International Journal of Instruction *October 2026 ● Vol.19, No.4*
e-ISSN: 1308-1470 ● www.e-iji.net *p-ISSN: 1694-609X*
pp. 345-370

Article submission code: 20260103114522
Received: 03/01/2026 Revision: 12/04/2026
Accepted: 21/04/2026 OnlineFirst: 01/07/2026

Effectiveness of Game-Based Learning (GBL) Approach on Students' Learning Outcomes in Mathematics Education: A Meta-Analysis

Da Tien Nguyen
Hanoi Metropolitan University, 98 Duong Quang Ham, Hanoi, Vietnam, *ntda@hnmu.edu.vn*

This study conducted a meta-analysis of 23 empirical studies (k = 23) (during the period 2010-2025) to evaluate the effectiveness of game-based learning (GBL) on student learning outcomes in Mathematics. The pooled results indicate that GBL has a positive and significant impact compared to traditional teaching methods, yielding a substantial overall effect size of Hedges's g = 0.651 (95% CI: 0.304 to 0.998, $p < 0.001$). This effect is especially evident when GBL is systematically integrated and aligned with learning objectives. In addition to improving academic scores and problem-solving skills, the findings suggest that GBL enhances learners' motivation and engagement. However, the study identified several important moderating factors, including game type, student age, and pedagogical implementation. This analysis provides strong empirical evidence supporting GBL as an effective tool in mathematics education and offers strategic recommendations for educators and developers to optimize its impact.

Keywords: effectiveness, game-based learning (GBL), mathematics education, meta-analysis, students learning outcomes

INTRODUCTION

In the era of innovative education, traditional mathematics instruction-characterized by one-way lectures and repetitive exercises-struggles to maintain student motivation and interest (Lagmay et al., 2024). This pedagogical limitation often results in a widespread paradox: significant "math anxiety" and a lack of confidence that hinders knowledge absorption and yields uneven learning outcomes (Ming Hung et al., 2014; Zapata et al., 2024). Consequently, the transition toward more effective methodologies has become urgent (Lagmay et al., 2024).

Game-Based Learning (GBL) has emerged as a prominent alternative, leveraging interaction, challenge, and instantaneous feedback to create dynamic engagement (Debrenti, 2024). By transforming abstract concepts into immersive experiences, GBL not only facilitates natural knowledge acquisition but also serves as a catalyst for 21st-century competencies. Specifically, GBL is proven to enhance critical thinking, problem-solving skills (Siew et al., 2017), and collaborative teamwork (İlhan, 2021),

Citation: Nguyen, D. T. (2026). Effectiveness of game-based learning (GBL) approach on students' learning outcomes in mathematics education: A meta-analysis. *International Journal of Instruction*, *19*(4), 345-370. https://doi.org/10.29333/iji.2026.19418a

providing students with the essential cognitive and social tools required for future success.

The need for GBL in the modern educational context has been reinforced by a large number of empirical studies suggesting its enhanced effectiveness compared to traditional methods (McLaren et al., 2017; Letsa-Agbozo et al., 2023; Bhadawkar & Gupta, 2021). These studies have shown that GBL has the potential to have a multidimensional impact on students' learning. Specifically, Bhadawkar and Gupta (2021) provided evidence indicating that GBL has a positive and significant impact on various aspects of students. Not only does it associated with improvements in academic performance in terms of grades, GBL also plays an important role in increasing students' confidence and reducing their anxiety levels in Mathematics (Ming Hung et al., 2014)–a major psychological barrier in learning. Furthermore, GBL has also been recognized as an effective pedagogical approach for training higher-level cognitive skills, such as improving logical thinking ability in Algebra (Siew et al., 2017), and enhancing computational performance (Jay et al., 2019). Another important finding is the potential of GBL to support disadvantaged student groups. Bang et al. (2023) specifically emphasized that GBL can be flexibly adapted to support students with low initial proficiency, thereby positively contributing to narrowing achievement gaps between student groups, promoting equity in education. It is worth noting that the effectiveness of GBL is not limited to complex digital forms. Research literature shows that GBL includes digital games such as "My Math Academy" (Bang et al., 2023) or "Decimal Point" (McLaren et al., 2017), but also extends to familiar non-digital forms such as card games and board games (Chin & Zakaria, 2015; Szilágyi et al., 2025). This diversity underscores the flexibility and wide applicability of GBL in many different educational environments and conditions.

While empirical evidence for GBL appears promising, existing research is often constrained by small-scale trials and narrow topical focuses, which limit generalizability (McLaren et al., 2017). Furthermore, inconsistent results stemming from divergent research designs highlight a significant knowledge gap (McLaren et al., 2017). To address this heterogeneity, this meta-analysis provides a systematic, quantitative assessment of GBL's effectiveness in Mathematics education, aiming to move beyond individual study limitations to offer a scientifically grounded synthesis of "when, where, and how" GBL optimizes learning. Specifically, this study addresses the following Research Questions (RQs):

RQ1: What is the overall impact of GBL on students' mathematics learning outcomes, specifically regarding cognitive achievement, attitudes, and practical skills?

RQ2: To what extent do the mathematical domain (e.g., arithmetic, algebra, geometry) and grade level moderate GBL's effectiveness?

RQ3: Which game-specific factors, including type, design architecture, and intervention duration, most significantly influence learning outcomes?

The success of GBL in promoting mathematical competencies is fundamentally anchored in the Circular Learning Framework (Figure 1), which operationalizes Experiential Learning Theory into a coherent pedagogical cycle that directly informs the research questions of this study. This framework serves as more than a descriptive model; it provides the structural foundation for RQ1 by mapping how the transition from "Concrete Experience" (game interaction) to "Reflective Observation" (analysis of outcomes) impacts cognitive achievement, attitudes, and practical skills. Furthermore, the stage of "Abstract Conceptualization"—where learners generalize game rules into formal mathematical principles—provides the theoretical basis for RQ2, as the ease of this conceptual transition may be moderated by specific mathematics domains, such as algebra or geometry, and the developmental grade level of the student. Finally, the cycle concludes with "Active Experimentation", where students apply knowledge to novel tasks, a process influenced by the game's type, design, and duration, which forms the core of RQ3.

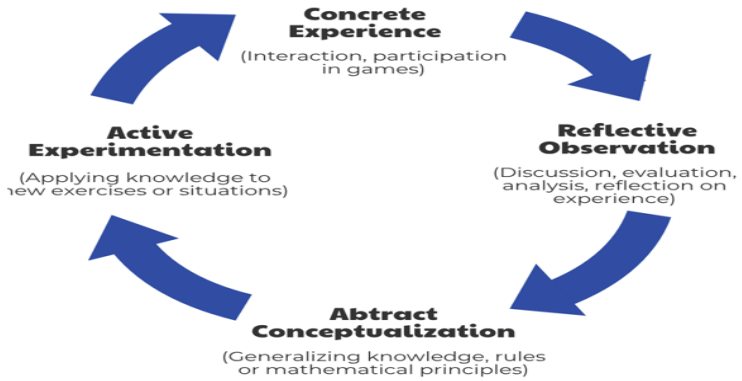


Figure 1
Experiential learning theory (Kolb, 1984)

LITERATURE REVIEW

Definitions of GBL

As a modern pedagogical shift, GBL redefines the educational experience by transitioning from passive reception to an active, learner-centered model where knowledge is constructed through engagement (Abdallah & Wardat, 2021; Letsa-Agbozo et al., 2023). Researchers emphasize that GBL leverages the intrinsic motivational and social elements of games—such as competition, rules, and entertainment-to significantly enhance learning quality (Jarrah et al., 2022; Ukwueze, 2024). By applying core gaming principles to practical educational contexts, GBL strategies foster sustained motivation and engagement (Anastasiadis et al., 2018; Bhadawkar H.R et al., 2021). According to Rebecca Teed, the efficacy of GBL is rooted in three key features: the creation of meaningful challenges to drive perseverance, the integration of compelling narratives or fantasy elements to immerse learners, and a safe

environment that encourages learning from mistakes, allowing students to refine their competencies through iterative play.

Previous survey studies on GBL

Empirical research consistently demonstrates that Game-Based Learning (GBL) significantly enhances learning outcomes, particularly in STEM fields (Conmy, 2023; Ming-Hsiu et al., 2019). By intentionally embedding instructional objectives within game mechanics, GBL fosters a learner-centered environment that promotes active problem-solving and knowledge construction (Hung et al., 2014; Jarrah et al., 2022). This approach not only improves cognitive functions such as concentration and critical thinking (Hung et al., 2014; Ming-Hsiu et al., 2019) but also elevates student motivation and engagement (Chin & Effandi, 2015; Ukwueze, 2024). Notably, studies in Vietnam and Hungary indicate that GBL empowers lower-attaining students to participate more confidently and aligns with student preferences for practical problem-solving over abstract definitions (Nguyen et al., 2022; Szilvia Szilágyi et al., 2025). Consequently, the successful implementation of GBL necessitates strategic lesson design, enhanced teacher technological proficiency, and the flexible use of blended learning platforms (Nguyen et al., 2022).

The growing adoption of GBL in higher education reflects a paradigm shift from passive instruction to active knowledge exploration, leveraging immersive virtual environments to facilitate cost-effective and evaluative learning experiences (Omar Al-Jamili et al., 2024; Anupama et al., 2021). Beyond stimulating intrinsic motivation and collaborative depth, meta-analytical evidence confirms that GBL significantly outperforms traditional methods in improving mathematics outcomes (Anupama et al., 2021; Umit et al., 2019). Specifically, a comprehensive review by Conmy et al. (2023) demonstrated significant gains across algebra, fractions, and problem-solving, with arithmetic yielding an exceptional effect size of 3.10, underscoring GBL's unique potency for computational and logical tasks. However, the successful implementation of this innovative shift remains hindered by systemic challenges, including technical constraints and the difficulty of aligning game mechanics with learning objectives. Consequently, maximizing GBL's educational potential is contingent upon rigorous game design, standardized pedagogical frameworks, and active teacher intervention—empirical prerequisites for bridging the gap between immersive gameplay and formal mathematical mastery (Conmy et al., 2023; Umit et al., 2019).

