Vol. 3 No. 4 (2025)

ISSN 3007-3170(O)
ISSN :3007-3162(P)

Social Sciences & Humanity Research Review

Gamification and Learning Efficacy: A Meta Analysis

Dr Rabia Tabassum[1], Dr Bushra Yasin[2], Dr Mahvish Fatima Kashif[3]

[1]Lecturer, Department of Research and Policy Studies,Faculty of Education,Lahore College for Women University, Lahore, Email: rabia786@lcwu.edu.pk,
[2]Assistant Professor, Department of Education,University of Jhang
[3]Assistant Professor, Department of STEM Education,Lahore College for Women University, Lahore

ARTICLE INFO

Keywords:
Gamification in Education, Learning Efficacy, Meta-Analysis, Moderator Analysis, Competitive Environments

Corresponding Author:
Dr Rabia Tabassum,
Lecturer, Department of Research and Policy Studies,Faculty of Education,Lahore College for Women University, Lahore,
Email: rabia786@lcwu.edu.pk

ABSTRACT

Gamification is becoming increasingly popular in field of education. However, literature indicates an inconsistent pattern of their impact on learning contexts. This meta-analysis synthesizes evidence of gamification on learning efficacy. Moderating variables included conceptual approaches, contextual and situational factors, intervention duration, and methodological rigor. Sample of study included 40 research papers which included studies on gamification along with either of selected moderating variables. The data analysis included correlation and David Wilson moderation analysis using SPSS. This was done to disentangle the conditions under which gamification is most effective. Findings of the study indicated that collaborative gaming had weak and inconsistent associations with learning efficacy but this may indicate a moderate positive effect when combined with the competitive environments of blended social interaction game designs. As regard to time constraint, ultra-short applications of one day were negatively correlated with learning efficacy, while sustained interventions of one month or longer yielded moderate and consistent positive effects. On the other hand, methodological rigor, the experimental designs produced reliable moderate effects. Measurement instruments introduced notable variation; standardized tools reported the strongest effects, though based on limited evidence, while self-developed instruments consistently demonstrated moderate positive effects across a larger sample, albeit with high heterogeneity. It was concluded that gamification's impact on learning efficacy is conditional rather than universal. This study contributes to the growing body of evidence by clarifying the contextual and methodological conditions that moderate gamification's educational impact, offering guidance for both theory development and instructional practice.

Introduction

Gamification of learning is an emerging hot topic that aims at supplementing the learning process through increased motivation and interest by the learners. The gamification passion is meant to promote learning effectiveness and success in most of the literature (Molero et al., 2021; Rosa-Castillo et al., 2022). Nevertheless, the inconsistencies in the literature can be witnessed in this context as well i.e. some studies indicate significant engagement progress (Nguyen-Viet & Nguyen-Viet, 2023) and the rest indicate insignificant results (Pitoyo & Asib, 2020) . This fact sheds light on the fact that systematic synthesis is necessary to explain the contribution of gamification to the effectiveness of learning at what and how.There is new research that indicates that the effects of gamification are moderated by various factors. The different contextual factors that define gamification, as in the case of collaboration or competition, are differentiated by context (Eliyas & Ranjana, 2022a). These two factors appeal to the different students in different ways depending on their character and quality. Student interaction is subject to conditions of context like collaborative and competitive activity (Eliyas & Ranjana, 2022b). The length of intervention also seems to be a key point, as brief interventions do not always support the learning results, whereas longer interventions are more consistent.Conversely, methodological observations and rigors of the research which are, in turn, also dependent on the various working stages. Experimental designs or standardized measures, as the case of self-developed measures, would provide more reliable evidence (Wirani et al., 2022a) . However, the universal advantages and disadvantage of any of research methods can not be denied.To deal with these complexities, the current study performs a meta-analysis of empirical research on gamification in higher education. The study is going to segregate factors related to context, duration of intervention, and rigor of the methodology under moderation to unravel the conditions in which gamification can be the most effective and enhance theoretical insight and provide practical advice on instructional design (Wirani et al., 2022b).

Research Objectives

To systematically synthesize empirical evidence on the impact of gamification on learning efficacy within higher education, while examining the influence of key moderating variables such as contextual factors, duration of intervention, and methodological rigor.

Research Significance

This meta-analysis of Gamification and Learning Efficacy adds to the growing body of literature explaining how the design aspects of a game can be used to improve the learning process. To instructional designers and educators, the results highlight the necessity of hybrid competitive-collaborative models that wisely strike the balance between particular performance contingency and collective objective and peer support and thus develop not only personal responsibility but also group involvement. The findings also stress the fact that gamification should be implemented as a long-term, programmatic intervention that is in line with feedback processes, autonomy facilitation, and curricular evaluation, and not as a thin veneer. The methodological rigor is obligatory: prior registration of findings and complementary methods of measurements are suggested (standardized tests combined with validated and researcher-constructed scales) to reduce the criterion contamination and to strengthen the credibility. In the policy-and-practice sense, the research highlights the need to develop institutional structures that foster the adoption of gamified platforms in the long term, professional learning, and the infrastructure of scalable platforms. The policy-makers and administrators are encouraged to encourage monitoring of fidelity, multiple testing, and inter-disciplinary work to achieve stability of the effects and be replicable. Together, these facts put gamification as a purposeful, evidence-supported, pedagogical approach that has the potential to drive learning effectiveness in a wide range of educational settings and guide decision-making at the institutional and policy-making levels.

