Attractive : Innovative Education Journal
Vol. 8 No. 4, December 2026

Gradual, Not Instant: Learning Trajectories and Heterogeneity in Multiplication Learning with GASING and Congklak

Andika Monalisa [1*], Yunus Abidin [1], Yeni Yuniarti [1]
[1] *Universitas Pendidikan Indonesia, Indonesia*

andikamonalisa@upi.edu*

ABSTRACT

Evaluations of elementary mathematics instruction usually ask only whether scores rise, without examining the shape of the gain, how evenly it is distributed, and how meaningful it is. This study analyzes the learning trajectory, heterogeneity dynamics, and practical significance of multiplication concept mastery in GASING (Easy, Fun, and Enjoyable) instruction using the traditional game congklak. A single-group quasi-experimental repeated-measures design (one baseline and two intervention-phase measurements) involved 132 third-grade students from six elementary schools in Banyuwangi Regency, Indonesia. Essay-test data were analyzed using descriptive statistics, interval gains, the coefficient of variation, effect sizes, normalized gain, segmented regression, and sensitivity analysis. Mean scores rose gradually from 21.34 to 22.52 and 23.95, with the second-interval gain (1.43 points) slightly exceeding the first (1.18 points). The standard deviation widened (18.93 to 20.46) while the coefficient of variation decreased (88.7% to 85.4%), indicating that relative inequality did not worsen although heterogeneity remained very high. The overall effect was small (Cohen's d = 0.13), normalized gain was low (0.033), and the between-time test was not significant (F = 0.573; p = 0.564); sensitivity analysis showed that the overall gain becomes significant only when the test–retest correlation exceeds 0.71. The findings suggest that GASING with congklak operates as gradual reinforcement rather than an immediate jump, and they underline the need for differentiated support for low-achieving students and for evaluations with more measurement points.

ARTICLE INFO
Article history:
Received
August 06, 2026
Revised
October 27, 2026
Accepted
November 18, 2026

Keywords: GASING, Congklak, Learning Trajectory, Ability Heterogeneity, Multiplication

Journal Homepage https://attractivejournal.com/index.php/aj/
This is an open access article under the CC BY SA license
https://creativecommons.org/licenses/by-sa/4.0/
@ 2026 by the authors
Published by CV. Creative Tugu Pena

INTRODUCTION

Multiplication marks students' transition from additive to multiplicative thinking, so the quality of their understanding in Grade 3 has long-term consequences for division, fractions, ratio, and word-problem solving. This understanding does not form in a single leap; it grows through a series of stages, from one-by-one counting, repeated addition, and equal grouping to relational reasoning based on arrays and the properties of operations (Park & Nunes, 2001; Huang, 2014). The development of multiplicative concepts is therefore better viewed as a learning trajectory that unfolds gradually than as a momentary change in scores (Simon, 1995; Clements & Sarama, 2004). In the Indonesian context, a learning trajectory for multiplication that begins from real-world contexts has been shown to help rural students construct the meaning of the operation (Hendriana et al., 2019).

In many classrooms, multiplication is still taught with an emphasis on memorizing tables and answering quickly. This approach produces students who can state products but struggle to explain their meaning, choose representations, or apply them in new contexts (Mills, 2019). Conceptual and procedural knowledge in fact reinforce each other, so procedural practice

without meaning tends to yield fragile understanding (Rittle-Johnson et al., 2015). Research on array representations shows that visual models help students organize the spatial structure of multiplication and strengthen their reasoning (Barmby et al., 2009; Shanty & Wijaya, 2012). The concrete–representational–abstract sequence also improves both skills and understanding of multiplication, especially when students are asked to explain their strategies (Camp, 2018; Flores & Hinton, 2019; Milton et al., 2019), and a meta-analysis indicates that concrete manipulatives yield positive effects when their use is explicitly directed toward concepts (Carbonneau et al., 2013).

The GASING Method (from the Indonesian Gampang, Asyik, dan Menyenangkan, or Easy, Fun, and Enjoyable) is an approach developed in Indonesia on similar principles: concepts are built from concrete experience, mastered through critical points, and then abstracted into symbols and mental calculation (Kusuma & Sulistiawati, 2014; Siregar et al., 2014). Several studies report contributions of GASING to multiplication learning outcomes (Armianti et al., 2016), conceptual understanding of two-digit multiplication (Maulida et al., 2024), problem-solving ability (Hayati et al., 2024), and arithmetic fluency (Sihombing et al., 2026). Recent studies have also combined GASING with games, for example a dice game (Yusuf et al., 2026) and the traditional game mong-mongan (Sariningsih et al., 2025). Congklak is a strong candidate for a similar combination because it naturally involves repeatedly distributing seeds into holes in equal amounts. Ethnomathematical studies show that congklak embodies arithmetic, addition, modular concepts, and the ideas of multiplication, division, and the greatest common divisor, which can serve as a bridge to mathematical symbols (Khasanah et al., 2023; Mulyadi & Putra, 2025; Ninggar et al., 2024; Tusolihah et al., 2022). Recent work points in the same direction: explicitly verbalizing the structure of equal groups is associated with stronger conceptual understanding of multiplication (Hankeln & Prediger, 2025), and a mancala-type game can lead young children from one-by-one counting toward treating groups of counters as a unit (Thiel, 2025).

Nevertheless, the evidence for this kind of approach still leaves three gaps. First, a systematic review of GASING research notes a predominance of small samples, the absence of control groups, and short-term evaluation (Gesiradja et al., 2025), while most studies use only momentary quantitative indicators (Pratiwi et al., 2026). A two-point (pretest–posttest) design shows only the final difference, so the shape of the developmental trajectory remains invisible. Repeated measurement within an interrupted time series (ITS) framework offers a richer picture of changes in level and trend (Hallberg et al., 2018; Lopez Bernal et al., 2017; Wagner et al., 2002). Second, effectiveness reports rarely examine whether gains in the mean are shared evenly, even though the effects of concrete media may differ according to students' initial ability (Ahmad & Siller, 2024), and population data show that students with different starting points grow at different rates (Larsen & Little, 2023). Third, many studies report statistical significance without effect sizes and without assessing how much score variation can be explained, although both pieces of information determine whether an intervention is suitable for scaling up (Kraft, 2020; Lakens, 2013).