Some studies on GBL in Mathematics Education

In the era of digital transformation, GBL has emerged as a multidimensional pedagogical framework that transcends traditional instruction by simultaneously addressing academic achievement, psychological factors, and core competencies. Synthesis of recent literature reveals that GBL significantly mitigates math anxiety while bolstering self-efficacy and intrinsic motivation (Hung et al., 2014), particularly when supported by adaptive technologies that foster self-regulation (S-RAL) (Bang et al., 2023). This efficacy extends to fundamental computational skills, where specialized tools like ABACUS and game-based arithmetic instruction enhance calculation speed, accuracy, and the mastery of complex concepts such as fractions (Jarrah et al., 2022;

Jay et al., 2019). Beyond digital platforms, physical motor games and 3D simulations catalyze higher-order competencies including mathematical modeling (MathMM), spatial reasoning, and social collaboration (MathCC) (Bhadawkar et al., 2021; Christopoulos et al., 2024; İlhan, 2021). However, optimizing these benefits requires a strategic architectural balance between academic rigor and "conceptual scaffolding" (Barzilai et al., 2014)-a necessity that extends to higher education for equipping pre-service teachers with essential technological pedagogical content knowledge (Debrenti, 2024). Ultimately, GBL facilitates a transition from "learning to know" to "playing to understand," though the variance in individual study outcomes warrants a systematic meta-analysis to provide a robust, quantitative synthesis of its overall impact.

METHOD

This study employed a meta-analysis design to synthesize and quantify empirical evidence regarding the effectiveness of GBL in mathematics education. The methodology followed a rigorous process, beginning with a systematic literature search across reputable scientific databases, guided by clearly defined inclusion and exclusion criteria to ensure the selection of high-quality, relevant studies. To ensure the integrity of the coding procedures, data were independently coded and subsequently verified by an independent researcher to establish inter-rater reliability. This verification process, measured by Cohen's Kappa (> 0.8), yielded a high level of agreement, thereby mitigating potential bias from single-author coding. Finally, data analysis was conducted using a random-effects model to estimate the pooled effect size (Hedges's g) on learning outcomes and motivation, while moderator analyses were performed to explore the underlying factors contributing to the observed heterogeneity across the included studies.

Research Design

The study used a meta-analysis design to quantitatively synthesize the results of experimental studies that have examined the effectiveness of GBL methods for teaching and learning Mathematics. Meta-analysis is a systematic research method that allows the integration of individual results from many previous studies, thereby estimating the overall impact of a pedagogical measure, testing the stability of the effect, and analyzing modulating factors that may affect the results (Glass, 1976; Borenstein, Hedges, Higgins & Rothstein, 2010). In this study, the results of students' achievement and interest in the context of teaching Mathematics when applying GBL were systematized, coded and analyzed according to a standard statistical framework. This approach allows to draw general conclusions and compare the effectiveness of GBL under different conditions.

Search Procedures

The literature search process was conducted systematically to collect a full range of research works related to the effectiveness of Game-Based Learning (GBL) in teaching Mathematics. The research team used major electronic databases, including: ERIC, Google Scholar, Scopus, Web of Science, ScienceDirect, and SpringerLink, with the support of academic search engines.

The keywords used in the search process were a combination of the phrases: "game-based learning", "digital game", "mathematics education", "learning outcomes", "academic achievement", and "motivation". The search scope was limited to articles published from 2010 to 2025, ensuring the timeliness and relevance to the research objectives. In addition, the research team manually reviewed the reference lists of relevant articles, as well as a number of conference proceedings to add potential studies not indexed in the main database. Duplicates were removed by cross-referencing information on an electronic spreadsheet, followed by screening by title and abstract. Studies that met the initial criteria were read in full text to assess their potential for inclusion in the analysis.

The study selection process followed the PRISMA protocol guidelines, and the results were presented visually using a flow diagram showing the number of studies at each screening step. Flow diagram of the literature search using PRISMA (Moher et al., 2009) protocol (Figure 2):

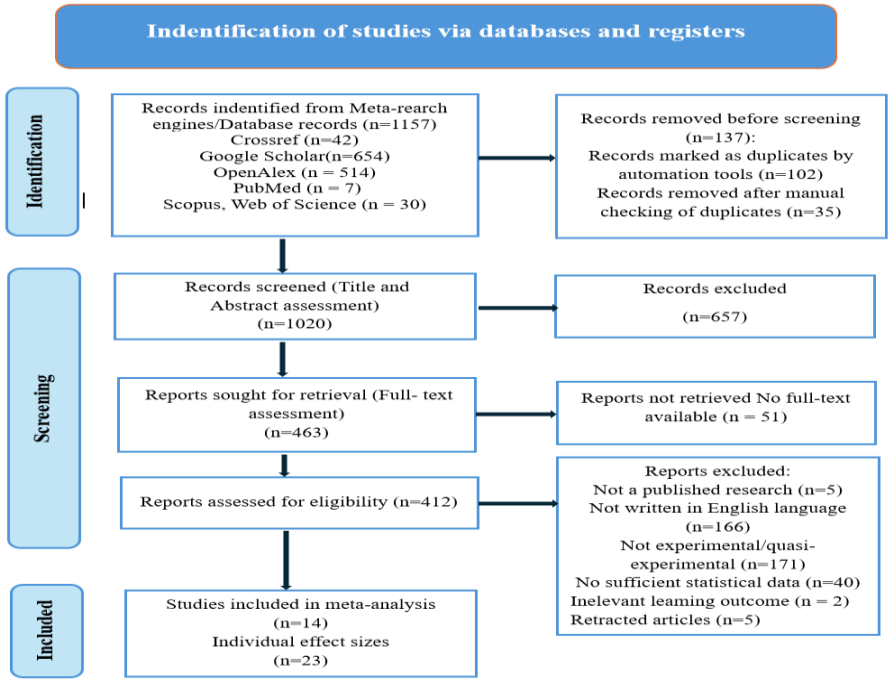


Figure 2
Flow diagram of the literature search using PRISMA protocol

Inclusion and Exclusion Criteria

Inclusion Criteria:

a) Full-text available online.
b) Written in English.
c) Experimental.

d) The systematic search encompassed studies published between January 1, 2010, and August 30, 2025. The final database retrieval was conducted on September 15, 2025, ensuring the inclusion of the most recent empirical evidence available up to the point of analysis.

e) The title or abstract must clearly mention Game-Based Learning (GBL).

f) Use an experimental or quasi-experimental design with at least one control group (using traditional or non-GBL methods) and one experimental group (using GBL methods).

g) Provide sufficient statistics or qualitative data necessary for analysis (e.g., mean, standard deviation, sample size).

h) Measure at least one of the following learning outcomes as the dependent variable: learning attitudes or learning achievement.

i) Conducted within the grade level from Primary to University.

j) Focus on the field of Mathematics education.

Exclusion Criteria:

a) Studies without full text.

b) Not experimental or lacking a control group.

c) Not providing sufficient data for calculation or analysis.

d) Not directly related to the application of GBL in Mathematics teaching.

To ensure consistency and reliability, this meta-analysis applied rigorous inclusion and exclusion criteria to select high-quality studies. Eligible works were required to: (1) be full-text English publications; (2) utilize experimental or quasi-experimental designs with both control and experimental groups; (3) be published between 2010–2025; (4) explicitly address GBL in mathematics from primary to university levels; and (5) provide sufficient data to measure learning attitudes or achievements. Studies lacking full-text access, experimental methodology, or direct relevance were excluded. This stringent screening process filtered out hundreds of candidates, retaining only the studies that fulfilled established academic standards for final synthesis.

Coding Procedures

Eligible studies were systematically extracted and coded using a standardized framework encompassing publication metadata, geographical context, educational variables (level, domain, discipline), GBL variations, and comparative statistical data (Mean, Standard Deviation, Sample size). Adopting the classification of Kraiger, Ford, and Sals (1993), learning outcomes were categorized into cognitive, affective, and psychomotor domains. To ensure reliability, two independent coders processed 20% of the sample, achieving a Kappa coefficient above 0.80. The final selection of 23 articles underwent rigorous supervision and collaborative review to rectify discrepancies and standardize notation. Ambiguous or missing variables were strictly coded as "Not Reported" (NR) to maintain analytical transparency and data integrity.

Data Analysis

Data analysis was performed using a random-effects statistical model, with the assumption that the 23 effect sizes from the 23 included studies reflected a random

sample of the total population of possible studies on the effectiveness of game-based learning (GBL) in teaching Mathematics (Borenstein, Hedges, Higgins & Rothstein, 2010).

The effect size for each study was calculated using Hedges' g to measure the standardized difference between the GBL and control groups, a metric preferred over Cohen's d for its superior correction of small sample size bias (Hedges & Olkin, 1985). According to Cohen's (1988) classification, effect sizes were categorized into four levels: large ($g \geq 0.80$), medium ($0.50 \leq g < 0.79$), small ($0.20 \leq g < 0.49$), and insignificant ($g < 0.19$), where a higher positive value indicates superior learning outcomes for GBL over traditional methods. Statistical analyses were performed using Comprehensive Meta-Analysis (CMA) version 4 (Borenstein et al., 2022), incorporating a mixed-effects model for moderator and subgroup analyses across factors such as (a) research region, (b) school level, (c) learning outcome domain, (d) GBL variant, and (e) implementation duration. All tests were conducted at a 95% confidence interval ($p < 0.05$), while potential publication bias was rigorously evaluated through Begg-Mazumdar rank correlation, Egger regression, and both Orwin and traditional fail-safe N tests (Begg & Mazumdar, 1994; Egger et al., 1997), with bias confirmed if $\rho < 0.05$.

