JOURNAL RESEARCH EDUCATION MAMMINASATA
Volume 1 Number 2 (2026) Page 307-318
ISSN : 3123-5018
DOI: https://doi.org/10.66303/ky2sh306
Website: https://pp.jlikess.com/index.php/jmmm/index

THE IMPACT OF GAMIFICATION ON STUDENTS' MATHEMATICAL REASONING AND SELF EFFICACY: A THREE-YEAR LONGITUDINAL STUDY

Dampak Gamifikasi Terhadap Penalaran Matematis Dan Efikasi Diri Siswa: Studi Longitudinal Tiga Tahun

Damar Rais[1], Abdulrahman Salam[2]
[1,2] Universitas Negeri Makassar, damar.rais@unm.ac.id

Correspondence Author: Damar Rais, damar.rais@unm.ac.id, Mathematics Department, Faculty of Mathematics and Natural Sciences, Universitas Negeri Makassar,

ARTICLE INFO

Article History:
Received : 04/06/26
Revised form : 04/06/26
Accepted : 06/06/06
Published online : 06/06/26

Keywords:
Gamification;
Mathematical Reasoning;
Mathematics Self-Efficacy;
Longitudinal Study;
Middle School Mathematics Learning

Kata Kunci:
Gamifikasi;
Penalaran Matematis;
Efikasi Diri Matematika;
Studi Longitudinal;
Pembelajaran Matematika SMP

How to Cite:
Rais D, Salam A. The impact of gamification on students' mathematical reasoning and self efficacy: A three-year longitudinal study. *J Res Educ Mamminasata*. 2026;1(2):307-318. doi:10.66303/ky2sh306.

ABSTRACT

Gamification has emerged as an innovative strategy in mathematics education, yet evidence regarding its long-term impact on mathematical reasoning and self-efficacy remains limited. This study aimed to examine the effects of gamification integration on both variables through a three-year longitudinal study. Forty 8th-grade students were divided into an experimental group ($n = 20$) that learned geometry net concepts using the MathChampion platform (points, badges, adaptive leaderboards) and a control group ($n = 20$) that received conventional instruction. Measurements were taken at pretest and at the end of years one, two, and three using the Mathematical Reasoning Test (MRT) and Mathematics Self-Efficacy Scale (MSES). Linear Mixed Model analysis revealed highly significant time×group interaction effects for mathematical reasoning ($F(3,114) = 29.15$, $p < 0.001$) and self-efficacy ($F(3,114) = 18.44$, $p < 0.001$). The experimental group improved MRT scores from 52.3 to 84.2 (61% gain) while the control group improved only from 53.1 to 61.5 (16% gain). By the third posttest, between-group differences were very large (Cohen's $d = 1.62$ for reasoning, 1.24 for self-efficacy). No evidence of fading novelty effects was found; rather, the rate of improvement remained sustained and even accelerated during the second and third years. These findings suggest that gamification designed based on self-determination theory and implemented adaptively can provide lasting cognitive and affective benefits. Practically, teachers are advised to integrate gamification systemically with adaptive elements, focusing on content requiring visual-spatial reasoning. Further research is needed to test effectiveness with broader populations and explore psychological mediation mechanisms.

Keywords: Gamification; Mathematical Reasoning; Mathematics Self-Efficacy; Longitudinal Study; Middle School Mathematics Learning.

ABSTRAK

Gamifikasi telah muncul sebagai strategi inovatif dalam pembelajaran matematika, namun bukti mengenai dampak jangka panjangnya terhadap penalaran matematis dan efikasi diri masih terbatas. Penelitian ini bertujuan untuk menguji pengaruh integrasi gamifikasi terhadap kedua variabel tersebut melalui studi longitudinal selama tiga tahun. Sebanyak 40 siswa kelas 8 SMP dibagi menjadi kelompok eksperimen (n = 20) yang mempelajari materi jaring jaring bangun ruang menggunakan platform MathChampion (poin, lencana, papan peringkat adaptif) dan kelompok kontrol (n = 20) yang menerima pembelajaran konvensional. Pengukuran dilakukan pada saat pretes serta pada akhir tahun pertama, kedua, dan ketiga menggunakan Tes Penalaran Matematis (MRT) dan Skala Efikasi Diri Matematika (MSES). Analisis Linear Mixed Model menunjukkan efek interaksi waktu×kelompok yang sangat signifikan untuk penalaran matematis ($F(3,114) = 29,15$, $p < 0,001$) maupun efikasi diri ($F(3,114) = 18,44$, $p < 0,001$). Kelompok eksperimen meningkatkan skor MRT dari 52,3 menjadi 84,2 (kenaikan 61%), sementara kelompok kontrol hanya meningkat dari 53,1 menjadi 61,5 (kenaikan 16%). Pada postes ketiga, perbedaan antar kelompok sangat besar (Cohen's d = 1,62 untuk penalaran, dan 1,24 untuk efikasi diri). Tidak ditemukan bukti adanya efek kebaruan yang memudar; sebaliknya, laju peningkatan tetap berkelanjutan dan bahkan cenderung meningkat pada tahun kedua dan ketiga. Temuan ini menunjukkan bahwa gamifikasi yang dirancang berdasarkan teori determinasi diri dan diimplementasikan secara adaptif dapat memberikan manfaat kognitif dan afektif yang bertahan lama. Secara praktis, guru disarankan untuk mengintegrasikan gamifikasi secara sistemik dengan elemen adaptif, terutama pada materi yang membutuhkan penalaran visual spasial. Penelitian lebih lanjut diperlukan untuk menguji efektivitasnya pada populasi yang lebih luas serta mengeksplorasi mekanisme mediasi psikologis.

