Effects of reward strategies in gamified learning on academic performance: a systematic review

Shurui Bai, Yingxue Liu, Khe Foon Hew, Michael Sailer, Ya Xiao

Angaben zur Veröffentlichung / Publication details:

Bai, Shurui, Yingxue Liu, Khe Foon Hew, Michael Sailer, and Ya Xiao. 2026. "Effects of reward strategies in gamified learning on academic performance: a systematic review." Educational Research Review 50: 100766. https://doi.org/10.1016/j.edurev.2026.100766.

Educational Research Review 50 (2026) 100766

Contents lists available at ScienceDirect

Educational Research Review

journal homepage: www.elsevier.com/locate/edurev

Review

Effects of reward strategies in gamified learning on academic performance: A systematic review

Shurui Bai [a,*], Yingxue Liu [b], Khe Foon Hew [c], Michael Sailer [d], Ya Xiao [e]

[a] Department of Mathematics and Information Technology, The Education University of Hong Kong, China
[b] HT Nanjing Impact Academy, Jiangsu, China
[c] Faculty of Education, The University of Hong Kong, China
[d] Learning Analytics and Educational Data Mining, University of Augsburg, Augsburg, Germany
[e] School of Education Science, Nanjing Normal University, China

ARTICLE INFO

Keywords:
Gamification
Tangible rewards
Reward contingency
Cognitive evaluation theory
Academic performance

ABSTRACT

Tangible rewards, such as class participation scores, are often incorporated into gamified learning to motivate students. This systematic review synthesised evidence from 16 studies with 18 independent interventions to examine how tangible rewards have been designed and implemented in gamified learning and how they relate to academic performance. Across the studies, tangible rewards varied widely in form, contingency and stakes, ranging from small tokens and badges to extra course grades. The result found that the "exceeding a norm" reward contingency with "low-stakes rewards" associates with more positive academic performance in gamified classes. Based on qualitative insights, good reward administration practices include giving students autonomy in choosing rewards and recognizing learning progress with playful elements (e.g., Lego bricks). However, due to the limited number of interventions, more empirical evidence is needed to clarify whether and under what conditions tangible rewards can enhance academic performance in gamified learning.

In gamified learning, intangible rewards such as virtual points and badges are commonly used to engage students (Bai et al., 2021, pp. 899–904). Unlike tangible rewards, intangible rewards do not offer material goods to recipients (Meder et al., 2018). Although intangible rewards can create enjoyable experiences in gamified courses, their value may diminish over time for students who prefer tangible resources, such as extra course grades (Huang & Hew, 2018) and book discount vouchers (Zichermann & Cunningham, 2011). Although gamification has been widely implemented using intangible virtual game elements, the added complexity of tangible rewards has not been thoroughly investigated.

In gamified learning, tangible rewards are often combined with game elements in two distinct approaches. The first approach involves simultaneously providing game elements and tangible rewards through different earning rules. For example, Çakıroğlu and Güler, 2021 awarded badges and ranked students on a leaderboard based on the activities they completed while distributing small real-world gifts to those who answered questions correctly. The second approach allows students to accumulate game elements and then exchange them for tangible rewards. For instance, Metwally et al. (2021) enabled students to exchange points and badges earned by completing activities for additional course grades.

This review synthesised evidence on how tangible rewards in gamified learning relate to students' academic performance. We

* Corresponding author.
E-mail address: tstbai@eduhk.hk (S. Bai).

https://doi.org/10.1016/j.edurev.2026.100766
Received 11 May 2024; Received in revised form 11 January 2026; Accepted 14 January 2026
Available online 16 January 2026
1747-938X/

examined two comparisons: (1) gamified learning with tangible rewards versus gamified learning without them, and (2) gamified learning with tangible rewards versus non-gamified, no-reward conditions. Academic performance was defined as students' ability to complete academic tasks, measured by objective indicators such as final grades and grade point averages.

We categorised tangible rewards along two dimensions: contingency and stakes. Contingency refers to the conditions required to earn a reward, such as completing a task, surpassing a score, or exceeding a norm. Stakes refer to the real-world value or significance of the reward, such as whether the reward is a small token or has a substantial impact like extra course grades. This classification allowed us to examine patterns in how contingency and stakes may influence academic performance in gamified learning environments Understanding these patterns can inform the design of reward strategies that better align with motivational theory and educational goals. While the evidence is limited, this review provides a foundation for more evidence-informed and contextually appropriate gamification practices.

In the following sections, we discuss the theoretical framework used in this study to analyse the type of reward contingency employed in previous empirical research. Next, we describe our research aims and questions, followed by the method section, which explains how we searched for relevant empirical studies, coded, and analysed the data.

1. Theoretical framework: Cognitive Evaluation Theory

Cognitive evaluation theory (CET) explores the impact of reward contingency on human perceptions of competence and autonomy, which then influence intrinsic motivation (Deci, 1975). This theory is crucial to our study as intrinsic motivation is a key factor determining students' academic performance. Numerous studies have highlighted the positive impact of intrinsic motivation on academic outcomes (e.g., Liu et al., 2020; Ryan & Deci, 2020; Taylor et al., 2014). For example, a meta-analysis by Taylor et al. (2014) revealed a positive relationship between intrinsic motivation and academic performance among high school and college students. This means that students with higher intrinsic motivation usually perform better.

In educational settings, CET offers a valuable theoretical model for understanding how different reward types affect student motivation and, in turn, their academic performance. CET suggests that the effect of a tangible reward on motivation depends on whether the reward is perceived as controlling (forcing task completion) or informational (indicating personal competence performance; Deci et al., 2001). Controlling rewards are believed to decrease motivation by shifting attention from task enjoyment to the reward itself, whereas informational rewards can enhance motivation by providing positive feedback on one's competence (Deci et al., 2001). These controlling and informational aspects can be implemented through reward contingency, which concerns the conditions under which rewards are assigned.

CET generally shows that tangible rewards undermine student motivation, based largely on studies conducted in laboratory settings (Deci et al., 1999, 2001). However, other researchers have argued that tangible rewards do not necessarily harm motivation (Cameron et al., 2001; Eisenberger & Cameron, 1996). This discrepancy may arise from how tangible rewards are categorised into different reward contingencies. For example, Cameron et al. (2001) argued that CET's reward contingency framework is too broad, making it difficult to draw definitive conclusions about the effects of rewards on motivation. To address this, Cameron (2001) divided the four reward contingencies into seven categories (see Table 1 for details), providing a more nuanced understanding of the potential effects of tangible rewards.