FINDINGS

General Study Characteristics

A total of 23 studies with various samples of game-based learning (GBL) were included in this analysis, including 3327 learners (GBL group 1696, Non-GBL group 1631) in various educational contexts. The studies were published between 2010 and 2025, with a significant increase in the period 2021-2025. Geographically, the majority of studies were conducted in Asia (Malaysia, Vietnam, Philippines, Turkey, UAE, etc.), while others were from Europe (Hungary, Romania, Spain, Scotland, UK) and North America (USA), as well as South America (Ecuador) and Africa (Ghana, Nigeria). This shows the global nature of GBL research and application. At the same time, it reflects the widespread interest of researchers and educators in the effectiveness of the GBL method in teaching mathematics, not only in terms of improving academic performance but also in creating motivation, interest and developing comprehensive capacity for learners.

In terms of learning domains, 13 studies focused on cognitive abilities (Cognitive), 7 studies explored both cognition and attitudes simultaneously, while a few primarily focused on the domain of attitudes or learning behaviors (Affective). The levels of education ranged from preschool (PreS), primary school (PS), junior high school (JHS), senior high school (SHS), to university (HE), suggesting that GBL can be applied across is applicable across various levels of education. Of these, junior high and senior high school students were the most studied groups. In terms of science domains, all studies were conducted in Mathematics, covering a diverse range of topics such as arithmetic, fractions, algebra, spatial geometry, and applied mathematics skills.

Regarding the research design, the majority were quasi-experimental (Quasi), accounting for 14/23 studies. A few applied randomized controlled trials (RCT) (Miller et al., 2011; Barzilai et al., 2014; McLaren et al., 2017; Bang et al., 2022), or mixed

designs and action research (Letsa-Agbozo et al., 2023). Statistics from the experimental and control groups showed statistically significant improvements associated with GBL, especially in improving mathematics achievement and positive learning attitudes. The negative results in some studies may reflect difficulties in integrating complex mathematical concepts into game mechanics, or a lack of teacher guidance (Scaffolding) in these studies

Thus, the aggregated data show that GBL is not only international and applicable across all levels of education, but also diverse in implementation and research design. This provides a solid foundation for further exploration of GBL as a promising teaching method in the context of mathematics education reform.

Effect Sizes of GBL approach in mathematics education

When analyzed using the fixed model (**Table 2**), the pooled estimate for Hedges' g was 0.403 (SE = 0.037), 95% confidence interval [0.331, 0.474]; the z test showed that this effect was statistically significant ($z = 11.032$, $p < 0.001$). However, given the heterogeneity across studies, the random model was employed: under the random model, Hedges' $g = 0.651$ (SE = 0.177), 95% confidence interval [0.304, 0.998] ($z = 3.674$, $p < 0.001$). The observation that the random model estimate was larger than the fixed model suggests that when accounting for differences across studies, the average effect of GBL yielded a more pronounced average effect size. Heterogeneity analysis pointed to significant between-study heterogeneity: Q statistic = 440.588 ($df = 22$, $p < 0.001$) and $I^2 \approx 95.0\%$, suggesting that a substantial proportion of the observed variance stems from genuine differences between studies rather than sampling error. The tau-squared (τ^2) values ≈ 0.654 and tau ≈ 0.809 indicated substantial between-study variance. These indices suggest that the pooled effect should be interpreted cautiously due to high heterogeneity. Forest plots illustrate that the majority of studies reported positive effects associated with GBL interventions, some had large effect sizes (e.g., several studies with $g > 1.0$), while a few studies had insignificant or negative effects. The presence of significant outliers in the data warrants further investigation to determine if they stem from measurement disparities, data entry errors, or unique contextual factors. Therefore, sensitivity and subgroup analyses should be performed to understand the source of heterogeneity.

Overall, the findings provide empirical support for the positive impact of GBL on mathematics learning outcomes. The random model estimates mean $g \approx 0.65$, which is a medium-large effect size empirically. This suggests that, on average, GBL interventions – if designed and implemented appropriately-have the potential to improve learning outcomes compared to traditional approaches. The very high heterogeneity ($I^2 \approx 95\%$) suggests that the results are not uniform: the effectiveness of GBL depends strongly on many contextual factors such as: type of GBL (digital/non-digital; single-player game/game-based design), intervention duration, level of teacher-student interaction, grade level, content of mathematics taught, and quality of study design (RCT vs quasi). Therefore, although the evidence supports GBL in general, it is not advisable to recommend uniform application of all forms of GBL in all contexts, but rather to consider design philosophy and school context.

To assess the robustness of the findings and the potential impact of extreme outliers, a leave-one-out sensitivity analysis was conducted. By iteratively removing one study at a time and re-calculating the pooled effect size, the results remained remarkably stable, with Hedges's g fluctuating within a narrow range between 0.592 and 0.699. This consistency indicates that the primary result (g = 0.651) was not disproportionately influenced by any single outlier, confirming the robustness and reliability of the synthesized evidence.

Moderator Analysis of Studies Using GBL Instruction

The meta-analysis revealed a significant positive effect of GBL on students' mathematics learning outcomes. Based on the random-effects model, the overall effect size across the 23 included studies was 0.651 (95% CI [0.304, 0.998], $p < 0.001$), suggesting that GBL is associated with a moderate to strong positive impact on mathematics learning. However, the heterogeneity among the studies was extremely high (I^2 = 95%), points to substantial variability in effect sizes. This implies that the effectiveness of GBL may be influenced by contextual factors such as educational level, implementation approach, subject focus, or geographic context.

In the moderator analysis, the following variables were included in the evaluation: study locale, educational level, learning domain outcomes, GBL approach variation, and implementation time (Table 4). Results indicate that the largest effect size on learner learning outcomes was observed in studies in South America (g = 3.246), followed by studies in Asia (g = 1.041) and Europe (g = 0.249). The between-group heterogeneity test confirmed significant disparities across study locations (Q_b = 68.908, ρ = 0.000). In terms of educational level, GBL teaching method had a larger impact on junior high school students (g = 2.325) than on primary school students (g = 1.376), and the lowest effect was recorded at the high school level (g = -0.352). The effect sizes of the included studies differed significantly (Q-value = 78.020, ρ = 0.000). Further analysis of the effect sizes of the studies when classified by learning domain is also needed. Affective skills (g = 1.505), cognitive skills (g = 0.096), combined cognitive and affective (g = 1.123) are the three learning domains that were mainly affected by GBL teaching method. Contrary to the initial observation, significant differences were found across these three domains (Q_b = 10.589, ρ = 0.000). In addition, when it comes to the GBL approach variation variable, serious games (g = 1.706), digital (g = 0.853) and non-digital (g = 0.799) demonstrated substantial effect sizes, whereas game-based and action games recorded lower or negative results (Q = 179.008, $p < 0.001$). Finally, regarding implementation time, notable effect sizes were identified for durations of 16 hours (g = 2.325), 3 weeks (g = 1.952), and 2–3 months (g = 1.949), with significant variability also noted across this variable (Q = 338.109, $p < 0.001$).

Statistical analysis by continent

When analyzed by region (see "Locale" in **Table 4** and see **Table 5**), the effectiveness of GBL also varied significantly. Notably, studies from South America showed the largest effect size (g = 3.246), followed by Africa with g = 1.031. Although these figures were statistically significant, heterogeneity was very high (I^2 > 79%). In contrast, Europe had a significantly smaller effect size (0.138) and p-value = 0.025,

suggesting that the effect, while statistically significant, was not as robust as in other regions. This difference reinforces the importance of considering cultural and educational context when applying GBL. Mixed effects analysis also confirmed the overall positive effect (effect size 0.614, $p < 0.001$) and indicated that between-group heterogeneity factors played an important role in explaining this difference.

Statistical analysis by educational level

The meta-analysis shows that the effectiveness of GBL on mathematics learning outcomes varies significantly by grade level (see Table 6). When analyzed by grade level, Pre-S, Primary Ed, SS, and JHS showed positive and significant effect sizes, with the Secondary level having the highest effect (2.325, $p < 0.001$). In contrast, at the Higher education level, the effect size was significantly smaller, recording a negative result of $g = 0.352$, but was still significantly significant ($p = 0.044$). Notably, both JHS (Junior High School) and SHS (Senior High School) also had positive effects of 0.277 and 0.993, respectively. These results suggest that GBL may be particularly effective at the primary and early childhood education levels, where students may be more receptive to game elements. Contrary, mixed effects analysis also confirmed an overall positive effect (effect size 0.655) and highlighted that between-group heterogeneity plays an important role in explaining the heterogeneity of results.

Statistical analysis by GBL Approach variation

Meta-analysis results indicate that the impact of GBL is highly heterogeneous, varying significantly by method variation. Under both fixed-effect and mixed-effect models, Serious Games demonstrated the most robust and superior impact ($g = 1.647$ to 1.706, $p < 0.001$), followed by the strong, stable performance of Non-digital GBL ($g = 0.799$, $p < 0.001$) and an improved effect for Digital GBL in the mixed-effect model ($g = 0.853$, $p < 0.001$). Conversely, general Game-based approaches yielded statistically insignificant results ($p > 0.05$), while Action Games exhibited a consistent negative effect ($g = -5.799$, $p < 0.001$), suggesting they may be counterproductive in educational settings. These findings underscore that GBL should not be treated as a monolithic entity; rather, Serious Games and Digital GBL represent promising pathways for pedagogical innovation, whereas the integration of Action games into mathematics curricula warrants extreme caution due to potential inherent risks to learning efficacy.

