

Cross-national effects of a gamified algebra learning intervention: A mixed-design analysis of country and gender differences in Indonesia and Malaysia

Ramadoni¹ , Mazlini Adnan^{1,2*} 

¹ National Research and Innovation Agency, Jakarta, INDONESIA

² Universiti Pendidikan Sultan Idris, Tanjung Malim, MALAYSIA

Received 23 February 2026 • Accepted 13 April 2026

Abstract

Algebra achievement remains a persistent challenge in secondary mathematics education, and although gamified instruction has shown promise, evidence for its effectiveness across national contexts remains limited. This study investigated the cross-national effectiveness of AlgeUno, a classroom-based gamified algebra intervention, among secondary school students in Indonesia and Malaysia, two comparable Southeast Asian educational contexts selected because they share regional and curricular similarities while differing in instructional and classroom conditions that may shape intervention responsiveness. AlgeUno is a classroom-based gamification approach that adapts the mechanics of the UNO card game into structured algebra problem-solving activities. A quasi-experimental pre-/post-test design was implemented with 62 secondary students (Indonesia: $n = 31$; Malaysia: $n = 31$). A 2 (time: pre-test vs. post-test) \times 2 (country) \times 2 (gender) mixed-design ANOVA and subsequent ANCOVA were conducted to evaluate intervention effects while controlling baseline differences. Results revealed a highly significant main effect of time, $F(1, 58) = 695.69$, $p < .001$, $\eta^2p = .923$, indicating substantial improvement in algebra achievement following the intervention. Significant time \times country and time \times gender interactions suggested moderate variations in gain magnitude across contexts and gender groups. After controlling for pre-test scores, country differences remained significant, $F(1, 57) = 5.65$, $p = .021$, $\eta^2p = .095$, while gender differences were not significant. However, a significant country \times gender interaction indicated context-specific gender dynamics. The findings show that AlgeUno is a robust classroom gamification intervention for improving algebra learning across both settings, while also indicating that its effectiveness is conditional on baseline readiness and contextual implementation factors. These results support the use of structured gamified algebra instruction as a promising cross-context strategy, with evidence that structured classroom gamification can produce substantial learning gains while remaining sensitive to contextual and demographic factors.

Keywords: algebra learning, gamification, cross-national comparison, mixed-design ANOVA, gender differences, Southeast Asia

INTRODUCTION

Improving algebra achievement remains a persistent and globally recognized challenge in secondary mathematics education (Tuong et al., 2023). Algebra functions as a critical gateway to advanced mathematical reasoning (Dewi et al., 2026), STEM participation (Abd Jalil et al., 2024), and long-term academic attainment (Kizhukarakkatu, 2025). Students' early mastery of

algebraic concepts strongly predicts later success in higher-level mathematics and science disciplines. However, empirical evidence consistently shows that many students struggle with symbolic manipulation (Mutodi & Mosimege, 2021), structural understanding and the transition from arithmetic to generalized reasoning (Beck & da Silva, 2020). These difficulties are often characterized by procedural rigidity and

Contribution to the literature

- This study extends the literature on gamified mathematics instruction by providing cross-national validation beyond single-site implementation.
- It advances methodological rigor in educational intervention research by simultaneously modeling within-subject change and adjusted between-group differences.
- It contributes theoretically by framing instructional effectiveness within an ecological-interactionist perspective, emphasizing conditional robustness rather than universal portability of educational innovations.

fragmented conceptual knowledge (Stewart & Thomas, 2007).

International large-scale assessments further highlight substantial variability in mathematical achievement across countries, including within Southeast Asia (e.g., TIMSS and PISA reports). Cross-national disparities suggest that algebra performance is shaped not only by individual cognitive factors but also by curricular alignment (Suter, 2017), instructional practices (Bokhove, 2022), and sociocultural educational norms (Lau et al., 2022; Yang Hansen et al., 2024). While Southeast Asian countries such as Indonesia and Malaysia share regional proximity and certain curricular similarities, differences in pedagogical traditions, assessment orientation, and classroom ecology may produce distinct learning trajectories (Yunos et al., 2019).

Gamification has been increasingly proposed as a mechanism to address persistent challenges in algebra learning, aiming to enhance both conceptual understanding (Vanacore et al., 2023) and student engagement (García-López et al., 2023; Ratinho et al., 2026). The evidence from various studies supports the positive impact of gamification in educational settings, particularly in mathematics (Akgül & Güler, 2025). Structured gamification environments may promote cognitive engagement through rule-based interaction, competitive dynamics, and iterative problem exposure (Baah et al., 2024). Moreover, gamification can support multiple representations of algebraic structures, which are essential for conceptual development (Krath et al., 2021).

Although gamified and game-based instructional approaches have shown promising effects in mathematics education, their effectiveness is not uniform across contexts (Pan et al., 2025). As emphasized by Pathania et al. (2025) and Zabala-Vargas et al. (2021), instructional outcomes depend on implementation quality, learner readiness, and contextual alignment. Even when interventions are pedagogically well-designed, their impact may vary depending on curricular sequencing, classroom norms, and students' prior conceptual readiness (Running et al., 2023). A substantial body of cognitive research further demonstrates that prior knowledge serves as a central determinant of knowledge integration, conceptual change, and responsiveness to instruction in

mathematics learning (Bittermann et al., 2023). Most intervention studies are conducted within single-country settings, limiting the external validity of reported effect sizes. Cross-national evaluations of classroom-based gamification remain comparatively scarce, particularly in middle-income Southeast Asian contexts. Without such comparative evidence, it is difficult to determine whether observed gains reflect universally robust instructional mechanisms or context-specific advantages (Álvarez Barroso et al., 2025; Sambo et al., 2025).

Beyond contextual differences, gender-related patterns in mathematics achievement continue to generate scholarly debate. Large-scale meta-analyses suggest that overall gender differences in mathematics performance are small or negligible (Else-Quest et al., 2010), yet variability across countries indicates that sociocultural and educational conditions moderate gender gaps (Moon et al., 2022). More recent research emphasizes that gender disparities are domain-specific and context-dependent rather than biologically determined (Tang & Zhao, 2024). Prior studies on gamified mathematics learning have generally focused on overall effectiveness, student engagement, or single-context implementation, while relatively few have explicitly examined whether gender-related learning gains vary across national settings. Thus, although the broader literature suggests that both contextual and gender-related factors may shape mathematics outcomes, limited evidence is available on how these dimensions interact within cross-national gamified algebra interventions. This gap is especially important because gamified instructional designs may either reduce or unintentionally reproduce inequities depending on how participation, competition, and feedback is structured. Consequently, examining whether gamified instructional interventions differentially benefit male and female students across national contexts is both theoretically important and practically urgent for evaluating equity in classroom innovation.

Taken together, the literature reveals three critical gaps. First, although gamified learning interventions demonstrate promising effects, cross-national validation in comparable educational systems remains limited. Second, the extent to which country-level differences

persist after controlling baseline readiness has not been sufficiently examined in quasi-experimental algebra studies. Third, few studies simultaneously investigate time effects, country differences, and gender moderation within a unified mixed-design analytical framework. Accordingly, the novelty of the present study lies in integrating cross-national comparison, gender moderation, and covariance-adjusted analysis within a single classroom-based gamified algebra intervention. The urgency of this investigation stems from the need to determine whether a promising instructional innovation such as AlgeUno demonstrates robust effectiveness across comparable educational contexts or whether its benefits remain contingent on local instructional and learner-related conditions.

The present study addresses these gaps by investigating the cross-national effectiveness of a gamified algebra learning intervention (AlgeUno) implemented in Indonesia and Malaysia. Rather than functioning solely as a general review of gamified learning, this study is designed to evaluate a specific instructional problem: whether a structured classroom-based gamification model can produce meaningful algebra gains across contexts while remaining equitable across gender groups and responsive to differences in prior knowledge. In this sense, the study seeks not only to test effectiveness, but also to clarify the boundary conditions under which that effectiveness is maintained. Specifically, the study was guided by the following research questions:

1. Whether the intervention significantly improves algebra achievement over time.
2. Whether the magnitude of improvement differs between Indonesia and Malaysia.
3. Whether gender moderates learning gains within and across national contexts.
4. Whether country effects remain statistically significant after controlling for baseline performance through ANCOVA.