In this study, we used CET based on the revised categories of Cameron et al. (2001) to guide our coding scheme for tangible rewards based on reward contingency. This enabled us to systematically analyse how different types of rewards (i.e., high-stakes and low-stakes rewards; see the "High- and low-stakes rewards in the classroom" section in the next section) influence students' academic performance in gamified learning environments.

Table 1
Our Reward Contingency Coding Scheme Adapted from the Seven Categories of Cameron et al. (2001, p. 12).

Reward contingency	Description
Task noncontingent	Reward is offered for agreeing to participate, for coming to study, for joining the experiment. Offer of reward is unrelated to the level of engagement in the task.
Rewards offered for doing well	Reward is offered for doing well on the task or for doing a good job. No specification is given as to what it means to do a good job or to do well. For instance, it can refer to producing quality work or be based on subjective description of the quality criteria without prior specification.
Rewards offered for doing a task	Reward is offered to engage in the experimental activity. No instructions are given about how well participants must perform or whether they must complete the task.
Rewards offered for finishing or completing a task	Reward is offered to finish an activity, to complete a task, or to get to a certain point on the task. The reward is not related to quality of performance. Completion is necessary, but quality performance is not required.
Rewards offered for solving each unit	Reward is offered for solving each unit, puzzle, problem, etc., that is solved. Missions can be achieved progressively, or with varied difficulty levels.
Rewards offered for surpassing a score	Reward is offered for surpassing a particular specified score (absolute standard). In some cases, the better the score, the higher the reward. Objective description of the quality criteria is provided.
Rewards offered for exceeding a norm	Reward is offered to meet or exceed the performance of others on the task (relative standard). For instance, being the first, outperforming others, or receiving a limited number of competitive rewards.

1.1. High- and low-stakes rewards in the classroom

Although not the focus of gamification research, the stakes associated with rewards could be an important factor to consider in advance. The literature has argued that high stakes refer to the significant consequences associated with the outcome of a particular event (Amrein & Berliner, 2002), such as a college admission test, as a high-stakes test (Stenlund et al., 2017). It is well known that schools and teachers evaluate almost everything a student does using grades (Ryan & Deci, 2020).

Although high-stakes rewards, such as grades, can significantly control student behaviour, their motivational quality warrants closer examination. These rewards are typically tied to consequential outcomes, such as passing a course or gaining admission to a university are often used to drive performance. However, their effectiveness depends on the type of motivation they elicit. According to self-determination theory, motivation can be broadly categorised as intrinsic (driven by interest or enjoyment) or extrinsic (driven by external incentives). High-stakes rewards tend to foster extrinsic motivation, where students engage in learning to obtain rewards or avoid punishment (Ryan & Weinstein, 2009). While extrinsic motivation may prompt short-term compliance, it is considered lower in quality because it may undermine students' sense of autonomy and intrinsic interest in learning (Deci et al., 1999). Despite this, some educational strategies continue to use external incentives to improve academic achievement (Levitt et al., 2016). These strategies operate on the assumption that external incentives can enhance performance, but they may not support deeper engagement or long-term learning outcomes.

In gamified learning, rewards are integral components to enhance students' motivation and academic performance. We define high-stakes rewards in gamified learning as rewards that directly affect students' course grades (e.g., Kwon & Özpolat, 2021; Metwally et al., 2021). In contrast, low-stakes rewards do not directly affect course grades, such as small gifts and trophies (Sánchez et al., 2021).

1.2. Reward usage in gamified learning

To date, several meta-analyses have examined the effect of gamification on students' academic performance. Gamification has been found to have significant, positive moderate effect sizes on students' cognitive learning outcomes, namely Hedges' $g = 0.49$ in Sailer and Homner's (2020) meta-analysis, Hedges' $g = 0.56$ in Yıldırım and Şen's (2021) meta-analysis, and Hedges' $g = 0.50$ in Bai et al.'s (2020) meta-analysis. Several systematic reviews have also summarised the underlying theories, gamified learning platforms, and game elements used in gamified classrooms. However, studies have not considered additional tangible rewards and reward contingencies (e.g., Dicheva et al., 2015; Kim & Castelli, 2021; Subhash & Cudney, 2018; Zainuddin et al., 2020), instead focusing primarily on game elements such as points and badges. This leaves open the question of how tangible and performance-contingent rewards might influence learning within gamified environments.

To begin addressing this gap, it is important to examine how specific tangible reward or reward contingencies have been implemented and evaluated in empirical studies. For example, one study evaluated the effects of reward contingency on students' intrinsic motivation to learn (Li et al., 2024). They discovered that the largest positive effect size (Hedges'$g = 0.63$) occurred when rewards were given for a combination of factors: "solving each unit," "surpassing a score," and "surpassing a norm." Although Li et al. (2024) compared gamified and non-gamified learning groups, they did not directly consider the conditions of gamified learning with and without tangible rewards. Furthermore, their study did not explore the association between reward contingency and students' academic performance.

1.3. Research aims and questions

The original scope of the primary studies was to describe how tangible reward schemes were implemented in gamified classes and to compare academic performance outcomes with and without such rewards. By categorising tangible rewards based on reward contingency and stakes, and examining their relationship with academic performance, we aim to offer educators a clearer understanding of how such rewards function within gamified learning contexts. However, after conducting a primary screening and eligibility check of potential studies using the preferred reporting items for systematic reviews and meta-analyses (PRISMA) statement (Moher et al., 2009; see the Study exclusion criteria section), only two empirical studies (i.e., Bai et al., 2021, pp. 899–904; Xiao & Hew, 2023) directly compared gamified learning with and without tangible rewards. Hence, we also included studies examining the effects of *gamified learning with tangible rewards* versus *non-gamified, no-reward groups*.

In addition, despite the widespread use of high-stakes rewards such as grades in educational settings, there is insufficient evidence that using course grades as rewards would enhance students' academic performance; on the contrary, there are indications that they may have negative effects (Ryan & Deci, 2020). It is therefore necessary to explore this issue in the context of gamified learning. We categorised tangible rewards into seven reward contingencies (see Table 1) and two real-world stakes (see the "High- and low-stakes rewards in the classroom" section). Next, we summarised the quantitative results of students' academic performance with calculated effect sizes. Finally, we suggested some tangible reward administration schemes that may help enhance students' academic performance. The following four research questions guided this study.

RQ1: What patterns in reward contingencies emerge across studies using tangible rewards in gamified learning, and how are these patterns associated with academic performance?
RQ2: What patterns in real-world stakes emerge across studies using tangible rewards in gamified learning, and how are these patterns associated with academic performance?