This study addresses these three gaps by examining repeated-measures data from GASING instruction using congklak in terms of trajectory, evenness of gains, and practical significance. Its novelty lies in combining an analysis of interval gain rates, the coefficient of variation as an indicator of relative inequality, effect sizes interpreted against both general and education-intervention benchmarks, and a sensitivity analysis of the assumed correlation between measurements. The focal variable is multiplication concept mastery, measured with the same essay test at three time points in 132 third-grade students from six schools. Because a single-group design precludes causal inference, the analysis is exploratory and does not test a directional hypothesis about effectiveness. On this basis, the research questions (RQs) are as follows.

1. What is the shape of the developmental trajectory of third-grade students' multiplication concept mastery across three measurement points in GASING instruction using congklak, in terms of the magnitude and rate of gain between intervals?

2. How do the dynamics of students' ability heterogeneity unfold during instruction; that is, does the gap between students narrow, remain stable, or widen?
3. How large is the practical significance of the change in multiplication concept mastery, and how strong is that change statistically under various analytic assumptions?

Theoretically, this study extends the evidence on culture-based multiplication learning by positioning conceptual development as a trajectory. Practically, the findings on heterogeneity and effect sizes are expected to help teachers and policymakers set realistic expectations and design differentiated support.

METHOD

This study used a quantitative approach with a single-group quasi-experimental repeated-measures design (Shadish et al., 2002). Multiplication concept mastery was measured at three points: T1 as the baseline before the intervention, and T2 and T3 during the intervention. The interrupted time series (ITS) framework was used as a complementary analytic tool for reading changes in level and rate, not as a basis for causal claims, because the baseline phase contained only one measurement point (Linden, 2015). This design was chosen because instruction took place in intact classes in which randomization was not possible.

The data in this study are part of the authors' research project on the implementation of GASING with congklak in multiplication instruction. This article specifically addresses the research questions on learning trajectory, ability heterogeneity, and practical significance.

The study was conducted from 24 November 2025 to 4 February 2026 in six public elementary schools in Banyuwangi Regency. The participants were 132 third-grade students selected through cluster sampling, with intact classes as the sampling unit. All students in the selected classes were included because they took part in the instruction and in all three measurements. The distribution of participants is presented in Table 1.

Table 1. Distribution of Research Participants

No	School	Third-Grade Students (n)	Proportion (%)
1	SDN 2 Pakis	25	18.9
2	SDN Bakungan	20	15.2
3	SDN 4 Ketapang	19	14.4
4	SDN 2 Alasrejo	25	18.9
5	SDN 1 Kalibaru Wetan	32	24.2
6	SDN 5 Kaligung	11	8.3
	Total	**132**	**100.0**

This study obtained permission from the Banyuwangi District Education Office and written consent from students' parents or guardians before data collection. Student identities were anonymized in data processing and reporting.

The research procedure is presented in Figure 1. Multiplication instruction followed the steps of the GASING Method from concrete to abstract, with congklak as the main medium. Students played congklak in pairs or groups, and the teacher then directed their attention to the number of holes and the equal number of seeds placed in each hole, for example, four holes with three seeds each as a representation of 4 × 3. This experience was then translated into semi-concrete representations (pictures, tables, arrays) and finally into mathematical sentences. The intervention comprised six meetings of 35 minutes each, with T2 administered at the fifth meeting and T3 at the sixth. Implementation was guided by the same teaching module and student worksheets in all six schools.

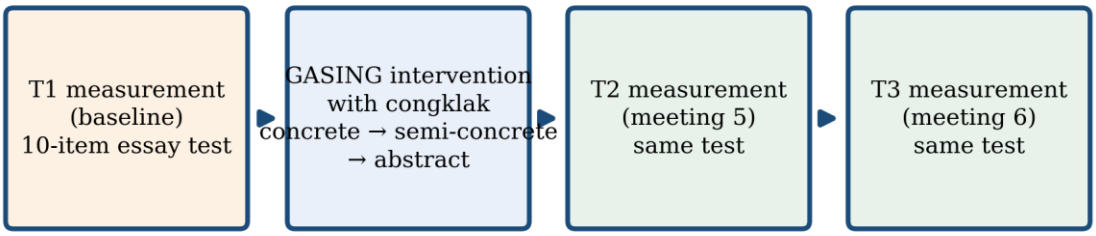


Figure 1. Research Procedure

The main instrument was a 10-item essay test of multiplication concept mastery. The test measured five indicators: (1) the meaning of multiplication as repeated addition, equal groups, or arrays; (2) the use of representations; (3) the selection and explanation of strategies; (4) the use of the properties of multiplication; and (5) the solution of word problems. Each item was scored from 0 to 3 using an analytic rubric (0 = no answer or irrelevant answer; 1 = incorrect concept; 2 = correct concept but incomplete explanation; 3 = complete concept, steps, and explanation). Raw scores of 0–30 were converted to a 0–100 scale for analysis. Content validity was assessed by two experts using Aiken's V index.

The analysis was organized around the three RQs, as summarized in Table 2.