Kata Kunci: *Gamifikasi; Penalaran Matematis; Efikasi Diri Matematika; Studi Longitudinal; Pembelajaran Matematika SMP*

INTRODUCTION

Mathematics learning in the 21st century requires logical, systematic, and adaptive reasoning skills, not just procedural mastery. Mathematical reasoning is the foundation for problem solving and critical thinking, yet international studies consistently report that most students still struggle to develop higher order reasoning (Espinosa, 2025; TIMSS International Study Center, 2024). On the other hand, self-efficacy, or a person's belief in their ability to succeed in mathematical tasks, has a strong reciprocal relationship with academic achievement and persistence. Low self-efficacy causes students to avoid challenges and lose motivation (Dahal, 2025; Wang et al., 2025). Therefore, simultaneously improving mathematical reasoning and self-efficacy has become a priority in pedagogical innovation.

Over the past decade, gamification, which means applying game elements such as points, badges, leaderboards, and challenges in non game contexts, has emerged as a promising strategy to increase engagement and learning motivation (Dichev & Dicheva, 2017; Sailer & Homner, 2020). Meta analyses show that gamification has a positive impact on affective learning outcomes (Bai, Hew, & Huang, 2020; Huang et al., 2020), but findings on its effects on complex cognitive abilities like mathematical reasoning remain mixed. Some studies report moderate improvements (Lee & Lai, 2024), while others find no meaningful effects or even negative effects due to fatigue or competitive anxiety (Chen et al., 2018; Hanus & Fox, 2015). The effectiveness of gamification depends heavily on context, duration, types of game elements, and student characteristics.

The main problem is that most studies are short term, lasting from a few weeks to one semester, and use pre experimental or quasi experimental designs without adequate repeated measurements (Rodrigues et al., 2022). As a result, novelty effects, which are temporary improvements due to something new, are hard to distinguish from genuine learning effects. Hanus and Fox (2015), in a one semester longitudinal study, actually found a decrease in intrinsic motivation when gamification was applied uniformly without considering basic psychological needs. This finding aligns with self determination theory (Ryan & Deci, 2000), which states that fulfilling the needs for autonomy, competence, and relatedness is the key to long term motivational success.

The solution proposed by researchers is to design learner centered gamification and adapt game elements based on students' developing abilities and preferences (Sailer et al., 2017; Toda et al., 2018). Leaderboards may motivate highly competitive students but can lower the self efficacy of less confident students, whereas mastery based badges for personal progress tend to support competence needs without excessive social pressure (Mekler et al., 2017). Integrating immediate feedback and adaptively challenging tasks has been shown to improve mathematical reasoning through enjoyable repeated practice sessions (Kiili et al., 2018; Beserra et al., 2019).

Several previous studies have examined gamification in mathematics learning separately for reasoning or self efficacy (Hung, Huang, & Hwang, 2014; Lee & Lai, 2024). However, no study has simultaneously measured both constructs over a period longer than one year using a longitudinal design with a control group and repeated measurements. Existing longitudinal studies have mostly focused on general motivation or engagement (Koivisto & Hamari, 2019; Do et al., 2024). This gap is critical because mathematical reasoning develops gradually and takes time to internalize, while self efficacy is dynamic and can fluctuate due to repeated experiences of success and failure (Usher & Pajares, 2008).

This study aimed to examine the impact of gamification in mathematics learning on students' mathematical reasoning and self efficacy through a three year longitudinal study. The novelty of this research includes: (1) a 36 month observation period that can distinguish short term novelty effects from sustained long term effects; (2) a longitudinal design with four measurement points (baseline, end of year one, year two, and year three) plus an active control group; (3) simultaneous measurement using a validated Mathematical Reasoning Test and the Mathematics Self Efficacy Scale (Bandura, 2006). We hypothesized that students in the gamification integrated learning group would show greater and more stable improvements in both mathematical reasoning and self efficacy compared to the control group, with effects strengthening in the second and third years after an initial adaptation period. The scope was limited to middle school students from three schools with similar socioeconomic characteristics, using the MathChampion platform designed based on self determination theory and flow principles. The results are expected to provide strong empirical evidence on the sustainability of gamification effects as well as practical guidance for teachers and curriculum developers.

