Journal of Research in Instructional
e-ISSN: 2776-222X
Vol. 6(1) 2026, pp. 413 – 425
https://doi.org/10.30862/jri.v6i1.64

Gamification versus traditional teaching: Impacts on mathematics learning outcomes and engagement

Edy Sucipto*, Ibrahim Bafadal, Siti Mas'ula

Universitas Negeri Malang, Indonesia

Submitted: 25-09-2025

Accepted: 31-01-2026

Published: 30-06-2026

Abstract: This quasi-experimental study examined the effectiveness of low-tech gamified mathematics instruction compared with traditional teaching in an Indonesian elementary school. Intact classes at SD Negeri 001 Tana Tidung (Sesayap, North Kalimantan) were assigned to gamification (*n*=47) or traditional (*n*=46). Both groups completed a 25-item pretest and posttest across 4–6 sessions. The primary outcome was posttest achievement analyzed with ANCOVA (covariate: pretest); secondary indicators were N-Gain, Hedges' g, mastery (≥70), on-task engagement, and treatment fidelity. Results showed a significant group effect, $F(1,89)=11.0$, $p=0.0012$, partial $\eta^2=0.11$. Adjusted means favored gamification (79.2; 95% CI [77.0, 81.4]) over traditional (72.7; [70.4, 75.0]), Δ=6.5; g≈0.70. N-Gain was 0.53 (gamification) vs 0.37 (traditional). Mastery reached 83.0% vs 63.0% (OR=2.60, 95% CI [1.20, 5.60]). Engagement was higher under gamification (78.6% vs 68.1%) and positively predicted posttest (β=0.28, *p*=0.004). Fidelity averaged 88% (κ=0.79). Findings indicate that progress- and process-oriented gamification with immediate feedback improves learning outcomes and classroom engagement, offering a feasible, scalable option for resource-limited settings.

Keywords: Elementary mathematics, engagement, gamification, low-tech

Abstrak: Penelitian quasi-experimental ini menilai efektivitas pembelajaran matematika bergamifikasi berbasis *low-tech* dibanding pembelajaran tradisional di SD Negeri 001 Tana Tidung (Sesayap, Kalimantan Utara). Kelas utuh ditetapkan sebagai kelompok gamification (*n*=47) atau tradisional (*n*=46). Kedua kelompok mengerjakan pretest–posttest (25 butir) selama 4–6 sesi. Hasil utama dianalisis dengan ANCOVA (kovariat: pretest); indikator tambahan meliputi N-Gain, Hedges' g, ketuntasan (≥70), *on-task engagement*, dan *treatment fidelity*. Terdapat perbedaan kelompok yang signifikan, $F(1,89)=11{,}0$, $p=0{,}0012$, *partial* $\eta^2=0{,}11$. *Adjusted mean* lebih tinggi pada *gamification* (79,2; IK95% [77,0; 81,4]) dibanding tradisional (72,7; [70,4; 75,0]), Δ=6,5; g≈0,70. N-Gain mencapai 0,53 (gamification) vs 0,37 (tradisional). Ketuntasan 83,0% vs 63,0% (OR=2,60, IK95% [1,20; 5,60]). *Engagement* lebih tinggi pada *gamification* (78,6% vs 68,1%) dan memprediksi *posttest* (β=0,28, *p*=0,004). *Fidelity* rata-rata 88% (κ=0,79). Temuan menunjukkan gamifikasi berorientasi progres dan proses dengan umpan balik segera meningkatkan hasil belajar dan *engagement*, serta layak diterapkan di konteks sumber daya terbatas.

Kata kunci: Matematika sekolah dasar, keterlibatan, gamifikasi, teknologi sederhana

*Corresponding author: suciptoedy64@gmail.com

INTRODUCTION

Elementary school mathematics education requires pedagogical strategies that not only improve learning outcomes but also sustain student engagement while supporting the transition from concrete to symbolic representations. This transition is particularly critical at the elementary level, where students' cognitive development relies heavily on meaningful, process-oriented, and representationally rich learning experiences (Nasir et al., 2024; Runniarsiti et al., 2025, Siregar et al., 2025). One widely researched approach addressing these needs is gamification, defined as the application of game design elements such as points, badges, levels, leaderboards, narrative, and immediate feedback into non-game contexts to enhance learning experiences and motivation (Deterding et al., 2011).

Quantitative meta-analytic evidence indicates that gamification has a small-to-medium positive effect on cognitive achievement, motivation, and learning behavior, making it a promising pedagogical intervention in elementary school settings (Sailer & Homner, 2020).