Statistical analysis by learning domain

Analysis using the Fixed-effect model reveals significant disparities across learning domains, where the Affective group demonstrated a robust effect ($g = 1.366$, $p < 0.001$), starkly contrasting with the lower impacts observed in the Cognitive ($g = 0.205$) and combined Cognitive/Affective ($g = 0.433$) domains. While the pooled effect size was $g = 0.403$ (95% CI [0.331; 0.474]), extreme heterogeneity indices ($Q = 440.588$, $I^2 = 95.007\%$, $\tau^2 = 0.654$) necessitate a Mixed-effects analysis to account for inter-study variance. Under this more robust model, the overall effect size increased to 0.585, reaffirming the dominant and stable influence of the Affective factor ($g = 1.505$, $p = 0.001$) and the Cognitive/Affective synergy ($g = 1.123$, $p = 0.001$). Notably, the Cognitive group lost statistical significance ($g = 0.096$, $p = 0.687$) in the Mixed-effects

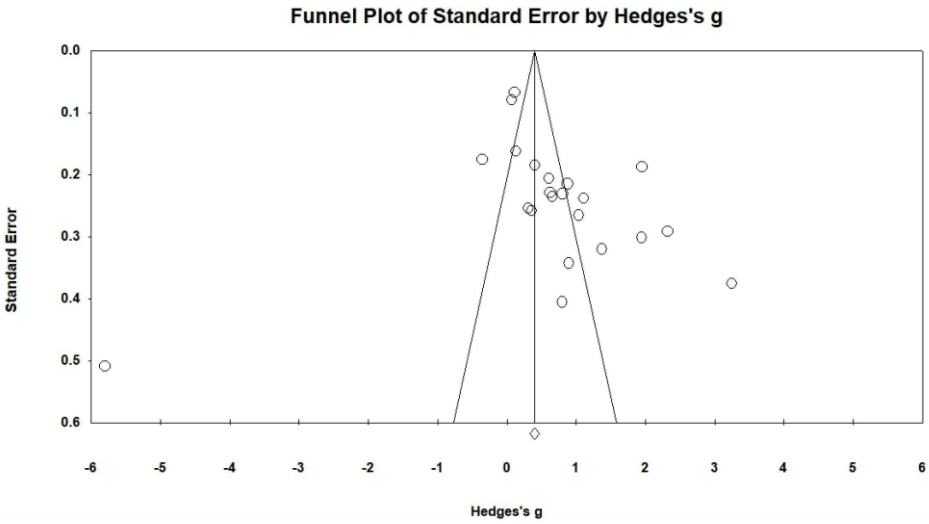
model, highlighting that while emotional and attitudinal outcomes remain consistent, cognitive gains are highly contingent upon specific research contexts and implementation conditions.

Statistical analysis by duration

The meta-analysis identifies intervention duration as a critical moderator, with significant between-group differences. While the overall pooled effect size remained small-to-medium across both Fixed ($g = 0.403$) and Mixed-effects models ($g = 0.369$), granular analysis reveals that GBL effectiveness follows a non-linear trend. Extremely brief interventions (< 1 week) were either statistically insignificant ($p = 0.220$) or yielded negative outliers ($g = -5.799$), whereas intermediate durations produced the most robust gains, specifically at 3 weeks ($g = 1.952$), 6 weeks ($g = 0.805$), and 2–3 months ($g = 1.949$, $p < 0.001$). Conversely, extended implementations, such as 9-week ($g = 0.074$) or 4-month ($g = 0.112$) programs, failed to reach significance, suggesting a "diminishing returns" effect where prolonged exposure without novelty may reduce motivation. These results indicate that a 3-to-12-week window is the optimal "temporal sweet spot" for GBL, providing sufficient time for cognitive adaptation while maintaining the engagement necessary to foster significant learning outcomes.

Publication Bias

The funnel plot (Figure 3) illustrates a low risk of publication bias, as the 23 studies exhibit a relatively symmetrical distribution around a cluster of positive effect values ($g \approx 0.5–1.5$). While small-sample studies ($SE > 0.4$) show characteristic dispersion at the plot's base, large-sample studies ($SE < 0.2$) remain concentrated at the apex, consistently reporting moderate to high positive effects. Despite the presence of a few extreme outliers, the overall stability of the findings is confirmed by Orwin's and classical fail-safe N tests; specifically, the classical N-test indicates that 1,065 additional null-effect studies (over 40 times the current sample) would be required to render the overall p-value insignificant. With a Z-value of 13.47562 and a p-value of 0.000, the results demonstrate exceptional statistical robustness, suggesting that the observed positive impact of GBL is reliable and highly unlikely to be reversed by unpublished research.



Z-value for observed studies	13.47562	Tails	2.000000	Number of missing studies that would bring p > α (Fail-safe N)	1065.000
p-value for observed studies	0.000000	Z for alpha	1.959960		
Alpha	0.050000	Number of observed studies	23.00000		

Figure 3
Funnel plot of standard error of publication bias

GBL Approach Variations and Additional Strategies Used

This meta-analysis categorizes GBL applications in mathematics into five distinct models (see Table 10). Blended GBL is the most prevalent (44%), integrating digital and physical tools (e.g., Minecraft, GeoGebra, Hula Hoop) to bridge theory and practice. This is followed by Immersive GBL (30%), which emphasizes role-playing and collaboration (Christopoulos et al., 2024), and Simulation-based GBL (13%), utilizing realistic scenarios for decision-making and reflection (Barzilai & Blau, 2014). Augmented GBL (9%) leverages tools like Kahoot and Quizizz for post-class reinforcement, while Pure GBL (4%) delivers entire instructional content through traditional games without formal lectures (Bhadwakar & Gupta, 2021).

DISCUSSION

General assessment

This meta-analysis of 23 experimental studies demonstrates that GBL exerts a moderate to strong positive influence on mathematics achievement, evidenced by a significant overall effect size of Hedges' g = 0.651 ($p < 0.001$). These findings align with prior research highlighting GBL's pedagogical advantages in enhancing academic performance (Ming Hung et al., 2014), student engagement, logical reasoning in algebra (Siew et al., 2017), and computational proficiency (Jay et al., 2019). However, the high

heterogeneity (I^2 = 95%) underscores that effectiveness is contingent upon contextual moderators such as game design, learning environments, and implementation duration. Geographically, a stark disparity exists between the substantially higher effect sizes in South America (g = 3.246) and Asia (g = 1.041) compared to North America (g = 0.229) and Europe (g = 0.249), suggesting that GBL's success is deeply intertwined with cultural and pedagogical settings. In Asian contexts-where teacher-centered instruction is often prevalent-the introduction of interactive GBL provides a significant novelty effect that dramatically bolsters motivation and achievement (Ramli et al., 2022; Siew et al., 2017); conversely, in regions where learner-centered pedagogies are already standard, the incremental benefits of GBL appear less pronounced as students are already accustomed to active learning methodologies.

The effectiveness of GBL is significantly moderated by learner level and game type. Regarding learner level, GBL yielded the highest effect sizes in K-12 education-particularly middle school (g = 2.325) and preschool (g = 1.376)-where visual imagery and playful exploration align with younger students' developmental needs (Chin & Zakaria, 2015). Conversely, higher education showed lower effectiveness (g = 0.878), suggesting that adult learners require more sophisticated games closely integrated with specialized, high-level cognitive content (İlhan, 2021). In terms of GBL type, intentional design proved crucial: Serious games (g = 1.706) and structured digital/non-digital formats outperformed general action games, which showed unstable or even negative effects (g = -5.799). These findings emphasize that GBL is most impactful when game mechanics are strictly aligned with learning objectives through instructional scaffolding and formative feedback (Debrenti, 2024; Christopoulos et al., 2024). Ultimately, the success of GBL depends on balancing pedagogical structure with age-appropriate challenges.

Analysis of learning domains reveals significant variations in Mathematics, where the Affective domain yielded the strongest effect (g = 1.505), followed by the integrated cognitive-affective domain (g = 1.123), while the Cognitive domain alone recorded only a minor impact (g = 0.096). These findings suggest that GBL's primary contribution lies in fostering emotional engagement and reducing anxiety, which indirectly facilitates knowledge acquisition (Christopoulos et al., 2024; Jarrah et al., 2022; Debrenti, 2024). Furthermore, implementation duration significantly moderates effectiveness; medium-term interventions-specifically those lasting approximately three weeks (g = 1.952), sixteen hours (g = 2.325), or two to three months (g = 1.949)-proved most beneficial. This suggests that while extreme durations may diminish novelty or motivation, moderate exposure optimizes the balance between adaptation and sustained engagement (Ramli et al., 2022; Hung et al., 2014), necessitating careful temporal planning to maximize educational gains.

Thus, these moderators show that the effectiveness of GBL varies significantly depending on the context and pedagogical design. Specifically, learning domains determine the types of cognitive or emotional outcomes that respond most strongly to GBL; grade level determines the relevance of the game activity to the learner's development; while geography influences how students perceive and engage with the game-based learning activity. In addition, variation and implementation duration also

impact how learners perceive motivation, challenge, and feedback during the learning process. Overall, these results suggest that the success of GBL depends on maintaining a flexible balance between learner readiness, pedagogical goals, and game design quality, thereby ensuring that play is not only entertaining but also contributes to improving mathematical understanding and sustaining long-term learning engagement (Hung et al., 2014; Debrenti, 2024; Christopoulos et al., 2024).

IMPLICATIONS

Theoretically, this study provides robust synthetic evidence for GBL's effectiveness, enhancing its external validity as an active learning method while overcoming the sample size and methodological limitations of individual studies. By highlighting performance variances across grade levels and GBL formats, these findings elucidate the interaction between game design elements and cognitive mechanisms like working memory and critical thinking.