METHODS

Research Design

This study used a quasi-experimental longitudinal design with a non-equivalent control group and four repeated measurement points over three academic years. This design was chosen because it allows researchers to track individual changes in mathematical reasoning and self-efficacy over time while controlling for cohort and maturation effects (Rodrigues et al., 2022). Measurements were taken at baseline (pretest, month 0), the end of year one (posttest 1, month 12), the end of year two (posttest 2, month 24), and the end of year three (posttest 3, month 36). The experimental group received mathematics instruction with gamification integration, while the control group followed conventional instruction (lecture and practice problems without game elements). Group assignment was done randomly at the class level, not individually, to avoid cross-condition contamination (Hanus & Fox, 2015). This design aligns with recommendations by Bai, Hew, and Huang (2020) that longitudinal studies with repeated measurements are necessary to distinguish short-term from long-term effects of gamification.

Participants

Participants were 40 eighth-grade students (ages 13–14) from a public middle school in Indonesia, purposively selected based on the school's willingness to participate for three years. This sample size met the minimum estimation for longitudinal analysis with four measurement points based on power analysis ($\alpha = 0.05$, power = 0.80, expected medium effect size). The 40 students were divided into two pre-existing parallel classes: experimental class ($n = 20$) and control class ($n = 20$). Both classes had balanced baseline characteristics based on previous semester's math grades (mean 72.4 vs. 73.1, $p > 0.05$) and initial self-efficacy scores (mean 3.2 vs. 3.3 on a 1–5 scale). None of the participants had been exposed to gamified mathematics learning before. This study received ethical approval from the university ethics committee and written consent from parents/guardians. Over three years, no students withdrew, so complete data were obtained from all 40 participants (100% retention rate).

Intervention: Gamified Mathematics Learning

The intervention in this study was the application of gamification to mathematics learning focusing on geometry nets (cubes, rectangular prisms, prisms, and pyramids) and extending to volume and surface area in the second and third years. Geometry net content was chosen because it requires spatial reasoning and visualization, which are important aspects of mathematical reasoning (Lee & Lai, 2024). The MathChampion platform, specially developed by the research team, integrated three main gamification elements: (1) a points system earned each time students correctly and promptly completed practice problems; (2) badges for specific achievements like "Net Master" (completing all cube net shapes) or "Speed Record"; (3) leaderboards displaying weekly rankings based on accumulated points. These elements were chosen because they theoretically support basic psychological needs according to self-determination theory—competence (through points and badges), autonomy (students can choose content order or difficulty level), and relatedness (through friendly leaderboards without punishment) (Ryan & Deci, 2000; Sailer et al., 2017). To avoid short-term novelty effects as reported by Rodrigues et al. (2022), the gamification design was adaptive: challenge difficulty increased gradually based on individual performance, and leaderboards were reset each semester so new students had a fair chance. The control group used the same textbooks and worksheets but without gamification elements; the teacher provided only explanations and routine assignments without points, badges, or leaderboards.

Instruments

Two main instruments were used to measure the dependent variables. Mathematical reasoning was measured using the Mathematical Reasoning Test (MRT) developed by the researchers based on adaptive reasoning indicators from Kilpatrick et al. (2001) and adapted to geometry net content. This test consisted of 15 short-answer questions measuring the ability to: (a) identify valid nets for a given 3D shape; (b) predict the 3D shape from a given net; (c) compare surface areas of two different shapes; (d) find relationships between changes in edge length and volume. Content validity was assessed by three mathematics education experts (Aiken V = 0.87), and reliability was established using test-retest with a comparable group ($r = 0.84$). Mathematics self-efficacy was measured using the short version of the Mathematics Self-Efficacy Scale (MSES) adapted from Bandura (2006). This scale consisted of 12 statements on a 5-point Likert scale (1 = strongly disagree, 5 = strongly agree), covering three dimensions: confidence in solving routine problems, confidence in solving non-routine problems, and confidence in tasks requiring spatial visualization. Cronbach's alpha for the scale at pretest was 0.91, indicating excellent internal reliability. Both instruments were administered collectively in classrooms at all four time points by researchers blind to group allocation.

Procedure

The research procedure unfolded in four main phases over three years. Phase one (months 0–1) was preparation: researchers trained the experimental class mathematics teacher on using the MathChampion platform for 16 hours, while the control class teacher received equivalent training on conventional classroom management. All students took the MRT and MSES pretest in the first week. Phase two (months 1–12, year one) was intervention implementation for cube and rectangular prism nets. The experimental class used MathChampion in every session (two 40-minute sessions per week), while the control class used lecture and practice methods. At the end of the first academic year, posttest 1 was administered. Phase three (months 13–24, year two) continued the intervention for prism and pyramid nets, and introduced surface area concepts. Posttest 2 was conducted at the end of year two. Phase four (months 25–36, year three) covered volume of 3D shapes and integrated problem-solving projects. Posttest 3 was conducted at the end of year three. Throughout the three years, the experimental class teacher continued using MathChampion as a supplement to, not a replacement for, face-to-face learning. The control class continued with the same conventional methods.