Table 2. Research Questions, Analytic Indicators, and Interpretation Criteria

Research Question	Analytic Indicators	Interpretation Criteria
RQ1: Developmental trajectory	Mean and 95% confidence interval at each measurement point; absolute and relative gain between intervals; post-intervention change in level and rate from segmented regression	Second-interval gain > first-interval gain = strengthening trajectory; equal = linear; smaller = decelerating
RQ2: Ability heterogeneity	Standard deviation (SD), variance, T3/T1 variance ratio, coefficient of variation CV = (SD/mean) × 100%	CV decreases = relative inequality narrows; CV > 30% = high heterogeneity
RQ3: Practical significance and statistical strength	Cohen's *d*; Hake's normalized gain (N-gain); between-time test (*F*); sensitivity analysis of *dz* and paired *t*-test for *r* = 0.50–0.90; Durbin–Watson	*d*: Cohen's benchmarks (0.2 small; 0.5 medium; 0.8 large) and Kraft's benchmarks (< 0.05 small; 0.05–0.20 medium; ≥ 0.20 large); N-gain < 0.3 low; $p < 0.05$

Effect sizes between measurement points were computed with Equation (1) using the mean SD of the two points (Lakens, 2013), whereas N-gain was computed with Equation (2) on a 0–100 scale (Hake, 1998).

$$d = (\bar{X}_2 - \bar{X}_1) / [(SD_1 + SD_2)/2] \quad (1)$$

$$g = (\bar{X}final - \bar{X}initial) / (100 - \bar{X}initial) \quad (2)$$

Because the three measurements were taken from the same students, the statistical power for detecting score change depends on the correlation between measurements (*r*). To assess the robustness of the conclusions to this assumption, a sensitivity analysis was conducted by computing the standard deviation of the difference, the paired effect size *dz*, and the paired *t* statistic for *r* = 0.50 to 0.90 using Equations (3) and (4).

$$SDdiff = \sqrt{(SD_1^2 + SD_3^2 - 2r \cdot SD_1 \cdot SD_3)} \quad (3)$$

$$dz = (\bar{X}_3 - \bar{X}_1) / SDdiff\,; \quad t(n-1) = dz \cdot \sqrt{n} \quad (4)$$

As a complement, the segmented regression model is specified in Equation (5), where β1 is the baseline trend, β2 is the change in level when the intervention begins, and β3 is the change in trend afterward (Penfold & Zhang, 2013; Wagner et al., 2002). The post-intervention rate was computed as β1 + β3, and residual autocorrelation was checked with the Durbin–Watson statistic.

$$Y_t = \beta_0 + \beta_1 \cdot Time_t + \beta_2 \cdot Intervention_t + \beta_3 \cdot TimeAfter_t + \varepsilon_t \quad (5)$$

RESULT AND DISCUSSION

For RQ1, mean scores rose at every measurement point, from 21.34 at T1 to 22.52 at T2 and 23.95 at T3 (Table 3). The second-interval gain (1.43 points) was larger than the first (1.18 points), so the trajectory not only rose but also strengthened slightly. Cumulatively, the T3 mean was 2.61 points, or 12.2%, above baseline. The 95% confidence intervals of the three measurement points still overlapped (Figure 2), so the increase should be read as a tendency rather than a clear-cut difference.

Table 3. Trajectory of Mean Multiplication Concept Mastery Scores (0–100 Scale)

Time Point	Phase	N	Mean	SD	95% Confidence Interval	Gain from Previous Point	Cumulative Gain from T1
T1	Baseline	132	21.34	18.93	18.08–24.60	–	–
T2	Intervention	132	22.52	20.18	19.05–25.99	+1.18 (5.5%)	+1.18 (5.5%)
T3	Intervention	132	23.95	20.46	20.43–27.47	+1.43 (6.3%)	+2.61 (12.2%)

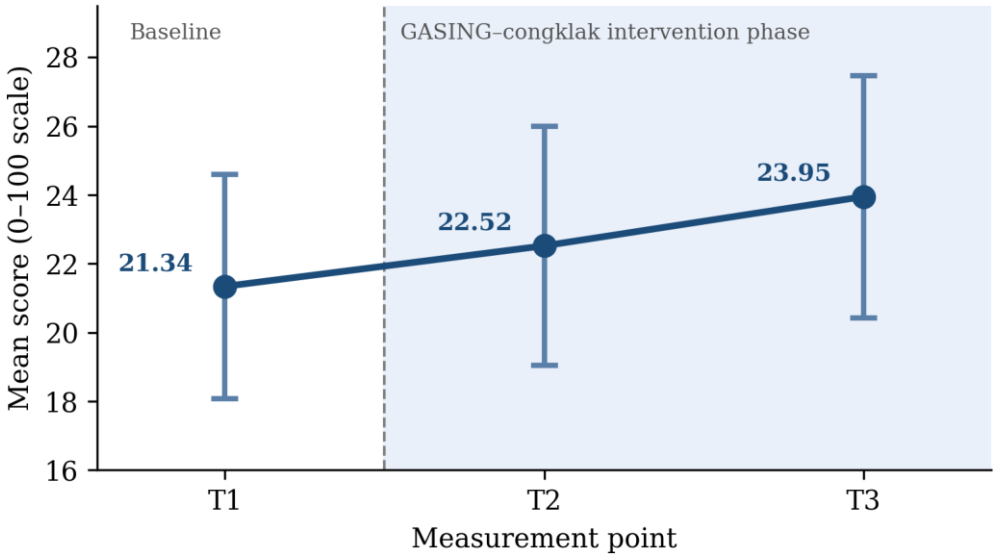


Figure 2. Mean Multiplication Concept Mastery Scores with 95% Confidence Intervals

Segmented regression gave a picture consistent with the descriptive pattern. The level-change coefficient was close to zero ($B = -0.250$; $p = 0.953$), meaning there was no jump in scores at the moment congklak began to be used. The post-intervention rate of 7.591 + (−6.159) = 1.432 points per measurement point equals the empirical T2→T3 gain. Thus, change appeared as a gradual increase across measurement occasions, not as an initial jump.

