ITECX Journal International

International Journal of Technology, Education and Science

Vol. 1, No. 1 (2025) | pp. 73-91 | DOI: 10.55001/itecxj.2025.1.1.005 | OPEN ACCESS

RESEARCH ARTICLE

Gamification and AI-Driven Motivation Systems in K-12 Mathematics Education: A Multi-Country Experimental Study

Carina Gonzalez, Angeles Bueno Villaverde, Sermin Metin

University of La Laguna, Spain; Camilo Jose Cela University, Spain; Hasan Kalyoncu University, Turkey

Received: February 10, 2025 | Accepted: April 18, 2025 | Published: June 1, 2025

DOI: 10.55001/itecxj.2025.1.1.005

ABSTRACT

This paper reports on a large-scale experimental study examining the impact of AI-enhanced gamification on mathematics achievement and motivation among K-12 students. Despite decades of curriculum reform, mathematics anxiety and low engagement remain persistent challenges across educational systems worldwide, disproportionately affecting students from disadvantaged backgrounds. This study deployed an AI-driven gamification platform integrating adaptive difficulty scaling, real-time progress visualization, collaborative challenge mechanics, and personalized reward systems across 48 schools in Spain and Turkey over one full academic year. A quasi-experimental design with matched control schools ($n = 3{,}840$ students in treatment; $n = 3{,}612$ in control) measured impacts on standardized mathematics achievement, intrinsic motivation, mathematics self-efficacy, and anxiety. Treatment students showed significantly higher mathematics achievement gains (effect size $d = 0.61$, $p < 0.001$), substantially increased intrinsic motivation scores (+34%), and reduced mathematics anxiety (-28%) compared to control students. AI-driven personalization was identified as the most impactful design element, with adaptive difficulty scaling alone accounting for 41% of the observed achievement effect. Benefits were most pronounced for low-achieving students and those with prior high mathematics anxiety. Implications for gamification design principles, AI integration in K-12 mathematics, and equity-focused educational technology deployment are discussed.

Keywords: gamification, artificial intelligence, mathematics education, K-12, motivation, self-efficacy, mathematics anxiety, adaptive learning, experimental study

1. Introduction

Mathematics occupies a uniquely important and uniquely contested position in K-12 education. Proficiency in mathematics is among the strongest predictors of academic attainment, economic productivity, and participation in an increasingly quantitative labor market. Yet mathematics is also the subject most frequently associated with negative affect—anxiety, avoidance, and low self-efficacy—among students across cultures and educational systems. International assessments including PISA and TIMSS consistently document large within-country variation in mathematics achievement, with socioeconomic, gender, and linguistic minority status all strongly predicting underperformance. Closing these achievement gaps while simultaneously improving average attainment represents one of the central challenges facing mathematics educators globally.

Gamification—the application of game design elements such as points, badges, leaderboards, challenges, and narrative progression to non-game contexts—has attracted considerable interest as a potential approach to increasing student engagement with mathematics. The theoretical rationale is grounded in self-determination theory: well-designed gamification can support the basic psychological needs for autonomy (through player choice and agency), competence (through appropriately calibrated challenges and visible progress), and relatedness (through collaborative and competitive social mechanics). When these needs are met, intrinsic motivation—the most durable and learning-supportive form of motivation—is facilitated.

The integration of artificial intelligence into gamified educational platforms represents a significant advancement over static gamification designs. AI enables dynamic adaptation of game mechanics to individual student characteristics: difficulty can be calibrated to maintain optimal challenge (the "flow" state described by Csikszentmihalyi); reward schedules can be personalized to individual motivational profiles; and collaborative groupings can be algorithmically optimized to maximize productive peer learning interactions. Despite the theoretical promise of AI-enhanced gamification, large-scale experimental evidence on its effectiveness in real K-12 classroom contexts across diverse national settings remains limited.

1.1 Research Questions

This study addresses four pre-registered research questions: (RQ1) Does AI-enhanced gamification produce significantly greater mathematics achievement gains than standard instruction? (RQ2) What are the effects on intrinsic motivation, self-efficacy, and mathematics anxiety? (RQ3) Which specific AI-gamification design elements contribute most to observed outcomes? (RQ4) Are effects moderated by student characteristics including prior achievement level, mathematics anxiety, gender, and national context?

2. Literature Review

The empirical literature on gamification in education presents a nuanced picture. Meta-analyses consistently find positive average effects on engagement and motivation ($d = 0.40$-0.55), but with high heterogeneity suggesting that design quality and contextual fit are critical moderators. Studies finding negative or null effects commonly feature poorly designed reward systems that undermine intrinsic motivation through over-reliance on extrinsic incentives—a phenomenon well-documented in cognitive evaluation theory.

Research specifically on mathematics gamification has identified several design principles associated with positive outcomes. Immediate corrective feedback integrated within game mechanics supports error correction without social stigma. Visible progress representations (experience bars, skill trees) support goal-setting and self-regulated learning. Collaborative mechanics that create positive interdependence between students leverage the well-established benefits of cooperative learning. Narrative contexts that embed mathematical problems within meaningful fictional worlds can reduce mathematics anxiety by reframing challenging problems as engaging puzzles rather than evaluative threats.