Within the domain of mathematics education for young learners, classroom-based studies suggest that gamification designs combining competitive, collaborative, and adaptive mechanics can improve both academic performance and student engagement. However, these effects are highly contingent on instructional design quality, representational support, and classroom dynamics (Jagušt et al., 2018). Longitudinal evidence further cautions that gamification designs that overemphasize competition and extrinsic rewards may lead to a decline in students' intrinsic motivation over time (Hanus & Fox, 2015). Consequently, the critical issue in gamified learning is not merely whether gamification is effective, but how it is pedagogically designed—particularly whether elements such as progress-based leaderboards, process-oriented badges, and immediate feedback are aligned with elementary students' cognitive and motivational needs.

Field observations conducted at SD Negeri 001 Tana Tidung, Sesayap District, Tana Tidung Regency (Kalimantan), revealed several practical challenges in mathematics instruction. Teaching practices tended to be teacher-centered, dominated by individual drills; feedback was frequently delayed, limiting opportunities for immediate formative responses; multi-representational practices supporting the transition from concrete to symbolic understanding were inconsistently implemented; student engagement declined during independent work, with some students displaying off-task behavior; game elements, when used, were ad hoc and predominantly competitive; and limited access to digital devices necessitated low-tech instructional solutions. These contextual challenges underscore the need for a feasible gamification design that emphasizes learning progress and process quality, provides rapid feedback, and systematically supports concrete-to-symbolic transitions. Recent empirical evidence in gamified mathematics learning further highlights the central role of multi-representational scaffolding and immediate feedback in maximizing instructional impact (Chen et al., 2023).

Gamification was selected as the intervention of choice for addressing the mathematics learning problems observed at SD Negeri 001 Tana Tidung because its core design elements directly target the specific instructional gaps identified during field observations, namely delayed feedback, teacher-centered drilling, and inconsistent multi-representational support. Shute (2008) established that formative feedback delivered promptly after a response is one of the most powerful mechanisms for correcting misconceptions and sustaining learning momentum, which directly addresses the delayed-feedback problem observed in conventional instruction at the school. Landers (2014) further theorized that game attributes do not influence learning outcomes directly but operate indirectly through learning-relevant behaviors such as sustained time-on-task and attention, offering a testable mechanism for why gamification might improve achievement beyond simply making lessons more enjoyable.

Plass et al. (2015) added that well-designed game-based learning environments support cognitive engagement by embedding content within meaningful narrative and representational structures, which is particularly relevant for elementary students transitioning from concrete to symbolic mathematical understanding. Deterding et al. (2011) and Handaryani et al. (2026) provided the foundational definition of gamification as the application of game design elements, such as points, badges, levels, and immediate feedback, in non-game contexts, a definition that anchors the specific elements selected for this study's low-tech design. Taken together, these theoretical perspectives suggest that gamification is not merely a motivational add-on but a structured pedagogical mechanism capable of addressing the precise instructional deficits documented in the field observations.

Beyond addressing feedback and representational gaps, gamification is expected to improve learning outcomes through its capacity to increase student engagement, which in turn supports deeper cognitive processing of mathematical content. Sailer and Homner (2020) meta-analysis demonstrated small-to-medium positive effects of gamification on cognitive, motivational, and behavioral learning outcomes across a wide range of educational contexts, providing empirical grounding for expecting a measurable achievement advantage in the present study. Hamari et al. (2014) and Ahmad et al. (2026) found that gamification most consistently improves engagement and enjoyment, which are theorized to function as proximal drivers of the more distal outcome of academic achievement. Mekler et al. (2017) showed that individual game elements such as points, levels, and leaderboards operate by satisfying distinct psychological needs rather than through a single uniform mechanism, which justifies the deliberate selection of progress-oriented rather than purely competitive elements in the present design.

Sailer et al. (2017) similarly demonstrated experimentally that specific game design elements differentially satisfy needs for competence, autonomy, and relatedness, and that this need satisfaction subsequently predicts intrinsic motivation and behavioral engagement. Building on this evidence, the gamification design implemented at SD Negeri 001 Tana Tidung emphasized progress-based leaderboards, process-oriented badges, tiered challenges, and immediate feedback, with the explicit expectation that these elements would increase on-task engagement and, through this engagement, produce measurable gains in mathematics learning outcomes relative to traditional instruction. This study aims to examine the effectiveness of gamification-based mathematics instruction compared to traditional teaching in improving elementary students' learning outcomes while controlling for prior knowledge.