By integrating mixed-design ANOVA and covariance adjustment techniques, this study provides a rigorous evaluation of both within-subject change and adjusted between-group differences. The significance of the study lies in three contributions:

- (1) extending the evidence base on gamified mathematics instruction beyond single-country implementation,
- (2) clarifying whether cross-national differences persist after accounting for baseline readiness, and
- (3) providing an equity-oriented perspective by examining whether gender-related response patterns vary across contexts.

Collectively, these contributions strengthen the external validity and practical relevance of classroom-

based gamified algebra interventions in comparable educational systems.

METHOD

Research Design

This study employed a quasi-experimental pre-/post-test design to examine the effectiveness of the AlgeUno gamified instructional intervention across two national contexts and gender groups. Because random assignment at the individual level was not feasible within intact classroom settings, a non-equivalent group design was implemented. The analytical framework followed a 2 (time: pre-test vs. post-test) \times 2 (country: Indonesia vs. Malaysia) \times 2 (gender: male vs. female) mixed factorial structure.

Time was treated as a within-subjects factor, allowing for the examination of learning gains across the instructional period, while country and gender functioned as between-subjects factors. This design enabled the investigation of overall intervention effects, cross-national differences, gender differences, and potential interaction effects among time, country, and gender. Ethical approval was obtained from the respective institutional authorities, and informed consent was secured from schools and students participating prior to data collection.

Participants

The sample consisted of 62 secondary school students drawn from two Southeast Asian countries: Indonesia ($n = 31$) and Malaysia ($n = 31$). Participants were recruited from intact classrooms in public secondary schools selected through purposive sampling based on curricular comparability and institutional consent. Students were enrolled in comparable grade levels (approximately 14-15 years old) and were studying introductory algebra topics aligned with their respective national curricula.

Gender distribution was relatively balanced within each country, enabling moderation analysis. Prior to the intervention, both groups had completed foundational arithmetic instruction and introductory symbolic manipulation units, ensuring comparable baseline curricular exposure. Participation was voluntary, and all students completed both pre- and post-test assessments under standardized testing conditions.

The AlgeUno Gamification Model

AlgeUno is a classroom-based gamification model that adapts the mechanics of the traditional UNO card game into structured algebra learning activities. Rather than functioning as a digital platform, AlgeUno operates as a physical card-based instructional tool designed to integrate algebraic problem-solving into competitive and rule-governed gameplay. The core mechanics of



Figure 1. Representative AlgeUno algebra and action cards integrating symbolic manipulation tasks (Source: Authors’ own elaboration)

UNO, including turn-taking, matching rules, special action cards, and strategic play, are systematically aligned with algebraic tasks such as equation solving, symbolic manipulation, and representational interpretation.

Each card contains algebraic expressions, equations, or conceptual prompts that students must solve or interpret in order to proceed in the game. Standard UNO rules are modified so that successful problem completion becomes a prerequisite for valid card placement. Special action cards are embedded with higher-order algebra tasks that require multi-step reasoning. This design ensures that mathematical engagement is not peripheral but structurally embedded within the game progression.

Pedagogically, AlgeUno integrates repetition, immediate peer feedback, and distributed practice within a socially interactive learning environment. The competitive yet collaborative dynamics of gameplay are intended to increase engagement while reinforcing conceptual connections. By embedding algebraic reasoning within familiar game mechanics, AlgeUno seeks to reduce mathematics anxiety, increase participation, and promote deeper symbolic understanding.

Importantly, AlgeUno does not rely on digital automation or adaptive algorithms. Its effectiveness depends on structured rule alignment, teacher facilitation, and curricular integration. As such, it represents a low-cost, scalable gamification approach that can be implemented in diverse educational contexts without technological infrastructure.

The structural design of AlgeUno is illustrated in Figure 1. Figure 1 presents representative examples of algebraic cards and action cards adapted from traditional UNO mechanics. Each card embeds algebraic tasks requiring symbolic manipulation prior to placement, thereby integrating procedural and

conceptual engagement within competitive gameplay dynamics.

Figure 2 illustrates the instructional gameplay sequence implemented during classroom sessions. Students were required to draw a card, solve the embedded algebraic task, receive peer verification, and place the card according to modified UNO rules. This cycle was repeated across structured rounds to reinforce iterative problem-solving practice.

All gameplay rules, card classifications, scoring procedures, and instructional sequences were explicitly documented on the AlgeUno game board to ensure procedural clarity during classroom implementation. To enhance accessibility and standardization across sites, a QR code was embedded on the board, allowing teachers and students to access a complete digital rulebook and implementation guide. This feature was designed to minimize instructional variability and strengthen cross-national implementation fidelity.

Instrument

Algebra achievement was measured using a standardized researcher-developed test consisting of 10 structured items designed to assess both procedural fluency and conceptual understanding. Items included equation solving, symbolic manipulation, and representational translation tasks. The instrument combined short constructed-response and structured problem-solving formats to capture both accuracy and reasoning processes. Total scores ranged from 0 to 20, with higher scores indicating greater algebra proficiency.

Content validity was established through expert review by mathematics education specialists to ensure alignment with curricular objectives and cognitive demands. A scoring rubric was developed to ensure consistency across administrations. Internal consistency reliability analysis indicated good reliability at pre-test

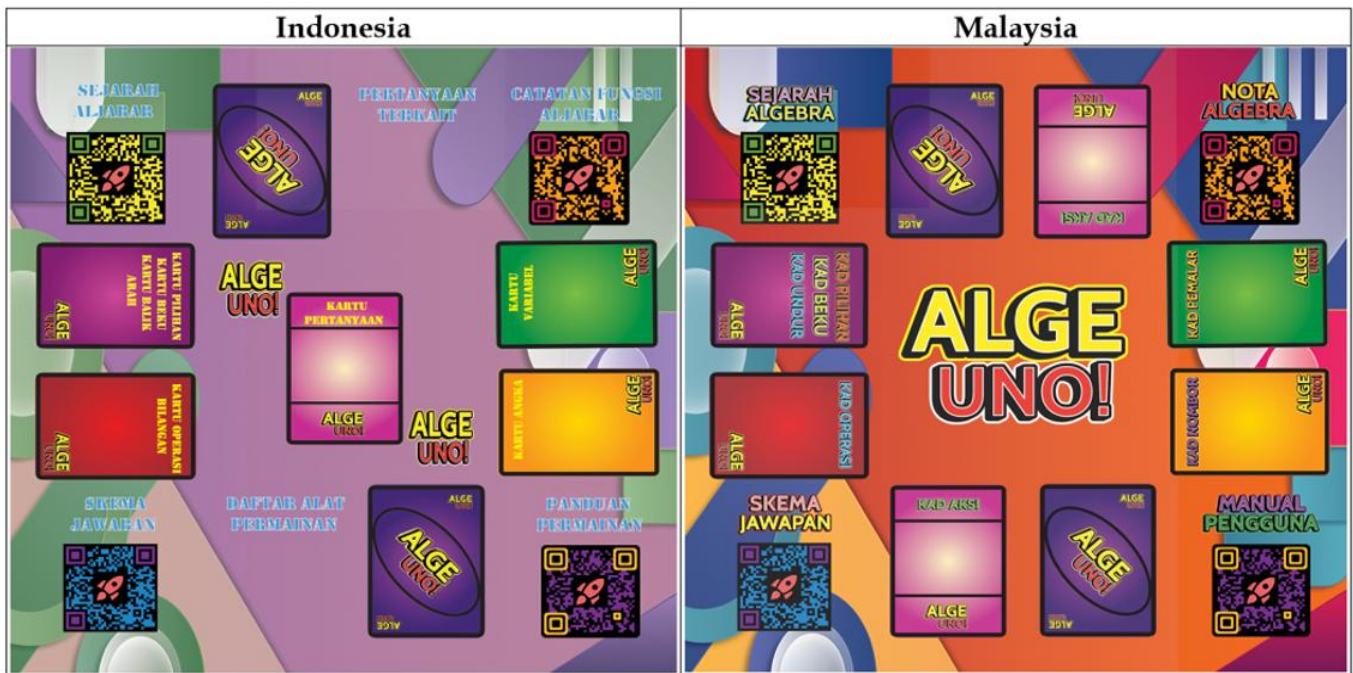


Figure 2. AlgeUno instructional gameplay cycle (Source: Authors' own elaboration)

(Cronbach's $\alpha = .82$) and strong reliability at post-test (Cronbach's $\alpha = .88$), demonstrating stable measurement properties across administrations. These coefficients exceed the commonly accepted threshold of .70 for research instruments, supporting the adequacy of the instrument for inferential analysis.