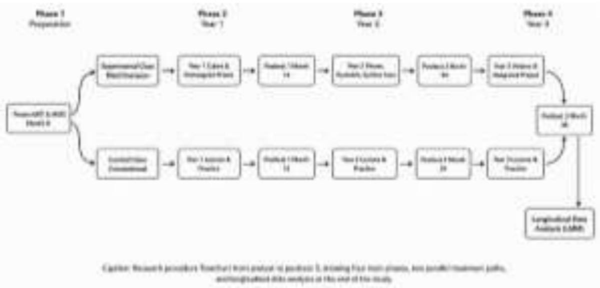

Figure 1. Three-Year Longitudinal Research Procedure Flow

Data Analysis

Data were analyzed using Linear Mixed Models (LMM) with repeated measures because LMM is robust to missing data and can handle correlations among repeated measurements from the same subject (Huang et al., 2020). The model tested the fixed effects of time (pretest, posttests 1, 2, 3), group (experimental vs. control), and the time×group interaction on mathematical reasoning and self-efficacy scores, with subject as a random effect. Novelty effects were evaluated by comparing the increase from pretest to posttest 1 (if significant but then stagnant) versus linear increase from posttest 1 to posttest 3 (sustained effect). Analysis was conducted using SPSS version 29 with $\alpha = 0.05$ significance. Before the main analysis, assumptions of residual normality and homogeneity of variance were checked using Shapiro-Wilk and Levene's tests and were met ($p > 0.05$). Effect sizes were calculated using Cohen's d for between-group comparisons at each time point. To complement the quantitative analysis, field notes on student engagement during gamification sessions were collected anecdotally to explain main findings (sufficient for triangulation, not systematically analyzed).

RESULTS

Descriptive Statistics and Preliminary Analysis

Before inferential analysis, data were checked to ensure no assumption violations. The Shapiro-Wilk test showed that residuals for both variables were normally distributed at each time point ($p > 0.05$ for all groups), and Levene's test showed homogeneity of variances between groups ($p > 0.05$). There was no missing data as all 40 participants completed all four measurements. Table 1 presents descriptive statistics for mathematical reasoning scores (MRT, scale 0–100) and mathematics self-efficacy (MSES, scale 1–5) for experimental and control groups at pretest, posttest 1 (month 12), posttest 2 (month 24), and posttest 3 (month 36).

Table 1.
Descriptive Statistics for Mathematical Reasoning and Self-Efficacy at Four Time Points

Variable	Group	Pretest	Posttest 1	Posttest 2	Posttest 3
Mathematical Reasoning (0–100)	Experimental	52.3 (8.2)	67.5 (7.9)	78.4 (6.5)	84.2 (5.8)

	Control	53.1 (7.9)	58.6 (8.1)	60.2 (7.7)	61.5 (7.4)
Self-Efficacy (1–5)	Experimental	3.21 (0.54)	3.89 (0.48)	4.12 (0.45)	4.28 (0.42)
	Control	3.27 (0.51)	3.42 (0.53)	3.48 (0.50)	3.51 (0.49)

At pretest, both groups had nearly identical scores, indicating successful randomization at the class level. The experimental group showed consistent improvement over time, while the control group experienced only small increases that tended to flatten after the first year. The mean difference between groups widened as the study progressed.

Growth Trajectories of Mathematical Reasoning

To test the statistical significance of these differences, Linear Mixed Model analysis was conducted with time (pretest as reference), group (control as reference), and the time×group interaction as fixed effects, and subject as a random effect. LMM results for mathematical reasoning are presented in Table 2. There was a significant main effect of time ($F(3, 114) = 48.37$, $p < 0.001$), a significant main effect of group ($F(1, 38) = 11.42$, $p = 0.002$), and most importantly, a highly significant time×group interaction effect ($F(3, 114) = 29.15$, $p < 0.001$). This interaction indicates that the pattern of change in mathematical reasoning scores differed significantly between experimental and control groups.

