**M A T H L I N E**
**JURNAL MATEMATIKA DAN PENDIDIKAN MATEMATIKA**
**ISSN 2502-5872 (Print)**
**ISSN 2622-3627 (Elektronik)**

This is an open access article under the CC BY-SA license

Volume 11 Number 4, November 2026, 1041-1056

# THE GASING METHOD AND A MULTIPLICATION DICE GAME TO IMPROVE ELEMENTARY STUDENTS' MATHEMATICS LEARNING MOTIVATION IN AMBON CITY

**Darlin Adam[1], Heri Yusuf Muslihin[2], Ika Fitri Apriani[3]***
[1,2,3]Elementary School Teacher Education Study Program, Graduate School, Universitas Pendidikan Indonesia, Tasikmalaya Campus, Tasikmalaya, West Java 46115, Indonesia

*Correspondence: apriani25@upi.edu

Received 25 October 2026; Revised 7 November 2026; Accepted 11 November 2026

***ABSTRACT***

Elementary students' motivation to learn mathematics may be suppressed by rote-oriented instruction, a concern that merits empirical investigation in Ambon City, Maluku. This quasi-experimental study investigated whether the GASING method (an Indonesian acronym for Easy, Fun, and Enjoyable), a locally developed constructivist approach, integrated with a multiplication dice game, was associated with a greater improvement in the learning motivation of fourth-grade students in Ambon City. A nonequivalent control group pretest–posttest design (the same motivation questionnaire before and after) involved 40 students from two intact classes of a public elementary school (experimental, $n = 20$, taught with GASING and the dice game; control, $n = 20$, taught with conventional textbook-based instruction). Motivation was measured with a validated 21-item, four-point Likert questionnaire (Cronbach's $\alpha = 0.932$) adapted from an established Indonesian motivation framework. Pretest equivalence was confirmed by a Mann–Whitney $U$ test ($U = 170.0$, $p = 0.422$). The experimental class obtained a mean normalized gain (N-Gain) of 0.614 (95% CI [0.556, 0.672]; medium category), far higher than the control ($M = 0.188$, 95% CI [0.136, 0.239]; low category). The difference was statistically significant ($t(38) = 11.535$, $p < 0.001$; N-Gain difference = 0.426). The GASING-integrated dice game appears to be a low-cost, locally preparable, and motivationally promising approach for teaching multiplication, although the combined delivery of both components, the single teacher per condition, and the small sample prevent attributing the gain to the method alone and call for cautious interpretation and replication in Maluku.

**Keywords:** Elementary Mathematics Education, Game-Based Learning, Normalized Gain, Quasi-Experiment, Self-Determination Theory.

**How to Cite:** Adam, D., Muslihin, H. Y., & Apriani, I. F. (2026). The GASING Method and A Multiplication Dice Game to Improve Elementary Students' Mathematics Learning Motivation in Ambon City. *Mathline: Jurnal Matematika dan Pendidikan Matematika, 11*(4), 1041-1056. http://doi.org/10.31943/mathline.v11i4.1240

## PRELIMINARY

Elementary school mathematics builds foundational competence, and fluency in multiplication supports later reasoning about fractions, ratios, and algebra (McNeil et al., 2025). Ambon City, the capital of Maluku Province, is an archipelagic region whose schools may face limited facilities and learning resources, so instructional approaches that

are inexpensive, easy to prepare locally, and independent of technology are highly relevant. Low motivation is a recognized obstacle when instruction does not engage students (Schukajlow et al., 2023), and students with stronger intrinsic motivation show greater persistence and better outcomes (Howard et al., 2021; Schunk & DiBenedetto, 2020), yet in teacher-centered classrooms motivation is often suppressed by rote memorization and repetitive drill detached from students' everyday experience (Boaler et al., 2021). Empirical documentation of this pattern in Ambon City is still lacking, so it is treated as a plausible concern, not an established local fact.

Motivation in mathematics encompasses intrinsic interest, perceived competence, task value, and emotional engagement (Wigfield & Eccles, 2020). In the Indonesian context, Uno (2023) proposes six dimensions of motivation: the desire and wish to succeed, the drive and need to learn, hopes and aspirations, rewards in learning, interesting learning activities, and a conducive learning environment; this framework underlies the instrument. Self-Determination Theory (SDT) describes a continuum from external regulation to intrinsic motivation that is sustained by satisfaction of three needs: autonomy (experiencing one's actions as self-endorsed), competence (feeling effective in meeting challenges), and relatedness (feeling connected to others) (Ryan & Deci, 2017, 2020). The six dimensions can be aligned with these needs in a many-to-many way: the drive and need to learn corresponds mainly to autonomy; the desire to succeed and hopes and aspirations mainly to competence (Wigfield & Eccles, 2020); interesting learning activities to autonomy and competence; and a conducive learning environment to relatedness and autonomy-supportive conditions (Guay, 2022). Rewards in learning is the least straightforward dimension: rewards support competence when informational but undermine autonomy when controlling (Ryan & Deci, 2020). The mapping is theoretical and untested, so the total score is interpreted as general learning motivation, not as separate SDT needs.