RQ3: What are the performance differences between gamified learning with tangible rewards to non-gamified, no-reward conditions, and how do they vary by reward contingency?
RQ4: What are the performance differences between gamified learning with tangible rewards to non-gamified, no-reward conditions, and how do they vary by real-world stakes of the rewards?

2. Methods

To ensure a focused and conceptually coherent review, we designed our search strategy to identify studies that explicitly conceptualise their interventions as gamified learning. We searched multiple academic databases, including ACM Digital Library, Science Direct, Web of Science, and IEEE Xplore, from their inception until May 31, 2025. To supplement this, we used Publish or Perish software (version 7) to capture additional studies indexed in Google Scholar and manually reviewed the reference lists of included studies. The search string used was *gamif* AND (educat* OR class* OR course OR learn* OR performance OR outcome OR engag* OR effect*),* which targeted studies using gamification-related terminology. As a result, studies exploring reward mechanisms, such as alternative grading strategies or incentive systems, which do not frame their interventions within the context of gamification may not have been included. This was a deliberate decision to maintain the integrity of this review's scope, which centres on tangible rewards embedded in gamified learning environments.

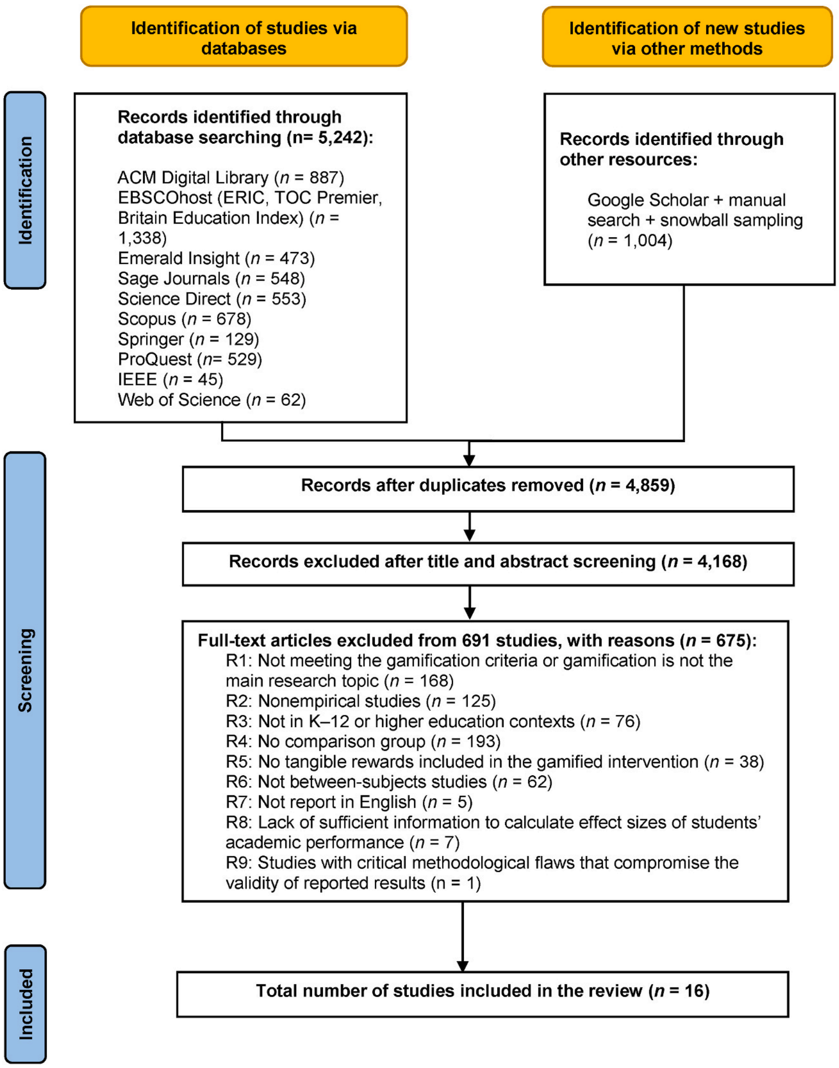

Fig. 1. PRISMA Flowchart of the Article Screening Procedure
Note. R = reason.

2.1. Study exclusion criteria

There are 691 full-text studies assessed for eligibility. To facilitate study selection, we complied with the guidance of the PRISMA statement (Moher et al., 2009, Fig. 1). We identified the following eight exclusion criteria to determine the eligibility of potential studies: (1) not meeting the gamification criteria or gamification is not the main research topic ($n = 168$); (2) nonempirical studies ($n = 125$); (3) not in K–12 or higher education contexts ($n = 76$); (4) no comparison group ($n = 193$); (5) no tangible rewards included in the gamified intervention ($n = 38$); (6) not between-subjects studies ($n = 62$); (7) not reported in English ($n = 5$); (8) lack of sufficient information to calculate effect sizes for students' academic performance ($n = 7$) and (9) Studies with critical methodological flaws that compromise the validity of reported results (n = 1). To maintain methodological rigour within this systematic review, we report the effect size for each intervention, thereby offering clear insights into the direction and magnitude of the impact of gamified learning with tangible rewards. This approach facilitates a nuanced understanding and supports the development of evidence-based strategies for future reward administration in educational contexts.

Finally, we identified 16 relevant primary studies with 18 independent interventions. Of these, two primary studies (Bai et al., 2021, pp. 899–904; Xiao & Hew, 2023) compared *gamified classes with tangible rewards* to *gamified classes without tangible rewards*, while 14 primary studies (comprising 16 independent interventions) compared *gamified classes with tangible rewards* to *non-gamified, no-reward classes* (e.g., Mese & Dursun, 2019; Metwally et al., 2021). To avoid selection bias, the two authors independently searched and screened the initial set of studies before reading the full text. Two authors separately read the 691 full-text studies to select those that meet the inclusion criteria. The two authors coded all the included studies independently.

2.2. Data extraction and coding scheme

We coded the basic information of an intervention based on previous systematic reviews on gamification (e.g., Kalogiannakis et al., 2021; Zainuddin et al., 2020), such as educational context, discipline, and gamified components (Appendix A). We coded the reward scheme by (1) reward name (if any), (2) reward administration, (3) reward contingency, (4) reward utilitarian or redeemable stakes, and (5) reward stakes (Table 2). All the authors agreed that "rewards offered for solving each unit" could be interpreted as the rewards are granted progressively for different levels of difficulty in the research context. "Rewards offered for doing well" could be interpreted as performance-contingent rewards without explicitly specifying the criteria for good performance in advance.