Procedure

The study was conducted over a two-week instructional period integrated into regular mathematics lessons. At baseline, students completed the algebra pre-test under standardized classroom conditions. Following pre-test, participants engaged in the AlgeUno intervention during scheduled algebra sessions.

AlgeUno is a classroom-based gamification activity that adapts the mechanics of the UNO card game into algebraic problem-solving tasks. Students played in small groups of four to five, following modified UNO rules in which each card contained an algebra problem that had to be correctly solved before placement. Special action cards require multi-step equation solving or conceptual explanation. Gameplay was structured to ensure repeated exposure to algebraic expressions and immediate peer verification of answers.

Teachers in both countries received standardized implementation guidelines, including lesson sequencing, gameplay rules, facilitation strategies, and time allocation. Each session lasted approximately 60 minutes, and students participated in eight structured gameplay sessions over the intervention period. This standardization was intended to ensure procedural equivalence across national contexts. At the conclusion

of the intervention, students completed the post-test under conditions comparable to the pre-test.

Data Analysis

Data analysis proceeded in several stages using SPSS statistical software. First, descriptive statistics were computed to summarize means (M), standard deviations (SD), and normalized gain scores across groups. Normalized gain analysis was used to estimate relative learning improvement while accounting for baseline differences.

Second, a mixed-design ANOVA was conducted to examine the main effects of time, country, and gender, as well as their interaction effects. Assumptions of normality (Shapiro-Wilk test), homogeneity of variance (Levene's test), and sphericity were evaluated prior to inferential testing. Additionally, homogeneity of regression slopes was examined to determine the appropriateness of covariate adjustment.

To control for initial performance differences, an ANCOVA was performed using pre-test scores as a covariate. Adjusted marginal means were computed to estimate post-intervention differences while statistically controlling baseline variation. Pairwise comparisons were conducted using Bonferroni correction to reduce Type I error inflation.

Effect sizes were interpreted using partial eta squared (η^2p) benchmarks, with values of .01, .06, and .14 representing small, medium, and large effects, respectively. Statistical significance was evaluated at the $\alpha = .05$ level.

Table 1. Descriptive statistics of pre- and post-test scores by country

Country	n	M (SD)		
		Pre-test	Post-test	Normalized gain
Indonesia	31	10.52 (3.31)	16.55 (2.28)	0.66 (0.17)
Malaysia	31	14.16 (1.70)	19.03 (1.33)	0.86 (0.14)

Table 2. Cross-national mean differences (Malaysia-Indonesia)

Variable	Mean difference
Pre-test	+3.64
Post-test	+2.48
Normalized gain	+0.20

RESULTS

Descriptive Statistics

Descriptive statistics were computed to examine students' algebra achievement before and after the implementation of AlgeUno in Indonesia and Malaysia. **Table 1** presents the M scores, SDs, and normalized gains for both countries.

As shown in **Table 1**, substantial improvements were observed in both countries following the intervention. Indonesian students improved from a pre-test mean of 10.52 (SD = 3.31) to a post-test mean of 16.55 (SD = 2.28), yielding a normalized gain of 0.66 (SD = 0.17), which reflects a medium-to-high level of instructional effectiveness.

Malaysian students demonstrated a higher baseline performance (M = 14.16, SD = 1.70) and achieved a post-test mean of 19.03 (SD = 1.33). Their normalized gain (g = 0.86, SD = 0.14) indicates a high level of learning improvement.

To further examine cross-national differences, mean differences between Malaysia and Indonesia were calculated.

Table 2 shows that Malaysia outperformed Indonesia at baseline (mean difference = 3.64 points). Following the intervention, the gap narrowed to 2.48 points, indicating a partial reduction of 1.16 points in achievement disparity.

In terms of distributional patterns, Indonesian students exhibited greater variability at both measurement points, whereas Malaysian students showed lower dispersion, particularly at post-test. The

Table 3. Descriptive statistics by time, country, and gender

Group	M (SE)		Gain
	Pre-test	Post-test	
Overall	12.42 (.35)	17.77 (.24)	5.35
Indonesia	10.61 (.49)	16.43 (.34)	5.83
Malaysia	14.23 (.50)	19.11 (.35)	4.88
Male	12.73 (.55)	17.64 (.39)	4.91

relatively small post-test SD in Malaysia suggests a potential ceiling effect.

To examine patterns across demographic groups, descriptive statistics were further disaggregated by time, country, and gender.

As shown in **Table 3**, overall mean scores increased substantially from pre- (M = 12.42, standard error [SE] = .35) to post-test (M = 17.77, SE = .24), yielding an average gain of 5.35 points.

Although Malaysian students demonstrated higher baseline and post-intervention scores, Indonesian students exhibited a slightly larger improvement magnitude.

Gender-based patterns indicated comparable baseline performance between male and female students. However, female students demonstrated a slightly greater improvement (5.79) compared to male students (4.91).

These descriptive findings provide preliminary evidence of meaningful learning gains across contexts and demographic groups. To test these patterns inferentially, a mixed-design ANOVA was conducted.

Mixed-Design ANOVA

A 2 (time: pre-test vs. post-test) × 2 (country: Indonesia vs. Malaysia) × 2 (gender) mixed-design ANOVA was performed to examine within- and between-subject effects (**Table 4**).

A highly significant main effect of time was observed, $F(1, 58) = 695.69, p < .001, \eta^2_p = .923$, indicating a substantial overall improvement from pre-test to post test.

Significant time × country and time × gender interactions suggest that improvement magnitude differ modestly across countries and gender groups. The significant three-way interaction indicates that gender-related differences in gains varied somewhat across national contexts, although effect sizes remained moderate.

Table 4. Mixed-design ANOVA results

Effect	F	df	p	Partial η^2	Effect size interpretation
Time	695.69	(1,58)	< .001	0.923	Extremely large
Time × country	5.45	(1,58)	0.023	0.086	Moderate
Time × gender	4.71	(1,58)	0.034	0.075	Moderate
Time × country × gender	5.36	(1,58)	0.024	0.085	Moderate
Country (between)	30.89	(1,58)	< .001	0.347	Large
Gender (between)	0.10	(1,58)	0.75	0.002	Negligible

Table 5. Homogeneity of regression slopes test

Effect	F	df	p	Partial η^2
Pre-test	26.42	(1,54)	< .001	0.328
Country \times pre-test	1.19	(1,54)	0.280	0.022
Gender \times pre-test	3.33	(1,54)	0.074	0.058
Country \times gender \times pre-test	4.15	(1,54)	0.047	0.071

Table 6. ANCOVA summary table

Effect	F	df	p	Partial η^2
Pre-test (covariate)	119.47	(1,57)	< .001	0.677
Country	5.65	(1,57)	0.021	0.095
Gender	0.08	(1,57)	0.780	0.001
Country \times gender	9.65	(1,57)	0.003	0.145

Note. Model $R^2 = .805$ (adjusted $R^2 = .791$)

Regarding between-subjects effects, Malaysian students scored higher overall than Indonesian students, while no overall gender difference was detected.