One approach that aims to make mathematics accessible and enjoyable is the GASING method (*Gampang, Asyik, dan Menyenangkan*; Easy, Fun, and Enjoyable), developed by Yohanes Surya as an Indonesian response to overly abstract mathematics instruction (Kusuma et al., 2019). It is conceptually consistent with constructivism and Vygotsky's sociocultural theory (Eun, 2019) and moves from concrete objects, through pictorial representations, to abstract notation. Evaluations report positive effects on achievement and engagement, but largely from single-group and pre-experimental designs (Kusuma et al., 2019); evidence on motivation with validated instruments is scarce, as

recent quasi-experimental studies center on cognitive outcomes such as addition skills (Maharudin et al., 2026) and motivation has been examined mainly alongside problem solving (Arifin et al., 2020).

Integrating a dice game is a logical extension that builds on research on game-based learning (GBL) (Mayer, 2019; Barz et al., 2024). A physical dice game offers kinesthetic engagement, collaboration, immediate feedback, and graded challenge, which speak to the SDT needs (Ryan & Deci, 2017, 2020). Cooperative game formats show the strongest motivational effects in K-12 mathematics (Pan et al., 2022), and non-digital mathematics games generally show positive effects (Russo et al., 2024), although few studies used validated measurement or verified group equivalence. Collaborative non-digital games may also reduce mathematics anxiety more than digital ones (Dondio et al., 2023), yet the non-digital evidence base remains dominated by single-group designs (Russo et al., 2024). A recent study in this journal, using a nonequivalent control group design, reported that pizza-shaped game media improved fifth-grade students' fraction learning outcomes (Pratiwi & Yamin, 2026); however, it measured achievement rather than motivation.

Autonomous motivation grows when the three needs are satisfied (Guay, 2022), and relatedness is the need most directly shaped by local culture: the needs are regarded as universal, but their expression is culturally patterned; autonomy, for example, means volitional endorsement of one's actions, not individualism (Chirkov et al., 2003; Vansteenkiste et al., 2020), and culturally responsive pedagogy argues that students engage more deeply when instruction connects with their cultural frames of reference (Gay, 2018; Ladson-Billings, 1995). In Ambon City, social life is shaped by *basudara* (brotherhood), voiced in *ale rasa, beta rasa* (what you feel, I feel) and *sagu salempeng dibage dua* (a piece of sago bread shared by two), by the bonds of *pela* and *gandong* between villages, and by everyday practices such as *masohi* (mutual help) and *ma'anu* (cooperation) (Wakano, 2019). These values are conceptually aligned with the need for relatedness and with the positive interdependence on which cooperative learning is built (Johnson & Johnson, 2009). The dice game follows the same logic: groups of four, shared team score sheets, and collaborative assessment emphasize the team's shared outcome rather than individual ranking. This cultural fit is a conceptual proposition, not tested here because relatedness and cultural identification were not measured, and it does not imply that all students hold these values uniformly.

Taken together, this literature leaves three gaps: the motivational effects of GASING measured with a validated instrument remain largely unexplored (Maharudin et al., 2026) and no comparable evidence from elementary schools in Ambon City is known; research on non-digital games relies largely on single-group designs, and Indonesian intervention research often lacks psychometric validation, control groups, and effect size reporting; and, to the authors' knowledge, the motivational effect of combining a dice game with GASING has not been tested with a validated instrument and verified pretest equivalence. The novelty of this study therefore lies in (1) testing the motivational effect, rather than the achievement effect, of GASING combined with a cooperative dice game; (2) using a psychometrically examined instrument and a nonequivalent control group design with verified pretest equivalence and effect size reporting; and (3) providing, to the authors' knowledge, the first such evidence from elementary schools in Ambon City. The research question is: is the GASING method integrated with a multiplication dice game associated with a significantly greater improvement in learning motivation among fourth-grade students in Ambon City than conventional instruction, as measured by a validated questionnaire?.

## METHODS

### Design and Participants

This study used a nonequivalent control group pretest–posttest quasi-experimental design because intact classes could not be re-randomized without disrupting school administration (Reichardt, 2019). "Pretest" and "posttest" refer to two administrations of the same learning-motivation questionnaire (see Instrument and Validation), one before the first session and one after the last; no achievement test was administered, because motivation is a non-cognitive construct measured by self-report. Participants were 40 fourth-grade students aged 9 to 10 years from two intact classes at a public elementary school in Ambon City, selected purposively: Class A ($n$ = 20) as the experimental group and Class B ($n$ = 20) as the control group. The researcher determined which class received the intervention, so selection bias cannot be excluded and pretest equivalence was examined. Fourth grade (Phase B of the Merdeka Curriculum) covers the basic multiplication content used here, and children aged 9 to 10 are still in the concrete operational stage (Piaget & Inhelder, 1969), so instruction should begin with concrete experience, as the GASING sequence does (Kusuma et al., 2019). A separate sample of 30 fourth-grade students from another school in Ambon City served as the try-out group. An *a*

*priori* power analysis with G*Power 3.1 (Kang, 2021), assuming a large effect ($d$ = 1.0; Pan et al., 2022), $\alpha$ = 0.05, and power = 0.80, indicated a minimum of $n$ = 17 per group, which the sample satisfied. Ethical procedures included written consent from the school principal and parents or guardians and identity coding; research permission was granted by Universitas Pendidikan Indonesia (reference number B-5563/UN40.C3.3.1/PK.03.04/2026, dated 13 February 2026). The control group received the intervention after the study ended.