We extracted information on research outcomes regarding (1) measures, (2) data for effect size calculation, (3) effect size, (4) quantitative findings, and (5) qualitative, additional, or overall findings (Table 2). Only objective measures of academic performance (e.g., extra course grades, exam scores) were included. The two authors conducted data extraction in the first round and then cross-checked the data between them. A Cohen's Kappa (κ) = 0.77 (substantial agreement) was obtained. The two authors discussed and resolved any discrepancies.

For the moderator analysis aimed at analysing information on heterogeneity, two types of categorical moderators were coded: *reward-related* and *non-reward-related variables*. The reward-related variables were (1) reward contingency (see Table 1), (2) reward stakes (high-stakes, low-stakes rewards), and (3) reward combination (e.g., rewards offered for "finishing or completing a task" + high-stakes rewards). The non-reward-related variables were (1) educational context (K–12 or tertiary), (2) course subject, (3) intervention duration, (4) gamified component, and (5) theoretical perspectives. We used "not reported" and "not applicable" for coding if a study did not provide moderator information and if this information could not be categorised based on the coding scheme, respectively. For instance, in Ge (2018), the tangible rewards implemented in continuing education were neither K–12 nor tertiary and were thus coded as not applicable. Furthermore, the minimum number of studies in each level for moderator analysis was two. Because it may be difficult to obtain meaningful findings if the number of interventions is too small (Fu et al., 2011). Hence, levels with $k \leq 1$ were excluded from these analyses.

2.3. Data analysis

To answer the research questions, we summarised the quantitative data on students' academic performance and reported the effect sizes for each reward contingency and reward stakes condition. We calculated the effect sizes by comparing a gamification group with tangible rewards (GG in Table 2) and a gamification group without tangible rewards (CG' in Table 2) on students' academic performance. We also computed effect sizes by comparing a gamification group with tangible rewards (GG in Table 2) and a non-gamified, no-reward group (CG in Table 2). The effect size can inform the overall direction and magnitude of different tangible reward administration conditions across studies (Becker, 1994). Hedges' g was selected as the effect size index to correct for bias in small sample sizes. Effect sizes of ±0.20, ±0.50, and ±0.80 are considered small, moderate, and large effects, respectively (Cohen, 1988).

A random-effects model was applied to calculate the effect sizes to account for heterogeneity in research results (Lipsey & Wilson, 2001). An effect size was calculated for each eligible intervention to meet the assumption of independence of effect sizes for independent samples of students (Scammacca et al., 2014). If a study reported effect sizes for separate groups of students (e.g., workshop I and workshop II in Putz et al., 2020), an independent effect size was calculated. If a study reported multiple assessments of a single course subject of the same groups of participants (e.g., Xiao & Hew, 2023), we selected the most summative assessment, such as final course grades, rather than weekly quizzes.

Based on the random-effects model, heterogeneity can be detected and interpreted using several metrics, such as Cochran's Q (Cochran, 1954) and I^2 (Higgins & Thompson, 2002). A significant Q value suggests the presence of heterogeneity (Cochran, 1954). The I^2 statistic is the percentage of the difference explained by heterogeneity between studies (Higgins & Thompson, 2002). The rule of

Table 2
Reward information, academic performance data, and research findings.

Author/s (year)	Reward name	Reward scheme: Reward administration	Reward scheme: Reward contingency	Reward scheme: Reward utilitarian/ redeemable stakes	Reward scheme: Reward stakes	Measures	Data for effect size calculation	Effect size (Hedges' g)	95 % CI	Quantitative findings	Qualitative, additional, or overall findings
Gamified learning with tangible rewards versus gamified learning without tangible rewards (two interventions)											
(1) Bai et al. (2021)	High-quality assignment samples	Individual: The top five students in the weekly leaderboard and the five students with the most progress each week can receive high-quality assignment samples.	Rewards offered for exceeding a norm	Students earn gamification points by completing certain learning activities. The ranking of the leaderboard depends on their points.	Low-stakes reward	Self-developed test: Final course grade	GG (*M* = 77.82, *SD* = 9.89, *n* = 26), CG' (*M* = 78.57, *SD* = 13.85, *n* = 26) [note1]	−0.06	[−0.61, 0.48]	The use of expected tangible rewards had no significant impact on learning performance.	The use of tangible rewards could encourage students to participate in knowledge construction in online discussions, with more posts being created and replied to. (+)
(2) Xiao & Hew (2023)	Extra course materials	Individual: Earn points based on completion and accuracy rate.	Rewards offered for surpassing a score	At least 160 points (approximately 80 % of the maximum number of points) are needed to redeem the tangible reward (e.g., extra course materials) each week.	Low-stakes reward	Self-developed test: Final exam score	GG (*M* = 100.679, *SD* = 15.011, *n* = 28), CG' (*M* = 87.231, *SD* = 22.004, *n* = 29) [note1]	0.71	[0.17, 1.24]	The redeemable tangible rewards had a significant positive influence on students' final exam scores. (+)	Tangible rewards were more effective in increasing students' task interest, encouraging them to complete class tasks and improving their final exam performance. (+) Although the difference between GG and CG was not statistically significant in earlier weeks, it can still be concluded that rewarding students with redeemable tangible rewards significantly increased their behavioural engagement, as measured by the completion rate of preclass tasks and in-class quizzes over time. (+)
Gamified learning with tangible rewards versus non-gamified, no-reward group (16 interventions)											
(1) Kwon & Özpolat (2021)	Team tickets + individual extra credit + individual grade (a)	Individual: Submit an additional individual report.	Rewards offered for completing a task	Tickets lead to "team power-ups," such as textbooks, calculators, and handwritten notes.	High-stakes reward	Self-developed test: Final exam score	GG (*M* = 70.71, *SD* = 6.59, *n* = 33), CG (*M* = 76.41, *SD*	−0.8	[−1.32, −0.28]	Students in the gamified group achieved significantly lower scores in the exam	Gamification can hurt content knowledge and student perceptions when applied only

(*continued on next page*)

Table 2 (*continued*)