Given the baseline differences between countries, further analysis using ANCOVA was conducted to control pre-test scores.

Assumption Testing for ANCOVA

Before performing ANCOVA, the homogeneity of regression slopes assumption was tested (Table 5).

The country \times pre-test interaction was not significant, indicating general slope homogeneity across countries. However, a significant three-way interaction ($p = .047$) suggested partial heterogeneity. Therefore, ANCOVA results are interpreted with caution.

ANCOVA Results

An ANCOVA was conducted with post-test scores as the dependent variable, country and gender as fixed factors, and pre-test as a covariate (Table 6).

The covariate (pre-test) significantly predicted post-test performance, accounting for a large proportion of variance. After controlling baseline differences, a significant main effect of country remained, while gender was not significant. However, a significant country \times gender interaction indicates that gender-related performance differences varied across national contexts.

The overall model explained 80.5% of the variance in post-test scores, demonstrating strong explanatory power. To clarify the country effect, adjusted marginal means were examined.

Adjusted Marginal Means

To better understand the significant country effect, adjusted marginal means were examined (Table 7).

After adjustment for baseline differences, Malaysian students demonstrated significantly higher post-test scores than Indonesian students (mean difference = 1.09, $p = .021$).

Table 7. Adjusted post-test means (controlling pre-test)

Group	Adjusted mean	SE	95% confidence interval
Indonesia	17.40	0.22	[16.97, 17.84]
Malaysia	18.49	0.40	[17.68, 19.30]

The significant country \times gender interaction suggests that gender differences were not uniform across countries, supporting the interpretation that contextual factors may moderate learning outcomes.

Taken together, these findings indicate that baseline knowledge strongly predicts post-intervention achievement, and that country-level differences remain statistically significant after adjustment. However, interaction effects suggest that learning outcomes are shaped by contextual and demographic dynamics rather than uniform treatment effects.

The next section discusses these findings in relation to cross-national educational research and theoretical perspectives on contextual and gender influences in learning outcomes.

DISCUSSION

This study investigated the cross-national effectiveness of the AlgeUno gamified algebra intervention across Indonesia and Malaysia within a mixed factorial framework. By integrating within-subject change, between-country comparison, gender moderation, and covariance adjustment, the analysis offers a nuanced understanding of both general instructional effectiveness and contextually embedded variation. More specifically, the principal findings indicate that

- (1) algebra achievement improved substantially from pre- to post-test across both countries,
- (2) country-level differences remained significant after adjusting for baseline readiness,
- (3) no overall gender difference was observed despite context-sensitive interaction patterns, and
- (4) pre-test performance significantly predicted post-test outcomes.

These findings collectively suggest that the intervention was broadly effective, but that its impact was shaped by both systemic and learner-level moderators. In line with the reviewers' concerns, the following discussion situates these findings in relation to prior research, clarifies the strengths and limitations of the present study, and highlights both theoretical and practical implications

Robust Within-Subject Learning Gains

The extremely large main effect of time indicates substantial improvement from pre- to post-test across both national contexts. This finding is consistent with meta-analytic evidence showing that well-structured game-based and interactive mathematics interventions

can produce moderate to large learning gains (Hunt et al., 2025). In this respect, the present findings are broadly consistent with prior studies reporting that gamified mathematics instruction can enhance both achievement and engagement when gameplay is aligned with curricular goals and supported by explicit instructional structure. This pattern is also consistent with prior mathematics education research showing that structured active-learning designs beyond conventional lecturing (Ramadoni & Mustofa, 2022). In particular, instructional designs that embed immediate feedback, repeated practice, and structured scaffolding are known to support both conceptual restructuring and procedural consolidation.

From a cognitive load perspective, structured gamified activities may optimize germane cognitive processing by directing learners' attention toward problem-relevant features while minimizing extraneous demands (Xu, 2025). Algebra learning requires coordination between symbolic manipulation and conceptual representation, and repeated engagement with algebraic expressions in rule-based gameplay can promote schema construction and representational flexibility (Lee & Ke, 2019). The magnitude of the time effect observed in this study suggests that the AlgeUno gamification structure effectively facilitated iterative problem-solving and reinforced algebraic reasoning processes.

Importantly, comparable gains across Indonesia and Malaysia indicate a degree of cross-contextual robustness. While prior research emphasizes that the effectiveness of instructional innovations depends on implementation fidelity and contextual alignment (Rahayu et al., 2011; Silver et al., 2019), the consistent improvement observed here suggests that the standardized gameplay protocol, including clearly articulated rules and QR-accessible implementation guidelines, supported procedural equivalence across settings.

At the same time, one clear strength of the present study is that it moves beyond a simple pre-/post-test evaluation by demonstrating that substantial learning gains were observable across two national settings rather than within a single classroom context. This cross-context implementation strengthens the external relevance of the findings compared with many prior gamification studies that remain confined to single-site designs. However, the very large magnitude of the time effect should also be interpreted with methodological caution, as repeated exposure to structurally similar algebraic tasks may have amplified short-term gains beyond what would be expected under delayed retention conditions.

Nevertheless, caution is warranted in interpreting the magnitude of the observed effect. The use of structurally parallel pre- and post-test instruments may partially

contribute to performance gains (Lindstromberg, 2025) and repeated exposure to similar algebraic structures during gameplay could intensify short-term improvement (Chen & Yang, 2020). Future studies employing delayed post-tests would be valuable in determining the durability of these gains (Chan et al., 2023).

Persistent Cross-National Differences

Despite strong overall gains, cross-national differences remained statistically significant after controlling baseline performance via ANCOVA. Malaysian students maintained higher adjusted post-test means, indicating that factors beyond initial readiness contributed to differential outcomes. This pattern suggests that instructional responsiveness may be partially shaped by systemic and contextual conditions rather than by the intervention alone (da Silva et al., 2020).

Cross-national variation in mathematics achievement has been consistently documented in international large-scale assessments such as PISA, which demonstrate that educational outcomes are influenced not only by curricular content but also by instructional norms, assessment cultures, and pedagogical traditions (Jerrim, 2013; Leung, 2014). Even countries with geographic proximity and broadly comparable curricular structures may differ in sequencing of algebra topics, exposure intensity, and classroom discourse practices. Such differences may affect how readily students integrate rule-based gamified activities into existing cognitive schemas (Ng et al., 2021).

Pozo (2022) further argue that sociocultural and educational structures operate as macro-level moderators of academic performance. The persistence of country effects after statistical adjustment therefore suggests that contextual affordances, such as prior symbolic exposure, teacher familiarity with algebraic reasoning, or curriculum pacing may enhance or constrain responsiveness to gamified instructional designs.

This finding both aligns with and extends prior research. It aligns with the broader literature showing that instructional technologies and gamified environments do not operate independently of local pedagogical ecologies, but it extends prior work by demonstrating that such contextual moderation remains evident even after statistically controlling for baseline readiness. In other words, the present study does not merely show that countries differ in raw performance; rather, it shows that cross-national differences persist in adjusted post-intervention outcomes. This constitutes an important contribution because many previous classroom-based gamification studies report overall gains without explicitly testing whether those gains

remain comparable after accounting for pre-existing differences.

At the same time, this finding should be interpreted cautiously. The relatively modest sample size and quasi-experimental structure limit strong causal inference regarding systemic differences (Martínez-Abad & León, 2023). Moreover, potential ceiling effects or subtle variations in implementation dynamics may have contributed to the observed pattern. Rather than diminishing the value of the intervention, these results reinforce ecological perspectives on learning, which conceptualize instructional effectiveness as contextually embedded rather than universally uniform (Markauskaite et al., 2023).

From a practical standpoint, this pattern suggests that the advantage of the AlgeUno intervention lies in its transferability across comparable settings, whereas a corresponding limitation is that transferability does not imply uniformity of effect. The intervention appears portable, but its magnitude remains sensitive to contextual differences in curriculum pacing, prior exposure, and classroom routines. This distinction is important for policymakers seeking to scale gamified interventions across schools or national systems.