**Intervention**

The intervention lasted six weeks (three 70-minute sessions per week; 18 sessions) in three two-week modules: single-digit facts, two-digit by one-digit, and two-digit by two-digit multiplication. The experimental class teacher attended a three-day (15-hour) professional development program on GASING principles and concrete-to-abstract scaffolding. Each session consisted of 45–50 minutes of GASING instruction followed by 20–25 minutes of dice play, always following a concrete–pictorial–abstract sequence. In Module 2, for example, 23 × 4 was shown as 4 × 23 (four boxes, each with two tens cards and three ones cards) using the GASING number cards (*kartu bilangan*), which are color-coded by place value, bundled sticks (*lidi*), and bottle caps; students then drew bars and dots, computed the partial products 4 × 20 and 4 × 3, and linked the drawings to the written algorithm (80 + 12 = 92).

Groups of four were formed at random, and one student who was stronger in multiplication was placed in each group as its leader, so that every group had a more capable peer who could scaffold the others within their zone of proximal development (Eun, 2019). In each round, one pair rolled two dice and all four members calculated the product and raced to state it. Each correct answer earned a point on the team score sheet, and an incorrect answer simply earned none. The dice carried 1 to 6 dots at the first level and up to 9 dots once students became fluent, and groups moved up a difficulty level as their members answered faster and more fluently; team scores and collaborative assessment reduced performance-avoidance dynamics. Implementation was monitored only through teacher-completed session logs reviewed by the researchers.

The control class teacher, who received neither GASING nor dice training, taught the same content for the same number of sessions using the Grade 4 mathematics textbook under the Merdeka Curriculum and conventional teacher-centered instruction (whiteboard demonstration, guided textbook elaboration, and individual worksheets), without manipulatives, games, or collaborative activities. Teachers are a major source of between-

classroom variance in student outcomes (Nye et al., 2004). Both were senior teachers with equivalent qualifications but differed in approach: the control teacher was teacher-centered, whereas the experimental teacher acted as a facilitator and students participated more proactively. Both received identical pacing guidelines and completed session logs. These measures reduce but cannot eliminate the confound: with one teacher per condition, teacher and method are perfectly aliased, a recognized threat to internal validity in quasi-experiments (Shadish et al., 2002). Teacher rotation or random assignment was not feasible with two intact classes in one school, so this remains the main limitation (see Limitations).

**Instrument and Validation**

Learning motivation was measured with a 21-item, four-point Likert questionnaire (1 = strongly disagree; 4 = strongly agree; maximum score = 84) with positively worded items, adapted from Nurfurqon (2017) and organized according to Uno's (2023) six dimensions: desire and wish to succeed (3 items), drive and need to learn (4), hopes and aspirations (3), rewards in learning (3), interesting learning activities (4), and a conducive learning environment (4). An example item is "I feel enthusiastic when mathematics lessons use activities that I can do together with my classmates." The most important modification was the change from a five-point to a four-point scale, which removes the neutral midpoint: respondents who do not wish to think carefully about an item tend to satisfice by choosing it (Krosnick, 1991), and a midpoint can also be used to avoid taking a position (Chyung et al., 2017). Removing it should yield more discriminating scores, at the cost of forcing a choice on undecided students. Items were also regrouped and reworded following recommended scale-development stages (DeVellis & Thorpe, 2021).

Content validation was carried out in September 2025 by one expert, a professor of educational psychology who heads the Educational Psychology study program at Universitas Pendidikan Indonesia. Using a structured sheet, the expert rated ten aspects on a four-point scale in three categories (construct, content, and language) and judged the instrument feasible for data collection with revisions, which were made before the try-out (for example, absolute wording such as "selalu" [always] was revised). With one expert and an overall judgment only, a content validity index could not be computed; at least three experts are generally recommended (Lynn, 1986). The questionnaire was then tried out on 30 students (mean age = 9.4 years) from another school. Item validity was assessed with corrected item-total correlations (critical $r = 0.361$ for $n = 30$, $\alpha = 0.05$) and reliability with Cronbach's alpha, computed in Python (SciPy 1.10) and cross-verified with IBM SPSS Statistics 26. Corrected correlations ranged from 0.336 to 0.764, and 19 of the 21 items

exceeded the critical value (Appendix 1). Items 5 ($r$ = 0.336) and 13 ($r$ = 0.349) were slightly below it but above the commonly used criterion of 0.30 (Field, 2018), and deleting either would have raised alpha by only 0.002 and 0.001, so no item was dropped. Factor analysis was not conducted because the try-out sample (about 1.4 participants per item) was far below the guidelines of 100–200 cases or 5–10 per item (Comrey & Lee, 1992; Costello & Osborne, 2005; Kyriazos, 2018); small samples suffice only when communalities are uniformly high (MacCallum et al., 1999), which cannot be assumed for a new instrument with as few as three items per dimension. Evidence for the structure is therefore limited to expert review, item-total correlations, and internal consistency.

**Outcome Measure and Data Analysis**

Change from pretest to posttest was expressed as normalized gain (N-Gain; Hake, 1998; Coletta & Steinert, 2020), the improvement as a proportion of the maximum gain each student could achieve from their own starting point:

$$N - Gain = \frac{S_{post} - S_{pre}}{S_{max} - S_{pre}} \quad (1)$$

where $S_{post}$ and $S_{pre}$ are the posttest and pretest scores and $S_{max}$ is the maximum possible score (84). N-Gain controls for individual pretest differences, which suits quasi-experimental designs (Coletta & Steinert, 2020); gains below 0.3 are categorized as low, 0.3 to 0.7 as medium, and 0.7 or higher as high (Hake, 1998). Because N-Gain was developed for physics concept inventories, its use with a Likert-type scale is an adaptation: the scale is treated as approximately interval, the lowest possible score (21) is not zero, and the Hake categories are descriptive conventions only. Analyses comprised descriptive statistics with 95% confidence intervals (CI), the Shapiro–Wilk test, the Mann–Whitney $U$ test for pretest equivalence (pretest scores were not normal), an independent-samples $t$ test on N-Gain for the main hypothesis, Cohen's $d$ with 95% CI, and the Wilcoxon signed-rank test for within-group change ($\alpha = 0.05$).