Literature Review

As presented above, gamification is a much debated subject because the past decades, empirical studies into gamification have grown quicker, and there are multiple meta-analyses that generalize its impact in various educational settings. (Legaki et al., 2020a) presented some preliminary results that gamification could have a positive effect on engagement and participation, however, results varied based on design quality. According to (Legaki et al., 2020b) , the positive impact on learning achievement was small-to-moderate but stated that the presence of meaningful integration of the game elements, but not the superficial application, is essential. More recent meta-analyses have supported these results, demonstrating that gamification positively affects cognitive outcomes including knowledge retention, motivational outcomes including self-efficacy, and behavioral outcomes including persistence (Sailer & Homner, 2020a) . Notably, these studies point out that combinations between mechanics (e.g., narrative and feedback) or badges and social interaction have a better tendency to produce more impact in comparison with single factors.Other moderators that define the effects of gamification are also found in the literature. Responsiveness depends on age and level of education with the younger learners usually demonstrating more motivation gains (Sailer & Homner, 2020b) . The issues of the subject domain are also important: more consistent improvements are observed in STEM and language learning contexts than in humanities, perhaps because the task structure is more apparent. The length of the intervention is also a variable; a short intervention is often able to capture the newness of the engagement, whereas a long intervention runs the risk of losing its effect unless the adaptive or narrative components of the implementation keep the interest alive (Gui et al., 2023) . Additionally, the quality of design may be a determining factor: while leaderboards may encourage competition, it becomes possible to lose lower-ranking students and badges and mastery paths are more efficient when they are associated with continuous feedback and advancement.Although the results are encouraging, there are still methodological shortcomings. Most of the studies have high self-reported indicators of motivation instead of objective measures of learning effectiveness, and there are concerns over validity (Slamet & Meng, 2025a) . The use of small sample sizes, convenience sample, and absence of blinding are typical, which adds to the bias. It has also been observed that there is publication bias on positive findings, which issues preregistered studies and standardized outcome measure (Slamet & Meng, 2025b) . Moreover, the majority of interventions are short-term, preventing the conclusion regarding the long-term retention and transfer of learning.Future research directions are towards adaptive gamification where learning challenges and rewards are configured to the learner profile and integration with learning analytics, which facilitates real-time feedback (Yu et al., 2024).

These methods demonstrate potentials in reducing heterogeneity of effects and maintaining engagement in the long run. Yet, there are still areas that are not well-known in regard to cross-cultural differences and scale in large classes or the relationship between extrinsic rewards and intrinsic motivation (Puspitasari & Arifin, 2023) . These gaps will be critical to be addressed in order to further evidence base and to make gamification a valuable part of the educational practice.Gamification has become a widely used approach to instruction in learning, basing it on the aspects of game design, including points, badges, leaderboards and storytelling to boost student motivation and achievement. The underlying principles of gamification tend to be placed within the Self-Determination Theory (SDT) assumptions that indicate that autonomy, competence, and relatedness are the key triggers of intrinsic motivation (Luarn et al., 2023). Trying to meet these psychological needs by introducing gameful qualities to the learning setting, educators can enhance the effectiveness of learning. There have been supporting views like Flow Theory which propose that gamification can be used to maintain engagement through a balance of challenge and skill, which creates an environment that promotes deep learning (O. S. Kaya & Ercag, 2023).

Moderating Variables

This meta-analysis takes into consideration the heterogeneity in the context of primary studies by considering multiple moderating variables that were identified in the prior research as having effects on evaluation of gamification impacts on learning outcomes. The presence of these moderators makes it possible to have a more complex picture of which conditions make gamification most useful, and it is also possible to bring the synthesis to another level of synthesis beyond aggregate effect sizes and even analyze contextual and conceptual heterogeneity.

1. Social Interactions

Inclusion of social interaction significantly moderates gamification effects, with interventions incorporating competitive, collaborative, or hybrid social dynamics generally outperforming individual-based gamified designs, particularly for behavioral and motivational outcomes. Competitive gamification, characterized by performance comparison and ranking mechanisms, has been shown to increase engagement and task persistence but may yield inconsistent learning benefits due to heightened pressure and individual differences in competitiveness (Tabassum et al., 2024) . In contrast, collaborative gamification supports learning through shared goal pursuit and peer interaction, fostering engagement and social presence, though its effectiveness depends on structured coordination and accountability. Importantly, evidence suggests that **hybrid approaches combining competition and collaboration** produce more stable and positive effects by balancing motivational stimulation with social support, thereby reducing disengagement risks (Dikcius et al., 2021) . Accordingly, social interaction contributes substantially to variability in reported effect sizes across gamification studies and should be systematically examined as a moderating factor to better explain heterogeneity in learning efficacy outcomes

2. Period of Time (Intervention Duration)

The duration of gamification interventions is a critical factor in distinguishing short-term novelty effects from sustained learning gains. While some studies suggest that gamification's motivational benefits diminish over time, others indicate that extended exposure can enhance retention and transfer (Permana et al., 2023) . By examining intervention length, this meta-analysis seeks to clarify whether gamification produces durable improvements in learning efficacy or primarily short-lived engagement.