Gender as a Contextual Moderator

Consistent with the gender similarities hypothesis (Cantley & McAllister, 2021), no significant overall gender difference in algebra achievement was detected. This finding aligns with cross-national meta-analytic evidence indicating that mean gender differences in mathematics performance are generally small or negligible (Reynolds et al., 2015). The absence of a main gender effect suggests that, at an aggregate level, both male and female students benefited comparably from the intervention.

However, the observed interaction patterns indicate that gender-related learning trajectories may be shaped by contextual conditions rather than reflecting stable inherent differences (Moon et al., 2022). Contemporary scholarship increasingly conceptualizes gender disparities in mathematics as socially mediated, influenced by sociocultural expectations, domain-specific confidence, mathematics anxiety, and classroom participation norms rather than biological determinants. Such factors may interact with instructional structures to shape engagement and persistence (Ramadoni & Chien, 2023; Tang & Zhao, 2024).

Gamified classroom environments, particularly those structured around rule-based participation, peer verification, and iterative problem-solving, may attenuate gender disparities when they promote equitable turn-taking, structured feedback, and shared accountability (Liaquat, 2024). Research suggests that instructional designs fostering autonomy and inclusive engagement can reduce achievement gaps when

competitive dynamics are balanced with collaborative norms (Ata, 2024). The absence of a pronounced gender gap in the present study therefore suggests that the AlgeUno gamification structure may provide relatively equitable cognitive and participatory affordances.

In relation to previous research, the present findings are largely consistent with studies showing minimal overall gender differences in mathematics performance, while also supporting more recent arguments that gender patterns become visible primarily through context-specific interactions rather than global mean differences. Thus, the results do not contradict the gender similarities literature; instead, they refine it by suggesting that apparent gender equity at the aggregate level may coexist with more localized contextual variation. This nuance represents an important analytical advantage of the present study, as many intervention studies report only main effects and therefore risk overlooking moderated patterns.

Nevertheless, the interaction effects indicate that gender patterns are not entirely uniform across contexts. This reinforces the view that gender dynamics in mathematics are context-sensitive and embedded within broader educational cultures (Giofrè et al., 2020). Future research incorporating measures of mathematics anxiety, self-efficacy, and classroom interaction patterns would further clarify the mechanisms underlying these moderated effects.

The applied implication is that teachers should not assume that the absence of an overall gender gap automatically guarantees fully equitable participation. Instead, classroom implementation of gamified activities should be monitored for subtle differences in turn-taking, confidence expression, and engagement behavior. Theoretically, these findings support a context-sensitive interpretation of gender in mathematics learning, in which gender is better understood as an interactional and situational variable than as a stable predictor of achievement.

Baseline Readiness and Responsiveness

Pre-test scores emerged as a significant predictor of post-test performance, underscoring the central role of baseline readiness in shaping instructional responsiveness. This finding is consistent with extensive research demonstrating that prior knowledge functions as a primary determinant of knowledge integration and conceptual change (Bittermann et al., 2023). Learners do not approach new algebraic tasks as blank slates; rather, they interpret and assimilate new representations through existing cognitive schemas (Running et al., 2023; Zabala-Vargas et al., 2021).

From a cognitive perspective, students with more developed foundational schemas may process symbolic manipulations more efficiently, allocate cognitive resources more strategically, and integrate feedback

more productively during gameplay (Wirzberger et al., 2018). In contrast, students with weaker prior knowledge may experience higher intrinsic cognitive load, requiring additional scaffolding to achieve comparable gains (Faber et al., 2024). This pattern is consistent with the so-called Matthew effect in education, whereby initial advantages tend to accumulate over time if instructional supports are not sufficiently differentiated.

The significant covariate effect observed in the present study therefore highlights the importance of incorporating baseline diagnostics when evaluating intervention impact. Without statistical adjustment, treatment effects may be confounded with initial proficiency differences. At a pedagogical level, these findings suggest that gamified algebra interventions may benefit from adaptive scaffolding or tiered support structures to ensure that students with lower entry-level knowledge are not disproportionately disadvantaged (Lindstedt et al., 2020).

This result is strongly aligned with prior research emphasizing learner readiness as a key moderator of instructional effectiveness, but it also adds a practical refinement by showing that baseline knowledge remains consequential even within an engaging and highly structured gamified environment. In other words, gamification does not eliminate the importance of prior knowledge; rather, it appears to interact with it. A major practical advantage of this finding is that it points directly toward actionable design improvements, such as diagnostic grouping, differentiated card difficulty, or graduated feedback intensity. At the same time, it also reveals a limitation of one-size-fits-all gamified implementation: without adaptive supports, students entering with lower readiness may benefit less efficiently from the same gameplay structure.

Theoretical Integration

Taken together, the findings support an ecological-interactionist perspective on instructional effectiveness. Although the AlgeUno gamified intervention demonstrated strong overall learning gains, its impact was not uniform across participants or contexts. Instead, improvement emerged through the interaction of instructional structure (time effect), systemic context (country differences), and learner-level characteristics (baseline readiness and moderated gender patterns). This multidimensional pattern underscores that instructional impact cannot be reduced to a singular treatment effect.

Contemporary educational theory increasingly emphasizes that learning outcomes arise from dynamic interactions between cognitive readiness, classroom structures, and broader sociocultural conditions (Bokhove, 2022; Lau et al., 2022; Suter, 2017; Yang Hansen et al., 2024). The present results align with this

view by showing that a structured gamified algebra environment can produce substantial gains while remaining sensitive to contextual and demographic moderators. Rather than suggesting universal effectiveness, the findings indicate conditional robustness: the intervention is broadly effective, yet its magnitude and expression depend on embedded educational ecosystems, consistent with prior evidence that gamified instructional effects vary across contexts (Pan et al., 2025). This interpretation is also compatible with broader technology adoption literature, which suggests that sustained use of an instructional or organizational technology depends not only on initial exposure but also on users' confirmation of its value and perceived usefulness within a specific operational context (Adhyka et al., 2025).

By integrating within-subject change, cross-national comparison, and covariance-adjusted analysis, this study advances the methodological and theoretical conversation on external validity in educational intervention research. The results reinforce the necessity of cross-contextual replication to evaluate whether instructional innovations retain effectiveness beyond single-site implementation (Álvarez Barroso et al., 2025; Sambo et al., 2025). In doing so, the study contributes to a growing body of evidence that educational effectiveness is best understood as relational, situated, and interaction-dependent rather than inherently portable across settings.

The present study offers several clear advantages. Methodologically, it combines mixed-design ANOVA with ANCOVA to distinguish between within-subject gains and adjusted between-group differences. Empirically, it evaluates a classroom-based gamified intervention across two national contexts rather than within a single site. Analytically, it incorporates gender and baseline readiness as moderators rather than treating intervention effects as uniform. These features strengthen the study's contribution to both the gamification literature and comparative mathematics education research. However, the study also has important limitations that temper interpretation, including the modest sample size, quasi-experimental design, potential ceiling effects, and the absence of delayed retention measures. Accordingly, the findings should be interpreted as strong evidence of contextualized instructional promise rather than definitive proof of universally stable causal effects.

The theoretical implication of the study is that gamified mathematics instruction should be conceptualized not as a universally portable treatment, but as a context-sensitive instructional system whose effectiveness emerges through interactions among design features, learner readiness, and educational ecology. The applied implication is that successful implementation of AlgeUno or similar interventions requires more than distributing game materials; it

requires structured teacher guidance, clear procedural rules, baseline diagnostic assessment, and adaptive scaffolding to maximize equity and transferability. For researchers, the results underscore the importance of designing future intervention studies that move beyond simple effectiveness testing toward multi-context, moderator-sensitive models of instructional evaluation.