To optimize mathematical competencies, educators and developers should prioritize a hybrid integration of Serious and Digital Games, ensuring pedagogical objectives remain central while leveraging AI for personalized feedback. GBL should be employed as a deliberate strategy-particularly at the secondary level for complex modeling-rather than a total replacement for traditional methods. Developers are encouraged to facilitate a state of "flow" through adaptive difficulty, just-in-time instructional support to correct misconceptions, and multiplayer features that balance collaborative and competitive motivation.

LIMITATIONS

While this review deepens the understanding of GBL in promoting academic achievement across cognitive and affective domains, its findings are constrained by several methodological limitations, including a small sample size due to strict inclusion criteria, potential publication bias, and inconsistencies in study design, measurement tools, and intervention durations. To address these gaps and mitigate the risk of overestimating GBL's effectiveness, future research should transition toward advanced quantitative modelling-such as measuring reaction times and in-game decision-making-to clarify the cognitive mechanisms governing working memory and selective attention, while simultaneously examining GBL’s role as a moderator for affective factors like math anxiety and intrinsic motivation. Furthermore, there is a critical need for longitudinal and intervention studies that investigate previously overlooked contextual variables, specifically comparing individual versus collaborative learning and evaluating the impact of intensive teacher training within a unified theoretical framework. Finally, comparative analyses between GBL and other active learning paradigms, such as Problem-Based or Blended Learning, are essential to delineate the unique pedagogical contributions of GBL within the modern educational landscape.

CONCLUSION

This meta-analysis provides robust statistical evidence that GBL exerts a medium-to-large positive effect on mathematical learning outcomes, establishing it as a reliable pedagogical intervention that addresses previous ambiguities regarding its

generalizability. The efficacy of GBL is driven by its ability to mitigate math anxiety by transforming abstract concepts into concrete, goal-oriented challenges, alongside two core theoretical mechanisms: enhanced intrinsic motivation and active knowledge construction. Through elements like competition and rewards, GBL increases students' time-on-task and persistence, while the gaming environment necessitates the immediate application of mathematical concepts to resolve in-game puzzles. This creates an immediate feedback loop-often absent in traditional methods-that allows learners to iteratively self-adjust their strategies and consolidate their understanding more effectively. Our results are consistent with previous meta-analyses in STEM and Science education, but provide a more nuanced view in the context of Mathematics. Analysis of moderator variables, such as grade level and game type (digital vs. non-digital), revealed differences in the magnitude of the effects. For example, we found that the effect may be greater in primary school (where play elements are more strongly perceived) than in secondary school. Furthermore, while both digital and non-digital games are effective, digital applications may offer advantages in terms of personalization and the collection of detailed learning data, thereby optimizing the experience for individual students. These findings emphasize that GBL is not a "one-size-fits-all" solution and that its effectiveness depends on pedagogical design and adaptation to learner characteristics.

The practical implications of this research for educators and mathematics curriculum designers are numerous. First, educational institutions should consider systematically integrating GBL into the curriculum as a powerful pedagogical supplement rather than a wholesale replacement for traditional instruction. This is especially critical for topics often perceived as difficult or dry. However, effective implementation requires investment in teacher training, helping educators understand not only how to operate game engines but also how to synergize game mechanics with traditional classroom discourse. The goal is to ensure games serve as a core vehicle for learning, not just a peripheral reward. Furthermore, developers must prioritize mathematical accuracy and adaptive difficulty to support a diverse range of student proficiencies.

Although this meta-analysis provides compelling evidence, it does have several limitations. First, there was considerable heterogeneity among the component studies in intervention duration, outcome measures, and GBL definitions, which may affect the ability to synthesize. Second, the inherent risk of publication bias remains a potential concern. To address these issues, future research should focus on: (1) Conducting longitudinal studies to assess the sustainability of GBL effectiveness; (2) Investigating the optimal balance between GBL and direct instruction by comparing or combining them with other innovative methods like the Flipped Classroom; and (3) Exploring the social and emotional aspects of GBL, particularly the role of collaborative play in enhancing students' long-term mathematical self-efficacy.

DECLARATIONS

Confict of Interests The authors declare no confict of interest.

Data Availability The datasets generated and/or analyzed during the current study are not publicly available due to privacy concerns of the participants, but are available from the corresponding author on reasonable request.

REFERENCES

Abdallah, R., & Wardat, Y. (2021). Teachers' perceptions on the effectiveness of professional development programs in improving the curriculum implementation at Jordanian schools. *Elementary Education Online, 20*(5), 4438-4449. https://doi.org/10.17051/ilkonline.2021.01.126

Anastasiadis, T., Lampropoulos, G., & Siakas, K. (2018). Digital game-based learning and serious games in education. *International Journal of Advances in Scientific Research and Engineering, 4*(12), 139-144. https://doi.org/10.31695/IJASRE.2018.33016

Bang, H. J., Li, L., & Flynn, K. (2023). Efficacy of an adaptive game-based math learning app to support personalized learning and improve early elementary school students' learning. *Early Childhood Education Journal, 51*(4), 717–732. https://doi.org/10.1007/s10643-022- 01332-3

Barzilai, S., & Blau, I. (2014). Scaffolding game-based learning: Impact on learning achievements, perceived learning, and game experiences. *Computers & Education, 70*, 65–79. https://doi.org/10.1016/j.compedu.2013.08.003

Begg, C. B., & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics, 50*(4), 1088–1101. https://doi.org/10.2307/2533446

Bhadawkar, H. R., & Gupta, M. K. (2021). A study of the effectiveness of game-based learning strategies for enhancing achievement and interest in mathematics among middle school students. https://svv-research-data.s3.ap-south-1.amazonaws.com/1

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2010). *Introduction to meta-analysis*. Wiley. https://doi.org/10.1002/9780470743386

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2022). *Comprehensive meta-analysis* (Version 4) [Computer software]. Biostat. https://www.meta-analysis.com

Chin, L. C., & Zakaria, E. (2015). Effect of game-based learning activities on children's positive learning and prosocial behaviours. *Jurnal Pendidikan Malaysia, 40*(2), 159–165. http://dx.doi.org/10.17576/JPEN-2015-4002-08

Christopoulos, A., Mystakidis, S., Kurczaba, J., Laakso, M. J., & Stylios, C. (2024). Is immersion in 3D virtual games associated with mathematical ability improvement in game-based learning? *International Journal of Science and Mathematics Education, 22*(7), 1479–1499. https://doi.org/10.1007/s10763-023-10440-4

Conmy, T. (2023). We are still playing: a meta-analysis of game-based learning in mathematics education**.** https://repository.usfca.edu/diss/679

Debrenti, E. (2024). Using digital game-based learning in mathematics education: A case study with teacher training students. *International Journal for Technology in Mathematics Education, 31*(3), 153–162. http://dx.doi.org/10.1564/tme_v31.3.06

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ, 315*(7109), 629–634. https://doi.org/10.1136/bmj.315.7109.629

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher, 5*(10), 3–8. https://doi.org/10.3102/0013189X005010003

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), 151–166. https://doi.org/10.1007/s40692-014-0008-8

İlhan, A. (2021). The impact of game-based, modeling, and collaborative learning methods on the achievements, motivations, and visual mathematical literacy perceptions. *SAGE Open,11*(1), 21582440211003567. https://doi.org/10.1177/21582440211003567

Jarrah, A. M., Almassri, H., Johnson, J. D., & Wardat, Y. (2022). Assessing the impact of digital games-based learning on students' performance in learning fractions using (ABACUS) software application. *Eurasia Journal of Mathematics, Science and Technology Education,18*(10), em2159. https://doi.org/10.29333/ejmste/12421

Jay, T., Habgood, J., Mees, M., & Howard-Jones, P. (2019, October). Game-based training to promote arithmetic fluency. In *Frontiers in Education*, *4,* 118. https://doi.org/10.3389/feduc.2019.00118

Kolb, D. A. (1984). Experiential learning: Experience as the source of learning and development. Englewood Cliffs, NJ: Prentice Hall.

Lagmay, S. B., Duldulao, S. M. D., Gallena, B. D., Pascua, M. T., & Cruz, N. T. (2024). Effects of game-based mathematics learning on students' academic achievement. *International Journal for Multidisciplinary Research, 6*(5).

Letsa-Agbozo, J. K., Susuoroka, G., & Donnoe, M. A. (2023). The effects of game-based mathematics learning on pupils' academic performance in junior high school. *ISRG Journal of Arts Humanities & Social Sciences, 1*(6). https://doi.org/10.5281/zenodo.10141391

Maat, S. M., & Khalid, F. (2023). Digital game-based learning and learning analytics in mathematics. *Pegem Journal of Education & Instruction, 13*(1), 168-176. https://doi.org/10.47750/pegegog.13.01.19

McLaren, B. M., Adams, D. M., Mayer, R. E., & Forlizzi, J. (2017). A computer-based game promotes mathematics learning more than a conventional approach. *International Journal of Game-Based Learning (IJGBL), 7*(1), 36–56. https://doi/10.4018/IJGBL.2017010103

Miller, D. J., & Robertson, D. P. (2011). Educational benefits of using game consoles in a primary classroom: A randomised controlled trial. *British Journal of Educational Technology, 42*(5), 850–864. https://doi.org/10.1111/j.1467-8535.2010.01114.x

Ming-Hsiu et al. (2019). Effects of Game-Based Instruction on the Results of Primary School Children Taking a Natural Science Course Education Sciences. *Educ. Sci, 9*(2),79. https://doi.org/10.3390/educsci9020079

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *BMJ, 339*, b2535. https://doi.org/10.1136/bmj.b2535

Nguyen, L. P., Oanh, N. P. P., Thao, N. P., Van De, T., & Triet, L. V. M. (2022). Activity theory as a framework for teaching mathematics: An experimental study. *Heliyon, 8*(10). https://doi.org/10.1016/j.heliyon.2022.e10789