These results indicate gradual strengthening rather than a jump. The absence of a level change and the slightly strengthening gain from the first to the second interval indicate that GASING with congklak tends to work through the accumulation of experience. This pattern accords with the view that multiplication concepts develop through mutually building stages, from repeated addition to equal groups and finally to array reasoning (Clements & Sarama, 2004; Hendriana et al., 2019; Park & Nunes, 2001). At the start of the intervention, students still

had to learn the rules of the game and connect the distribution of seeds to the idea of equal groups; only afterward could concrete experience be translated into representations and symbols. This finding is consistent with concrete–representational–abstract studies that treat the concrete stage as a foundation whose benefits become visible only after students practice transferring meaning to the representational and abstract stages (Flores & Hinton, 2019; Milton et al., 2019). The implication is that evaluations conducted too early risk concluding that the approach is ineffective.

The magnitude of gain in this study was much smaller than that reported in other GASING studies, for example, an N-gain of 0.59 for whole-number multiplication (Armianti et al., 2016), 0.77 for GASING combined with a dice game (Yusuf et al., 2026), and 0.77 for a 15-day GASING program (Sihombing et al., 2026). Three factors may explain this difference. First, the instrument in this study required conceptual explanation and representation rather than mere computational accuracy, making scores harder to raise. Second, the sample comprised six schools with very high heterogeneity rather than a single class. Third, two measurement points in the intervention phase may not have captured the full effect of the GASING stages, which end in mental calculation.

For RQ2, score dispersion was very wide at all measurement points, with the standard deviation nearly equal to the mean (Table 4). The absolute standard deviation widened from 18.93 to 20.46, and the T3 variance was 16.8% larger than the T1 variance (ratio = 1.168). However, the coefficient of variation (CV) fell from 88.7% at T1 to 85.4% at T3, after rising slightly at T2.

Table 4. Indicators of Ability Heterogeneity at Each Measurement Point

Time Point	Mean	SD	Variance	Coefficient of Variation (%)	Heterogeneity Category
T1	21.34	18.93	358.34	88.71	Very high
T2	22.52	20.18	407.23	89.61	Very high
T3	23.95	20.46	418.61	85.43	Very high

These two indicators need to be read together. The widening absolute standard deviation shows that the score distance between high- and low-achieving students increased in points. The falling CV shows that, relative to the rise in the mean, inequality did not worsen and began to narrow slightly at T3. A low mean combined with a large SD also indicates a right-skewed distribution: most students were still clustered at low scores, while a small number obtained much higher scores.

Although the mean rose, the gap was not closed. A coefficient of variation above 85% at all measurement points shows that the participating classes were highly diverse in their readiness for multiplication concepts. The widening absolute standard deviation signals that students who already had a stronger additive foundation probably benefited more quickly from congklak activities, because they could more easily see the equal-group structure behind the game. Ahmad and Siller (2024) showed that the benefits of manipulatives can indeed differ according to students' initial achievement level. Students who still count one by one need explicit modeling and probing questions so that the game does not remain merely an entertaining activity (Carbonneau et al., 2013; Mills, 2019). The decrease in CV at T3 offers an optimistic note that relative inequality did not worsen, but these data do not yet show real equalization. GASING with congklak should therefore be accompanied by flexible grouping, tiered tasks, and special assistance for students with low initial scores.

For RQ3, all changes between measurement points were positive, but their magnitude was small relative to score dispersion (Table 5). The T1→T3 effect size of d = 0.13 is small by Cohen's (1988) benchmark but medium by Kraft's (2020) benchmark for education interventions. Comparing the baseline with the mean of the intervention phase yielded d = 0.10 with a pooled SD of 19.85. All N-gain values were in the low category (0.015–0.033).

Table 5. Effect Sizes and N-gain of Changes in Multiplication Concept Mastery

Comparison	Δ Mean	Cohen's *d*	Cohen Category	Kraft Category	N-gain	N-gain Category
T1 → T2	1.18	0.06	Very small	Medium	0.015	Low
T2 → T3	1.43	0.07	Very small	Medium	0.018	Low
T1 → T3	2.61	0.13	Small	Medium	0.033	Low
Baseline (T1) → Intervention (T2–T3)	1.90	0.10	Very small	Medium	0.024	Low

The between-time test was not significant, $F(2, 393) = 0.573$; $p = 0.564$, with $R^2 = 0.003$ (Table 6). With only three measurement points, the segmented regression model is saturated: three predictors carry only two degrees of freedom of between-time information, so the model F test is identical to a comparison of three means. The test also treats the 396 observations as independent and is therefore conservative for paired data. The Durbin–Watson value of 1.503 lies at the lower limit of the 1.5–2.5 range commonly considered to indicate no serious autocorrelation.

Table 6. Summary of the Between-Time Test and Segmented Regression Model

Statistic	Value
R / R^2 / Adjusted R^2	0.054 / 0.003 / −0.002
Std. Error of the Estimate	19.866
Durbin–Watson	1.503
SS Between Time (df = 2)	452.227
SS Residual (df = 393)	155,094.318
F / Sig.	0.573 / 0.564

Dependent variable: multiplication concept mastery score. Predictors: (Constant), Time, Intervention, TimeAfter.

The segmented regression coefficients are presented in Table 7 for transparency. Because the baseline trend rests on a single measurement point, the Time and TimeAfter coefficients cannot be identified separately, so their individual significance values are not interpreted inferentially. The usable information from this table is the near-zero level change and the post-intervention rate of 1.432 points per measurement point.

Table 7. Segmented Regression Coefficients (Complementary)

Predictor	B	Std. Error	t	Sig.	Meaning
(Constant)	13.750	1.412	9.739	0.000	Model starting point
Time	7.591	0.998	7.604	0.000[a]	Baseline trend (β1)
Intervention	−0.250	4.235	−0.059	0.953	Level change (β2)
TimeAfter	−6.159	1.729	−3.562	0.000[a]	Trend change (β3)

[a] Not interpreted inferentially because β1 and β3 are not separately identified with a single baseline point. Post-intervention rate = β1 + β3 = 1.432 points per measurement point.