Author/s (year)	Reward name	Reward scheme				Measures	Data for effect size calculation	Effect size (Hedges' g)	95 % CI	Quantitative findings	Qualitative, additional, or overall findings
		Reward administration	Reward contingency	Reward utilitarian/ redeemable stakes	Reward stakes						
				These "power-ups" help improve final exam scores.			= 7.46, *n* = 29)			than those in the control group. (−)	to coursework assessment. (−) The long duration of the gamified experience, the simplistic reward system that rewards the team for individual work, and latent social competition, which led to lower intrinsic motivation, may be insufficient to have a positive impact on students' content knowledge. (−)
(2) Metwally et al. (2021)	Extra scores	Individual: Complete exercises (e.g., homework) under a set of conditions.	Rewards offered for finishing or completing a task	Points can be redeemed as final scores.	High-stakes reward	Self-developed test: Final score of homework exercises	GG (*M* = 464.43, *SD* = 283.99, *n* = 40), CG (*M* = 704.32, *SD* = 187.86, *n* = 44)	−1.00	[−1.45, −0.54]	Students in the gamified group achieved significantly lower final scores than those in the control group. (−)	Students are likely to be intrinsically motivated when homework is attributed to factors under their control, when they consider that they have the expertise needed to be successful learners, and when they are interested in doing homework to learn, not just to obtain high grades. (+)
(3) Çakıroğlu & Güler (2021)	Real gifts	Individual: Give all the correct answers to the award-winning question of the week.	Rewards offered for surpassing a score	No redemption of rewards	Low-stakes reward	Self-developed test: Final exam score	GG (*M* = 66.5, *SD* = 24.93, *n* = 20), CG (*M* = 45, *SD* = 28.68, *n* = 21)	0.78	[0.15, 1.42]	Students in the gamified group achieved significantly higher scores than those in the control group. (+)	Although gamifying the instructional process had a positive impact on students' statistical literacy among medium and high-score students, the influence on low-score students was not positive. Whereas medium and high achievers preferred real gifts

(*continued on next page*)

Table 2 (*continued*)

Author/s (year)	Reward name	Reward scheme				Measures	Data for effect size calculation	Effect size (Hedges' g)	95 % CI	Quantitative findings	Qualitative, additional, or overall findings
		Reward administration	Reward contingency	Reward utilitarian/ redeemable stakes	Reward stakes						
											or leaderboards as a gamification element to show their high-level cognitive abilities, virtual rewards (e.g., badges) were the main choice of low achievers because they were easy to obtain.
(4) Sánchez et al. (2021)	Real badges	Individual: Each student receives real badges for correctly answering all questions.	Rewards offered for surpassing a score	No redemption of rewards	Low-stakes reward	Standardised test: Reading comprehension test	GG (*M* = 5.71, *SD* = 0.17, *n* = 43), CG (*M* = 5.18, *SD* = 0.17, *n* = 43).	3.09	[2.46, 3.72]	Students in the gamified group achieved significantly higher reading comprehension scores than those in the control group. (+)	Gamification can provide positive results regarding interest in reading. (+)
	Trophies	Individual: The winner (with the most badges) receives a trophy.	Rewards offered for exceeding a norm	No redemption of rewards							
(5) Marín et al. (2018)	Medals and points	Individual: Each student who gives a correct answer receives a medal and some points, which are added to their rank in the course.	Rewards offered for surpassing a score	The medals are images with related content, such as an image of a Marvel superhero with their official profile, a top YouTube video, or a mini game. If students successfully complete a challenge, their prize is the ability to access specific online materials and short video lessons.	Low-stakes reward	Self-developed test: Final exam score	GG (*M* = 49.109, *SD* = 12.6238, *n* = 267); CG (*M* = 48.175, *SD* = 14.3950, *n* = 143).	0.07	[−0.13, 0.27]	There was no significant difference between the two groups.	/
(6) Putz et al. (2020) **(1)**	Rewards and badges	Team: The best team receives a prize, a group picture of the winners and a badge, which is sent after the workshop.	Rewards offered for exceeding a norm	No redemption of rewards	Low-stakes reward	Self-developed test: Knowledge retention test	Workshop I: GG (*M* = 5.83, *SD* = 1.7, *n* = 16), CG (*M* = 2.78, *SD* = 1.58, *n* = 18) [note2]	1.82	[1.01, 2.63]	In the short term, the students in the gamified group showed significantly higher knowledge retention than those in the control group. (+)	Gamification had a positive effect on students' knowledge retention, regardless of age and gender. (+) Adapting the workshops to the students' needs partially led to increased
(7) Putz et al. (2020) **(2)**	Rewards and badges	Team: The best team receives a prize, a group	Rewards offered for	No redemption of rewards	Low-stakes reward	Self-developed test: Knowledge retention test	Workshop II: GG (*M* = 6.57, *SD* =	0.82	[0.54, 1.10]		

(*continued on next page*)

Table 2 (*continued*)

Author/s (year)	Reward name	Reward scheme				Measures	Data for effect size calculation	Effect size (Hedges' g)	95 % CI	Quantitative findings	Qualitative, additional, or overall findings
		Reward administration	Reward contingency	Reward utilitarian/ redeemable stakes	Reward stakes						
		picture of the winners and a badge, which is sent after the workshop.	exceeding a norm				1.82, $n = 230$), CG ($M = 5.06$, $SD = 1.89$, $n = 67$) note2				knowledge retention. (+) Teaching knowledge and learning objectives must be adapted to the age and educational background of the participants to achieve an adequate level of difficulty (i.e., neither too high or too low).
(8) Tsay et al. (2018)	Engagement score	Individual: Students are required to complete weekly or biweekly essential learning activities before attending class, and their participation is a major part of their engagement score.	Rewards offered for finishing or completing a task	Rewards can be redeemed in the form of students' engagement score (10 % of the final total mark).	High-stakes reward	Self-developed test: Final course grade	GG ($M = 62.69$, $SD = 10.49$, $n = 160$), CG ($M = 58.67$, $SD = 10.96$, $n = 175$).	0.37	[0.16, 0.59]	The students in the gamified group achieved significantly higher course grades than those in the control group. (+)	A gamified course may not be equally effective for all students. For example, the female students participated significantly more in online learning activities than the male students, and the students with jobs engaged significantly more in online learning activities than those without jobs.
(9) Murillo-Zamorano et al. (2021)	Extra redeemable points	Team: Students accumulate points by outperforming other teams on tests (faster and more accurate). Teams receiving more points for performing better on advanced-level tests will receive extra points for the final exams.	Rewards offered for exceeding a norm	Points earned on advanced-level tests will allow teams to receive extra points for their final exam grade.	High-stakes reward	Self-developed test: Final course grade	GG ($M = 5.063$, $SD = 1.529$, $n = 67$), CG ($M = 4.664$, $SD = 1.648$, $n = 65$).	0.25	[−0.09, 0.59]	No significant difference was found between the groups.	The students in the gamified group significantly outperformed those in the control group in terms of perceived improvement in skills (ability to work in groups, listen to others' opinions, apply knowledge in practice, analyse, and synthesise, as well as self-learning ability). (+)
	Badges and award ceremony	Team: (a) Flipped classroom badges: the number of team members who	Rewards offered for exceeding a norm	The gold, silver, and bronze badges indicate 3, 2, and 1 reward points	Low-stakes reward						