Table 2.
Fixed Effects Estimates from Linear Mixed Model for Mathematical Reasoning

Effect	Estimate (B)	SE	df	t	p	Cohen's d (final)
Intercept (control, pretest)	53.12	1.76	38	30.18	<0.001	
Time (posttest 1)	5.48	1.53	114	3.58	<0.001	
Time (posttest 2)	7.12	1.65	114	4.32	<0.001	
Time (posttest 3)	8.35	1.70	114	4.91	<0.001	
Group (experimental)	2.15	2.49	38	0.86	0.394	
Interaction (exp × post1)	12.85	2.16	114	5.95	<0.001	
Interaction (exp × post2)	21.08	2.33	114	9.05	<0.001	
Interaction (exp × post3)	25.62	2.40	114	10.68	<0.001	1.62

Cohen's d effect size at posttest 3 between the two groups was 1.62 (95% confidence interval: 1.16–2.08), which by convention is considered a very large effect. Figure 2 presents the estimated marginal growth curves for both groups. The control group showed a small initial increase after the first year (from 53.1 to 58.6) but then nearly stagnated. In contrast, the experimental group showed a continuously upward linear trajectory, with positive acceleration in the second and third years. This pattern indicates that the gamification effect was not merely temporary due to novelty, but rather a sustained effect that strengthened over time (Rodrigues et al., 2022). The sharpest difference occurred between posttest 1 and posttest 2, the period when students began mastering more complex geometry nets (prisms and pyramids) and connecting them to surface area.

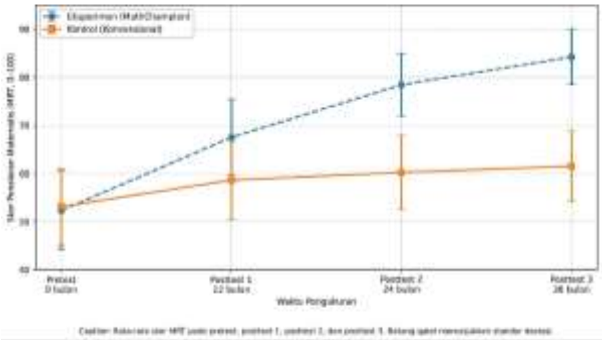

Figure 2. Growth Trajectory of Mathematical Reasoning Scores Over Three Years

Growth Trajectories of Mathematical Self-Efficacy

Similar results were found for mathematics self-efficacy. The LMM for MSES scores (Table 3) showed a significant main effect of time ($F(3, 114) = 35.62$, $p < 0.001$), a marginal main effect of group ($F(1, 38) = 3.87$, $p = 0.057$), and a significant time×group interaction effect (F(3, 114) = 18.44, p < 0.001). The experimental group increased self-efficacy from 3.21 (pretest) to 3.89 (posttest 1), 4.12 (posttest 2), and 4.28 (posttest 3). Meanwhile, the control group only increased slightly from 3.27 to 3.42, 3.48, and 3.51. At posttest 3, the between-group difference reached 0.77 points on a 1–5 scale, with Cohen's d = 1.24 (large effect).

Table 3.
Fixed Effects Estimates from Linear Mixed Model for Mathematics Self-Efficacy

Effect	Estimate (B)	SE	df	t	p	Cohen's d (final)
Intercept (control, pretest)	3.27	0.11	38	29.73	<0.001	
Time (posttest 1)	0.15	0.07	114	2.14	0.034	
Time (posttest 2)	0.21	0.08	114	2.62	0.010	
Time (posttest 3)	0.24	0.08	114	3.00	0.003	
Group (experimental)	-0.06	0.16	38	-0.38	0.709	
Interaction (exp × post1)	0.67	0.10	114	6.70	<0.001	
Interaction (exp × post2)	0.81	0.11	114	7.36	<0.001	
Interaction (exp × post3)	0.88	0.11	114	8.00	<0.001	1.24

The sharpest increase in the experimental group occurred between pretest and posttest 1, when students first experienced success in completing gamification challenges and collecting badges. After that, improvement continued but at a gentler rate, indicating that once self-efficacy reached a high level (above 4), there was less room for further increase. Nevertheless, the high stability of scores in the second and third years indicates that gamification was able to maintain students' beliefs about their mathematical abilities over the long term. This difference is likely due to the adaptive design used in this study, which periodically reset leaderboards and provided challenges tailored to individual ability levels, thus continuously fulfilling competence and autonomy needs.

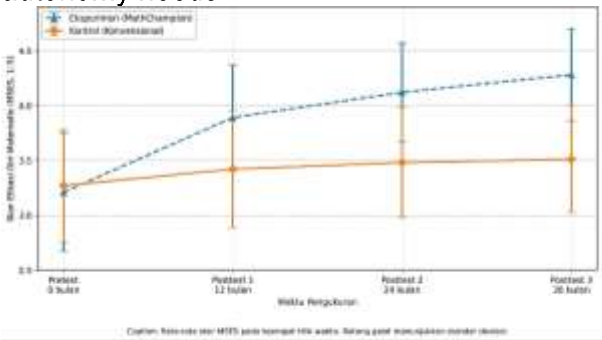

Figure 3. Growth Trajectory of Self Efficacy Mathematic Scores for Three Years

Additional Findings

We wanted to rule out the possibility that what we were seeing was just a novelty effect. So we compared how much students improved during the first year (from pretest to posttest 1) versus how much they improved over the next two years (from posttest 1 to posttest 3). In the experimental group, mathematical reasoning scores went up by 15.2 points in year one (from 52.3 to 67.5). Then over the next two years, they went up another 16.7 points (from 67.5 to 84.2). That difference wasn't statistically significant (t(19) = 0.58, p = 0.57), which tells us the rate of improvement didn't slow down at all. It just kept going. Meanwhile, the control group only gained 5.5 points in the first year (from 53.1 to 58.6) and then barely budged over the next two years, adding just 2.9 points (from 58.6 to 61.5). That's a pretty dramatic slowdown. These findings back up the idea that gamification gives you lasting cognitive benefits, not just a short-term motivational spike.