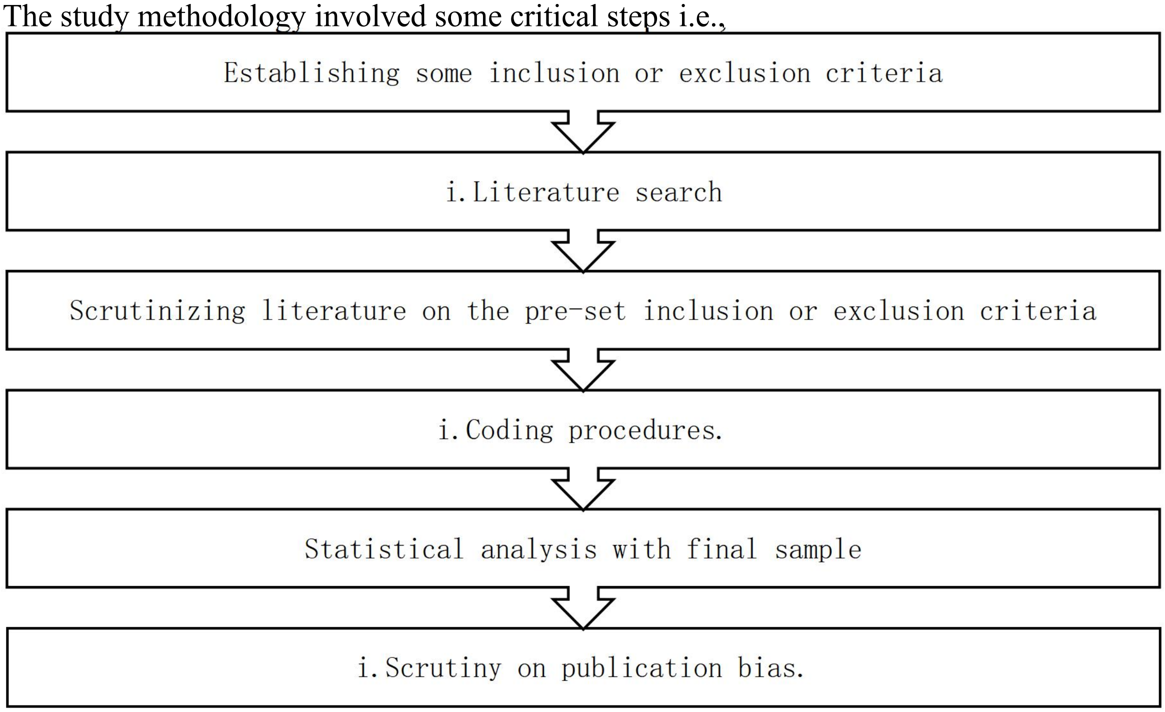
3. Methodological Rigor

Variability in study design and methodological quality can significantly influence reported outcomes. Factors such as randomization, use of pretests, and control for prior knowledge and motivation are essential indicators of rigor. Studies relying solely on posttest measures or lacking robust controls may overestimate gamification's effects (Candel et al., 2024) . Including methodological rigor as a moderator ensures that conclusions are based on reliable evidence and highlights the importance of design quality in interpreting gamification's impact.

By systematically analyzing conceptual differences, contextual and situational variables, research environment, intervention duration, and methodological rigor, this meta-analysis aims to identify the conditions under which gamification most effectively enhances learning efficacy (G. Kaya & Sagnak, 2022). This approach ensures that findings are not only statistically robust but also pedagogically meaningful, offering guidance for both theory development and instructional practice.

Method

The study methodology involved some critical steps i.e.,



1. Study Instruments

As regards a meta-analysis, published research is the only instrument of the study. However, in this regard, quality of the published work was considered carefully to overview previous research and to overcome challenges in previous work as well. Therefore, some inclusion or exclusion criteria were set for the research papers which should be included in the study i.e.,

The research methodology must include either validation process through psychometric properties, or the research should include standardized instruments.

This problem was addressed in the current investigation by using the type of instrument as a moderating variable. The term "gamification," which is defined as "the use of game design elements in non-game contexts," has drawn more attention and interest in academia and practice in recent years. One of the most popular areas of gamification research is education. Gamification is a particularly interesting approach for educational settings because of its purported motivational potential. But as gamification has grown in popularity, so too have critics who characterize it as "the latest buzzword and the next fad" (Boulet 2012, p. 1) or "Pavlovication" (Klabbers 2018, p. 232). However, what are the elements that make gamification successful, and how useful is it for learning?
This problem was addressed in the current investigation by using the type of instrument as a moderating factor.The term "gamification," which is defined as "the use of game design elements in non-game contexts," has drawn more attention and interest in academia and practice in recent years. One of the most popular areas of gamification research is education. Gamification is a particularly interesting approach for educational settings because of its purported motivational potential. But as gamification has grown in popularity, so too have critics who characterize it as "the latest buzzword and the next fad" (Boulet 2012, p. 1) or "Pavlovication" (Klabbers 2018, p. 232). However, what are the elements that make gamification successful, and how useful is it for learning?various study designs, research venues, and primary study methodological rigor. We evaluated publication bias and calculated subsplit analyses, which contained papers with high methodological rigor, to further examine the stability and robustness of results.

2. Literature Search

Literature search was made in October 2025. Till then latest known quality academic databases were searched such as SpringerLink, Elsvier, Taylor and Francis, JSTOR, PubMed, ACM Digital Library, ERIC, IEEE Xplore. Dimensions and Publish or Perish (version 8.19) was used to fetch the quick search results. These software tools allow all accessible search results to be automatically downloaded.

The data was searched using Boolean operators “gamification”, and “learning”. The downloaded data was exported to the Mendeley where the data was screened using various steps. At first step, the duplicates were removed. At this step 1500 papers were collected in the Mendeley. As each research Afterwards, the papers were scrutinized using the set criteria. As for example, different games are designed differently and cost different time and understanding of the students, and the literature indicate inconsistent results on learning efficacy, with very hard games. Therefore, such studies which included such game design elements, which claim to be exhaustive, were excluded from the analysis. Moreover, the game designs, which included such game design elements which may put analysis towards some specific direction, or elements, were also excluded from the analysis.