(*continued on next page*)

Table 2 (*continued*)

Author/s (year)	Reward name	Reward scheme				Measures	Data for effect size calculation	Effect size (Hedges' g)	95 % CI	Quantitative findings	Qualitative, additional, or overall findings
		Reward administration	Reward contingency	Reward utilitarian/ redeemable stakes	Reward stakes						
		completed all online questionnaires. (b) Cooperative learning badges: the number of team members who actively participated in co-creation activities (c) Gamification badges: teams with the most points in Kahoot quizzes in the grand finale		accumulated session after session by teams. Once all reward points are added up, a short awards ceremony is held where third and second place are announced and the winner is proclaimed the champion team of the year.							
(10) Zvarych et al. (2019)	Lego pieces	Individual: Game moderators reward the quality of tasks completed by students at each step with a certain number of Lego pieces.	Rewards offered for doing well	No redemption of rewards	Low-stakes reward	Self-developed test: Achievement test	GG (*M* = 54.22, *SD* = 16.14, *n* = 29), CG (*M* = 57.62, *SD* = 12.97, *n* = 27).	−0.23	[−0.75, 0.30]	No significant difference was found between the groups.	Most of the students in the gamified group reported improvements in their communication skills, foreign language skills, speed of thinking, flexibility in problem solving, and teamwork skills. (+)
(11) Domínguez et al. (2013)	Participation score	Individual: Students can earn up to 5 % of their final score based on their participation in activities in the classroom and on the e-learning platform.	Rewards offered for doing a task	Up to 5 % of the final score.	High-stakes reward	Self-developed test: Final examination score	GG (*M* = 58.05, *SD* = 14.218, *n* = 106), CG (*M* = 64.12, *SD* = 13.67, *n* = 68)	−0.43	[−0.74, −0.12]	The students in the gamified group scored significantly lower in the final exam than those in the control group. (−)	Although gamification can be a powerful driver of intrinsic motivation, several problems were reported concerning extrinsic motivation. First, participants may feel manipulated. Second, little or no transfer may occur if students' behaviour is only driven by rewards. Finally, if the reward vanishes, so does the behaviour. It can be conjectured that the

(*continued on next page*)

Table 2 (*continued*)

Author/s (year)	Reward name	Reward scheme				Measures	Data for effect size calculation	Effect size (Hedges' g)	95 % CI	Quantitative findings	Qualitative, additional, or overall findings
		Reward administration	Reward contingency	Reward utilitarian/ redeemable stakes	Reward stakes						
											students who did not follow the gamified approach or who quit were partly not attracted by the reward mechanisms implemented. (−)
(12) Ge (2018) **(1)**	Final course grade added to or deducted from points earned or lost	Individual: Each participant must answer multiple-choice questions of varying levels of difficulty. If the answer is correct and the explanation is reasonable, the points associated with the question will be added to the participant's final course grade. However, if the participant gives a wrong answer or the explanation is unreasonable, the points for the question will be lost. Accordingly, the same number of points will be deducted from the participant's final course grade.	Rewards offered for solving each unit	Points can be redeemed as extra course grades.	High-stakes reward	Self-developed test: Immediate knowledge test	GG [note3] (forfeit-or-prize reward scheme) (*M* = 3.85, *SD* = 1.132, *n* = 60), CG (*M* = 3.15, *SD* = 0.899, *n* = 60)	0.68	[0.31,1.05]	The gamified patterns significantly affected e-learners' learning performance more than the non-gamified pattern. (+)	The results showed that the forfeit-or-prize pattern and the prize-only pattern had a significantly better impact on the e-learners' learning than the no-prize-no-forfeit pattern. Additionally, the forfeit-or-prize pattern resulted in better knowledge retention than the prize-only pattern. The questionnaires revealed that the forfeit-or-prize and prize-only patterns could boost students' motivation to learn. However, a high level of anxiety was experienced by the subjects assigned to the forfeit-or-prize pattern.
(13) Ge (2018) **(2)**	Final course grade added to or deducted from points earned or lost	Individual: The procedure is basically the same as above. The main difference is that students will not lose the points associated with the questions even if they give wrong answers or unreasonable explanations.	Rewards offered for solving each unit	Points can be redeemed as course grade.	High-stakes reward	Self-developed test: Immediate knowledge test	GG [note3] (prize-only reward scheme) (*M* = 3.77, *SD* = 1.031, *n* = 60), CG (*M* = 3.15, *SD* = 0.899, *n* = 60)	0.64	[0.27, 1.00]		

(*continued on next page*)

Table 2 (*continued*)

Author/s (year)	Reward name	Reward scheme				Measures	Data for effect size calculation	Effect size (Hedges' g)	95 % CI	Quantitative findings	Qualitative, additional, or overall findings
		Reward administration	Reward contingency	Reward utilitarian/ redeemable stakes	Reward stakes						
(14) Schöbel et al. (2023)	Trophy badges	Individual: Complete each unit.	Rewards offered for solving each unit	No redemption of rewards	Low-stakes reward	Self-developed test: Problem-solving skills test	GG (*M* = 7.25, *SD* = 3.03, *n* = 36), CG (*M* = 6.02, *SD* = 3.05, *n* = 32)	0.40	[−0.08, 0.88]	No significant difference was found between the groups.	/
	Regular badges	Individual: View supplementary materials.	Rewards offered for finishing or completing a task								
(15) Riedmann et al. (2024)	A certificate of participation	Individual: Users received a certificate of participation as a physical reward, which displayed their name, score, and all earned badges.	Rewards offered for finishing or completing a task	No redemption of rewards	Low-stakes reward	Self-developed test: Immediate knowledge test	GG (mean = 12.8, SD = 1.45, n = 30), CG (mean = 13.15, SD = 1.28, n = 34)	−0.25	[-0.75,0.24]	No significant difference was found between the groups.	/
(16) Maimaiti & Hew (2025)	Redeemable badges	Individual: Complete tasks to earn achievement and extra learning material.	Rewards offered for finishing or completing a task	No redemption of rewards	Low-stakes reward	Adapted from a standardized test	GG (mean = 54.88, SD = 8.83, n = 76), CG (mean = 51.91, SD = 8.91, n = 75)	0.33	[0.01, 0.65]	Students in the gamified group achieved significant higher scores in English language learning than those in the control group. (+)	Students in the experimental group (gamified self-regulated flipped learning) engaged significantly more in self-regulated learning behaviours. The gamified intervention consistently supported students monitoring activities regardless of time, likely due to the continuous informational feedback provided by gamified elements, including redeemable rewards.