## RESULTS AND DISCUSSION

**Instrument Validity and Pretest Equivalence**

Corrected item-total correlations (0.336 to 0.764; Appendix 1) led to no item being removed, so all six dimensions were retained. Cronbach's α was 0.932, a very high level, although Taber (2018) cautions that alpha alone is limited evidence of instrument quality; the six-dimension structure still requires confirmatory factor analysis. Descriptive statistics are presented in Table 1. Raw pretest scores were not normally distributed in either the

experimental class ($W$ = 0.798; $p$ = 0.001) or the control class ($W$ = 0.811; $p$ = 0.001), so equivalence was tested with the Mann–Whitney $U$ test, which showed no significant difference between the experimental ($Mdn$ = 58.0) and control ($Mdn$ = 58.5) classes ($U$ = 170.0; $p$ = 0.422), with homogeneous variances (Levene's $F$ = 0.154; $p$ = 0.697). Both classes thus began with equivalent motivation, a basic requirement for causal inference in this design (Reichardt, 2019). At posttest, the experimental class reached 86.8% of the ideal score (from 65.5%) and the control class 73.5% (from 67.0%).

**Table 1.** Descriptive statistics with 95% confidence intervals for pretest and posttest scores

| Variable | *n* | Mean | *SD* | Min | Max | 95% CI | Shapiro–Wilk *p* |
|---|---|---|---|---|---|---|---|
| Experimental pretest | 20 | 55.05 | 6.79 | 32 | 63 | [51.87, 58.23] | 0.001 |
| Experimental posttest | 20 | 72.90 | 3.82 | 68 | 79 | [71.11, 74.69] | 0.061 |
| Control pretest | 20 | 56.30 | 5.80 | 38 | 63 | [53.59, 59.01] | 0.001 |
| Control posttest | 20 | 61.75 | 4.38 | 48 | 68 | [59.70, 63.80] | < 0.001 |

*Note. SD* = standard deviation; CI = confidence interval; ideal score = 84. Shapiro–Wilk $p < 0.05$ indicates a non-normal distribution.

## N-Gain Results and Hypothesis Test

N-Gain scores were normally distributed in both classes (experimental $W$ = 0.928; $p$ = 0.142; control $W$ = 0.956; $p$ = 0.465) with homogeneous variances ($F$ = 0.435; $p$ = 0.514), so the assumptions of the independent-samples $t$ test were met. As shown in Table 2 and Figure 1, the experimental class obtained a mean N-Gain of 0.614 (medium category; 5 students high, 15 medium, none low), compared with 0.188 in the control class (low category; 2 students medium, 18 low). The difference was statistically significant, $t(38) = 11.535$; $p < 0.001$, with an N-Gain difference of 0.426 (95% CI [0.355, 0.497]) and Cohen's $d$ = 3.648 (95% CI [2.620, 4.676]).

**Table 2.** N-Gain scores and category distribution by class

| Class | Mean | *SD* | 95% CI | Category | High / Medium / Low |
|---|---|---|---|---|---|
| Experimental (GASING + dice game) | 0.614 | 0.124 | [0.556, 0.672] | Medium | 5 / 15 / 0 |
| Control (conventional instruction) | 0.188 | 0.110 | [0.136, 0.239] | Low | 0 / 2 / 18 |

*Note.* Categories: High ≥ 0.70; Medium 0.30–0.70; Low < 0.30 (Hake, 1998). 95% CI is based on the $t$ distribution ($df$ = 19). Independent-samples $t$ test: $t(38) = 11.535$; $p < 0.001$; N-Gain difference = 0.426; 95% CI [0.355, 0.497]; Cohen's $d$ = 3.648; 95% CI [2.620, 4.676].