METHOD

Research design

The study uses a quasi-experimental, non-equivalent control group design with a pretest–posttest. The unit of assignment is the learning group (class), while the unit of analysis is the student. The experimental group receives gamification-based mathematics instruction (using game design elements such as points, badges, leaderboards, challenges, and immediate feedback to increase engagement), while the control group receives traditional instruction (expository and drills). The gamification framework refers to the classic definition that gamification is the use of game design elements in non-game contexts to enhance user experience and engagement (Chen et al., 2023; Deterding et al., 2011; Jagušt et al., 2018).

Design and analysis justification

In a non-equivalent pretest–posttest design, ANCOVA with the pretest as a covariate was selected as the primary analytical approach to increase statistical power and control for initial group differences, compared to analyses based solely on change scores. Oakes and Feldman (2001) demonstrated an integrated power test formulation for this design and recommended the use of ANCOVA. Breukelen (2006) noted that although ANCOVA is more powerful in randomized studies, both ANCOVA and change-score analyses may be biased in non-randomized designs if statistical assumptions are violated. Accordingly, this study explicitly tested the assumptions of linearity and homogeneity of regression slopes prior to conducting ANCOVA.

Effectiveness in this study is operationally defined as the difference in adjusted mean posttest scores between groups after controlling for pretest performance using ANCOVA. As complementary indicators, normalized gain (N-Gain) was calculated to describe relative pre-to-post improvement in a form that is communicative for practitioners (Hake, 1998), and effect size was reported using Hedges' g, which provides a less biased estimate for small-to-moderate sample sizes (Lakens, 2013).

The selection of gamification as the intervention is grounded in meta-review evidence indicating a small-to-medium positive impact on cognitive and motivational outcomes, with effectiveness varying according to the configuration of game mechanics (e.g., competitive versus collaborative designs and task adaptivity) and the developmental context of young learners. Empirical studies in elementary mathematics classrooms consistently show that well-designed gamification improves both learning performance and motivation compared to non-gamified or traditional instructional conditions (Chen et al., 2023; Jagušt et al., 2018; Sailer & Homner, 2020).

Participants and setting

The study was conducted with 5th-grade students at SD Negeri 001 Tana Tidung, Indonesia, within the context of routine mathematics instruction. Two relatively equivalent classes were selected as the sample through cluster/purposive sampling based on class availability and curriculum suitability. When more than two classes were available, the researcher first performed a simple matching using the average pretest scores or previous semester's mathematics report card grades, then assigned one class to the experimental group and the other to the control group. All students who participated in at least 80% of the intervention sessions, completed the pretest and posttest, and had obtained written permission from a parent or guardian were included in the analysis. Participants with individual curriculum adaptation needs that caused a significant difference in material, or who were absent on the day the posttest was administered, were excluded from the main analysis. The sample size was considered a priori: to detect an effect size of a medium magnitude (Cohen's $d \approx 0.50$) at a significance level of 0.05 and a power of 0.80, an independent t-test approach would require about 64 students per group. However, since the main analysis uses ANCOVA with a conservative pre-to-post correlation of about 0.60, the sample need can be reduced to around 41 students per group. Considering potential attrition, this study targeted at least 45 students per group to ensure adequate statistical power.

Intervention procedure: Gamification group

The intervention for the experimental group was designed to improve learning outcomes in arithmetic topics, particularly integers, fractions, decimals, measurements, and perimeter and area, by increasing engagement and providing quick feedback. Gamification is understood as the application of game design elements in a non-game context. In this study, the elements used included points, badges, levels, a narrative/story, immediate feedback, a progress bar, tiered challenges, and a "no-stakes retries" mechanism that allowed students to reattempt without a report card grade penalty. The intervention lasted for four to six meetings, each lasting 70 to 90 minutes. The media used were low-tech to be easily replicable: printed point and badge cards, a progress board in the classroom, mission sheets, and, when available, simple digital support for scoring. The learning flow began with an introductory story about "Pasar Nusantara" (The Archipelago Market), which presented a problem context involving calculating prices and change. In this session, students completed tiered missions and received immediate feedback via answer cards so their progress was immediately visible on the progress bar.

