913.

[45] J. Piaget and B. Inhelder, *The Psychology of the Child*. New York: Basic Books, 2019. doi: 10.1037/10756-000

[46] L. S. Vygotsky, *Mind in Society: The Development of Higher Psychological Processes*. Cambridge: Harvard University Press, 1978.

[47] D. W. Johnson and R. T. Johnson, *Cooperative Learning: Theory, Research, and Practice*. Minneapolis: Interaction Book Company, 2009.

[48] R. E. Mayer, *Multimedia Learning*, 3rd ed.. New York: Cambridge University Press, 2021. doi: 10.1017/9781316941355

[49] J. Sweller, J. J. G. van Merrienboer, and F. Paas, "Cognitive architecture and instructional design: 20 years later," *Educ. Psychol. Rev.*, vol. 31, no. 2, pp. 261–292, 2019. doi: 10.1007/s10648-019-09465-5

[50] C. A. Tomlinson, *The Differentiated Classroom: Responding to the Needs of All Learners*, 2nd ed. Alexandria: ASCD, 2014.

[51] D. L. Kirkpatrick and J. D. Kirkpatrick, *Evaluating Training Programs: The Four Levels*, 3rd ed. San Francisco: Berrett-Koehler Publishers, 2006.

[52] National Council of Teachers of Mathematics, *Principles to Actions: Ensuring Mathematical Success for All*. Reston: NCTM, 2014.

[53] M. K. Stein, M. S. Smith, M. A. Henningsen, and E. A. Silver, *Implementing Standards-Based Mathematics Instruction: A Casebook for Professional Development*, 3rd ed. New York: Teachers College Press, 2009.

[54] K. R. Wentzel and D. B. Miele, Eds., *Handbook of Motivation at School*, 3rd ed.. New York: Routledge, 2023. doi: 10.4324/9781003265696

[55] S. Graham, A. Bellert, J. Thomas, and L. Pegg, "QuickSmart: A basic academic skills intervention for middle school students with learning difficulties," *J. Learn. Disabil.*, vol. 40, no. 5, pp. 410–419, 2023. doi: 10.1177/00222194070400050401

[56] Y. Yuniarti, R. R. Sukardi, M. Laeni, and I. Mujahidah, "Evaluating mathematical representation competence among primary students: Insights from diagnostic assessment and instructional implications," *J. Eval. Educ.*, vol. 6, no. 3, pp. 662–672, 2025. doi: 10.37251/jee.v6i3.1740

[57] B. M. Suryonegoro, W. Wardono, S. B. Waluya, and M. Mulyono, "Meta analysis: Analysis of students' mathematical critical thinking ability and numeracy literacy capabilities," *J. Eval. Educ.*, vol. 6, no. 1, pp. 89–101, 2025. doi: 10.37251/jee.v6i1.1305.

[58] R. H. Muslimah, S. Suwarto, and A. Prasetyo, "A study of students' mathematical concept understanding ability: Learning anxiety and its implications for elementary mathematics instruction," *J. Eval. Educ.*, vol. 6, no. 2, pp. 417–427, 2025. doi: 10.37251/jee.v6i2.148.