Note. CI = confidence interval.
Note [1:] For Bai et al. (2021, pp. 899–904) and Xiao and Hew (2023), CG' refers to the gamified group without tangible rewards, and GG refers to the gamified group with tangible rewards. For all other studies, CG refers to the non-gamified, no-reward group.
Note [2:] Putz et al. (2020) reported the results of two groups of participants, so we treated this research as two independent interventions and computed two effect sizes. A knowledge assessment was conducted at three time points: immediately before (A1), 20 min after (A2), and 2 weeks after (A3) the workshop. We chose A2 as the posttest mean and standard deviation values as they were the immediate posttest results.
Note [3]: Ge (2018) reported the results of two types of tangible reward implementation schemes, so we treated this research as two independent interventions and computed two effect sizes.

thumb is that a I^2 value less than 0.25 indicates low heterogeneity, a I^2 value of 0.5 indicates moderate heterogeneity, and a I^2 value greater than 0.75 indicates high heterogeneity (Higgins & Thompson, 2002).

Research with statistically significant results is more likely to be published than research with nonsignificant results (Rosenthal, 1979). We used several estimates to address potential publication bias. First, we provide a funnel plot of all effect size calculations. A roughly symmetrical funnel plot suggests an absence of publication bias (Egger et al., 1997). Second, we ran Egger's regression test to quantify the degree of asymmetry in the funnel plots. A nonsignificant p-value ($p > .05$) for this test indicates an absence of publication bias (Egger et al., 1997). Third, we computed the Begg and Mazumdar rank correlation (Kendall's tau with continuity correction), which can "quantify the amount of bias captured by the funnel plot" (Borenstein, 2005, p. 195). A significant positive Kendall's tau suggests that smaller studies or those with higher variance report larger or more significant effects, which could be a sign of publication bias.

3. Results

3.1. RQ1: what patterns in reward contingencies emerge across studies using tangible rewards in gamified learning, and how are these patterns associated with academic performance?

We found two experimental studies that directly compared the effects of tangible rewards on students' academic performance in gamified learning, and the results were mixed. Specifically, Bai et al. (2021, pp. 899–904) offered rewards for "exceeding a norm" because only the top five students on the leaderboard and the five students who made the most significant learning progress over the previous week were eligible for the reward. In that study, behavioural engagement increased (more posts and replies), but there was no statistically significant difference in academic performance (Hedges' $g = -0.06$, $p > .05$). By contrast, in another study, Xiao and Hew (2023) offered rewards for "surpassing a score." They reported a statistically significant advantage for the rewarded group on final exam scores (Hedges' $g = 0.71$, $p = .035$). Given there are only two studies with differing contingencies, these findings should be interpreted as preliminary.

3.2. RQ2: what patterns in real-world stakes emerge across studies using tangible rewards in gamified learning, and how are these patterns associated with academic performance?

The two aforementioned primary studies used low-stakes rewards, so we cannot draw generalisable conclusions about stakes from this comparison alone. Bai et al. (2021, pp. 899–904) used course assignment samples as low-stakes tangible rewards and observed no statistically significant difference in academic performance. Xiao and Hew (2023) offered additional course materials as low-stakes rewards to the students who performed well on quizzes in an online flipped gamified classroom. The students could only redeem the reward after reaching 80 % of the total quiz points. They observed a significant positive effect on final exam scores in that specific setting. Overall, with only two low-stakes studies available, the evidence remains limited.

3.3. RQ3: what are the performance differences between gamified learning with tangible rewards to non-gamified, no-reward conditions, and how do they vary by reward contingency?

We found eight types of reward contingency that were applied in the 16 interventions comparing gamified learning with tangible rewards to non-gamified, no-reward conditions. They are "finishing or completing a task" ($k = 5$), "exceeding a norm" ($k = 3$), "surpassing a score" ($k = 2$), "solving each unit" ($k = 2$), "doing a task" ($k = 1$), "doing well" ($k = 1$), "a combination of solving each unit and finishing or completing a task" ($k = 1$), and "a combination of surpassing a score and exceeding a norm" ($k = 1$). The interventions under each category are presented in Table 3. Here, we present two studies that use two reward contingencies. In Schöbel et al. (2023), tangible rewards in their gamification are two types of badges, namely trophy badges (given when students completed each learning unit) and regular badges (given when students viewed supplementary materials). Thus, this study involves both rewards offered for "solving each unit" and "finishing or completing a task". In Sánchez et al. (2021), each student received real badges for correctly answering all questions ("surpassing a score"). At the same time, the one with the most badges receives a trophy ("exceeding a norm").

We examined the effect size for each independent intervention (see Table 2) and summarised the overall findings, highlighting whether the reward contingency was associated with positive, mixed, or negative effects on academic performance.

3.3.1. Observed positive effects of reward contingency on academic performance

We found two reward contingencies reported general positive effects: solving each unit and exceeding a norm.

(1) Two studies reported "solving each unit", which aims to encourage continuous engagement and progressive mastery. In Ge (2018) (1), each participant must answer multiple-choice questions of varying levels of difficulty. If the answer is correct and the explanation is reasonable, the points associated with the question will be added to the participant's final course grade. However, if the participant gives a wrong answer or the explanation is unreasonable, the points for the question will be lost. Accordingly, the same number of points will be deducted from the participant's final course grade. In Ge (2018) (2), the procedure is basically the same as above. The main difference is that students will not lose the points associated with the questions

even if they give wrong answers or unreasonable explanations. Both studies reported improved immediate post-test performance when students earned points for solving progressively difficult questions in an online English course.

(2) Three interventions introduced "exceeding a norm" competitive elements, where reward is offered to meet or exceed the performance of others on the task (relative standard). For example, in Murillo-Zamorano et al. (2021), this kind of rewards were implemented through two mechanisms. First, teams earned extra redeemable points by outperforming other teams on timed tests (higher speed and accuracy), with bonus multipliers applied to advanced-level items; these points were converted into extra credit for the final exam. Second, badges and an award ceremony recognized teams that surpassed participation and achievement thresholds: (a) Flipped classroom badges for full-team completion of all online questionnaires, (b) Cooperative learning badges for high rates of active co-creation among team members, and (c) Gamification badges for the top-scoring teams in Kahoot during the grand finale. The descriptive statistics showed that the gamified group had a higher final course grade than the non-gamified group. Putz et al. (2020) (1) and (2) reported improved short-term knowledge retention (within 14 days) in sustainability workshops in secondary and tertiary education settings where top teams earned group photos and badges. The findings suggest that competitive rewards may help motivate students in short-term, collaborative contexts, but may not generalize across settings. For example, we are not sure of the long-term effects (e.g., longer than a semester) of the rewarding contingency approach as none of the three studies examined the long-term effects of using such rewards.