CONCLUSION

This study set out to examine the cross-national effectiveness of the AlgeUno gamified algebra intervention among lower secondary students in Indonesia and Malaysia, while also evaluating whether learning outcomes varied by country, gender, and baseline readiness. The findings indicate that the research objectives were achieved and provide clear evidence that the intervention was effective, while also revealing important contextual and learner-level variations in its impact.

First, the study confirmed that students' algebra achievement improved significantly from pre- to post-test across both national contexts. This directly addresses the primary objective of evaluating whether the AlgeUno intervention could enhance algebra learning outcomes. The substantial within-subject gains indicate that structured rule-based gamification can support meaningful improvement in algebra performance. These findings suggest that repeated engagement with symbolic reasoning through iterative gameplay can strengthen students' algebraic understanding across comparable instructional settings.

Second, the study found that cross-national differences remained significant after controlling pre-test scores. This addresses the objective of determining whether the intervention functioned similarly across countries. Although students in both Indonesia and Malaysia benefited from the intervention, the persistence of adjusted country-level differences indicates that the magnitude of improvement was not fully equivalent across settings. Thus, the AlgeUno intervention demonstrated cross-national applicability, but not identical effects across educational contexts.

Third, no significant overall gender difference was observed in algebra achievement, indicating that male and female students benefited comparably at the aggregate level. This addresses the objective of examining whether gender functioned as a major determinant of intervention outcomes. However, moderated interaction patterns suggest that gender-related learning trajectories may still vary across contexts. Accordingly, gender did not emerge as a dominant main-effect predictor, but it remained relevant as a context-sensitive moderator.

Fourth, pre-test scores significantly predicted post-test performance, confirming that baseline readiness played an important role in shaping responsiveness to

the intervention. This directly answers the objective of assessing whether prior knowledge influenced post-intervention achievement. Students entering the intervention with stronger initial understanding tended to achieve higher post-test outcomes, highlighting the importance of prior conceptual readiness in gamified algebra learning.

Taken together, the findings show that the AlgeUno gamified algebra intervention was broadly effective across the two participating countries, but that its impact was conditioned by educational context and learner readiness. Therefore, the study concludes that AlgeUno is a promising and transferable instructional approach for improving algebra achievement in lower secondary classrooms, provided that implementation is supported by contextual adaptation and attention to students' baseline knowledge. Future research should extend this line of inquiry through larger multi-site samples, longitudinal retention measures, and more detailed analysis of implementation fidelity in order to clarify the mechanisms underlying cross-contextual variation in intervention effectiveness.

Practical Implications

The findings carry several practical implications for educators and policymakers seeking to implement gamified algebra interventions across diverse educational settings. Because the study showed that the intervention was effective but not uniformly so across all learners and contexts, implementation should be guided by both instructional structure and contextual readiness.

First, the significant predictive role of pre-test scores underscores the importance of conducting baseline diagnostic assessments prior to implementation. Identifying students' initial conceptual readiness allows instructional strategies to be more precisely aligned with learner needs. This means that teachers should not assume that all students will benefit equally from the same game-based activity without prior readiness mapping.

Second, instructional scaffolding should be calibrated to contextual and learner readiness rather than applied uniformly across sites. The persistence of cross-national differences after statistical adjustment suggests that curriculum pacing, prior exposure to symbolic reasoning, and pedagogical routines may influence how students respond to rule-based gameplay structures. Accordingly, tiered problem sequencing, structured feedback cycles, and differentiated facilitation strategies may help improve equity and effectiveness during implementation.

Third, although no overall gender gap was observed, the moderated interaction patterns indicate that gender dynamics should still be monitored within specific educational contexts. Gamified classroom environments can promote equitable participation when turn-taking

structures, feedback mechanisms, and peer interactions are deliberately designed to foster inclusion. Teachers should therefore remain attentive to participation balance, confidence expression, and engagement patterns so that gamified instruction functions as an inclusive rather than stratifying classroom practice.

The practical implication of this study is that AlgeUno should be implemented not merely as a game resource, but as a structured pedagogical intervention supported by diagnostic assessment, adaptive scaffolding, and context-sensitive classroom facilitation.

Limitations

Several limitations should be considered when interpreting the findings. First, the sample size was modest and drawn from two schools across two national contexts, which may constrain statistical power and limit generalizability. Although significant effects were detected, larger multi-site samples would allow more stable estimation of higher-order interaction effects and more refined subgroup analyses. Consequently, the cross-national patterns observed here should be interpreted with caution pending replication in broader populations.

Second, the quasi-experimental design, while appropriate for intact classroom implementation, does not permit causal inference comparable to randomized controlled trials. The absence of random assignment and a parallel control group limits the ability to attribute observed gains exclusively to the intervention. Although statistical controls such as ANCOVA were employed to adjust for baseline differences, unmeasured confounding variables may still have influenced the results.

Third, the use of structurally parallel pre- and post-test instruments may have contributed to short-term performance gains, particularly given the relatively brief intervention window. Additionally, potential ceiling effects—especially within the Malaysian sample—may have constrained observable post-test variability and attenuated estimates of differential growth. Future studies incorporating delayed post-tests would be valuable in assessing retention stability and longer-term conceptual consolidation.

Finally, although implementation protocols were standardized across sites, subtle variations in classroom dynamics, teacher facilitation styles, or peer interaction patterns may have influenced responsiveness to the gamified structure. These contextual micro-variations are difficult to eliminate entirely in naturalistic educational research. Taken together, these limitations suggest that the findings should be interpreted as strong associative evidence rather than definitive causal conclusions.

Future Research

Future research should extend this investigation through larger multi-country samples encompassing broader curricular and sociocultural variation. Expanding the geographic scope would enhance external validity and enable multilevel modeling approaches capable of disentangling classroom-, school-, and country-level influences on gamified algebra intervention effectiveness. Such designs would provide more precise estimates of contextual moderation and strengthen cross-national generalizability.

Longitudinal designs are also necessary to examine retention beyond immediate post-intervention outcomes. Short-term achievement gains do not automatically translate into durable conceptual understanding. Tracking students across subsequent instructional phases would clarify whether gamified algebra experiences promote sustained schema restructuring or primarily short-term performance enhancement. Incorporating delayed post-tests and growth modeling techniques would further illuminate the stability and trajectory of learning gains.

In addition, future studies should integrate classroom-level process variables, including instructional discourse patterns, teacher facilitation strategies, peer interaction dynamics, and implementation fidelity measures. Such data would allow researchers to move beyond outcome evaluation toward mechanism-oriented analysis, clarifying how contextual and instructional processes interact to shape responsiveness to gamified learning environments.

Author contributions: **R:** conceptualization, methodology, instructional design development and content review, classroom implementation and data coordination, statistical analysis and interpretation, formal analysis, writing – original draft, review & editing; **MA:** conceptualization, methodology, instructional design development and content review, classroom implementation and data coordination, supervision, review & editing. Both authors agreed with the results and conclusions.

Funding: This research was supported by the Directorate of Talent Management, National Research and Innovation Agency (BRIN, Indonesia), and Universiti Pendidikan Sultan Idris (Malaysia).

Ethical statement: The authors stated that the study was approved by the Ethics Committee on Social Studies and Humanities at National Research and Innovation Agency on 1 August 2025 with approval number 27062025000001. Written informed consents were obtained from the participants.

AI statement: The authors stated that they used DeepL (version 25.1) for translation, ChatGPT (GPT-4o), and Grammarly (version 1.103) for editing and correcting grammatical, syntactic, and spelling errors. All AI-generated suggestions were carefully reviewed and revised to ensure accuracy and consistency with the authors' intended meaning.

Declaration of interest: No conflict of interest is declared by the authors.

Data sharing statement: Data supporting the findings and conclusions are available upon request from the corresponding author.