The next sessions focused on fraction representation through "Cake Cutting" learning stations that guided the transition from concrete to symbolic models; the challenge of finding a common denominator was framed as a level to be completed with the support of immediate feedback. In the third session, fraction operations (addition and subtraction) were packaged in tiered challenges, ranging from basic to intermediate levels, and a co-op quest that encouraged pairs to strategize problem-solving. Badges were awarded based on accuracy and the use of appropriate representations. The fourth session targeted fraction-to-decimal conversion and rounding through repetitive mini-games that allowed for retries without grade consequences, so students felt safe taking cognitive risks by trying new strategies. The fifth session shifted to geometry and measurement through an "Architectural Garden" scenario, where students calculated the perimeter and area of various designs. Groups with consistent progress earned process badges that emphasized the quality of the process, not just the final outcome.

The intervention concluded with an integrated review session framed as a "boss battle," a mixed-question package with a reasonable time limit, followed by a reflection on strategies and classroom feedback to connect the game experience with key concepts. Throughout the intervention, the displayed leaderboard was progress-based (percentage of missions completed) to minimize excessive competition and highlight individual growth. The effective learning time (time-on-task) was matched with the control group so that outcome differences would not be influenced by the duration of learning exposure.

Comparison condition: Traditional group

The control group received traditional instruction with a brief expository pattern followed by examples and individual practice using textbooks or Student Activity Sheets. Feedback was provided after classroom correction, and the material coverage, topic sequence, and duration of each meeting were aligned with the gamification group. Thus, the primary difference between the groups lay in the presentation strategy and reinforcement mechanisms, not in the content or learning time.

Teacher training and treatment fidelity

Before implementation, participating teachers received a two-to-three-hour briefing on the intervention's objectives, the flow of each session, techniques for providing immediate feedback, and how to use the gamification tools (point cards, badges, progress boards, and mission sheets). Implementation fidelity was monitored through periodic classroom observations by two independent raters on two random visits per class, with a target inter-rater reliability (Cohen's κ) of at least 0.75. The aspects of fidelity observed included consistency in delivering the narrative, clarity of level objectives, accuracy in awarding points or badges according to set indicators, speed of feedback (ideally not exceeding 30 seconds after a student's response), student comprehension of the game rules, and sufficient time for meaningful practice. All observation findings were recorded for process audit and reporting purposes.

Instruments

The main instrument was a Learning Outcomes Test consisting of 25 multiple-choice items constructed based on a competency grid for arithmetic (integers, fractions–decimals, measurement, and perimeter–area). Content validity was established through expert judgment using Aiken's V coefficient, with all items meeting the minimum criterion of ≥ 0.80. The pilot testing yielded a Cronbach's α coefficient exceeding the minimum acceptable threshold of 0.70, indicating satisfactory internal consistency of the instrument**.** To minimize practice effects in the pretest–posttest design, two equivalent forms were developed: Form A (pretest) and Form B (posttest), sharing 7–10 identical anchor items. This anchoring strategy ensured score comparability while enabling item analysis across forms. Scoring was dichotomous (correct = 1; incorrect/blank = 0), with raw scores converted to a 0–100 scale to facilitate pedagogical interpretation and reporting of effect sizes (Lakens, 2013).

As a supplementary process measure, an Engagement Observation Sheet was used to document on-task behavior during instruction. Interval time sampling was conducted every two minutes (approximately 20 intervals per session) to estimate the proportion of time students remained on task, as indicated by attention to instruction, engagement in problem-solving, and participation in relevant discussion. Two independent observers applied the same rubric, and inter-rater reliability was evaluated using Cohen's κ with a target value of ≥ 0.75. Evidence from prior gamification research in mathematics supports the use of such engagement indicators for interpreting mechanisms underlying learning gains (Chen et al., 2023; Jagušt et al., 2018).

Research procedure

The pre-implementation phase included obtaining ethical approval and school permission, socializing the study to parents/guardians, training teachers on the intervention flow and immediate feedback techniques, and pilot testing the instruments to ensure validity and reliability and to conduct a distractor analysis before full use. If more than two classes were available, a simple matching was performed based on average mathematics scores or pretest scores to reduce initial group imbalance (Breukelen, 2006; Oakes & Feldman, 2001).

In Week 0, both groups simultaneously took the pretest (Form A) for 45 minutes under the same supervision. Weeks 1–3 consisted of four to six intervention sessions, each lasting 70–90 minutes; time-on-task was recorded to be equivalent across conditions. The experimental group received gamification-based instruction with elements of points, badges, levels, narrative, a progress bar, tiered challenges, and no-stakes retries, designed to increase engagement and provide immediate feedback. The control group followed a traditional approach (expository–example–practice) with matched material coverage, topic sequence, and duration. During this period, treatment fidelity was observed at least twice per class by trained observers. In Week 3/4, both groups took the posttest (Form B) on the same day (45 minutes). Afterward, additional data were collected, including attendance records, notes on implementation challenges, and a summary of fidelity findings for process auditing. The data were then managed through double data entry, outlier examination ($|z| > 3$ criterion), and handling missing data: <5% were removed listwise, while ≥5% were handled with multiple imputation (MI) on 20 datasets for sensitivity analysis.