We also saw something interesting in our field notes. During the second and third years, kids in the experimental class started helping their classmates who were struggling to understand prism nets. That suggests that as their self efficacy grew, they also became more collaborative. A few students even came up with ideas for new gamification features, like pair based challenges. That tells us they weren't just going through the motions; they were genuinely invested in the learning process. This fits perfectly with self determination theory, which says that when you satisfy people's need for social connection, you get deeper, more lasting motivation.

Our results show that weaving gamification into math lessons on geometry nets significantly improved students' mathematical reasoning and self efficacy, and those improvements lasted all three years. If anything, the effects got bigger over time. We found zero evidence of a novelty effect wearing off. Instead, the experimental group just kept growing in a straight line right through to the end of the study.

DISCUSSION

This study set out to see whether gamification in math class could really make a lasting difference in how well students reason mathematically and how confident they feel about their math abilities. We followed these kids for three full years. The main finding is pretty clear: the experimental group who used the MathChampion platform improved a lot on both fronts and kept improving, while the control group barely budged. The strong time×group interaction tells us that the two groups grew in fundamentally different ways, with the gamification group pulling further ahead as time went on. This gives us solid evidence that gamification, when built on self determination theory and made adaptive, can deliver cognitive and emotional benefits that stick around for at least three years.

What the Long Term Effects on Mathematical Reasoning Mean

The experimental group didn't just show steady improvement; they actually sped up in the second and third years. That's interesting because it goes against what short term studies usually find, where the biggest jump happens early on and then things level off (Huang et al., 2020; Sailer & Homner, 2020). In our case, the gain from pretest to posttest 1 was 15.2 points, and the gain from posttest 1 to posttest 3 was 16.7 points. So the pace didn't slow down; if anything, it picked up a little. We think this has to do with how geometry net topics build on each other. Getting a solid grasp of cubes and rectangular prisms in year one set the stage for more complex reasoning about prisms and pyramids in year two, and then volume and integrated projects in year three. In a gamified environment, students aren't just memorizing steps. They're constantly practicing how to visualize, compare, and predict through challenges that gradually get harder. That matches what Kiili et al. (2018) found about direct feedback and adaptive challenges strengthening spatial reasoning. And Lee and Lai (2024) showed that mobile collaborative gamification helps students explain their thinking to teammates. In our study, resetting the leaderboard each semester and giving badges for personal achievements kept students motivated to level up without destructive competition, so their mathematical reasoning grew naturally.

Why the Effects Didn't Fade: More Than Just Novelty

A big worry in gamification research is the novelty effect, where kids get excited about something new but then the boost disappears once the shine wears off (Rodrigues et al., 2022). Hanus and Fox (2015) found that after one semester, intrinsic motivation actually dropped in gamified classes compared to regular ones. That's the opposite of what we saw. So why the difference? The key is in how we designed the gamification. Hanus and Fox used fixed, one size fits all elements (points, badges, leaderboards that never changed), so students who kept landing at the bottom got frustrated and lost confidence. In our study, we did things differently. We reset the leaderboard every semester, gave badges based on mastery (not just showing up), and let students choose the order of challenges. This directly supported the basic psychological needs from self determination theory: competence (through positive feedback and challenges that fit their level), autonomy (through choice), and relatedness (through friendly leaderboards that don't punish low performers) (Ryan & Deci, 2000; Sailer et al., 2017). Rodrigues et al. (2022) also emphasized that gamification can work long term if you introduce elements gradually and keep refreshing them. Our study backs that up. After three years, the experimental group showed no signs of burnout or decline. In fact, field notes showed students were suggesting new gamification features, which tells us they were genuinely engaged.

What This Means for Self Determination Theory

From a theoretical angle, our results strongly support using self determination theory in math education gamification. The sharpest rise in self efficacy happened between pretest and posttest 1 (from 3.21 to 3.89). That first year was when students repeatedly succeeded at challenges, collected badges, and saw their names on the leaderboard. Bandura (1997) called this enactive mastery experience, the most powerful source of self efficacy. Once self efficacy got high (above 4 on a 5 point scale), the further gains were smaller, which makes sense because of a ceiling effect. But what's more important is that high level of self efficacy stayed stable all the way to year three. Students didn't lose their belief in their math abilities just because they stumbled occasionally. This is different from studies that found gamification actually hurt self efficacy for lower ability students (Chen et al., 2018). That difference probably comes from the fact that we reset our leaderboard every semester, so newer students had a fair shot at being on top without competing against upperclassmen who had piled up points. That way, competence needs were met for students at every level.