REFERENCES

- Abd Jalil, A. H., Abdullah, A. H., & Wani Wan Fakhruddin, W. F. (2024). STEM education and mathematics performance among Orang Asli Primary School students in Johor, Malaysia: Challenges and recommendations. In *Proceedings of the 2024 IEEE Frontiers in Education Conference* (pp. 1-9). IEEE. <https://doi.org/10.1109/FIE61694.2024.10893311>
- Adhyka, N., Rahmaddian, T., Yurizali, B., Ramadoni, R., & Wulandani, Y. P. (2025). A conceptual model of sustainable technology use: The role of confirmation and perceived usefulness in the hospital × management information system in Padang. *International Journal of Statistics in Medical Research*, 14, 76-85. <https://doi.org/10.6000/1929-6029.2025.14.08>
- Akgül, Ö., & Güler, M. (2025). Individual versus collaborative learning: Impact of gamification on achievement and student engagement. *Journal of Educational Research*. <https://doi.org/10.1080/00220671.2025.2548576>
- Álvarez Barroso, C., Fernández Galeote, D., Gertrudix Barrio, M., & Hamari, J. (2025). State-of-the-art in climate resilience gamification for coastal environments: A systematic literature review. *Environmental Research Letters*, 20(11), Article 113007. <https://doi.org/10.1088/1748-9326/ae151d>
- Ata, M., Irum, S., Azam, F., Shaheen, A., Riyaz, S., & Rabbani, M. W. (2024). Gamifying assessments in medical education: Impact of gamification on student's test scores. *Liaquat National Journal of Primary Care*, 7(2), 130-134. <https://doi.org/10.37184/lnjpc.2707-3521.7.30>
- Baah, C., Govender, I., & Subramaniam, P. R. (2024). Enhancing learning engagement: A study on gamification's influence on motivation and cognitive load. *Education Sciences*, 14(10), Article 1115. <https://doi.org/10.3390/educsci14101115>
- Beck, V. C., & da Silva, J. A. (2020). Operative invariants of algebraic pattern present in the strategies of students of the 3rd grade of elementary school. *Acta Scientiae*, 22(3), 48-64. <https://doi.org/10.17648/ACTA.SCIENTIAE.5735>
- Bittermann, A., McNamara, D., Simonsmeier, B. A., & Schneider, M. (2023). The landscape of research on prior knowledge and learning: A bibliometric analysis. *Educational Psychology Review*, 35(2), Article 58. <https://doi.org/10.1007/s10648-023-09775-9>
- Bokhove, C. (2022). Are instructional practices different between East and West? An analysis of grade 8 TIMSS 2019 data. *Asian Journal for Mathematics Education*, 1(2), 221-241. <https://doi.org/10.1177/27527263221109752>
- Cantley, I., & McAllister, J. (2021). The gender similarities hypothesis: Insights from a multilevel analysis of high-stakes examination results in mathematics. *Sex Roles*, 85(7-8), 481-496. <https://doi.org/10.1007/s11199-021-01234-5>
- Chan, J. Y., Closser, A. H., Ngo, V., Smith, H., Liu, A. S., & Ottmar, E. (2023). Examining shifts in conceptual knowledge, procedural knowledge and procedural flexibility in the context of two game-based technologies. *Journal of Computer Assisted Learning*, 39(4), 1274-1289. <https://doi.org/10.1111/jcal.12798>
- Chen, H., & Yang, J. (2020). Multiple exposures enhance both item memory and contextual memory over time. *Frontiers in Psychology*, 11. <https://doi.org/10.3389/fpsyg.2020.565169>
- da Silva, T., Shahrudin, F., Chan, L., Midin, M., & Ravindran, A. (2020). A cross-national, cross-cultural comparison of predictors of academic performance among Canadian and Malaysian university students. *International Journal of Noncommunicable Diseases*, 5(3), Article 114. https://doi.org/10.4103/jncd.jncd_32_19
- Dewi, N., Sumantri, M. S., & Wibowo, F. C. (2026). Gamification in STEM education: The impact of escape room-based problem-solving skills on elementary students. *Multidisciplinary Science Journal*, 8(4), Article 2026278. <https://doi.org/10.31893/multiscience.2026278>
- Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: A meta-analysis. *Psychological Bulletin*, 136(1), 103-127. <https://doi.org/10.1037/a0018053>
- Faber, T. J. E., Dankbaar, M. E. W., van den Broek, W. W., Bruinink, L. J., Hogeveen, M., & van Merriënboer, J. J. G. (2024). Effects of adaptive scaffolding on performance, cognitive load and engagement in game-based learning: A randomized controlled trial. *BMC Medical Education*, 24(1), Article 943. <https://doi.org/10.1186/s12909-024-05698-3>
- García-López, I. M., Acosta-Gonzaga, E., & Ruiz-Ledesma, E. F. (2023). Investigating the impact of gamification on student motivation, engagement, and performance. *Education Sciences*, 13(8), Article 813. <https://doi.org/10.3390/educsci13080813>
- Giofrè, D., Cornoldi, C., Martini, A., & Toffalini, E. (2020). A population level analysis of the gender gap in mathematics: Results on over 13 million children using the INVALSI dataset. *Intelligence*, 81, Article 101467. <https://doi.org/10.1016/j.intell.2020.101467>
- Hunt, J. H., Taub, M., Marino, M., Holman, K., & Womack-Adams, K. (2025). Increasing student engagement, fraction knowledge, and STEM interest through game-based intervention. *Journal*

- of *Special Education Technology*, 40(4), 480-492. <https://doi.org/10.1177/01626434251314014>
- Jerrim, J. (2013). The reliability of trends over time in international education test scores: Is the performance of England's secondary school pupils really in relative decline? *Journal of Social Policy*, 42(2), 259-279. <https://doi.org/10.1017/S0047279412000827>
- Kizhukarakkatu, J. A. (2025). Pedagogical approaches in STEAM education: Fostering innovation and creativity in the classroom. In S. Behera, A. Sorayyaei Azar, S. Curle, & J. Dials (Eds.), *Transformative approaches to STEAM integration in modern education* (pp. 79-102). IGI Global Scientific Publishing. <https://doi.org/10.4018/979-8-3693-7408-5.ch004>
- Krath, J., Schürmann, L., & von Korfflesch, H. F. O. (2021). Revealing the theoretical basis of gamification: A systematic review and analysis of theory in research on gamification, serious games and game-based learning. *Computers in Human Behavior*, 125, Article 106963. <https://doi.org/10.1016/j.chb.2021.106963>
- Lau, N. T. T., Hawes, Z., Tremblay, P., & Ansari, D. (2022). Disentangling the individual and contextual effects of math anxiety: A global perspective. *PNAS*, 119(7), Article e2115855119. <https://doi.org/10.1073/pnas.2115855119>
- Lee, S., & Ke, F. (2019). The format of problem representation for in-game learning supports. *Journal of Computer Assisted Learning*, 35(3), 390-406. <https://doi.org/10.1111/jcal.12345>
- Leung, F. K. S. (2014). What can and should we learn from international studies of mathematics achievement? *Mathematics Education Research Journal*, 26(3), 579-605. <https://doi.org/10.1007/s13394-013-0109-0>
- Lindstedt, A., Koskinen, A., McMullen, J., Ninaus, M., & Kiili, K. (2020). Flow experience and situational interest in an adaptive math game In I. Marfisi-Schottman, F. Bellotti, L. Hamon, & R. Klemke (eds.), *Games and learning alliance. GALA 2020. Lecture notes in computer science()*, vol 12517 (pp. 221-231). Springer. https://doi.org/10.1007/978-3-030-63464-3_21
- Lindstromberg, S. (2025). Eight reasons not to test for baseline group equivalence in a parallel groups pretest-posttest study. *Research Methods in Applied Linguistics*, 4(3), Article 100254. <https://doi.org/10.1016/j.rmal.2025.100254>
- Markauskaite, L., Carvalho, L., & Damsa, C. (2023). Ecological perspectives on learning and methodological implications for research. In C. Damsa, A. Rajala, G. Ritella, & J. Brouwer (Eds.), *Re-theorising learning and research methods in learning research* (pp. 30-46). Routledge. <https://doi.org/10.4324/9781003205838-3>
- Martínez-Abad, F., & León, J. (2023). Inferencia causal en investigación educativa: Análisis de la causalidad en estudios observacionales de carácter transversal [Causal inference in educational research: Analysis of causality in cross-sectional observational studies]. *RELIEVE-Revista Electrónica de Investigación y Evaluación Educativa*, 29(2). <https://doi.org/10.30827/relieve.v29i2.26843>
- Moon, C. Y., Larke, P., & James, M. (2022). Examining mathematics achievement: An analysis of fourth and eighth grade TIMSS U.S. data by ethnicity, gender, and sociocultural variables. *Journal of Ethnic and Cultural Studies*, 9(1), 226-243. <https://doi.org/10.29333/ejecs/942>
- Mutodi, P., & Mosimege, M. (2021). Learning mathematical symbolization: Conceptual challenges and instructional strategies in secondary schools. *Bolema Mathematics Education Bulletin*, 35(70), 1180-1199. <https://doi.org/10.1590/1980-4415v35n70a29>
- Ng, O.-L., Cheng, W. K., Ni, Y., & Shi, L. (2021). How linguistic features and patterns of discourse moves influence authority structures in the mathematics classroom. *Journal of Mathematics Teacher Education*, 24(6), 587-612. <https://doi.org/10.1007/s10857-020-09475-z>
- Pan, Z., Zhu, J., & Kling, J. (2025). The power of gamification. *Phi Delta Kappan*, 107(3-4), 33-36. <https://doi.org/10.1177/00317217251405519>
- Pathania, M., Singh, C. P., Kaur, D. P., & Mantri, A. (2025). Effects of self-adaptive approach of iterative game based learning on performance and satisfaction of elementary school students in mathematics: An action research field experiment. *SN Computer Science*, 6(5), Article 518. <https://doi.org/10.1007/s42979-025-04071-5>
- Pozo Burgos, E. J., Burbano Pulles, M. R., Vidal Chica, J. I., & Revelo Salgado, G. E. (2022). Sociocultural and demographic factors that influence academic performance: The pre-university case of the Universidad Politécnica Estatal del Carchi. *Journal of Technology and Science Education*, 12(1), Article 147. <https://doi.org/10.3926/jotse.1359>
- Rahayu, S., Chandrasegaran, A. L., Treagust, D. F., Kita, M., & Ibnu, S. (2011). Understanding acid-base concepts: Evaluating the efficacy of a senior high school student-centred instructional program in Indonesia. *International Journal of Science and Mathematics Education*, 9(6), 1439-1458. <https://doi.org/10.1007/s10763-010-9272-x>
- Ramadoni, & Mustofa. (2022). Enhancing flipped classroom with peer teaching to promote students' conceptual understanding and self-efficacy in calculus courses. *Pegem Journal of Education and*