Data analysis

Data analysis began with assumption testing for the primary model, including normality of residuals (Shapiro–Wilk test) and homogeneity of variances (Levene's test). Baseline equivalence between groups was assessed using an independent *t*-test on pretest scores, and the standardized mean difference (baseline Hedges' *g*) was reported to describe initial group similarity (Breukelen, 2006; Oakes & Feldman, 2001). The primary inferential analysis employed ANCOVA, with posttest scores as the dependent variable, instructional group (gamification vs. traditional) as the fixed factor, and pretest scores as the covariate. Results are reported using adjusted means with 95% confidence intervals, F values, p values, and partial η^2. Partial η^2 values were subsequently converted to Hedges' *g* to support pedagogical interpretation of effect magnitude (Lakens, 2013).

Additional analyses were conducted to support interpretation, not as alternative main tests**.** These included calculation of N-Gain scores for each student and group-level averages, analysis of mastery attainment (scores ≥ 70) using logistic regression with pretest scores as a covariate, and analysis of engagement data to examine process differences between groups. All analyses reported in this section correspond directly to those presented in the Results section**.** Effect sizes are reported using Hedges' *g* and N-Gain classifications (low < 0.30; medium 0.30–0.70; high > 0.70) following conventions established in interactive engagement research (Hake, 1998; Lakens, 2013).

Validity and bias reduction

Internal validity in a non-equivalent control group is supported by a combination of a pretest–posttest design and statistical control using ANCOVA, which helps manage initial differences between groups (Breukelen, 2006; Oakes & Feldman, 2001). Selection bias is minimized through simple matching before the intervention and the use of the pretest as a covariate. Performance bias is reduced by providing a written teaching script and aligning time-on-task between both conditions. Detection bias is mitigated through partial blinding during scoring (teachers do not grade their own class's posttests) and the use of a uniform scoring template. Instrumentation bias is minimized by using equivalent Form A/B with anchor items, accompanied by a post-pilot item analysis to ensure reasonable distractor function. Attrition bias is noted by documenting the reasons for absences and is handled in reporting by presenting ITT and PP results side-by-side.

Ethical considerations

All procedures comply with the principles of informed consent and child safeguarding. Ethical permission was obtained from the research ethics committee, and operational permission was secured from the school administration. Parents/guardians signed a consent form after receiving an information sheet that explained the study's purpose, benefits, minimal potential risks, and the right to withdraw at any time without academic consequences. Student identities are de-identified in all databases, and answer sheets are stored separately from name lists. Following the study, both groups were provided with equal remedial opportunities to ensure pedagogical equity.

RESULTS AND DISCUSSION

A total of 93 students were included in the final analysis, comprising 47 students in the gamification group and 46 students in the traditional group. Initial baseline equivalence between the two groups was examined using pretest mathematics achievement scores. The descriptive statistics indicated that both groups started from a relatively similar level of prior knowledge, with Mgam = 56.3 (SD = 9.4) and Mtrad = 55.8 (SD = 8.7). An independent samples t-test confirmed the absence of significant differences at baseline, $t(91) = 0.30$, $p = 0.76$, and the corresponding Hedges' g value showed a very small effect size ($g = 0.05$). No missing data were found in the primary analysis variables, and all sessions retained complete documentation of time-on-task and treatment fidelity. These results indicate that pre-intervention conditions were sufficiently balanced to proceed with comparative analysis across groups. The descriptive results for pretest and posttest performance for both groups are presented in Table 1.