Omar Al-Jamili et al. (2024). Evaluating the efficacy of computer games-based learning intervention in enhancing English speaking proficiency. *Science Direct*. https://doi.org/10.1016/j.heliyon.2024.e36440

Siew, N. M., Geofrey, J., & Lee, B. N. (2016). Students' algebraic thinking and attitudes towards algebra: The effects of game-based learning using Dragonbox 12+ app. *Electronic Journal of Mathematics & Technology, 10*(1).

https://ejmt.mathandtech.org/Contents/eJMT_v10n1p1.pdf

Sors, O., Puig, A., & Rodriguez, I. (2025). Analytical and practical study of digital game-based learning under the lens of activity theory: A geometry game case study. *International Journal of Serious Games, 12*(2), 91–113. https://doi.org/10.17083/ijsg.v12i2.916

Szilágyi, S., Palencsár, E., Körei, A., & Török, Z. (2025). Examining the effectiveness of non-digital game-based learning among university computer science students on the topic of improper integrals. *Education Sciences, 15*(2), 132. https://doi.org/10.3390/educsci15020132

Telner, D., Bujas-Bobanovic, M., Chan, D., Chester, B., Marlow, B., Meuser, J., ... & Harvey, B. (2010). Game-based versus traditional case-based learning: Comparing effectiveness in stroke continuing medical education. *Canadian Family Physician, 56*(9), e345–e351. https://www.cfp.ca/content/56/9/e345

Ukwueze, T. C., Ugwuanyi, C. C., & Odo, I. O. (2024). Efficacy of game-based learning (GBL) instructional approach on students' task persistence in number bases. *International Journal of Research and Innovation in Social Science (IJRISS), 8*(08), 3228–3236. https://dx.doi.org/10.47772/IJRISS.2024.8080237

Umit et al. (2019). Effects of game-based learning on students' mathematics achievement: A meta-analysis. *Journal of Computer Assisted Learning*. https://doi.org/10.1111/jcal.12347

Zapata, M., Ramos-Galarza, C., Valencia-Aragón, K., & Guachi, L. (2024). Enhancing mathematics learning with 3D augmented reality escape room. *International Journal of Educational Research Open, 7*, 100389. https://doi.org/10.1016/j.ijedro.2024.100389

Zhang, M., Trussell, R. P., Gallegos, B., & Asam, R. R. (2015). Using math apps for improving student learning: An exploratory study in an inclusive fourth grade classroom. *TechTrends, 59*(2), 32–39. https://doi.org/10.1007/s11528-015-0837-y

APPENDİCES

Table 1
Descriptive features of the included studies that investigated the effectiveness of GBL approach

No	Author	Year	Locale	Learning domain	Educational level	Scientific discipline	GBL Approach variation	Duration	Research design	GBL Group Statistical Data			Non - GBL Group Statistical Data		
										Mean	SD	n	Mean	SD	n
1	Ukwueze	2024	Nsukka	Affective	JHS	Mathematics	SG	3 weeks	Quasi	72.64	6.27	81	55.84	10.25	87
2	Lu Chung Chin & Effandi Zakaria	2015	Malaysia	Affective	PreS	Mathematics	NDGBL	NR	Quasi	38.74	5.5	23	23	14.79	24
3	Adeeb M. Jarrah et al.	2022	UAE	Cognitive	PS	Mathematics	DGBL	NR	Quasi	11.93	1.74	41	9.9	1.88	39
4	David J. Miller et al.	2011	Scotland	Cognitive, Affective	JHS	Mathematics	DGBL	9 weeks	Randomized interventional pretest-posttest	83.04	16.86	326	81.65	20.8	309
5	Tim Jay et al.	2019	UK	Cognitive	PS	Mathematics	SG	2 weeks	Quasi	10.65	7.4	49	6.2	7.17	48
6	Sarit Barzilai et al.	2014	Irasel	Cognitive	PS	Mathematics, finance	SGBL	5 months	RCT	3.01	1.915	130	2.58	1.89	52
7	Chun-Ming Hung et al.	2014	Taiwan	Affective, Cognitive	PS	Mathematics	DGBL	7 hours	Quasi	89.935	9.13	46	87.22	7.47	23
8	Szilvia Szilágyi et al.	2025	Hungary	Cognitive	HE	Mathematics	NDGBL	2 years	Quasi	3.03	1.78	31	1.78	1.34	32
9	Bhadawkar H.R et al.	2021	India	Cognitive	SHS	Mathematics	NDGBL	6 weeks	Quasi	29.57	5.39	40	25.20	5.36	40
10	Nyet Moi Siew et al.	2017	Malaysia	Affective	SS	Mathematics education	SG	16 hours	Quasi	71.1	9.5	39	49.6	8.8	39
11	Ramli et al	2022	Malaysia	Cognitive, Affective	PS	Mathematics	DGBL	2-3 months	Quasi	87.513	13.097	32	52.356	21.524	32
12	Athanasios Christopoulos et al.	2024	Greece	Cognitive, Affective	JHS	Mathematics	DGBL	4 weeks	Quasi	51.56	7.42	59	48.79	7.18	59
13	Mireya Zapata et al.	2024	South America (Ecuador)	Cognitive, Affective	SHS	Mathematics	DGBL	NR	Mixed	7.78	2.18	32	0.618	2.18	33
14	Edith Debrenti	2024	Europe (Romania)	Cognitive	HE	Mathematics	DGBL	1 year	Quasi	74.89	11.25	47	79.63	14.23	108
15	Saniata B. Lagmay	2024	Asia (Philippines)	Cognitive	JHS	Mathematics	GBL	4 weeks	Quasi	21.3	5.72	41	18.1	4.19	38
16	Nguyen Phu Loc et al.	2022	Vietnam	Cognitive, affective	SHS	Mathematics	NDGBL	5 months	Experimental	6.24	1.04	37	5.34	1.6	38
17	Oriol Sors et al.	2025	Spain	Cognitive	JHS	Mathematics	NDGBL	NR	Quasi	6.95	2.33	11	5.03	2.33	14
18	Aziz İlhan	2021	Turkey	Cognitive	PS	Mathematics	DGBL	NR	Post-test experimental	17.11	1.31	35	16.39	2.61	26
19	Meilan Zhang et al.	2015	USA	Cognitive	PS	Mathematics	DGBL	8 weeks	Quasi	13.77	3.19	18	10.27	4.36	18

No	Author	Year	Locale	Learning domain	Educational level	Scientific discipline	GBL Approach variation	Duration	Research design	GBL Group Statistical Data			Non - GBL Group Statistical Data		
										Mean	SD	n	Mean	SD	n
20	Bruce M. McLaren et al.	2017	USA	Cognitive	JHS	Mathematics	DGBL	2 weeks	RCT	38.6	9.6	70	30.1	10.5	83
21	Deanna Telner et al.	2010	Canada	Cognitive, Affective	Postgraduate	Education	GBL	3 months	RCT	27.4	4.5	16	27.7	3	15
22	Hee Jin Bang et al.	2022	USA	Cognitive	PS	Mathematics	GBL	4 months	RCT	20.97	8.01	452	20.08	7.84	434
23	Joseph Kafui Letsa-Agbozo et al.	2023	Wa Ghana	Cognitive	JHS	Mathematics	GBL	5 days	Action research. mixed methods	40.4	13.18	40	98.0	4.45	40

Note. *JHS* = Junior High School; *SHS* = Senior High School; *PS*=Primary School; *PreS* = Pre School; *SS* = Secondary School; *HE* = Higher Education; *GBL* = Game-Based Learning; DGBL = Digital Game-Based Learning; SG = Serious Games; NDGBL = Non - Digital Game-Based Learning; *NR* = Not Reported; *Quasi* = Quasi Experimental Design; *Mixed* = Mixed Methods Design; *Pure* = Pure Experimental Design; *SD* = Standard Deviation; *n* = Sample size

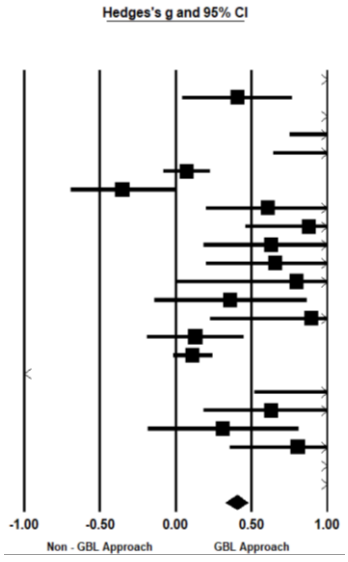
Table 2
Overall effect size and heterogeneity analysis

Model		Effect size and 95% confidence interval					Test of null (2-Tail)		Heterogeneity			Tau-squared				
Model	Number studies	Point estimate	St. error	Variance	Lower limit	Upper limit	Z-value	P-value	Q-value	df(Q)	P-value	I-squared	Tau squared	St. error	Variance	Tau
Fixed	23	0.403	0.037	0.001	0.331	0.474	11.032	0.000	440.588	22	0.000	95.007	0.654	0.343	0.117	0.809
Random	23	0.651	0.177	0.031	0.304	0.998	3.674	0.000								

Note. Std=Standard error; Significant at $\rho < 0.05$

Table 3
Effect sizes distribution and forest plot of studies that used GBL approach