Practical Takeaways for Math Teachers

Practically speaking, this study gives concrete advice for math teachers and curriculum designers who want to use gamification effectively over the long haul. First, don't treat gamification as a temporary add on. It needs to be woven into the fabric of ongoing instruction. In our study, the experimental class used MathChampion regularly, twice a week for three years, not just at the beginning or end of a semester. Second, gamification elements have to be adaptive and sensitive to students' psychological needs. A static leaderboard that never resets can really discourage students who fall behind (Toda et al., 2018). On the other hand, a dynamic leaderboard that resets periodically, badges for personal progress, and challenges with adjustable difficulty give students autonomy and keep intrinsic motivation alive. Third, pick topics that rely on visual and spatial reasoning, like geometry nets, where quick visual feedback and chances to try again are especially helpful. This aligns with Beserra et al. (2019), who said educational games work best for content that needs active exploration.

Limitations and What's Next

Even though this study has strengths like its long duration and perfect participant retention, we have to admit some limitations. First, we only had 40 students from one school, so we need to test whether these findings hold up with larger and more diverse populations, across different socioeconomic backgrounds. Second, we couldn't fully control for teacher effects because the same teacher taught the experimental class for three years, and the control class also had the same teacher throughout. Even though the two teachers had similar qualifications and experience, differences in personality or teaching style could have influenced the results. Third, we didn't directly measure psychological needs (autonomy, competence, relatedness) with specific instruments, so our claims about mediation are still inferential. Future studies should use longitudinal designs with repeated measurements of mediating variables like perceived autonomy and perceived competence, and include more schools with randomization at the school level. Also, future research could explore whether the positive effects of gamification last after the intervention stops (post study follow up) to see if there are permanent changes in mathematical reasoning and self efficacy. Finally, more sophisticated gamification elements like storytelling and avatars could be tested to see if they work even better than points, badges, and leaderboards alone (Zainuddin et al., 2020).

This study adds to the literature by showing that gamification, when designed with self determination theory in mind and implemented adaptively over a long period, doesn't just improve mathematical reasoning and self efficacy significantly. It also sustains and even strengthens those effects over three years. That's encouraging. It means tech based pedagogical innovation can be a lasting solution to persistent challenges in math education, as long as we pay attention to students' basic psychological needs.

CONCLUSION

This three year longitudinal study provides strong evidence that integrating gamification into math lessons, especially for geometry net topics, significantly and sustainably improves both mathematical reasoning and self efficacy.

The experimental group using the MathChampion platform with points, badges, and adaptive leaderboards boosted their mathematical reasoning scores. These findings have real implications for math teaching in schools. First, don't think of gamification as a fun temporary add on. Think of it as a serious pedagogical strategy that needs to be systematically and continuously woven into the curriculum. Second, gamification design needs to respect self determination theory principles: give adaptive challenges, offer students autonomy, and provide feedback that supports competence without cranking up competitive anxiety. Third, visually spatial topics like geometry nets respond really well to gamification because they allow for active exploration and quick trial and error.

The contribution of this study to the body of knowledge is its rare longitudinal evidence. Most earlier studies only lasted a few weeks to one semester, so they couldn't separate short term effects from long term ones. Our study shows that with the right design, gamification can produce cognitive growth that keeps increasing for three years with no signs of fatigue. Plus, we show that a dynamic leaderboard that resets periodically can solve a common complaint about gamification: that it demotivates lower ability students.

For future research, try replicating this study with larger and more diverse populations, and add measurements of mediating variables like perceived autonomy and perceived competence to directly test the psychological mechanisms behind gamification's effects. Follow up studies after the intervention ends are also needed to see whether the gains are permanent. Finally, explore other gamification elements like storytelling, avatars, and collaborative missions to find the best combination for maximum impact.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest regarding the publication of this paper.

REFERENCES

Bai, S., Hew, K. F., & Huang, B. (2020). Does gamification improve student learning outcome? Evidence from a meta-analysis and synthesis of qualitative data in educational contexts. *Educational Research Review*, *30*, 100322. https://doi.org/10.1016/j.edurev.2020.100322

Bandura, A. (2006). Guide for constructing self-efficacy scales. In F. Pajares & T. Urdan (Eds.), *Self-efficacy beliefs of adolescents* (pp. 307–337). Information Age Publishing.