Table 1. Pretest–posttest descriptives per group

Group	n	Pretest M (SD)	Posttest M (SD)
Gamification	47	56,3 (9,4)	79,6 (8,8)
Traditional	46	55,8 (8,7)	72,3 (9,1)

Following the descriptive evidence (Table 1), further inferential testing was conducted to evaluate the main effect of the intervention. Preliminary assumption checks for the ANCOVA model indicated that the residuals were reasonably normally distributed (Shapiro–Wilk $W = 0.98$, $p = 0.17$), homogeneity of variances was met (Levene's $p = 0.28$), and the assumption of homogeneity of regression slopes was satisfied (pretest × group interaction: $F(1,88) = 0.26$, $p = 0.61$). These findings confirm that ANCOVA was appropriate for the dataset. After controlling for pretest differences, ANCOVA results showed a significant effect of instructional method on posttest scores: $F(1,89) = 11.0$, $p = 0.0012$, partial $\eta^2 = 0.11$. The adjusted means indicated that the gamification group outperformed the traditional group, with Madj(gam) = 79.2 (95% CI [77.0; 81.4]) and Madj(trad) = 72.7 (95% CI [70.4; 75.0]). This 6.5-point difference represents a medium practical effect size (gANCOVA ≈ 0.70), confirming a meaningful learning advantage under gamification. The full ANCOVA output is summarized in Table 2.

Table 2. ANCOVA (posttest controlled by pretest)

Source	df	F	p	partial η^2	$M_a dj$ (95% CI)
Group (gam vs trad)	1	11,0	0,0012	0,11	Gam: 79,2 [77,0; 81,4] — Trad: 72,7 [70,4; 75,0]

In addition to posttest performance, gain-score analysis was conducted to examine relative improvement across groups. The average gain score for the gamification group was 23.3 points, compared to 16.5 points in the traditional group, and the difference was statistically significant (Welch's $t \approx 4.07$, $p < 0.001$). The normalized gain analysis (N-Gain) yielded medium-category scores for both groups, although gamification demonstrated a clear advantage (0.53 vs. 0.37). This pattern reinforces the outcome of the ANCOVA results and suggests that gamification not only improved end-point achievement but also supported stronger relative learning growth. These results are shown in Table 3.

Table 3. Gain & n-gain

Group	Gain M (SD)	N-Gain M (SD)	Category
Gamification	23,3	0,53	Medium
Tradisional	16,5	0,37	Medium

Mastery-level analysis further supported the impact of the intervention. Using ≥70 as the mastery threshold, 83.0% (39/47) of students in the gamification group met the criterion, compared with 63.0% (29/46) in the traditional group. Logistic regression controlling for pretest differences showed that students in the gamification group were 2.6 times more likely to reach mastery (OR = 2.60, 95% CI [1.20; 5.60], $p = 0.015$), reflecting a meaningful shift in proficiency outcomes attributable to the instructional design. Mastery comparisons are presented in Table 4.

Table 4. Proportion of achievements ≥ 70 and logistic analysis (pretest control)

Indicator	Gamification	Traditional	OR (95% CI)	p
≥ 70 point	83,0% (39/47)	63,0% (29/46)	2,60 [1,20; 5,60]	0,015

Process indicators collected during the intervention provide insight into classroom mechanisms associated with learning improvements. On-task engagement in the gamification condition was notably higher (78.6%) than in the traditional group (68.1%), with the difference of +10.5 percentage points reaching statistical significance ($t(91) = 5.12$, $p < 0.001$). Hierarchical regression analysis further indicated that engagement positively predicted posttest achievement ($\beta = 0.28$, $p = 0.004$), suggesting partial mediation via increased learning focus. Treatment fidelity was high, with 88% of indicators completed as intended and strong inter-rater reliability ($\kappa = 0.79$). These findings are summarized in Table 5.

Table 5. Process indicator

Indicator	Gamification	Traditional	Statistic
On-task engagement (%)	78,6 (8,9)	68,1 (10,2)	$t(91) = 5{,}12$, $p < 0{,}001$
Treatment fidelity	88%	—	$\kappa = 0{,}79$

Collectively, these results demonstrate that gamification-based learning produced stronger mathematical achievement outcomes than traditional instruction, even after adjusting for initial ability. Improvements were observed consistently across posttest scores, gain scores, and mastery attainment, indicating that the advantage of gamification was not confined to a single indicator but was reproduced across complementary ways of measuring learning.

The process indicators in Table 5 help explain why this achievement advantage emerged. On-task engagement was 10.5 percentage points higher in the gamification group (78.6% vs. 68.1%), and engagement positively predicted posttest achievement ($\beta = 0.28$, $p = 0.004$), indicating that a meaningful part of the learning advantage operated through students remaining more consistently focused on the mathematical task rather than through the gamified material being inherently easier or covering different content. Because material coverage, topic sequence, and time-on-task were held constant between groups, this engagement gap is best attributed to the presentation strategy itself, that is, to the points, badges, tiered challenges, and immediate feedback used in the gamification condition, rather than to differences in instructional time or content. Equally important, treatment fidelity was high (88% of indicators completed as intended, $\kappa = 0.79$), which indicates that the intervention was implemented largely as designed in both conditions; this reduces the plausibility of an alternative explanation in which the achievement gap reflects inconsistent or poorly delivered instruction rather than the instructional design itself.