Study name	Statistics for each study						
	Hedges's g	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value
Ukwueze (2024)	1.952	0.187	0.035	1.586	2.319	10.440	0.000
Christopoulos et al. (2024)	0.405	0.185	0.034	0.043	0.767	2.192	0.028
Zapata et al. (2024)	3.246	0.376	0.141	2.510	3.982	8.640	0.000
Chin & Zakaria (2015)	1.376	0.320	0.102	0.748	2.003	4.298	0.000
Jay et al. (2019)	1.111	0.238	0.057	0.644	1.578	4.662	0.000
Miller et al. (2011)	0.074	0.079	0.006	-0.082	0.229	0.927	0.354
Debrenti (2024)	-0.352	0.175	0.031	-0.695	-0.009	-2.010	0.044
Barzilai et al. (2014a)	0.606	0.206	0.042	0.202	1.010	2.939	0.003
Telner et al. (2010a)	0.878	0.214	0.046	0.458	1.298	4.095	0.000
Bhadawkar et al. (2021)	0.628	0.229	0.052	0.180	1.076	2.750	0.006
Nguyen et al. (2022)	0.658	0.235	0.055	0.198	1.119	2.803	0.005
Sors et al. (2025)	0.797	0.406	0.165	0.002	1.592	1.965	0.049
Aziz Ihan (2021)	0.361	0.258	0.066	-0.144	0.866	1.401	0.161
Zhang et al. (2015)	0.896	0.343	0.117	0.224	1.567	2.615	0.009
McLaren et al. (2017)	0.129	0.162	0.026	-0.187	0.446	0.801	0.423
Bang et al. (2022)	0.112	0.067	0.005	-0.020	0.244	1.669	0.095
Letsa-Agbozo et al. (2023)	-5.799	0.509	0.259	-6.797	-4.801	-11.390	0.000



Szilágyi et al. (2025)	1.037	0.265	0.070	0.516	1.557	3.905	0.000
Lagmay (2024)	0.628	0.229	0.052	0.180	1.076	2.750	0.006
Hung et al. (2014)	0.311	0.254	0.064	-0.186	0.809	1.227	0.220
Bhadawkar et al. (2021b)	0.805	0.230	0.053	0.354	1.257	3.494	0.000
Siew et al. (2017)	2.325	0.291	0.085	1.754	2.896	7.978	0.000
Ramli et al. (2022)	1.949	0.301	0.091	1.359	2.540	6.474	0.000
	0.651	0.177	0.031	0.304	0.998	3.674	0.000

Table 4
Moderator and subgroup analyses of studies that investigated the effectiveness of GBL approach

Moderator	Subgroups	k	Test for Effect				Test for Heterogeneity		
			Hedges' g	Std error	Variance	95% CI	Q-Value	Df (Q)	Q - Value
Locale	South America	1	3.246	0.376	0.141	2.520, 3.982	68.908	4	0.000
	Asia	11	1.041	0.166	0.027	0.716, 1.366			
	Europe	6	0.249	0.149	0.022	-0.043, 0.540			
	America	3	0.229	0.148	0.022	-0.062, 0.520			
	Africa	2	-1.909	3.876	15.022	-9.505, 5.687			
Educational level	Secondary	1	2.325	0.291	0.085	1.754, 2.896	78.020	6	0.000
	Pre School	1	1.376	0.320	0.102	0.748, 2.003			
	Senior High	5	1.229	0.371	0.138	0.502, 1.956			
	Postgraduate	1	0.878	0.214	0.046	0.468, 1.298			
	Primary School	7	0.734	0.234	0.055	0.276, 1.192			
	Junior High	7	-0.152	0.436	0.190	-1.006, 0.702			
	High Education	1	-0.352	0.175	0.031	-0.695, -0.009			
	Senior High	5	1.229	0.371	0.138	0.502, 1.956			
Learning domain	Affective	4	1.505	0.466	0.218	0.591. 2.419	10.589	2	0.000
	Cognitive, affective	7	1.123	0.346	0.119	0.446, 1.800			
	Cognitive	12	0.096	0.237	0.056	-0.369, 0.560			
GBL approach variation	Serious Games	3	1.706	0.424	0.180	0.875, 2.537	179.008	4	0.000
	Digital	10	0.853	0.227	0.052	0.408, 1.298			
	Non digital	7	0.799	0.097	0.009	0.608, 0.990			
	Game-based	2	-0.092	0.230	0.053	-0.543, 0.360			
	Action	1	-5.799	0.509	0.259	-6.797, -4.801			
Duration of implementation	16 hours	1	2.325	0.291	0.085	1.754, 2.896	338.109	13	0.000
	3 weeks	1	1.952	0.187	0.035	1.586, 2.319			
	2-3 months	1	1.949	0.301	0.091	1.350, 2.540			
	2 years	1	1.037	0.265	0.070	0.516, 1.557			
	NR	7	1.032	0.386	0.149	0.275, 1.789			
	8 weeks	1	0.896	0.343	0.117	0.224, 1.567			
	6 weeks	1	0.805	0.230	0.053	0.354, 1.257			
	5 month	1	0.658	0.235	0.055	0.198, 1.119			
	4 weeks	3	0.532	0.122	0.015	0.293, 0.770			
	2 weeks	2	0.350	0.238	0.056	-0.115, 0.816			
	7 hours	1	0.311	0.254	0.064	-0.186, 0.809			
	4 months	1	0.112	0.067	0.005	-0.020, 0.244			
	9 weeks	1	0.074	0.079	0.006	-0.082, 0.229			
	5 days	1	-5.799	0.509	0.259	-6.797, -4.801			

Significant at $\rho < .05$ error; CI

Table 5
Results of the Regional Effect Meta-Analysis

Groups		Effect size and 95% confidence interval					Test of null (2-Tail)		Heterogeneity				Tau-squared			
Group	Number studies	Point estimate	Std error	Variance	Lower limit	Upper limit	Z-value	P-value	Q-value	df (Q)	P-value	I-squared	Tau squared	Std error	Variance	Tau
Fixed effect analysis																

Africa	2	1.031	0.176	0.031	0.687	1.375	5.874	0.000	204.237	1	0.000	99.510	29.897	42.488	1805.248	5.468
America	3	0.140	0.061	0.004	0.020	0.259	2.285	0.022	5.043	2	0.080	60.342	0.038	0.069	0.005	0.195
Asia	11	0.970	0.075	0.006	0.823	1.118	12.908	0.000	47.868	10	0.000	79.109	0.237	0.136	0.018	0.487
Europe	6	0.138	0.061	0.004	0.018	0.258	2.246	0.025	19.124	5	0.002	73.855	0.088	0.087	0.008	0.297
South America	1	3.246	0.376	0.141	2.510	3.982	8.640	0.000	0.000	0	0.000	0.000	0.000	0.000	0.000	0.000
Total within									276.273	18	0.000					
Total between									164.316	4	0.000					
Overall	23	0.403	0.037	0.001	0.331	0.474	11.032	0.000	440.588	22	0.000	95.007	0.654	0.343	0.117	0.809
Mixed effect analysis																
Africa	2	-1.909	3.876	15.022	−9.505	5.687	−0.493	0.622								
America	3	0.229	0.148	0.022	−0.062	0.520	1.543	0.123								
Asia	11	1.041	0.166	0.027	0.716	1.366	6.282	0.000								
Europe	6	0.249	0.149	0.022	−0.043	0.540	1.673	0.094								
South America	1	3.246	0.376	0.141	2.510	3.982	8.640	0.000								
Total between									68.908	4	0.000					
Overall	23	0.614	0.086	0.007	0.445	0.783	7.116	0.000								

Table 6
Results of the Meta-Analysis of Effects by Education Level

Groups		Effect size and 95% confidence interval					Test of null (2-Tail)		Heterogeneity				Tau-squared			
Group	Number studies	Point estimate	Std error	Vari-ance	Lower limit	Upper limit	Z-value	P-value	Q-value	df (Q)	P-value	I-squared	Tau squared	Std error	Vari-ance	Tau
Fixed effect analysis																
High Education	1	−0.352	0.175	0.031	−0.695	−0.009	−2.010	0.044	0.000	0	1.000	0.000	0.000	0.000	0.000	0.000
Junior High	7	0.277	0.059	0.004	0.161	0.394	4.677	0.000	234.588	6	0.000	97.442	1.251	0.985	0.970	1.118
Postgraduate	1	0.878	0.214	0.046	0.458	1.298	4.095	0.000	0.000	0	1.000	0.000	0.000	0.000	0.000	0.000
Pre School	1	1.376	0.320	0.102	0.748	2.003	4.298	0.000	0.000	0	1.000	0.000	0.000	0.000	0.000	0.000
Primary School	7	0.314	0.057	0.003	0.203	0.424	5.546	0.000	54.622	6	0.000	89.015	0.322	0.271	0.073	0.567
Secondary	1	2.325	0.291	0.085	1.754	2.896	7.978	0.000	0.000	0	1.000	0.000	0.000	0.000	0.000	0.000
Senior High	5	0.993	0.114	0.013	0.770	1.216	8.735	0.000	41.236	4	0.000	90.300	0.615	0.497	0.247	0.784
Total within									330.446	16	0.000					
Total between									110.142	6	0.000					
Overall	23	0.403	0.037	0.001	0.331	0.474	11.032	0.000	440.588	22	0.000	95.007	0.654	0.343	0.117	0.809
Mixed effect analysis																
High Education	1	−0.352	0.175	0.031	−0.695	−0.009	−2.010	0.044								
Junior High	7	−0.152	0.436	0.190	−1.006	0.702	−0.349	0.727								
Postgraduate	1	0.878	0.214	0.046	0.458	1.298	4.095	0.000								
Pre School	1	1.376	0.320	0.102	0.748	2.003	4.298	0.000								
Primary School	7	0.734	0.234	0.055	0.276	1.192	3.141	0.002								
Secondary	1	2.325	0.291	0.085	1.754	2.896	7.978	0.000								
Senior High	5	1.229	0.371	0.138	0.502	1.956	3.314	0.001								
Total between									78.020	6	0.000					