Beserra, V., Nussbaum, M., & Oteo, M. (2019). On-task and off-task behavior in the classroom: A study on mathematics learning with educational video games. *Journal of Educational Computing Research*, *56*(8), 1361–1383. https://doi.org/10.1177/0735633117744346

Chen, C. H., Law, V., & Chen, W. Y. (2018). The effects of peer competition-based science learning game on students' learning engagement, performance, and competition anxiety. *Computers & Education*, *125*, 145–160. https://doi.org/10.1016/j.compedu.2018.06.008

Dahal, P. P. (2025). Perception of mathematics self-efficacy and its role in academic achievement: A systematic review. *Mathematics Education Forum Chitwan*, 10(1), 52–68. https://doi.org/10.3126/mefc.v10i1.90995

Dichev, C., & Dicheva, D. (2017). Gamifying education: What is known, what is believed and what remains uncertain: A critical review. *International Journal of Educational Technology in Higher Education*, *14*(1), 1–36. https://doi.org/10.1186/s41239-017-0042-5

Do, N., Jin, T., Priest, R., Meredith, L. N., & Landers, R. N. (2024). A longitudinal quasi-experiment of leaderboard effectiveness on learner behaviors and course performance. *Learning and Individual Differences*, *116*, 104120. https://doi.org/10.1016/j.lindif.2024.104120

Espinosa, C. G. (2025). Learning with confidence: Self-efficacy as a catalyst for mathematical growth. *Education and Industry Review*, 1(1). https://doi.org/10.62596/eir.xx6nv094

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

Huang, B., & Hew, K. F. (2018). Implementing a theory-driven gamification model in higher education flipped courses: Effects on out-of-class activity completion and quality of artifacts. *Computers & Education*, *125*, 254–272. https://doi.org/10.1016/j.compedu.2018.06.018

Huang, R., Ritzhaupt, A. D., Sommer, M., Zhu, J., Stephen, A., Valle, N., … Li, J. (2020). The impact of gamification in educational settings on student learning outcomes: A meta-analysis. *Educational Technology Research and Development*, *68*(4), 1875–1901. https://doi.org/10.1007/s11423-020-09807-z

Hung, C. M., Huang, I., & Hwang, G. J. (2014). Effects of digital game-based learning on students' self-efficacy, motivation, anxiety, and achievements in learning mathematics. *Journal of Computers in Education*, *1*(2–3), 151–166. https://doi.org/10.1007/s40692-014-0008-8

Kiili, K., Moeller, K., & Ninaus, M. (2018). Evaluating the effectiveness of a game-based rational number training: In-game metrics as learning indicators. *Computers & Education*, *120*, 13–28. https://doi.org/10.1016/j.compedu.2018.01.012

Kilpatrick, J., Swafford, J., & Findell, B. (Eds.). (2001). *Adding it up: Helping children learn mathematics*. National Academy Press.

Koivisto, J., & Hamari, J. (2019). The rise of motivational information systems: A review of gamification research. *International Journal of Information Management*, *45*, 191–210. https://doi.org/10.1016/j.ijinfomgt.2018.10.013

Lee, W. C., & Lai, C. L. (2024). Facilitating mathematical argumentation by gamification: A gamified mobile collaborative learning approach for math courses. *International Journal of Science and Mathematics Education*, *22*(Suppl 1), 11–35. https://doi.org/10.1007/s10763-023-10350-5

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

Rodrigues, L., Pereira, F. D., Toda, A. M., Palomino, P. T., Pessoa, M., Carvalho, L. S. G., … Isotani, S. (2022). Gamification suffers from the novelty effect but benefits from the familiarization effect: Findings from a longitudinal study. *International Journal of Educational Technology in Higher Education*, *19*, Article 13. https://doi.org/10.1186/s41239-021-00314-6

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*(1), 68–78. https://doi.org/10.1037/0003-066X.55.1.68

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

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

TIMSS International Study Center. (2024). *TIMSS 2023 international results in mathematics and science*. Boston College, TIMSS & PIRLS International Study Center. https://doi.org/10.6017/lse.tpisc.timss.rs6460

Toda, A. M., Valle, P. H. D., & Isotani, S. (2018). The dark side of gamification: An overview of negative effects of gamification in education. In *Communications in Computer and Information Science* (Vol. 832, pp. 143–156). Springer. https://doi.org/10.1007/978-3-319-97934-2_9

Usher, E. L., & Pajares, F. (2008). Sources of self-efficacy in school: Critical review of the literature and future directions. *Review of Educational Research*, *78*(4), 751–796. https://doi.org/10.3102/0034654308321456

Wang, Y., Li, D., Li, J., & Bai, S. (2025). Exploring the relationship between self-efficacy, social support, academic anxiety, and academic outcomes: A meta-analysis structural equation modeling approach. *Frontiers in Psychology*, 16, 1714845. https://doi.org/10.3389/fpsyg.2025.1714845

Zainuddin, Z., Chu, S. K. W., Shujahat, M., & Perera, C. J. (2020). The impact of gamification on learning and instruction: A systematic review of empirical evidence. *Educational Research Review*, *30*, 100326. https://doi.org/10.1016/j.edurev.2020.100326