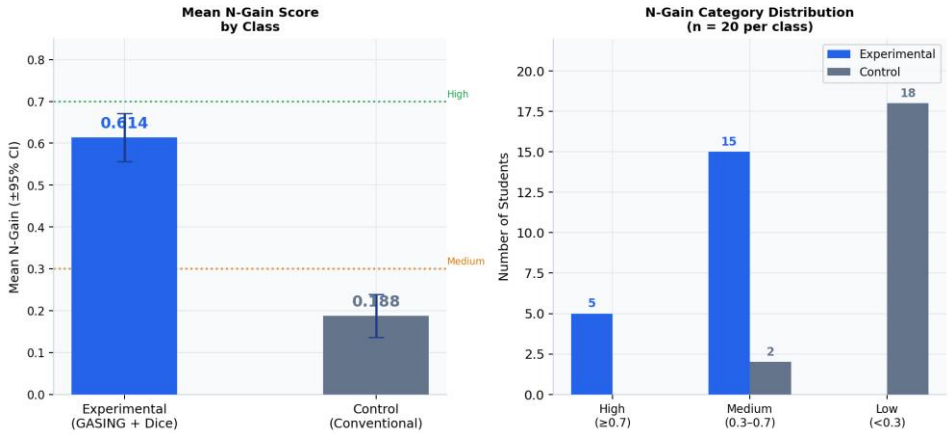



**Figure 1.** Mean N-Gain with 95% CI (left) and category distribution (right) by class

The Wilcoxon signed-rank test showed significant improvement in both classes: the experimental class from *Mdn* = 58.0 to 73.5 ($W = 0.0$; $p < 0.001$) and the control class from *Mdn* = 58.5 to 62.5 ($W = 1.0$; $p < 0.001$). The control-class improvement may reflect repeated exposure to the material and maturation, test–retest familiarity, sustained teacher attention (Schunk & DiBenedetto, 2020), and the Hawthorne effect and demand characteristics (Corneille & Lush, 2023). Because both groups were measured identically, the between-group N-Gain difference adjusts for these shared mechanisms to the extent that they affected both classes equally, so the *t* test is the primary inferential test. These mechanisms may also have operated more strongly in the experimental class (novel activity format, teacher training, more attention), so the difference is interpreted as associated with the intervention package, not as the effect of either component alone.

**Interpretation, Mechanisms, and Implications**

The GASING-dice class achieved a much greater increase in motivation: no experimental student fell into the low N-Gain category, compared with 90% of control students. The value of $d = 3.648$ far exceeds conventional benchmarks for a large effect (Schäfer & Schwarz, 2019) and the typical magnitude of field educational intervention effects (Kraft, 2020), so it should be read with caution for four reasons. First, *d* increases as within-group variability decreases; with small N-Gain standard deviations (0.124 and 0.110), a moderate mean difference yields a very large *d*. Second, small samples produce imprecise and, once reported, often inflated estimates (Button et al., 2013); the small-sample correction (Hedges' $g \approx 3.58$; Lakens, 2013) changes the value only slightly, so the size more likely reflects novelty effects, demand characteristics, and teacher enthusiasm in a single-teacher design than a correctable bias. Third, the estimate is far above those of comparable interventions: $g = 0.32$ for digital game-based learning on affective-

motivational outcomes (Barz et al., 2024) and $g = 0.36$ for gamification on motivational outcomes (Sailer & Homner, 2020), although these come from larger controlled studies and are not strictly comparable with a *d* computed from N-Gain scores. A GASING quasi-experiment with 45 third-grade students shows how within-group comparisons inflate standardized effects: within-group *d* exceeded 3.5 for both GASING and direct instruction, whereas the between-group contrast was much smaller (N-Gain 0.90 versus 0.73; $d = 1.23$; Maharudin et al., 2026), although it measured achievement. Fourth, the N-Gain difference (0.426; 95% CI [0.355, 0.497]) is the more informative effect metric: the experimental class closed about 61% of its attainable gap, compared with about 19% in the control class. The true effect is likely smaller than *d* suggests. The result is nevertheless consistent with findings that collaborative non-digital games reduce mathematics anxiety more than digital games (Dondio et al., 2023) and generally improve dispositions toward mathematics (Russo et al., 2024).

One plausible but untested explanation is simultaneous need satisfaction. Repetitive teacher-centered drill can thwart all three needs, offering no choice (autonomy), exposing students to failure through speed-based recall (competence), and isolating them in individual practice (relatedness) (Howard et al., 2021; Ryan & Deci, 2020); the intervention addressed all three in one activity. *Competence* is most plausibly supported by the GASING component: concrete-to-abstract scaffolding that begins with manipulatives follows the logic of concreteness fading (Fyfe et al., 2014; Kusuma et al., 2019) and operates within the zone of proximal development (Eun, 2019), while graded difficulty levels keep tasks challenging but attainable, and repeated success is the core source of competence satisfaction (Ryan & Deci, 2017). *Relatedness* is most plausibly supported by the cooperative game: groups of four, shared score sheets, and collaborative assessment create positive interdependence (Johnson & Johnson, 2009) and shift emphasis from individual ranking to the team's result; social elements such as teammates are tied to relatedness (Sailer et al., 2017), and gamification effects are moderated by social interaction (Sailer & Homner, 2020). Team play may also have resonated with *basudara* and *pela gandong* (Wakano, 2019), although this remains an interpretation. *Autonomy* is probably the most weakly supported: strategic choices in the game offer structured, bounded choice, a recognized form of autonomy support (Reeve & Cheon, 2021), and in a culture that emphasizes togetherness it may be experienced as volitional participation in a shared activity (Chirkov et al., 2003). Interventions that satisfy all three needs yield greater gains in autonomous motivation than those satisfying only some (Guay, 2022), consistent

with the more homogeneous posttest scores of the experimental class (*SD* = 3.82 versus 4.38) and with the lowest-scoring experimental student at pretest (32; 38.1% of the ideal score), who achieved an N-Gain of 0.693. Because need satisfaction was not measured, this remains a theoretical interpretation, testable with a three-group design (GASING only, dice game only, combined), subscale-level data, and a need-satisfaction measure.

Alternative explanations (combined delivery, the single-teacher confound, and test familiarity; see Limitations) mean that the results show association with the combined package, not a proven cause. The study adds to the non-digital GBL literature, which is underrepresented and largely single-group without validated motivation instruments (Russo et al., 2024), by presenting a two-group quasi-experiment from Ambon City with psychometric measurement and verified pretest equivalence. The findings do not show that physical games are superior to digital ones, but that a cooperative physical game is a feasible and motivationally promising option for resource-limited schools: dice and score sheets are made locally at minimal cost, and session logs indicated implementation as planned after 15 hours of teacher training.