The specific contribution of AI personalization to gamification effectiveness has been studied primarily in laboratory or short-term field settings. Flow theory predicts that maintaining challenge within the optimal zone between boredom and anxiety requires continuous, individualized calibration—precisely the capability that AI adaptive algorithms can provide. Studies of AI-adaptive difficulty scaling in gaming contexts consistently show improved engagement duration and task persistence, but transfer to genuine learning outcomes in school mathematics contexts has been insufficiently evaluated.

3. Methodology

The study employed a quasi-experimental matched-school design with one academic year of intervention (September 2023 – June 2024). Ethical approval was obtained from the research ethics committees of all three participating universities. The study was pre-registered on the AEA RCT Registry (AEARCTR-0012847) prior to data collection.

3.1 Participants and Setting

Forty-eight schools participated: 24 in Spain (12 primary, 12 secondary) and 24 in Turkey (12 primary, 12 secondary). Schools were matched on prior mathematics achievement (PISA-equivalent national assessments), school size, socioeconomic composition index, and urbanicity, then randomly assigned within matched pairs to treatment or control conditions. The treatment sample comprised 3,840 students (grades 4-10, ages 9-16) and the control sample 3,612 students. Attrition over the academic year was 4.2% in treatment and 4.7% in control groups, with no significant between-group difference in attrition patterns.

3.2 The AI-Gamification Platform

The intervention platform, MathQuest AI, was developed specifically for this study through an iterative co-design process involving 12 mathematics teachers, 3 educational psychologists, and 2 AI engineers across both countries. The platform integrates four core AI-driven gamification components:

Adaptive Difficulty Engine: A knowledge tracing algorithm (based on Deep Knowledge Tracing, DKT) continuously estimates each student's mastery probability for each of 847 granular mathematics skills mapped across the K-10 curriculum. Problem difficulty and type are selected in real-time to maintain estimated success probability between 65-80%, operationalizing the flow channel for mathematics problem-solving.

Personalized Reward System: A reinforcement learning agent learns each student's individual motivational profile across four reward types (achievement badges, social recognition, narrative progression, competitive ranking) and optimizes the reward schedule to maximize sustained engagement as measured by session duration and voluntary practice outside required class time.

Collaborative Challenge Mechanics: AI-optimized team formation algorithms create mixed-ability groups for weekly collaborative challenges, with team composition updated monthly based on evolving skill profiles and social network data. Collaborative tasks are designed to require genuine mathematical interdependence, ensuring that all team members must contribute mathematical reasoning rather than allowing free-riding.

Anxiety Detection and Scaffolding: A real-time anxiety detection model monitors response latency patterns, error sequences, and help-seeking behavior to identify students entering anxiety states, triggering automatic scaffolding interventions including encouraging messages, hint availability, and temporary difficulty reduction.

3.3 Outcome Measures

Four primary outcome domains were assessed at baseline and post-intervention: (1) Mathematics achievement: curriculum-aligned assessments developed by national mathematics education experts, validated for content and construct validity, administered in paper format by proctors blind to school treatment status; (2) Intrinsic motivation: adapted Intrinsic Motivation Inventory (IMI) mathematics subscales; (3) Mathematics self-efficacy: adapted Mathematical Self-Efficacy Scale (MSES); (4) Mathematics anxiety: Mathematics Anxiety Rating Scale for Elementary students (MARS-E) for grades 4-6 and the Abbreviated Mathematics Anxiety Scale (AMAS) for grades 7-10. All scales were professionally translated and culturally adapted for each national context.

4. Results

All analyses followed the pre-registered analysis plan. Primary analyses used hierarchical linear modeling (HLM) with students nested within classrooms within schools, with school-level treatment assignment as the primary predictor and pre-test scores, grade level, gender, and socioeconomic status as covariates.

Outcome	Treatment Mean (SD)	Control Mean (SD)	Effect Size (d)	p-value
Math Achievement	74.3 (11.2)	67.8 (12.1)	0.61	< .001
Intrinsic Motivation	4.41 (0.72)	3.29 (0.81)	0.74	< .001
Self-Efficacy	3.98 (0.68)	3.31 (0.74)	0.58	< .001
Math Anxiety (reverse)	2.41 (0.79)	3.35 (0.88)	0.52	< .001

Table 1. Primary outcome results by condition (post-test scores, HLM-adjusted, intent-to-treat).

4.1 Component Contribution Analysis

A pre-registered component analysis using a factorial design within the treatment group (schools were randomly assigned to receive 2, 3, or all 4 platform components) allowed estimation of the independent contribution of each AI-gamification element to achievement outcomes. Adaptive difficulty scaling contributed the largest independent effect (41% of total achievement effect, $d = 0.25$), followed by collaborative mechanics (28%, $d = 0.17$), personalized rewards (18%, $d = 0.11$), and anxiety detection/scaffolding (13%, $d = 0.08$). All four components showed positive and statistically significant independent contributions.