Three specific features of the present design plausibly explain why gamification outperformed traditional instruction in this study. First, feedback in the gamification condition was immediate, delivered through answer cards within the session, whereas feedback in the traditional condition followed classroom correction after the fact; this matches the delayed-feedback problem identified in the initial field observations and is consistent with formative-feedback theory, which holds that timely feedback allows students to correct misconceptions while the relevant problem is still cognitively active (Shute, 2008). Second, the leaderboard used in this study was progress-based rather than purely competitive, tracking the percentage of missions completed rather than ranking students against one another; this design choice was intended to preserve the motivational benefits of visible progress while avoiding the decline in intrinsic motivation that has been documented when gamification over-emphasizes competition and extrinsic reward (Hanus & Fox, 2015). Third, the intervention embedded multi-representational scaffolding directly into the game narrative, for example, the transition from concrete fraction-cutting stations to symbolic notation in the "Cake Cutting" sessions, which aligns with the representational-transition difficulty noted in the introduction and is consistent with evidence that combining representations with structured game elements supports stronger learning outcomes for young learners (Ainsworth, 2006; Plass et al., 2015).

These patterns are consistent with, and extend, prior evidence in the gamification literature. The magnitude of the effect observed here ($g \approx 0.70$) is at the upper end of the small-to-medium effects typically reported in broader meta-analytic reviews (Sailer & Homner, 2020), which may reflect the deliberate emphasis on progress and process quality rather than competition in the present design, in line with evidence that the specific configuration of game elements, rather than gamification in general, determines whether motivational benefits are realized (Hamari et al., 2014; Mekler et al., 2017; Sailer et al., 2017). The engagement-mediated pattern observed in this study is also consistent with theoretical accounts holding that game attributes influence achievement indirectly, through learning-relevant behaviors such as time-on-task, rather than through any direct effect on cognitive processing (Landers, 2014).

From a practical standpoint, the fact that these gains were achieved using low-tech components, such as printed point and badge cards, a classroom progress board, and mission sheets, indicates that meaningful gamification benefits do not require advanced digital infrastructure. This reinforces findings from both classroom-based and online studies showing that such adaptations are feasible and effective in resource-limited educational contexts (Domínguez et al., 2013; Looyestyn et al., 2017), and it directly supports the practical motivation for this study, namely identifying a scalable instructional option for schools such as SD Negeri 001 Tana Tidung that cannot rely on digital gamification platforms.

CONCLUSION

This quasi-experimental study demonstrates that gamification-based mathematics instruction was more effective than traditional teaching in improving elementary students' learning outcomes at SD Negeri 001 Tana Tidung, even after statistically accounting for students' prior ability. The gamification group achieved higher final scores, greater relative improvement from pretest to posttest, and a substantially larger proportion of students reaching the minimum mastery criterion, indicating that the advantage was consistent across several different ways of measuring learning rather than an artifact of any single indicator. Practically, these findings confirm that a progress-oriented, low-tech gamification design, one that emphasizes process quality, immediate feedback, and concrete-to-symbolic representational support, can be feasibly embedded in regular classroom routines without reliance on digital infrastructure, offering a scalable option for numeracy recovery in resource-limited settings. These results should nonetheless be interpreted in light of the study's limitations, including its reliance on intact classes rather than random assignment, a relatively short intervention period, and a single-school setting. Future research involving multiple schools, longer implementation periods, delayed posttests to assess retention of learning, and analytical approaches robust to classroom-level clustering would help strengthen confidence in the generalizability of these findings.

REFERENCES

Ahmad, Y., Gani, H. A., & Hakim, A. (2026). The development of a gamification-based evaluation tool for the creative projects and entrepreneurship subject in grade XI. *Journal of Research in Instructional*, *6*(1), 158–173. https://doi.org/10.30862/jri.v6i1.19

Ainsworth, S. (2006). DeFT: A conceptual framework for considering learning with multiple representations. *Learning and Instruction*, *16*(3), 183–198. https://doi.org/10.1016/j.learninstruc.2006.03.001

Breukelen, G. J. P. (2006). ANCOVA versus change from baseline: More power in randomized studies, more bias in nonrandomized studies. *Journal of Clinical Epidemiology*, *59*(9), 920–925. https://doi.org/10.1016/j.jclinepi.2006.02.007