Sensitivity analysis results are presented in Table 8. The T1→T3 gain was not significant when the correlation between measurements was 0.50–0.60, was borderline at $r = 0.70$, and was significant at $r = 0.80$ and above. The critical correlation at which the T1→T3 gain becomes significant at $\alpha = 0.05$ is $r = 0.71$. Gains per interval (T1→T2 and T2→T3) were not significant at any tested value of r ($p \geq 0.073$).

Table 8. Sensitivity Analysis of the T1→T3 Gain to the Correlation Between Measurements

T1–T3 Correlation (r)	SD of Difference	*dz*	t(131)	p	Decision (α = 0.05)
0.50	19.74	0.13	1.52	0.131	Not significant
0.60	17.67	0.15	1.70	0.092	Not significant
0.70	15.32	0.17	1.96	0.052	Borderline
0.80	12.54	0.21	2.39	0.018	Significant
0.90	8.93	0.29	3.36	0.001	Significant

Regarding practical significance, an effect that is small by Cohen's benchmark but medium by Kraft's shows the importance of context in judging effectiveness. Kraft's (2020) benchmark is derived from school-based education interventions, which generally yield smaller effects than controlled experiments, so d = 0.13 over roughly two and a half months is still worth considering, especially because the congklak medium is inexpensive and easy to replicate. However, the low N-gain and a T3 mean still below 25 on the 0–100 scale indicate that most of the potential for improvement was not realized. The sensitivity analysis also underscores that the statistical conclusion depends on the stability of scores across measurements: the T1→T3 gain is convincing only when the test–retest correlation is high (r > 0.71), which is commonly achieved by reliable achievement tests but must still be confirmed with individual-level data.

Compared with previous findings, the gradual pattern of gain in this study accords with several international studies. Dumas et al. (2019) showed that a mathematics intervention may produce no clear difference in final scores yet still yield steeper growth curves, so the rate of learning should be read separately from final attainment. A similar pattern appears in studies of mancala-type traditional games, close relatives of congklak. Thiel (2025) observed children over four sessions within one month and found that progress emerged slowly, from one-by-one counting toward treating groups of counters as a unit, and that each child's trajectory differed. Fouze and Amit (2018) likewise reported that folklore games integrated into instruction can develop mathematical thinking, and Gasteiger and Moeller (2021) found that seven board-game sessions over four weeks produced an advantage in recognizing and using structures that was still visible one year later. The high heterogeneity is in line with Brezovszky et al. (2019), who found that the effects of a game-based learning environment differed across grade levels and were stronger in Grade 5. Berrett and Carter (2018) also reported, in a post hoc exploration, that third-grade students who were more proficient before the intervention obtained the largest gains in multiplication fluency, a pattern in the same direction as the widening standard deviation in this study. In the context of multiplication concepts, Sanchez et al. (2025) reported that a multiplicative reasoning intervention for third graders at risk for mathematics difficulties outperformed business-as-usual instruction on multiplicative reasoning measures, although treatment was not a significant predictor of scores on a broader standardized test; Jitendra et al. (2023) reported preliminary evidence of the same intervention's potential in a feasibility study. The former pattern resembles the small effect on the conceptual essay test in this study.

The findings also differ from some earlier studies. In contrast to this study, Fraga-Varela et al. (2021) and Pan et al. (2025) reported significant improvements among elementary students learning with games, and Brezovszky et al. (2019) found an advantage for the experimental group over the control group. Several factors may explain these differences. First, the outcomes measured differed: those studies measured fluency or computational skill, which rises quickly through repeated practice with feedback, whereas the test in this study demanded explanations of concepts and representations. Hankeln and Prediger (2025) showed that explanations using meaning-related language, for example, "two groups of four," rarely appeared spontaneously among students and were strongly associated with higher scores, so it is reasonable that conceptual scores move more slowly. Second, the designs differed: some studies used control groups and more homogeneous classes, whereas this study used a single group across six

schools with a CV above 85%. Third, duration and dose differed: in Fraga-Varela et al. (2021), students used the software for about 27 days on average, whereas this study comprised only six 35-minute meetings. Fourth, the outcome domains differed beyond achievement: Freitas et al. (2024) reported that five sessions of exploratory teaching supported by an applet in Grade 2 improved arithmetic self-efficacy, an affective outcome not measured here. Fifth, the media differed: digital games give immediate feedback, whereas congklak depends on teacher mediation. Kromminga and Codding (2023) found that all groups of third-grade students, whatever the practice modality, grew significantly on proximal measures with few differences between groups, so a gain without a comparison group is difficult to attribute to a particular medium. Thiel (2025) stressed that the mathematical richness of traditional games becomes meaningful only when adults direct children's attention to their mathematical structure. Differences also appear in heterogeneity: Larsen and Little (2023) found a compensatory growth pattern in population data from Grades 3 to 9, that is, shrinking variance and students with lower initial achievement growing faster, whereas the SD in this study actually widened. The very short time span, the different grade level, and the absence of a comparison group may explain why a compensatory pattern is not yet visible.

The contribution of this article lies in shifting the evaluative question from "do scores rise?" to "how do scores grow, for whom, and how meaningfully?" To the authors' knowledge, existing GASING studies, including those combined with games, generally report pretest–posttest differences. This study adds an analysis of interval gain rates, the coefficient of variation as an indicator of relative inequality, effect sizes read against two benchmarks, and a sensitivity analysis of the correlation between measurements. The finding that inferential conclusions depend on the test–retest correlation ($r > 0.71$) shows that non-significance in aggregate data does not automatically mean the absence of benefit, and it reminds readers that claims of success must be accompanied by effect sizes and their analytic assumptions. This novelty is methodological and conceptual, not a claim of causal effectiveness.