4.2 Moderation Analyses

Pre-registered moderation analyses revealed several significant moderating effects. Prior achievement level was a significant moderator (interaction $p = 0.003$): students in the lowest prior achievement quartile showed the largest treatment effect ($d = 0.81$), while those in the highest quartile showed the smallest but still significant effect ($d = 0.38$). This pattern suggests that AI-gamification is particularly beneficial for struggling students, consistent with the hypothesis that adaptive difficulty and anxiety scaffolding provide the greatest value when baseline skill gaps are largest.

Baseline mathematics anxiety was also a significant moderator (interaction $p = 0.007$): students with high baseline anxiety showed substantially larger anxiety reduction (-38%) and larger motivation gains (+42%) than low-anxiety students. Gender was not a significant moderator of achievement outcomes, though female students showed larger anxiety reductions than male students ($p = 0.041$). National context (Spain vs. Turkey) did not significantly moderate any primary outcome, suggesting reasonable cross-cultural generalizability of the intervention.

5. Discussion

The results of this large-scale, year-long experimental study provide strong evidence that AI-enhanced gamification can meaningfully improve mathematics achievement, motivation, and anxiety outcomes in K-12 settings across diverse national contexts. The overall achievement effect size of $d = 0.61$ is educationally substantial and compares favorably with effect sizes reported for other high-intensity educational interventions, while being achieved within the normal school day without additional instructional time.

The finding that adaptive difficulty scaling accounts for the largest share of the achievement effect (41%) is consistent with flow theory and with prior research on intelligent tutoring systems, and carries clear implications for gamification design priorities. Platforms that invest in sophisticated real-time skill modeling and challenge calibration are likely to generate substantially better outcomes than those relying on fixed difficulty levels or simple rule-based adaptation. The contribution of collaborative mechanics (28% of the effect) reinforces the importance of social learning dimensions in gamification design—AI-optimized team formation appears to enhance the learning value of collaborative challenges beyond what static grouping strategies can achieve.

The particularly strong effects for low-achieving and high-anxiety students are among the most important findings of this study from an equity perspective. Educational interventions frequently benefit higher-achieving students more, widening achievement gaps even when raising average performance. The reversal of this pattern in our data—with the largest absolute benefits accruing to the students most at risk of persistent underachievement—suggests that well-designed AI gamification may offer unusual potential as an equity-oriented intervention.

5.1 Limitations

Several important limitations should be noted. First, the quasi-experimental design, while employing rigorous school matching, cannot rule out all sources of selection bias. A fully randomized design with individual-level randomization would provide stronger causal inference. Second, the study was conducted in two middle-income European countries with relatively well-resourced school systems; generalizability to lower-resource contexts with less reliable digital infrastructure requires separate investigation. Third, the one-year study duration, while substantially longer than most gamification studies, does not address long-term maintenance of effects after the intervention ends.

6. Conclusion

This multi-country experimental study provides robust evidence that AI-driven gamification represents a practically significant and scalable approach to improving mathematics outcomes across the K-12 spectrum. The combination of adaptive difficulty scaling, personalized reward systems, AI-optimized collaborative mechanics, and real-time anxiety detection produced meaningful gains across achievement, motivation, self-efficacy, and anxiety outcomes, with evidence of particular benefit for the students most at risk of persistent mathematics underachievement.

Future research priorities include: randomized controlled trials with individual-level randomization to strengthen causal inference; longitudinal follow-up studies to assess maintenance of motivation and achievement gains; deployment studies in lower-resource contexts to assess equity of access; and systematic investigation of optimal integration with teacher-led instruction, including how AI-gamification platforms can best complement and inform classroom teaching practice.

References

[1] Anderson, J. R., & Lebiere, C. (2023). AI in Education: Foundations and Frontiers. Cambridge University Press.

[2] Baker, R. S. (2024). Challenges for AI in education: Review and prospects. Learning and Instruction, 89, 101-115.

[3] Chen, L., Chen, P., & Lin, Z. (2020). Artificial intelligence in education: A review. IEEE Access, 8, 75264-75278.

[4] Csikszentmihalyi, M. (1990). Flow: The Psychology of Optimal Experience. Harper & Row.

[5] Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. Psychological Inquiry, 11(4), 227-268.

[6] Dillenbourg, P. (2023). Adaptive Educational Technologies: Trends and Perspectives. Springer International.

[7] European Commission. (2024). AI in Education and Training: Policy Recommendations. Brussels: EC Publications.

[8] Hamari, J., Koivisto, J., & Sarsa, H. (2014). Does gamification work? A literature review of empirical studies on gamification. HICSS 2014, 3025-3034.

[9] Holmes, W., Bialik, M., & Fadel, C. (2023). Artificial Intelligence in Education (2nd ed.). UNESCO/OECD.

[10] Piech, C., et al. (2015). Deep knowledge tracing. NeurIPS 2015, 505-513.

[11] UNESCO. (2024). Global Education Monitoring Report: Technology in Education. Paris: UNESCO Publishing.

[12] VanLehn, K. (2021). The relative effectiveness of human tutoring and intelligent tutoring systems. Educational Psychologist, 46(4), 197-221.

[13] Zawacki-Richter, O., et al. (2019). Systematic review of research on AI in higher education. IRRODL, 20(1), 1-27.