They study included only researches from recent years, so that latest trends among this field might be made clear. Further inclusion or exclusion criteria is given under following sections.

3. Inclusion and Exclusion Criteria

The following criteria were included in the research for this meta-analysis:What constitutes gamification? Gamification is defined as "the use of game design elements in non-game contexts" by Deterding et al. (2011, p. 9). This definition context was established as a major inclusion criterion because the study uses gamification as a major independent variable. In other words, studies that describe interventions that use the term gamification were excluded if the definition of gamification used in this meta-analysis did not apply to the intervention.

Learning Efficacy: Academic performance, motivation, and information retention are the three main categories of learning efficacy. A minimum of one learning efficacy component had to be present in the qualifying trials.

Motivation includes sentiments of confidence and self-efficacy, as well as inclinations, preferences, attitudes, involvement, and both internal and extrinsic motivation.

Knowledge Conceptual or application-oriented knowledge is referred to as retention of knowledge. Conversely, conceptual knowledge refers to the understanding of concepts, facts, and principles. On the other hand, application-oriented knowledge includes situational, strategic, and procedural knowledge (de Jong and Ferguson-Hessler 1996).

Academic performance, which is defined as the degree to which a student succeeds in their educational endeavors and is typically assessed through grades, test scores, coursework, and overall achievement across several courses, was the other component of learning efficacy. It shows how well a student has learned new information, honed their abilities, and used them in both theoretical and real-world situations. Academic success includes engagement in class, consistency in finishing assignments, and the capacity to show growth over time in addition to numerical outcomes. It acts as a crucial signal for educators, institutions, and society, as it not only highlights a student's mastery of academic content but also provides insight into their motivation, discipline, and readiness for future opportunities such as higher education or career advancement.

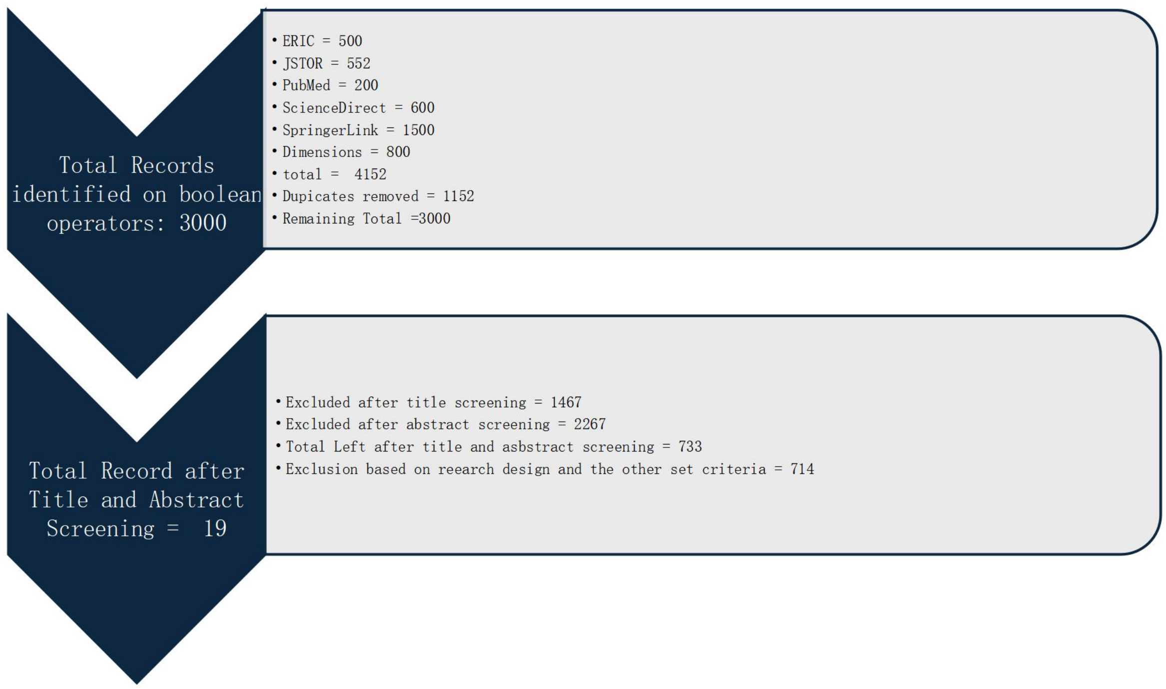
Language Eligible studies were required to be published in English.

Research Design Only primary studies applying quantitative statistical methods to examine samples of human participants were eligible. Furthermore, descriptive studies that did not compare different groups were excluded because the data obtained from such studies does not allow effect sizes to be calculated.

As the goal of this meta-analysis was to investigate the addition of game design elements to the study systems that did not already contain them, studies comparing gamification with fully fledged games were ineligible.

Statistical Data Studies were required to report sufficient statistical data to allow for the application of meta-analytic techniques.

Figure 1: Search Results and Exclusion in Each Step on Set Criteria.



4. *Coding Procedure*

Step by step coding procedure is stated as below:

1. Screening publication titles and eliminating the ineligible studies from the bud (e.g., publications which were not published English language).
2. Screening abstracts for eligibility on aforementioned criteria (An overview of the number of search results excluded per step is found in Figure. 1)
3. Coding the eligible publications.
4. Moderator coding for eligible studies was then performed by two independent coders on a random selection of 10 studies (approximately 25%), with interrater reliability ranging from κ = .76 to perfect agreement.

Lastly, information about moderator variables was taken out. The moderators that were previously mentioned that may have an impact on how well gamification affects learning outcomes were coded as follows.