The findings carry practical, academic, and policy implications. Practically, teachers should regard GASING with congklak as a medium-term reinforcement and plan for more than six meetings. Explicit mediation is important: teachers can encourage students to describe the equal-group structure in meaning-related language (Hankeln & Prediger, 2025) and prepare tiered tasks and flexible grouping for students who still count one by one (Thiel, 2025). Academically, reporting trajectories, CV, effect sizes with clear benchmarks, and sensitivity analyses deserves to become standard practice in evaluating elementary mathematics instruction. For policymakers, an inexpensive medium such as congklak can be expanded together with teacher training in managing heterogeneous classes and with realistic expectations of outcomes. Teacher mediation itself also needs to be designed and tested, because Pan et al. (2025) noted that postgame discussion does not necessarily improve retention. Future research should use comparison groups with comparative ITS, add measurement points in both the baseline and intervention phases, use parallel tests, analyze student-level data with multilevel or latent growth models, include follow-up measurements to test the durability of effects, given that the effects of educational interventions often fade (Bailey et al., 2017; Pan et al., 2025), and examine students' strategies qualitatively to explain the sources of heterogeneity.

Several limitations should be considered when interpreting the findings. First, the absence of a comparison group means that the score gains cannot be fully separated from maturation and regular learning outside the intervention (Shadish et al., 2002). Second, using the same test at all three measurement points opens the possibility of test-practice effects, so part of the gain may stem from students' familiarity with the items. Third, the baseline phase had only one measurement point, so the pre-intervention trend could not be estimated, whereas the ITS literature recommends several measurement points in each phase (Lopez Bernal et al., 2017; Wagner et al., 2002). Fourth, student data are nested within schools, so multilevel analysis is needed to separate between-student and between-school variation (Hallberg et al., 2018). These limitations do not negate the consistent pattern of gain, but they restrict the causal claims that can be made.

CONCLUSION

Multiplication concept mastery in GASING instruction with congklak grew gradually and with slight strengthening, without a jump at the start of the intervention, so the approach appears to work as cumulative reinforcement rather than an instant effect. Ability heterogeneity remained very high throughout: the absolute gap between students widened slightly while relative inequality did not worsen, which means that the gains are not yet shared evenly. The practical significance of the change was small by general benchmarks and medium by education-intervention benchmarks, and its statistical support depended on the stability of scores across measurements. These findings are therefore best read as a promising tendency, not as a proven effect. For teachers and schools, the findings imply implementing GASING with congklak over a sufficiently long period, judging it by its trajectory rather than by momentary results, and complementing it with flexible grouping, tiered tasks, and explicit modeling for students who still count one by one. Future research should add measurement points, comparison groups, and parallel tests, and should analyze student- and school-level data with multilevel or latent growth models to test whether the benefits differ by initial ability and school characteristics.

ABBREVIATIONS

CV, coefficient of variation; GASING, Gampang, Asyik, dan Menyenangkan (Easy, Fun, and Enjoyable); ITS, interrupted time series; N-gain, normalized gain; RQ, research question; SD, standard deviation.

ACKNOWLEDGMENTS

The authors thank the principals, classroom teachers, and third-grade students of SDN 2 Pakis, SDN Bakungan, SDN 4 Ketapang, SDN 2 Alasrejo, SDN 1 Kalibaru Wetan, and SDN 5 Kaligung in Banyuwangi Regency for their participation and support during the study. Appreciation is also extended to Yayasan Teknologi Indonesia Jaya and Prof. Yohanes Surya for their support, and to the supervisors and reviewers for their constructive feedback.

REFERENCES

Ahmad, S., & Siller, H.-S. (2024). Investigating the effect of manipulatives on mathematics achievement: The role of concrete and virtual manipulatives for diverse achievement level groups. *Journal on Mathematics Education, 15*(3), 979–1002. https://doi.org/10.22342/jme.v15i3.pp979-1002

Armianti, A., Yani, I., Widuri, K., & Sulistiawati, S. (2016). Pengaruh Matematika GASING (Gampang, ASyIk, dan menyenaNGkan) pada materi perkalian bilangan bulat terhadap hasil belajar peserta matrikulasi STKIP Surya [The effect of GASING (Easy, Fun, and Enjoyable) Mathematics on whole-number multiplication on the learning outcomes of STKIP Surya matriculation participants]. *Kreano: Jurnal Matematika Kreatif-Inovatif, 7*(1), 74–81. https://doi.org/10.15294/kreano.v7i1.5012

Bailey, D., Duncan, G. J., Odgers, C. L., & Yu, W. (2017). Persistence and fadeout in the impacts of child and adolescent interventions. *Journal of Research on Educational Effectiveness, 10*(1), 7–39. https://doi.org/10.1080/19345747.2016.1232459

Barmby, P., Harries, T., Higgins, S., & Suggate, J. (2009). The array representation and primary children's understanding and reasoning in multiplication. *Educational Studies in Mathematics, 70*(3), 217–241. https://doi.org/10.1007/s10649-008-9145-1

Berrett, A. N., & Carter, N. J. (2018). Imagine Math Facts improves multiplication fact fluency in third-grade students. *Journal of Behavioral Education, 27*, 223–239. https://doi.org/10.1007/s10864-017-9288-1

Brezovszky, B., McMullen, J., Veermans, K., Hannula-Sormunen, M. M., Rodríguez-Aflecht, G., Pongsakdi, N., Laakkonen, E., & Lehtinen, E. (2019). Effects of a mathematics game-based learning environment on primary school students' adaptive number knowledge. *Computers & Education, 128*, 63–74. https://doi.org/10.1016/j.compedu.2018.09.011

Camp, M. E. (2018). *Building conceptual understandings of multiplication: Integrating virtual manipulatives and think-alouds with repeated addition, arrays, and decomposing numbers* [Doctoral dissertation, University of South Carolina]. Scholar Commons. https://scholarcommons.sc.edu/cgi/viewcontent.cgi?article=5960&context=etd