Generalization requires caution: the study involved two classes in one public school, whereas resources, teacher qualifications, and community support vary across Maluku, especially in remote or island areas. Effects can differ across units, settings, and implementers (Shadish et al., 2002), and educational treatment effects are heterogeneous across contexts, so an estimate from one site cannot be assumed to transfer (Bryan et al., 2021). Future research should test the intervention across schools differing in location, resources, and teacher qualification, modeling these as moderators.

**Limitations**

Several limitations constrain causal claims. First, and most critical, is the teacher confound: with one teacher per condition, teacher effectiveness and instructional method cannot be separated. The contrast between a teacher-centered and a facilitating style is part of the intended treatment, but a facilitating style is itself autonomy-supportive (Reeve & Cheon, 2021), so part of the effect may reflect the teacher's style rather than GASING or the dice game; identical pacing guidelines and session-log review reduced but could not remove this confound, so future studies should randomly assign several teachers to conditions. Second, GASING and the dice game cannot be separated. Third, the sample ($n$ = 20 per group) met the *a priori* power requirement but produced wide confidence intervals, and two purposively chosen intact classes from one school leave selection bias possible (Shadish et al., 2002); replications should target $n \geq 50$ per group. Fourth,

implementation was verified only through teacher-completed session logs, without independent structured observation. Fifth, the six-dimension structure needs confirmatory factor analysis, and content validation relied on one expert (Lynn, 1986). Sixth, N-Gain was developed for concept inventories; future studies should add sensitivity analyses such as ANCOVA on posttest scores with the pretest as covariate. Seventh, the six-week span did not allow assessment of sustainability, so follow-ups at three and six months are needed. Finally, sole reliance on self-report is vulnerable to response biases (Podsakoff et al., 2003): young students may answer positively to please the teacher or researcher, or because they infer what the study expects (Corneille & Lush, 2023). This risk is probably greater here because the experimental class had just enjoyed a novel activity, students were not blind to condition, and all items were positively worded (allowing acquiescence), which could inflate the difference.

Because motivation is closely related to multidimensional engagement (Fredricks et al., 2004; Skinner et al., 2009), future studies can add observable indicators, such as time-sampled on-task behavior, persistence, voluntary participation, and peer explanation, with structured protocols and agreement between two trained observers checked with Cohen's kappa (McHugh, 2012); divergence from self-report would indicate response bias.

## CONCLUSION

This quasi-experimental study provides initial evidence that the GASING method integrated with a multiplication dice game is associated with a significantly greater improvement in learning motivation among fourth-grade students in Ambon City than conventional instruction (mean N-Gain 0.614, medium, versus 0.188, low; $t(38) = 11.535$; N-Gain difference = 0.426; $p < 0.001$; $d = 3.648$). These statistics should be interpreted cautiously given the small sample, the teacher confound, the combined delivery of the two components, and possible demand characteristics. Nevertheless, an indigenous Indonesian approach implemented with minimal resources and brief teacher training may be associated with meaningful motivational gains, so it merits consideration in Ambon City and, after further testing, elsewhere in Maluku.

Theoretically, the findings are consistent with an SDT-based explanation in which a cooperative, concrete-to-abstract activity satisfies competence and relatedness together, and suggest that values of togetherness (*basudara*, *pela gandong*) are a testable resource for relatedness in Maluku classrooms. Practically, teachers can add a cooperative game phase with locally made dice and team score sheets after GASING instruction, and schools

and education offices can consider short (about 15-hour) teacher training on GASING and cooperative game facilitation.

Suggestions for future research are: (1) cross-school and cross-teacher replication with random assignment of several teachers to each condition, larger samples ($n \geq 50$ per group), and multilevel modeling; (2) follow-up at three and six months; (3) a three-group design separating the GASING and game components, with a need-satisfaction measure; (4) content validation by at least three experts and factor analysis with a larger sample, with subscale-level analysis of SDT pathways; (5) measurement of relatedness and cultural identification to test the alignment with *basudara* and *pela gandong*; and (6) behavioral engagement measures and structured fidelity observation.

## DECLARATIONS


| | | |
|---|---|---|
| Author Contribution | : | D. Adam: Conceptualization, Methodology, Investigation, Formal Analysis, Writing – Original Draft, and Visualization. H. Y. Muslihin: Validation, Supervision, and Writing – Review & Editing. I. F. Apriani: Validation, Supervision, and Writing – Review & Editing. |
| Funding Statement | : | This research received no external funding. |
| Conflict of Interest | : | The authors declare no conflict of interest. |
| Additional Information | : | Additional information is available for this paper. |


## REFERENCES


Arifin, Z., Setyosari, P., Ulfa, S., & Kuswandi, D. (2020). The effect of GASING learning on problem solving ability and student motivation in elementary schools. *Jurnal Ilmiah Sekolah Dasar, 4*(4), 516–525. https://doi.org/10.23887/jisd.v4i4.28694

Barz, N., Benick, M., Dörrenbächer-Ulrich, L., & Perels, F. (2024). The effect of digital game-based learning interventions on cognitive, metacognitive, and affective-motivational learning outcomes in school: A meta-analysis. *Review of Educational Research, 94*(2), 193–227. https://doi.org/10.3102/00346543231167795