Social interaction studies were classified as collaborative when students worked together with nonplayer characters or each other during gamified interventions, and as competitive when students competed against one other. A study was given the code competitive-collaborative if it had both collaborative and competitive components. Studies where students completed a task on their own were classified as none. Interrater reliability for this moderator was κ =.83.

Duration The duration of the intervention was operationalized as **the time period** in which it took place. One day or less, one week or less (but longer than one day), one month or less (but longer than one week), half a year or less (but longer than one month), or more than half a year were the categories to which the research was assigned. Coders reached complete consensus for this moderator.

Methodological Rigor: Studies that randomly assigned participants to experimental and control groups were classified as experimental, whereas those that used nonrandom assignment were classified as quasi-experimental. Coders came to a wonderful agreement for this moderator.

Additionally, studies were categorized according to whether they performed a pretest (pre- and posttest) or simply employed posttest measures (posttest only). Coders reached complete consensus for this moderator.

Studies that measured the variables of interest using pre-existing, standardized instruments were labeled as standardized, while studies that used modified versions of standardized measures were coded as adapted. Studies were given the value "self-developed" if the authors created a new metric. Coders came to a wonderful agreement for this moderator.

5. Final Sample

Application of the exclusion criteria detailed above resulted in a final sample of 19 publications, published between 2015 and 2025.

The final sample of studies reporting cognitive learning outcomes comprised 19 primary studies and examined a total of 1760 participants. Considering motivational learning efficacy, the final sample of studies consisted of 16 primary studies which examined a total of 2246 participants. Finally, the final sample of studies examining academic performance consisted of nine primary studies consisting of a total of 951 participants.

Results

1. Correlation matrix

The resulting correlation matrix, including the number of studies for each cell, is shown in Table 1.

Table 1: Correlation between Learning Efficacy and Considered Moderators

		1.	2.	3.	4.	5.	6.	7	8	9	10	11
Learning	1. Learning Efficacy											
Social Interactions	2. Collaborative	.07(2)	-.12(1)									
	3. Competitive	-.13(9)	.45**(3)	-.10(5)								
	4. Competitive+Collaborative	.07(5)	-.09(2)	n/a(0)	n/a(0)							
Time Spent	5. Time Spent (1 day)	-.53**(5)	0(9)	-.10(1)	0(4)	.17(1)						
	6. Time Spent (1 month)	.28(3)	.09(2)	n/a(0)	n/a(0)	.01(1)	n/a(0)					
	7. Time Spent (6 months)	.31(8)	.22(7)	.03(1)	-.05(2)	n/a(0)	n/a(0)	n/a(0)				
Method	8. Experimental studies	.31(6)	.04(4)	.10(1)	.06(2)	.04(1)	.10(1)	n/a(0)	.20(4)			
	9. Standardize	-	.06(	n/a	.16	-	n/a	.33	.28	.01		

	d instruments	.13(2)	3)	(0)	(2)	.08(2)	(0)	(6)	(3)	(1)		
	10. Adaptive instruments	-.28(2)	.33(6)	.10(1)	.06(2)	n/a(0)	.22(5)	n/a(0)	-.01(2)	n/a(0)	n/a(0)	
	11. Self-developed instruments	.35*(17)	-.38(11)	-.07(2)	.14(7)	.28(7)	0(13)	.17(3)	-.25(6)	.11(6)	-.36(2)	-.57**(2)

Pearson correlation between each cell with number of study sample (k) in parentheses. Moderator levels 1 week or less and more than half a year were omitted because the sample size was 0.
n/a = not applicable because k=0
*p<.05; **p<.01

The correlation matrix presented in Table 1 summarizes the relationship between learning efficacy and several moderating variables, with the number of studies contributing to each correlation noted in parentheses. Comprehensively, the review shows that the effect of gamification on the efficacy of learning is greatly contextual, time-based and methodological.Concerning social interactions, the collaborative settings showed weak and inconsistent relationships with learning efficacy ($r = .07$, $k = 2$) indicating that collaboration does not have a significant beneficial effect on learning outcomes. Competitive nature, on the contrary, showed ambivalent outcome; the negative relationship with learning efficacy ($r = -.13$, $k = 9$) and the positive strong relationship with collaborative contexts ($r = .45$, $p < .01$, $k = 3$). This trend suggests that competition on its own can impair learning effectiveness but a combination of competition and collaboration can enhance the levels of engagement. In fact, hybrid competitivecollaborative setting was positively correlated with efficacy ($r = .07$, $k = 5$), indicating the possibility of hybrid solutions in reducing the disadvantages of pure competition. Another important modulating factor was found to be the length of gamified learning interventions. One-day interventions were found to have adverse effects on learning efficacy ($r = -.53$, $p = .01$, $k = 5$), which means that gamification did not produce positive results in ultra-brief interventions. Conversely, the positive effects of medium-term intervention, which is a one-month intervention, were of modest positive relationships ($r = .28$, k 3), whereas the long-term interventions that lasted up to 6 months gave higher positive effects ($r = .31$,k 8). These results indicate that gamification is more effective when undertaken in the long term as opposed to a short-term activity.

Lastly, methodological rigor and measuring instruments were also influential in determining results. Experimental research proved the positive correlation with learning efficacy ($r = .31$, $k=6$) that evidences the importance of rigorous research designs. There was however significant variation in the nature of the instrument employed to assess efficacy. Standardized tools had a low negative correlation ($r = -.13$, $k = 2$) and adaptive tools had a negative correlation as well ($r = -.28$, $k = 2$) which may indicate the incompatibility of these tools with motivational aspects of gamification. By contrast, self-developed instruments were associated with a significant positive relationship ($r = .35$, $p < .05$, $k = 17$), but also with large negative relationships with other moderators (e.g., $r = .57$, $p = .01$, $k = 2$ with adaptive instruments). This trend puts the issue of measurement bias in question because instruments created by the researcher might overrate the perceived prominence of gamification over more objective instruments.