Carbonneau, K. J., Marley, S. C., & Selig, J. P. (2013). A meta-analysis of the efficacy of teaching mathematics with concrete manipulatives. *Journal of Educational Psychology, 105*(2), 380–400. https://doi.org/10.1037/a0031084

Clements, D. H., & Sarama, J. (2004). Learning trajectories in mathematics education. *Mathematical Thinking and Learning, 6*(2), 81–89. https://doi.org/10.1207/s15327833mtl0602_1

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Dumas, D., McNeish, D., Sarama, J., & Clements, D. (2019). Preschool mathematics intervention can significantly improve student learning trajectories through elementary school. *AERA Open, 5*(4), 1–5. https://doi.org/10.1177/2332858419879446

Flores, M. M., & Hinton, V. M. (2019). Improvement in elementary students' multiplication skills and understanding after learning through the combination of the concrete-representational-abstract sequence and strategic instruction. *Education and Treatment of Children, 42*(3), 353–377. https://doi.org/10.1353/etc.2019.0017

Fouze, A. Q., & Amit, M. (2018). Development of mathematical thinking through integration of ethnomathematic folklore game in math instruction. *Eurasia Journal of Mathematics, Science and Technology Education, 14*(2), 617–630. https://doi.org/10.12973/ejmste/80626

Fraga-Varela, F., Vila-Couñago, E., & Martínez-Piñeiro, E. (2021). The impact of serious games in mathematics fluency: A study in primary education. *Comunicar, 29*(69), 115–125. https://doi.org/10.3916/C69-2021-10

Freitas, Y. A., Abbasi, M., Brito-Costa, S., Pinto, R., Rato, V., & Martins, F. (2024). Exploratory teaching: Integrating applet to teach arithmetic multiplication operation. *Eurasia Journal of Mathematics, Science and Technology Education, 20*(12), Article em2546. https://doi.org/10.29333/ejmste/15666

Gasteiger, H., & Moeller, K. (2021). Fostering early numerical competencies by playing conventional board games. *Journal of Experimental Child Psychology, 204*, Article 105060. https://doi.org/10.1016/j.jecp.2020.105060

Gesiradja, M. C., Sudiapermana, E., Lutfiansyach, D. Y., & Shantini, Y. (2025). Evaluating the effectiveness, challenges, and contextual applications of the Gasing Mathematics Method in Indonesia: A systematic literature review. *Pedagogia: Jurnal Ilmu Pendidikan, 23*(3). https://doi.org/10.17509/pdgia.v23i3.89961

Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics, 66*(1), 64–74. https://doi.org/10.1119/1.18809

Hallberg, K., Williams, R., Swanlund, A., & Eno, J. (2018). Short comparative interrupted time series using aggregate school-level data in education research. *Educational Researcher, 47*(5), 295–306. https://doi.org/10.3102/0013189X18769302

Hankeln, C., & Prediger, S. (2025). Language is essential for avoiding surface translations: Associations of students' spontaneous use of meaning-related phrases for explicating structures with conceptual understanding of multiplication. *Educational Studies in Mathematics, 120*(1), 57–79. https://doi.org/10.1007/s10649-025-10414-z

Hayati, Z., Satriani, N., Raharti, N., & Hijriati. (2024). Gasing mathematics instruction for enhancing problem-solving skills in elementary school students. *Genderang Asa: Journal of Primary Education, 5*(2), 25–36. https://doi.org/10.47766/jga.v5i2.3484

Hendriana, H., Prahmana, R. C. I., & Hidayat, W. (2019). The innovation of learning trajectory on multiplication operations for rural area students in Indonesia. *Journal on Mathematics Education, 10*(3), 397–408. https://doi.org/10.22342/jme.10.3.9257.397-408

Huang, H.-M. E. (2014). Third- to fourth-grade students' conceptions of multiplication and area measurement. *ZDM Mathematics Education, 46*, 449–463. https://doi.org/10.1007/s11858-014-0603-1

Jitendra, A. K., Dougherty, B., Sanchez, V., Harwell, M. R., & Harbour, S. (2023). Building conceptual understanding of multiplicative reasoning content in third graders struggling to learn mathematics: A feasibility study. *Learning Disabilities Research & Practice, 38*(4), 285–295. https://doi.org/10.1111/ldrp.12322

Khasanah, M., Khalil, I. A., & Prahmana, R. C. I. (2023). An inquiry into ethnomathematics within the framework of the traditional game of Congklak. *Journal of Honai Math, 6*(2), 175–188. https://doi.org/10.30862/jhm.v6i2.553

Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher, 49*(4), 241–253. https://doi.org/10.3102/0013189X20912798

Kromminga, K. R., & Codding, R. S. (2023). The impact of intervention modality on students' multiplication fact fluency. *Psychology in the Schools*. Advance online publication. https://doi.org/10.1002/pits.23054

Kusuma, J., & Sulistiawati, S. (2014). Teaching multiplication of numbers from 1 to 10 to STKIP Surya students using Matematika GASING. *Journal on Mathematics Education, 5*(1), 66–84. https://doi.org/10.22342/jme.5.1.1450.66-84

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology, 4*, 863. https://doi.org/10.3389/fpsyg.2013.00863

Larsen, S. A., & Little, C. W. (2023). Matthew effects in reading and mathematics: Examining developmental patterns in population data. *Contemporary Educational Psychology, 74*, 102201. https://doi.org/10.1016/j.cedpsych.2023.102201

Linden, A. (2015). Conducting interrupted time-series analysis for single- and multiple-group comparisons. *The Stata Journal, 15*(2), 480–500. https://doi.org/10.1177/1536867X1501500208

Lopez Bernal, J., Cummins, S., & Gasparrini, A. (2017). Interrupted time series regression for the evaluation of public health interventions: A tutorial. *International Journal of Epidemiology, 46*(1), 348–355. https://doi.org/10.1093/ije/dyw098