Overall	23	0.655	0.097	0.009	0.466	0.845	6.773	0.000

Table 7
Study Effects and Heterogeneity by Intervention Type

Groups		Effect size and 95% confidence interval					Test of null (2-Tail)		Heterogeneity			Tau-squared				
Group	Number studies	Point estimate	Std error	Vari-ance	Lower limit	Upper limit	Z-value	P-value	Q-value	df (Q)	P-value	I-squared	Tau squared	Std error	Vari-ance	Tau
Fixed effect analysis																
Action	1	-5.799	0.509	0.259	-6.797	-4.801	-11.390	0.000	0.000	0	1.000	0.000	0.000	0.000	0.000	0.000
Digital	10	0.384	0.055	0.003	0.275	0.493	6.920	0.000	115.657	9	0.000	92.218	0.454	0.310	0.096	0.674
Game-Based	2	0.053	0.063	0.004	-0.070	0.176	0.838	0.402	6.127	1	0.013	83.678	0.090	0.152	0.023	0.300
Non Digital	7	0.799	0.097	0.009	0.608	0.990	8.201	0.000	5.522	6	0.479	0.000	0.000	0.039	0.002	0.000
Serious Games	3	1.647	0.127	0.016	1.398	1.895	12.980	0.000	20.944	2	0.000	90.451	0.484	0.552	0.304	0.696
Total within									148.249	18	0.000		0.654	0.343	0.117	0.809
Total between									292.339	4	0.000					
Overall	23	0.403	0.037	0.001	0.331	0.474	11.032	0.000	440.588	22	0.000	95.007				
Mixed effect analysis																
Action	1	-5.799	0.509	0.259	-6.797	-4.801	-11.390	0.000								
Digital	10	0.853	0.227	0.052	0.408	1.298	3.759	0.000								
Game-Based	2	0.092	0.230	0.053	-0.543	0.360	-0.398	0.690								
Non Digital	7	0.799	0.097	0.009	0.608	0.990	8.201	0.000								
Serious Games	3	1.706	0.424	0.180	0.875	2.537	4.026	0.000								
Total between									179.008	4	0.000					
Overall	23	0.563	0.081	0.007	0.404	0.721	6.962	0.000								

Table 8
Summary of Meta-Analysis by Type of Outcome Measure

Groups		Effect size and 95% confidence interval					Test of null (2-Tail)		Heterogeneity			Tau-squared				
Group	Number studies	Point estimate	Std error	Vari-ance	Lower limit	Upper limit	Z-value	P-value	Q-value	df (Q)	P-value	I-squared	Tau squared	Std error	Vari-ance	Tau
Fixed effect analysis																
Affective	4	1.366	0.112	0.013	1.146	1.586	12.175	0.000	47.686	3	0.000	93.709	0.807	0.755	0.570	0.898
Cognitive	12	0.205	0.049	0.002	0.110	0.300	4.220	0.000	189.764	11	0.000	94.203	0.602	0.397	0.157	0.776
Cognitive, affective	7	0.433	0.064	0.004	0.308	0.558	6.798	0.000	112.586	6	0.000	94.671	0.770	0.608	0.370	0.877
Total within									350.036	20	0.000					
Total between									90.552	2	0.000					
Overall	23	0.403	0.037	0.001	0.331	0.474	11.032	0.000	440.588	22	0.000	95.007	0.654	0.343	0.117	0.809
Mixed effect analysis																
Affective	4	1.505	0.466	0.218	0.591	2.419	3.227	0.001								
Cognitive	12	0.096	0.237	0.056	-0.369	0.560	0.404	0.687								
Cognitive, affective	7	1.123	0.346	0.119	0.446	1.800	3.250	0.001								
Total between									10.589	2	0.005					

Overall	23	0.585	0.180	0.032	0.232	0.938	3.248	0.001

Table 9
Study Effects and Heterogeneity by duration

Groups		Effect size and 95% confidence interval					Test of null (2-Tail)		Heterogeneity			Tau-squared				
Group	Number studies	Point estimate	Std error	Vari-ance	Lower limit	Upper limit	Z-value	P-value	Q-value	df (Q)	P-value	I-squared	Tau squared	Standard error	Vari-ance	Tau
Fixed effect analysis																
16 hours	1	2.325	0.291	0.085	1.754	2.896	7.978	0.000	0.000	0	1.000	0.000	0.000	0.000	0.000	0.000
2 weeks	2	0.311	0.127	0.016	0.062	0.560	2.444	0.015	3.308	1	0.069	69.772	0.079	0.160	0.026	0.281
2 years	1	1.037	0.265	0.070	0.516	1.557	3.905	0.000	0.000	0	1.000	0.000	0.000	0.000	0.000	0.000
2–3 months	1	1.949	0.301	0.091	1.359	2.540	6.474	0.000	0.000	0	1.000	0.000	0.000	0.000	0.000	0.000
3 weeks	1	1.952	0.187	0.035	1.588	2.319	10.440	0.000	0.000	0	1.000	0.000	0.000	0.000	0.000	0.000
4 months	1	0.112	0.067	0.005	-0.020	0.244	1.669	0.095	0.000	0	1.000	0.000	0.000	0.000	0.000	0.000
4 weeks	3	0.532	0.122	0.015	0.293	0.770	4.371	0.000	0.826	2	0.662	0.000	0.000	0.045	0.002	0.000
5 days	1	-5.799	0.509	0.259	-6.797	-4.801	-11.390	0.000	0.000	0	1.000	0.000	0.000	0.000	0.000	0.000
5 months	1	0.658	0.235	0.055	0.198	1.119	2.803	0.005	0.000	0	1.000	0.000	0.000	0.000	0.000	0.000
6 weeks	1	0.805	0.230	0.053	0.354	1.257	3.494	0.000	0.000	0	1.000	0.000	0.000	0.000	0.000	0.000
7 hours	1	0.311	0.254	0.064	-0.186	0.809	1.227	0.220	0.000	0	1.000	0.000	0.000	0.000	0.000	0.000
8 weeks	1	0.896	0.343	0.117	0.224	1.567	2.615	0.009	0.000	0	1.000	0.000	0.000	0.000	0.000	0.000
9 weeks	1	0.074	0.079	0.006	-0.082	0.229	0.927	0.354	0.000	0	1.000	0.000	0.000	0.000	0.000	0.000
NR	7	0.672	0.095	0.009	0.485	0.859	7.045	0.000	91.862	6	0.000	93.468	0.960	0.655	0.429	0.980
Total within									95.996	9	0.000					
Total between									344.592	13	0.000					
Overall	23	0.403	0.037	0.001	0.331	0.474	11.032	0.000	440.588	22	0.000	95.007	0.654	0.343	0.117	0.809
Mixed effect analysis																
16 hours	1	2.325	0.291	0.085	1.754	2.896	7.978	0.000								
2 weeks	2	0.350	0.238	0.056	-0.115	0.816	1.475	0.140								
2 years	1	1.037	0.265	0.070	0.516	1.557	3.905	0.000								
2–3 months	1	1.949	0.301	0.091	1.359	2.540	6.474	0.000								
3 weeks	1	1.952	0.187	0.035	1.588	2.319	10.440	0.000								
4 months	1	0.112	0.067	0.005	-0.020	0.244	1.669	0.095								
4 weeks	3	0.532	0.122	0.015	0.293	0.770	4.371	0.000								
5 days	1	-5.799	0.509	0.259	-6.797	-4.801	-11.390	0.000								
5 months	1	0.658	0.235	0.055	0.198	1.119	2.803	0.005								
6 weeks	1	0.805	0.230	0.053	0.354	1.257	3.494	0.000								
7 hours	1	0.311	0.254	0.064	-0.186	0.809	1.227	0.220								
8 weeks	1	0.896	0.343	0.117	0.224	1.567	2.615	0.009								
9 weeks	1	0.074	0.079	0.006	-0.082	0.229	0.927	0.354								
NR	7	1.032	0.386	0.149	0.275	1.789	2.671	0.008								
Total between									338.109	13	0.000					
Overall	23	0.369	0.041	0.002	0.289	0.448	9.050	0.00								

Table 10
GBL variations and additional strategies used in science teaching

GBL Variants	k (number of studies)	Hedges' g (effect size)	Effect size (according to Cohen, 1988)	Additional Instructional Strategies Used	Study Exemplar
Pure GBL	1	-5.8	No effect	The entire teaching content is designed and implemented using games, without traditional lectures.	Bhadawkar & Gupta (2021) - Grade 6 students learn ratio, percentage, profit/loss using Ring Toss, Hula Hoop, Musical Ball, Passing the Ball
GBL add-on	2	0.05	Small	The teacher teaches theory first, then uses games to reinforce and test knowledge.	Studies using Kahoot, Quizizz as review exercises; Musical Ball used at the end of class to increase interest
GBL combined	10	0.83	Large	Combines digital games (Minecraft, GeoGebra) and non-digital games (Hula Hoop, Passing the Ball), interspersed with lectures	Research using Minecraft + GeoGebra to teach spatial geometry, then students practice with group games
GBL based on simulation	3	1.7	Large	Students participate in real-life simulations (Minecraft, business simulations, trading) to role-play, make decisions, then the teacher summarizes.	Barzilai & Blau (2014) - Scaffolding GBL; Minecraft Education research for students to build and calculate the volume of a cube
GBL Role Playing	7	0.8	Large	Students play roles in the game world (businessman/explorer/project manager), perform scenario tasks related to Math objectives; emphasize communication - cooperation - decision making; Teacher supports role coordination and debrief to draw mathematical principles and debrief to draw mathematical principles	Barzilai & Blau (2014): In My Money, students play the role of a financial manager, solving a business scenario. Christopoulos et al. (2024): role-playing in a 3D world; high levels of immersion are associated with improved mathematical performance in a GBL context.

Note: k = number of effect sizes; according to Cohen's (1988) criteria, the magnitude of effect sizes is classified as follows: Large: g = 0.80 or more/ Medium: g = 0.50 to 0.79/ Small: g = 0.20 to 0.49/ No effect: g < 0.19