Combined, these findings point to the fact that the effect of gamification on the effectiveness of learning is not universal. It seems to be best in the long term and collaborative settings with the help of stringent experimental designs, whereas in the short term and when using competitive conditions or non-standardized measures, it can perform poorly.

2. Moderator Analyses

As the homogeneity estimates showed a significant and substantial amount of heterogeneity for cognitive, motivational, and behavioral learning outcomes, moderator analyses were conducted to determine whether additional factors could account for the variance observed in the samples. Not all of the following comparisons contained all possible levels of the respective moderator because levels with $k \leq 1$ were excluded from these analyses.

Table 2: Results of the Moderator Analysis for Learning Efficacy

	Moderator Variable	g	95%CI	k	N	Q	df	I^2(%)	Fail-safe N
Contextual	1. Collaborative	.40	[-0.67, 1.48]	2	195	9.20**	1	89.13	-
	2. Competitive	.50**	[.34, 0.67]	9	922	14.03	8	35.36	134*
	3. Competitive+Collaborative	.63*	[0.06, 1.20]	5	275	18.87**	4	77.34	33
Time Spent	4. Time Spent (1 day)	.60**	[0.24,0.96]	5	344	5.85	4	51.70	37*
	5. Time Spent (1 month)	.52**	[0.29, 0.75]	3	297	1.01	2	0.00	13
	6. Time Spent (6 months)	.50**	[0.14, 0.86]	8	763	36.68**	7	85.01	80*
Methodological Rigor	7. Experimental studies	.51**	[0.16, 0.87]	6	423	9.76	5	43.69	31
	8. Standardized instruments	1.84**	[1.23, 2.44]	1	58	-	-	-	-
	9. Adaptive instruments	.14	[-0.20, 0.48]	1	88	-	-	-	-
	10. Self-developed instruments	.50**	[0.32, 0.69]	18	1630	44.38**	17	65.90	430*

g = posttest effect sizes adjusted for pretest scores between groups,
k = No. of study samples
N = total sample size
Q = homogeneity statistics
df = degree of freedom for Q test
I^2 = between study variance component

a= Not estimated due to a small number of studies
* in the fail-safe column indicates the failsafe number in robust (>5n+10)
*p<.05; **p<.01

This table presents the results of a meta-analysis examining how different moderator variables influence effect sizes across studies. In terms of contextual variables, collaborative settings produced a small, non-significant effect size (g = .40) with wide confidence intervals and high heterogeneity, suggesting inconsistent findings. Competitive contexts, however, yielded a moderate and statistically significant effect (g = .50, p < .01) across nine studies, with a robust fail-safe number of 134, indicating that the effect is unlikely to be overturned by unpublished null results. When competitive and collaborative elements were combined, the effect size increased to g = .63 (p < .05), though heterogeneity remained high, reflecting variability across studies.

The time spent in interventions also influenced outcomes. Short-term interventions lasting one day showed a moderate and significant effect (g = .60, p < .01), though with moderate heterogeneity. One-month interventions produced a similar effect (g = .52, p < .01) but with no heterogeneity, suggesting consistent findings across studies, albeit with a smaller fail-safe number. Longer interventions of six months maintained a moderate effect (g = .50, p < .01), but heterogeneity was very high, indicating substantial differences in study results despite a robust fail-safe number of 80.

Methodological rigor further shaped the findings. Experimental studies demonstrated a moderate and significant effect (g = .51, p < .01) with moderate heterogeneity. The use of standardized instruments produced the largest effect size (g = 1.84, p < .01), though this was based on a single study and cannot be generalized. Adaptive instruments showed a negligible, non-significant effect (g = .14), while self-developed instruments yielded a consistent moderate effect (g = .50, p < .01) across 18 studies, though with high heterogeneity. Importantly, the fail-safe number for self-developed instruments was very large (430), suggesting strong robustness despite variability.

Overall, the analysis indicates that competitive contexts, longer interventions, and methodological rigor, particularly the use of standardized or self-developed instruments—are associated with moderate and reliable effects. Collaborative-only contexts and adaptive instruments, by contrast, show weak or inconsistent results. This highlights the importance of study design and context in determining the strength and reliability of observed effects.

Discussion

This meta-analysis demonstrates that the effectiveness of gamification on learning efficacy is **conditional rather than universal**, varying systematically by **contextual configuration**, **exposure duration**, and **methodological rigor**. Across moderator analyses, competitive and hybrid (competitive + collaborative) contexts produced **moderate**, **statistically significant** effects (g = .50 and g = .63, respectively), whereas collaboration-only settings produced small, non-significant effects with wide uncertainty. Temporal analyses showed **consistently positive effects** for one-day, one-month, and six-month implementations (g ≈ .50–.60), albeit with heterogeneity that scaled with duration. Methodologically, experimental studies yielded moderate positive effects; standardized instruments produced the largest single-study estimate; and self-developed instruments yielded robust, moderate pooled effects while also introducing substantial between-study variance.