- Instruction*, 12(3), 154-168. <https://doi.org/10.47750/pegegog.12.03.17>
- Ramadoni, R., & Chien, K. T. (2023). Integrating peer tutoring video with flipped classroom in online statistics course to improve learning outcomes. *Infinity Journal*, 12(1), Article 13. <https://doi.org/10.22460/infinity.v12i1.p13-26>
- Ratinho, E., Figueiredo, M., Estêvão, D., Faísca, L., & Martins, C. (2026). Gamification on mathematics engagement and motivation in secondary school and higher education: A systematic review and meta-analysis. *Educational Psychology Review*, 38, Article 16. <https://doi.org/10.1007/s10648-025-10108-1>
- Reynolds, M. R., Scheiber, C., Hajovsky, D. B., Schwartz, B., & Kaufman, A. S. (2015). Gender differences in academic achievement: Is writing an exception to the gender similarities hypothesis? *The Journal of Genetic Psychology*, 176(4), 211-234. <https://doi.org/10.1080/00221325.2015.1036833>
- Running, K., Coddling, R. S., Varma, S., Rao, V. N. V., & Wackerle-Hollman, A. (2023). Comparing the effects of concepts-first and iterative fraction instruction sequences. *The Elementary School Journal*, 124(1), 85-108. <https://doi.org/10.1086/725730>
- Sambo, K. U., Shishah, W., & Isiaku, L. (2025). Barriers to gamification adoption in education: A systematic review of cultural, technological, and institutional challenges in developing countries. *Information Development*. <https://doi.org/10.1177/02666669251331268>
- Silver, R. E., Kogut, G., & Huynh, T. C. D. (2019). Learning “new” instructional strategies: Pedagogical innovation, teacher professional development, understanding and concerns. *Journal of Teacher Education*, 70(5), 552-566. <https://doi.org/10.1177/0022487119844712>
- Stewart, S., & Thomas, M. O. J. (2007). Embodied, symbolic and formal thinking in linear algebra. *International Journal of Mathematical Education in Science and Technology*, 38(7), 927-937. <https://doi.org/10.1080/00207390701573335>
- Suter, L. E. (2017). How international studies contributed to educational theory and methods through measurement of opportunity to learn mathematics. *Research in Comparative and International Education*, 12(2), 174-197. <https://doi.org/10.1177/1745499917711549>
- Tang, C., & Zhao, L. (2024). Gender social norms and gender Gap in math: Evidence and mechanisms. *Applied Economics*, 56(17), 2039-2057. <https://doi.org/10.1080/00036846.2023.2178631>
- Tuong, H. A., Nam, P. S., Hau, N. H., Tien, V. T. B., Lavicza, Z., & Houghton, T. (2023). Utilising stem-based practices to enhance mathematics teaching in Vietnam: Developing students’ real-world problem solving and 21st century skills. *Journal of Technology and Science Education*, 13(1), 73-91. <https://doi.org/10.3926/jotse.1790>
- Vanacore, K., Sales, A., Liu, A., & Ottmar, E. (2023). Benefit of gamification for persistent learners: Propensity to replay problems moderates algebra-game effectiveness. In *Proceedings of the 10th ACM Conference on Learning* (pp. 164-173). ACM. <https://doi.org/10.1145/3573051.3593395>
- Wirzberger, M., Herms, R., Esmaeili Bijarsari, S., Eibl, M., & Rey, G. D. (2018). Schema-related cognitive load influences performance, speech, and physiology in a dual-task setting: A continuous multi-measure approach. *Cognitive Research: Principles and Implications*, 3(1), Article 46. <https://doi.org/10.1186/s41235-018-0138-z>
- Xu, B. (2025). Impact of augmented reality gamified math task design on middle school students’ self-regulation. In S. Du, M. Sanmugam, N. Mohd Barkhaya, C. Chen, & J. Wang (Eds.), *Assessing educational digital transformation in the AI era* (pp. 85-114). IGI Global Scientific Publishing. <https://doi.org/10.4018/979-8-3373-7508-3.ch004>
- Yang Hansen, K., Rolfe, V., & Teig, N. (2024). Examining the role of teaching quality and assessment practice in reducing socioeconomic and ethnic inequities in mathematics achievement. In N. Teig, T. Nilsen, & K. Yang Hansen (Eds.), *Effective and equitable teacher practice in mathematics and science education. IEA research for education, vol 14* (pp. 251-276). Springer. https://doi.org/10.1007/978-3-031-49580-9_9
- Yunos, J. M., Rubani, S. N. K., Alias, M., Hamid, H., Yunus, F. A. N., Rizal, F., & Sumarwati, S. (2019). Vocational pedagogical decisions of Malaysian and Indonesian non-engineering TVET teachers. *Journal of Technical Education and Training*, 11(3), 1-8. <https://doi.org/10.30880/jtet.2019.11.03.001>
- Zabala-Vargas, S. A., García-Mora, L. H., Arciniegas-Hernandez, E., Reina-Medrano, J. I., De Benito-Crosetti, B., & Darder-Mésquida, A. (2021). Strengthening motivation in the mathematical engineering teaching processes—A proposal from gamification and game-based learning. *International Journal of Emerging Technologies in Learning*, 16(6), Article 4. <https://doi.org/10.3991/ijet.v16i06.16163>