Boaler, J., Dieckmann, J. A., LaMar, T., Leshin, M., Selbach-Allen, M., & Pérez-Núñez, G. (2021). The transformative impact of a mathematical mindset experience taught at scale. *Frontiers in Education, 6*, Article 784393. https://doi.org/10.3389/feduc.2021.784393

Bryan, C. J., Tipton, E., & Yeager, D. S. (2021). Behavioural science is unlikely to change the world without a heterogeneity revolution. *Nature Human Behaviour, 5*(8), 980–989. https://doi.org/10.1038/s41562-021-01143-3

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience, 14*(5), 365–376. https://doi.org/10.1038/nrn3475

Chirkov, V., Ryan, R. M., Kim, Y., & Kaplan, U. (2003). Differentiating autonomy from individualism and independence: A self-determination theory perspective on internalization of cultural orientations and well-being. *Journal of Personality and Social Psychology, 84*(1), 97–110. https://doi.org/10.1037/0022-3514.84.1.97

Chyung, S. Y., Barkin, J. R., & Shamsy, J. A. (2017). Evidence-based survey design: The use of a midpoint on the Likert scale. *Performance Improvement, 56*(10), 15–23. https://doi.org/10.1002/pfi.21727

Coletta, V. P., & Steinert, J. J. (2020). Why normalized gain should continue to be used in analyzing preinstruction and postinstruction scores on concept inventories. *Physical Review Physics Education Research, 16*(1), Article 010108. https://doi.org/10.1103/PhysRevPhysEducRes.16.010108

Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis* (2nd ed.). Lawrence Erlbaum Associates.

Corneille, O., & Lush, P. (2023). Sixty years after Orne's American Psychologist article: A conceptual framework for subjective experiences elicited by demand characteristics. *Personality and Social Psychology Review, 27*(1), 83–101. https://doi.org/10.1177/10888683221104368

Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research, and Evaluation, 10*(7), 1–9. https://doi.org/10.7275/jyj1-4868

DeVellis, R. F., & Thorpe, C. T. (2021). *Scale development: Theory and applications* (5th ed.). SAGE Publications.

Dondio, P., Gusev, V., & Rocha, M. (2023). Do games reduce maths anxiety? A meta-analysis. *Computers & Education, 194*, Article 104650. https://doi.org/10.1016/j.compedu.2022.104650

Eun, B. (2019). The zone of proximal development as an overarching concept: A framework for synthesizing Vygotsky's theories. *Educational Philosophy and Theory, 51*(1), 18–30. https://doi.org/10.1080/00131857.2017.1421941

Field, A. (2018). *Discovering statistics using IBM SPSS Statistics* (5th ed.). SAGE Publications.

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research, 74*(1), 59–109. https://doi.org/10.3102/00346543074001059

Fyfe, E. R., McNeil, N. M., Son, J. Y., & Goldstone, R. L. (2014). Concreteness fading in mathematics and science instruction: A systematic review. *Educational Psychology Review, 26*(1), 9–25. https://doi.org/10.1007/s10648-014-9249-3

Gay, G. (2018). *Culturally responsive teaching: Theory, research, and practice* (3rd ed.). Teachers College Press.

Guay, F. (2022). Applying self-determination theory to education: Regulations types, psychological needs, and autonomy supporting behaviors. *Canadian Journal of School Psychology, 37*(1), 75–92. https://doi.org/10.1177/08295735211055355

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

Howard, J. L., Bureau, J. S., Guay, F., Chong, J. X. Y., & Ryan, R. M. (2021). Student motivation and associated outcomes: A meta-analysis from self-determination theory. *Perspectives on Psychological Science, 16*(6), 1300–1323. https://doi.org/10.1177/1745691620966789

Johnson, D. W., & Johnson, R. T. (2009). An educational psychology success story: Social interdependence theory and cooperative learning. *Educational Researcher, 38*(5), 365–379. https://doi.org/10.3102/0013189X09339057

Kang, H. (2021). Sample size determination and power analysis using the G*Power software. *Journal of Educational Evaluation for Health Professions, 18*, Article 17. https://doi.org/10.3352/jeehp.2021.18.17

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

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology, 5*(3), 213–236. https://doi.org/10.1002/acp.2350050305

Kusuma, J., Kusumah, Y. S., Juandi, D., & Usdiyana, D. (2019). Effectiveness of GASING method on elementary students' mathematics learning outcomes: A meta-analysis. *Journal on Mathematics Education, 10*(3), 375–392. https://doi.org/10.22342/jme.10.3.7798.375-392

Kyriazos, T. A. (2018). Applied psychometrics: Sample size and sample power considerations in factor analysis (EFA, CFA) and SEM in general. *Psychology, 9*(8), 2207–2230. https://doi.org/10.4236/psych.2018.98126

Ladson-Billings, G. (1995). Toward a theory of culturally relevant pedagogy. *American Educational Research Journal, 32*(3), 465–491. https://doi.org/10.3102/00028312032003465

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

Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research, 35*(6), 382–385. https://doi.org/10.1097/00006199-198611000-00017

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods, 4*(1), 84–99. https://doi.org/10.1037/1082-989X.4.1.84