Maulida, S. A., Isrokatun, I., & Julia, J. (2024). Pengaruh pembelajaran kooperatif TGT dengan metode GASING terhadap pemahaman konsep matematis perkalian dua angka di SD [The effect of the TGT cooperative learning model with the GASING method on elementary students' conceptual understanding of two-digit multiplication]. *SJME (Supremum Journal of Mathematics Education), 8*(1), 101–113. https://doi.org/10.35706/sjme.v8i1.10789

Mills, J. (2019). Making multiplication meaningful: Teaching for conceptual understanding. *Teachers and Curriculum, 19*(1), 17–25. https://doi.org/10.15663/tandc.v19i1.334

Milton, J. H., Flores, M. M., Moore, A. J., Taylor, J. J., & Burton, M. E. (2019). Using the concrete–representational–abstract sequence to teach conceptual understanding of basic multiplication and division. *Learning Disability Quarterly, 42*(1), 32–45. https://doi.org/10.1177/0731948718790089

Mulyadi, M., & Putra, H. D. (2025). Ethnomathematics in action: Leveraging traditional congklak for meaningful mathematics learning. *Mathline: Jurnal Matematika dan Pendidikan Matematika, 10*(1), 71–83. https://doi.org/10.31943/mathline.v10i1.708

Ninggar, A. D., Restian, A., & Danawati, M. G. (2024). Development of congklak-based ethnomathematics media for teaching fractional numbers to grade V students. *Primary: Jurnal Pendidikan Guru Sekolah Dasar, 13*(4), 171–183. https://doi.org/10.33578/jpfkip-v13i4.p171-183

Pan, A.-J., Yuan, Y.-T., & Chou, P.-N. (2025). Incorporating PaGamO and postgame group discussions into an elementary mathematics course: An experimental study. *SAGE Open, 15*(1), Article 21582440251329698. https://doi.org/10.1177/21582440251329698

Park, J.-H., & Nunes, T. (2001). The development of the concept of multiplication. *Cognitive Development, 16*(3), 763–773. https://doi.org/10.1016/S0885-2014(01)00058-2

Penfold, R. B., & Zhang, F. (2013). Use of interrupted time series analysis in evaluating health care quality improvements. *Academic Pediatrics, 13*(6 Suppl.), S38–S44. https://doi.org/10.1016/j.acap.2013.08.002

Pratiwi, Y., Herlambang, Y. T., Yuniarti, Y., & Bakar, A. (2026). Transformation of elementary school students' mathematics learning outcomes using the GASING method in a critical pedagogical approach. *Journal of Educational Sciences, 10*(4), 195–206. https://doi.org/10.31258/jes.10.4.p.195-206

Rittle-Johnson, B., Schneider, M., & Star, J. R. (2015). Not a one-way street: Bidirectional relations between procedural and conceptual knowledge of mathematics. *Educational Psychology Review, 27*(4), 587–597. https://doi.org/10.1007/s10648-015-9302-x

Sanchez, V., Jitendra, A. K., Harwell, M., Dougherty, B., Pulse, M., & Zahn, D. (2025). Facilitating multiplicative reasoning among third-grade students with or at risk for mathematics difficulties: A pilot study. *Journal of Learning Disabilities*. Advance online publication. https://doi.org/10.1177/00222194251342207

Sariningsih, K. R., Parmiti, D. P., & Yudiana, I. K. E. (2025). The Gasing method with the traditional game-based mong-mongan approach on the mathematics learning outcomes of third grade elementary school students. *Jurnal Ilmiah Sekolah Dasar, 9*(2), 400–407. https://doi.org/10.23887/jisd.v9i2.96608

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.

Shanty, N. O., & Wijaya, S. (2012). Rectangular array model supporting students' spatial structuring in learning multiplication. *Journal on Mathematics Education, 3*(2), 175–186. https://doi.org/10.22342/jme.3.2.603.175-186

Sihombing, J. H., Agustin, M., & Muqodas, I. (2026). The GASING concrete-abstract-mental computation sequence: Learning outcomes and arithmetic fluency in elementary mathematics. *Jurnal Pembelajaran, Bimbingan, dan Pengelolaan Pendidikan, 6*(8), 7. https://doi.org/10.17977/um065.v6.i8.2026.7

Simon, M. A. (1995). Reconstructing mathematics pedagogy from a constructivist perspective. *Journal for Research in Mathematics Education, 26*(2), 114–145. https://doi.org/10.2307/749205

Siregar, J. H., Wiyanti, W., Wakhyuningsih, N. S., & Godjali, A. (2014). Learning the critical points for addition in Matematika GASING. *Journal on Mathematics Education, 5*(2), 160–169. https://doi.org/10.22342/jme.5.2.1500.160-169

Thiel, O. (2025). Playing Gebeta in preschool: Informal pathways to early numeracy through directionality and bundling. *Education Sciences, 15*(10), 1365. https://doi.org/10.3390/educsci15101365

Tusolihah, N., Misri, M. A., & Nursuprianah, I. (2022). Ethnomathematics exploration of the traditional game of congklak. *Eduma: Mathematics Education Learning and Teaching, 11*(2), 203–212. https://doi.org/10.24235/eduma.v11i2.10300

Wagner, A. K., Soumerai, S. B., Zhang, F., & Ross-Degnan, D. (2002). Segmented regression analysis of interrupted time series studies in medication use research. *Journal of Clinical Pharmacy and Therapeutics, 27*(4), 299–309. https://doi.org/10.1046/j.1365-2710.2002.00430.x

Yusuf, D. A., Muslihin, H. Y., & Apriani, I. F. (2026). Effectiveness of the GASING method based on dice games in improving multiplication skills of primary school students. *Indonesian Journal of Primary Education, 10*(1), 167–186.