The **hybrid competitive–collaborative** configuration outperformed other contextual arrangements (g = .63), suggesting that **balanced social contingencies**—where goal-directed competition coexists with peer support—may amplify engagement and persistence while mitigating the well-documented risks of zero-sum competition (e.g., disengagement among lower performers). In contrast, **collaboration-only** configurations (g = .40, non-significant) exhibited high heterogeneity and wide CIs, indicating **context sensitivity**: the benefits of collaboration appear to depend on complementary motivational structures (e.g., feedback, progress visibility, autonomy) rather than collaboration per se. The correlation matrix reinforces this pattern: competitive-only contexts showed a negative association with efficacy (r = –.13, k = 9), yet were **positively related** to collaborative contexts (r = .45, p < .01), implying that **co-occurrence and design integration** matter more than isolated features.The moderator analyses indicate **benefits across exposure windows**—from one-day (g = .60) to one-month (g = .52) to six-month (g = .50)—with **null heterogeneity at one month**, suggesting a **sweet spot of design stability and implementation fidelity**. The correlation pattern for one-day exposure (r = –.53, p < .01) appears superficially contradictory to the positive g estimate. This discrepancy can be reconciled by recognizing that: (i) **correlations reflect co-occurrence patterns** rather than pooled between-group effect magnitudes; (ii) the correlation cell sizes are small (k = 5), and (iii) the meta-analytic g accounts for pretest-adjusted between-group contrasts. In practice, **brief deployments can succeed** when tightly aligned with task demands and feedback cycles (as suggested by the moderate g), but are **more fragile** to contextual or design misfits (as implied by the negative correlation), particularly when "gamification" is appended to a learning task without adequate integration.Experimental studies exhibited **moderate and significant** effects (g = .51), supporting the **internal validity** of the observed benefits. The **instrumentation** pattern is more complex. The **single** standardized-instrument estimate (g = 1.84) is striking but **uninformative about generalizability**; it should be treated as hypothesis-generating rather than confirmatory. **Self-developed instruments** produced a **moderate pooled effect** (g = .50) with a **very large fail-safe number** (430), suggesting robustness to unpublished nulls. Yet the **high heterogeneity** ($I^2 \approx 66\%$) and the strong negative correlations with other instrument types (e.g., r = –.57** with adaptive instruments, k = 2) raise plausible concerns about **construct alignment and criterion inflation**—i.e., researcher-developed measures may map more directly onto the targeted gamified constructs (e.g., engagement, points/progression), thereby **inflating apparent efficacy** relative to broader, standardized outcomes. This underscores the importance of **convergent validation** and **triangulation** across **standardized**, **adaptive**, and **self-developed** measures.Heterogeneity levels were **non-trivial** (e.g., I^2 = 77–89% in some moderator strata), indicating **meaningful differences in implementation**, learner populations, disciplinary domains, platform features, and fidelity. Notably, **fail-safe numbers** for several strata (Competitive, One Day, Six Months, Self-developed instruments) suggest **robustness** against the file-drawer threat, though fail-safe metrics do not address **small-study effects**, **p-curve distortions**, or **selective reporting**. Given the observed pattern—larger effects where instrumentation is less standardized and where contextual complexity increases—reviewers will reasonably expect sensitivity checks (e.g., leave-one-out analyses; excluding k ≤ 2 strata; modeling outcome domain differences) in an extended appendix.

Limitations

There are a number of limitations that must be considered when discussing the results of this meta-analysis.

- First, small samples of cells (k 2) between diverse moderator levels (standardized and adaptive instruments) limit the accuracy of estimates and strength of inference, and some of the seemingly large effects are based on individual studies.
- Second, high heterogeneity in various strata (I 2 > 70) suggests that there are large between-study variations that meta-regression or sub-group analyses can only explain partly.
- Third, non-equivalence in measurement between standardized, adaptive and self-developed measures poses some risks of construct misalignment and may result in effects inflation especially where the measures are well-intervention-specific.
- Fourth, there was large observable fail-safe numbers but the chances of reporting and selection bias are not completely ruled out.
- Lastly, contextual confounds, such as content domain, age of learners and delivery modality, were not consistently coded and could be weak at multi-level modeling and hence the generalizability of findings.

Conclusion

This meta-analysis demonstrates that gamification can yield moderate and reliable improvements in learning efficacy, though such outcomes are contingent upon contextual design, implementation duration, and methodological rigor. The strongest evidence supports competitive and hybrid competitive–collaborative formats, particularly when sustained over periods ranging from one to six months, with experimental designs and well-aligned measurement strategies enhancing the detectability of effects. By contrast, collaboration-only contexts and assessments relying on adaptive instruments exhibit inconsistent or weaker impacts. Given the observed heterogeneity and measurement sensitivities, future research should employ pre-registered, multi-instrument designs that integrate validated standardized measures alongside theory-driven custom scales; disentangle the active ingredients of hybrid social structures—such as team-based competition and cooperative quests—through component-level trials; utilize multi-level meta-analytic models and meta-regressions to parse moderators related to domain, learner age, and platform features; and report fidelity, adherence, and implementation characteristics to reduce unexplained variance and strengthen reproducibility. In sum, the effectiveness of gamification in education is best understood as design- and context-dependent, with the most credible and impactful improvements emerging from the strategic alignment of social dynamics, sustained exposure, and rigorous measurement practices.