Chen, M.-F., Chen, Y.-C., Zuo, P.-Y., & Hou, H.-T. (2023). Design and evaluation of a remote synchronous gamified mathematics teaching activity that integrates multi-representational scaffolding and a mind tool for gamified learning. *Education and Information Technologies*, *28*(10), 13207–13233. https://doi.org/10.1007/s10639-023-11708-6

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: Defining gamification. In *CHI '11 Extended Abstracts on Human Factors in Computing Systems* (pp. 2425–2428). Association for Computing Machinery. https://doi.org/10.1145/1979742.1979575

Domínguez, A., Saenz-de-Navarrete, J., de-Marcos, L., Fernández-Sanz, L., Pagés, C., & Martínez-Herráiz, J.-J. (2013). Gamifying learning experiences: Practical implications and outcomes. *Computers & Education*, *63*, 380–392. https://doi.org/10.1016/j.compedu.2012.12.020

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

Hamari, J., Koivisto, J., & Sarsa, H. (2014). Does gamification work? A literature review of empirical studies on gamification. In *2014 47th Hawaii International Conference on System Sciences (HICSS)* (pp. 3025–3034). IEEE. https://doi.org/10.1109/HICSS.2014.377

Handaryani, H., Gani, H. A., & Muin, A. (2026). The influence of local wisdom-based gamification on students' literacy and numeracy in mathematics learning. *Journal of Research in Instructional*, *6*(1), 308–322. https://doi.org/10.30862/jri.v6i1.40

Hanus, M. D., & Fox, J. (2015). Assessing the effects of gamification in the classroom: A longitudinal study on intrinsic motivation, social comparison, satisfaction, effort, and academic performance. *Computers & Education*, *80*, 152–161. https://doi.org/10.1016/j.compedu.2014.08.019

Jagušt, T., Botički, I., & So, H.-J. (2018). Examining competitive, collaborative and adaptive gamification in young learners' math learning. *Computers & Education*, *125*, 444–457. https://doi.org/10.1016/j.compedu.2018.06.022

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

Landers, R. N. (2014). Developing a theory of gamified learning: Linking serious games and gamification of learning. *Simulation & Gaming*, *45*(6), 752–768. https://doi.org/10.1177/1046878114563660

Looyestyn, J., Kernot, J., Boshoff, K., Ryan, J., Edney, S., & Maher, C. (2017). Does gamification increase engagement with online programs? A systematic review. *PLOS ONE*, *12*(3), 0173403. https://doi.org/10.1371/journal.pone.0173403

Mekler, E. D., Brühlmann, F., Tuch, A. N., & Opwis, K. (2017). Towards understanding the effects of individual gamification elements on intrinsic motivation and performance. *Computers in Human Behavior*, *71*, 525–534. https://doi.org/10.1016/j.chb.2015.08.048

Nasir, N. I. R. F., Mahanal, S., Ekawati, R., Damopolii, I., Supriyono, S., & Rahayuningsih, S. (2024). Primary school students' knowledge about animal life cycle material: The survey study. *Journal of Research in Instructional*, *4*(1), 253–262.

https://doi.org/https://doi.org/10.30862/jri.v4i1.320
Runniarsiti, R., Arismunandar, A., & Febriati, F. (2025). Development of web-based learning multimedia to improve students understanding of mathematical concepts in elementary schools. *Journal of Research in Instructional*, *5*(2), 698–713. https://doi.org/10.30862/jri.v5i2.820

Oakes, J. M., & Feldman, H. A. (2001). Statistical power for nonequivalent pretest–posttest designs: The impact of change-score versus ANCOVA models. *Evaluation Review*, *25*(1), 3–28. https://doi.org/10.1177/0193841X0102500101

Plass, J. L., Homer, B. D., & Kinzer, C. K. (2015). Foundations of game-based learning. *Educational Psychologist*, *50*(4), 258–283. https://doi.org/10.1080/00461520.2015.1122533

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

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, *78*(1), 153–189. https://doi.org/10.3102/0034654307313795

Siregar, N. N., Firmansyah, F., Natasya, D., & Damopolii, I. (2025). Efektivitas Project Based Learning Berbantuan Wordwall untuk Meningkatkan Prestasi Belajar Ditinjau dari Rasa Ingin Tahu. *Cokroaminoto Journal of Primary Education*, *8*(1), 32–41. https://doi.org/10.30605/cjpe.812025.5338