Maharudin, L. O., Wahyudin, & Heryani, R. (2026). The influence of applying the GASING method on the addition skills of third-grade elementary school students in Batauga District, South Buton Regency. *Journal of Educational Sciences, 10*(2), 2245–2261. https://doi.org/10.31258/jes.10.2.p.2245-2261

Mayer, R. E. (2019). Computer games in education. *Annual Review of Psychology, 70*, 531–549. https://doi.org/10.1146/annurev-psych-010418-102917

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica, 22*(3), 276–282. https://doi.org/10.11613/BM.2012.031

McNeil, N. M., Jordan, N. C., Viegut, A. A., & Ansari, D. (2025). What the science of learning teaches us about arithmetic fluency. *Psychological Science in the Public Interest, 26*(1), 10–57. https://doi.org/10.1177/15291006241287726

Nurfurqon, F. F. (2017). *Pengaruh model value clarification technique terhadap motivasi dan hasil belajar matematika di sekolah dasar* [Unpublished undergraduate thesis].

Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis, 26*(3), 237–257. https://doi.org/10.3102/01623737026003237

Pan, Y., Ke, F., & Xu, X. (2022). A systematic review of the role of learning games in fostering mathematics education in K-12 settings. *Educational Research Review, 36*, Article 100448. https://doi.org/10.1016/j.edurev.2022.100448

Piaget, J., & Inhelder, B. (1969). *The psychology of the child*. Basic Books.

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*(5), 879–903. https://doi.org/10.1037/0021-9010.88.5.879

Pratiwi, S. Z., & Yamin, Y. (2026). The effect of using pizza-shaped game media in learning fractions on the mathematics learning outcomes of fifth-grade students at SDN Susukan 09 Pagi, East Jakarta. *Mathline: Jurnal Matematika dan Pendidikan Matematika, 11*(3), 771–786. https://doi.org/10.31943/mathline.v11i3.1188

Reeve, J., & Cheon, S. H. (2021). Autonomy-supportive teaching: Its malleability, benefits, and potential to improve educational practice. *Educational Psychologist, 56*(1), 54–77. https://doi.org/10.1080/00461520.2020.1862657

Reichardt, C. S. (2019). *Quasi-experimentation: A guide to design and analysis*. Guilford Press.

Russo, J., Kalogeropoulos, P., Bragg, L. A., & Heyeres, M. (2024). Non-digital games that promote mathematical learning in primary years students: A systematic review. *Education Sciences, 14*(2), Article 200. https://doi.org/10.3390/educsci14020200

Ryan, R. M., & Deci, E. L. (2017). *Self-determination theory: Basic psychological needs in motivation, development, and wellness*. Guilford Press.

Ryan, R. M., & Deci, E. L. (2020). Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions. *Contemporary Educational Psychology, 61*, Article 101860. https://doi.org/10.1016/j.cedpsych.2020.101860

Sailer, M., Hense, J. U., Mayr, S. K., & Mandl, H. (2017). How gamification motivates: An experimental study of the effects of specific game design elements on psychological need satisfaction. *Computers in Human Behavior, 69*, 371–380. https://doi.org/10.1016/j.chb.2016.12.033

Sailer, M., & Homner, L. (2020). The gamification of learning: A meta-analysis. *Educational Psychology Review, 32*(1), 77–112. https://doi.org/10.1007/s10648-019-09498-w

Schäfer, T., & Schwarz, M. A. (2019). The meaningfulness of effect sizes in psychological research: Differences between sub-disciplines and the impact of potential biases. *Frontiers in Psychology, 10*, Article 813. https://doi.org/10.3389/fpsyg.2019.00813

Schukajlow, S., Rakoczy, K., & Pekrun, R. (2023). Emotions and motivation in mathematics education: Where we are today and where we need to go. *ZDM – Mathematics Education, 55*(2), 249–267. https://doi.org/10.1007/s11858-022-01463-2

Schunk, D. H., & DiBenedetto, M. K. (2020). Motivation and social cognitive theory. *Contemporary Educational Psychology, 60*, Article 101832. https://doi.org/10.1016/j.cedpsych.2019.101832

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

Skinner, E. A., Kindermann, T. A., & Furrer, C. J. (2009). A motivational perspective on engagement and disaffection: Conceptualization and assessment of children's behavioral and emotional participation in academic activities in the classroom. *Educational and Psychological Measurement, 69*(3), 493–525. https://doi.org/10.1177/0013164408323233

Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education, 48*(6), 1273–1296. https://doi.org/10.1007/s11165-016-9602-2

Uno, H. B. (2023). *Teori motivasi dan pengukurannya: Analisis di bidang pendidikan* (Edisi ke-17). Bumi Aksara.

Vansteenkiste, M., Ryan, R. M., & Soenens, B. (2020). Basic psychological need theory: Advancements, critical themes, and future directions. *Motivation and Emotion, 44*(1), 1–31. https://doi.org/10.1007/s11031-019-09818-1

Wakano, A. (2019). Nilai-nilai pendidikan multikultural dalam kearifan lokal masyarakat Maluku. *al-Iltizam: Jurnal Pendidikan Agama Islam, 4*(2), 26–43. https://jurnal.iainambon.ac.id/index.php/ALT/article/view/1006

Wigfield, A., & Eccles, J. S. (2020). 35 years of research on students' subjective task values and motivation: A look back and a look forward. In A. J. Elliot (Ed.), *Advances in motivation science* (Vol. 7, pp. 161–198). Elsevier. https://doi.org/10.1016/bs.adms.2019.05.002