REFERENCES

Candel, E. C., de-la-Peña, C., & Yuste, B. C. (2024). Pre-service teachers' perception of active learning methodologies in history: Flipped classroom and gamification in an e-learning environment. *Education and Information* …. https://doi.org/10.1007/s10639-023-11924-0

Dikcius, V., Urbonavicius, S., & ... (2021). Learning marketing online: The role of social interactions and gamification rewards. *Journal of* …. https://doi.org/10.1177/0273475320968252

Eliyas, S., & Ranjana, P. (2022a). Improving Continuous Intention in E-Learning through the use of a Combined Gamification Model. *2022 IEEE International Conference on …*. https://ieeexplore.ieee.org/abstract/document/10080326/

Eliyas, S., & Ranjana, P. (2022b). Improving Continuous Intention in E-Learning through the use of a Combined Gamification Model. *2022 IEEE International Conference on …*. https://ieeexplore.ieee.org/abstract/document/10080326/

Gui, Y., Cai, Z., Yang, Y., Kong, L., Fan, X., & Tai, R. H. (2023). Effectiveness of digital educational game and game design in STEM learning: a meta-analytic review. *International Journal of STEM Education*, *10*(1). https://doi.org/10.1186/s40594-023-00424-9

Kaya, G., & Sagnak, H. C. (2022). Gamification in English as second language learning in secondary education aged between 11-18: A systematic review between 2013-2020. *International Journal of Game-Based Learning …*. https://www.igi-global.com/article/gamification-in-english-as-second-language-learning-in-secondary-education-aged-between-11-18/294010

Kaya, O. S., & Ercag, E. (2023). The impact of applying challenge-based gamification program on students' learning outcomes: Academic achievement, motivation and flow. *Education and Information Technologies*. https://doi.org/10.1007/s10639-023-11585-z

Legaki, N. Z., Xi, N., Hamari, J., Karpouzis, K., & ... (2020a). The effect of challenge-based gamification on learning: An experiment in the context of statistics education. *International Journal of …*. https://www.sciencedirect.com/science/article/pii/S1071581920300987

Legaki, N. Z., Xi, N., Hamari, J., Karpouzis, K., & ... (2020b). The effect of challenge-based gamification on learning: An experiment in the context of statistics education. *International Journal of …*. https://www.sciencedirect.com/science/article/pii/S1071581920300987

Luarn, P., Chen, C. C., & Chiu, Y. P. (2023). Enhancing intrinsic learning motivation through gamification: a self-determination theory perspective. *... Journal of Information and Learning …*. https://doi.org/10.1108/IJILT-07-2022-0145

Molero, D., Schez-Sobrino, S., Vallejo, D., & ... (2021). A novel approach to learning music and piano based on mixed reality and gamification. *Multimedia Tools and …*. https://doi.org/10.1007/s11042-020-09678-9

Nguyen-Viet, B., & Nguyen-Viet, B. (2023). Enhancing satisfaction among Vietnamese students through gamification: The mediating role of engagement and learning effectiveness. *Cogent Education*. https://doi.org/10.1080/2331186X.2023.2265276

Permana, P., Permatawati, I., & ... (2023). Foreign language learning gamification using quizizz: A systematic review based on students' perception. *... Jurnal Pendidikan Bahasa …*. https://www.researchgate.net/profile/Pepen-Permana/publication/375516510_Foreign_Language_Learning_Gamification_Using_Quizizz_A_Systematic_Review_Based_on_Students'_Perception/links/654cf9b2b1398a779d746617/Foreign-Language-Learning-Gamification-Using-Quizizz-A-Systematic-Review-Based-on-Students-Perception.pdf

Pitoyo, M. D., & Asib, A. (2020). Gamification-Based Assessment: The Washback Effect of Quizizz on Students' Learning in Higher Education. *International Journal of Language Education*. https://eric.ed.gov/?id=EJ1249875

Puspitasari, I., & Arifin, S. (2023). Implementation of gamification on learning motivation: A meta-analysis study. *International Journal of Progressive Sciences and …*.

Rosa-Castillo, A., García-Pañella, O., & ... (2022). Gamification on Instagram: Nursing students' degree of satisfaction with and perception of learning in an educational game. *Nurse Education …*. https://www.sciencedirect.com/science/article/pii/S0260691722002696

Sailer, M., & Homner, L. (2020a). The Gamification of Learning: a Meta-analysis. *Educational Psychology Review*, *32*(1), 77–112. https://doi.org/10.1007/s10648-019-09498-w

Sailer, M., & Homner, L. (2020b). The gamification of learning: A meta-analysis. *Educational Psychology Review*. https://doi.org/10.1007/S10648-019-09498-W

Slamet, T. I., & Meng, C. (2025a). Gamification in collaborative learning: Synthesizing evidence through meta-analysis. *Journal of Computers in Education*. https://doi.org/10.1007/s40692-024-00349-4

Slamet, T. I., & Meng, C. (2025b). Gamification in collaborative learning: Synthesizing evidence through meta-analysis. *Journal of Computers in Education*. https://doi.org/10.1007/s40692-024-00349-4

Tabassum, R., Fatima Kashif, M., Iqbal, S., Saad, J., Tahira, R., & Shahid, R. (2024). *A Study On Social Media Usage And Social Alienation Among University Students*. *21*(S11), 595–609. www.migrationletters.com

Wirani, Y., Nabarian, T., & Romadhon, M. S. (2022a). Evaluation of continued use on Kahoot! as a gamification-based learning platform from the perspective of Indonesia students. *Procedia Computer Science*. https://www.sciencedirect.com/science/article/pii/S1877050921023966

Wirani, Y., Nabarian, T., & Romadhon, M. S. (2022b). Evaluation of continued use on Kahoot! as a gamification-based learning platform from the perspective of Indonesia students. *Procedia Computer Science*. https://www.sciencedirect.com/science/article/pii/S1877050921023966

Yu, Q., Yu, K., & Li, B. (2024). Can gamification enhance online learning? Evidence from a meta-analysis. *Education and Information Technologies*. https://doi.org/10.1007/s10639-023-11977-1