3.3.2. Observed mixed effects of reward contingency on academic performance

We found two reward contingencies reported mixed effects: surpassing a score and finishing or completing a task.

(1) "Surpassing a score" ($k = 2$) designs typically reward students for meeting an absolute performance standard. In a Grade 7 statistics class, Çakıroğlu and Güler (2021) offered real gifts for students who submitted all correct answers to the question of a week, which was associated with significantly higher test scores. Conversely, Marín et al. (2018), in an undergraduate programming course, awarded medals and access to extra materials for students who submitted correct answers but found negligible differences between groups.

(2) "Finishing or completing a task" was the most common reward contingency ($k = 5$) and associated with mixed outcomes. "Finishing or completing a task" rewards students for completing a task. We observed that when this contingency was tied with high-stakes rewards, i.e., course grade, it tended to show less favourable outcomes. For example, Kwon and Özpolat (2021) tied rewards to additional reports and offered redeemable "power-ups" that could influence exam scores and found that students in the gamified group scored lower than those in the control group. Whereas, when the reward contingency was tied to low-stakes rewards (non-course grade-associated), it showed positive outcomes. In another study, Maimaiti and Hew (2025) reported positive outcomes when students earned redeemable badges and extra learning materials for completing tasks in a flipped classroom, indicating that low-stakes, informational rewards (such as recording the learning task completion progress) may mitigate negative effects.

3.3.3. Supplementary non-significant effect sizes of different reward contingencies on academic performance

When pooled, the interventions of gamified classes with tangible rewards yielded a small, non-significant overall effect (Hedges' $g = 0.39$, 95 % CI [–0.09, 0.88], $p = .114$) with high heterogeneity ($I^2 = 96.5$ %). Moderator analysis indicated that "solving each unit" (Hedges' $g = 0.66$, 95 % CI [−0.72, 2.03], $p = .35$) and "exceeding a norm" (Hedges' $g = 0.27$, 95 % CI [–1.52, 2.05], $p = .77$) trended positive, while "finishing a task" (Hedges' $g = −0.92$, 95 % CI [–2.55, 0.3], $p = .71$) and "surpassing a score" (Hedges' $g = −0.25$, 95 % CI [–2.2, 1.71], $p = .803$) trended negative, but none were statistically significant.

Fig. 2 presents the forest plot of effect sizes in favour of gamified classes with tangible rewards versus the non-gamified, no-reward group. Table 3 presents the effect sizes of students' academic performance for different tangible reward administration schemes.

3.4. RQ4: what are the performance differences between gamified learning with tangible rewards to non-gamified, no-reward conditions, and how do they vary by real-world stakes of the rewards?

Low-stakes rewards, such as badges, trophies, small gifts, and certificates were implemented in nine interventions and high-stakes rewards, i.e., course grades, were applied in seven interventions. "Low-stakes rewards" and "high-stakes rewards" were associated with mixed outcomes, and their effects depend on reward contingency.

3.4.1. Observed frequent positive effects of reward stake on academic performance

"Low-stakes rewards", such as small gifts and trophies (Sánchez et al., 2021), do not influence course grades directly but are intended to encourage student participation or improve performance in classroom activities. For example, Sánchez et al. (2021) found that the awarding of real badges to primary school students led to improvements in reading comprehension. By contrast, Riedmann et al. (2024) provided students with certificates of participation displaying their names, scores, and earned badges; this intervention produced neutral effects.

3.4.2. Observed mixed effects of reward stake on academic performance

"High-stakes rewards" are rewards that typically tie to extra course grades or redeemable course points. For example, Metwally et al. (2021) found negative effects in primary school English course when homework points could be converted into grades. Whereas, in the gamified class design of Tsay et al. (2018), students were required to complete weekly or biweekly essential learning activities before attending class, and their participation is a major part of their engagement score. A modest engagement score component was associated with higher course grades. The results found that the students in the gamified group obtained significantly higher course grades than those in the non-gamified control group.

3.4.3. *Supplementary non-significant effect sizes of different reward stakes on academic performance*

Taken together, we found that low-stakes rewards are more often associated with positive outcomes, while high-stakes rewards are associated with mixed outcomes. Pooled estimates supported these descriptive patterns but remained non-significant: low-stakes rewards trended positive (Hedges' $g = 0.812$, 95 % *CI* [–0.181, 1.806], $p = .109$) and high-stakes rewards trended negative (Hedges' $g = -0.081$, 95 % *CI* [–0.846, 0.683], $p = .835$). One study (Murillo-Zamorano et al., 2021) applying the combination of low and high-stakes reward trended positive (Hedges' $g = 0.332$, 95 % *CI* -1.684, 2.347], $p = .74$). These findings should be interpreted cautiously, given wide confidence intervals and substantial heterogeneity.

3.4.4. *Effects of the combination of reward contingency and reward stake on academic performance*

We observed some patterns in how combinations of reward contingencies and reward stakes influence academic outcomes, though the small number of studies in each group means these trends should be interpreted with caution. Here, we elaborate on the ten sets of combinations and their associated outcomes.

3.4.5. *Observed positive effects of the combination of reward contingency and reward stake on academic performance*

(1) “Exceeding a norm + low-stakes rewards” ($k = 2$) was associated with positive outcomes. In Putz et al. (2020) (1) and Putz et al. (2020) (2), group performance was rated based on correct answers or reasonable solutions to the questions, and the *best team* (with the highest number of correct answers) received a prize and a group picture of the winners. Their results suggested that the students in the two workshops appreciated the group rewards and competition. There were large positive effects on their short-term knowledge retention (Hedges' $g = 1.82$ for gamified workshop I and Hedges' $g = 0.82$ for gamified workshop II).

(2) “Solving each unit + high-stakes rewards” ($k = 2$) was associated with positive outcomes. In two interventions by Ge (2018), a forfeit-or-prize system was used where students could either gain or lose points on their final grades depending on how successfully they solved more challenging questions. Although the two interventions reported positive short-term effects on immediate knowledge tests, the effects on knowledge retention warrant investigation and the anxiety associated with forfeits raises concerns about long-term motivational costs.

(3) “Surpassing a score + low-stakes rewards” ($k = 2$) was associated with positive outcomes. In Çakıroğlu and Güler (2021), students were offered real gifts in exchange for submitting correct answers, resulting in significantly higher test scores among medium- and high-achieving students. The authors stated that although gamifying the instructional process had a positive impact on students' statistical literacy among medium and high-achieving students, the influence on low-achieving students was not positive. Whereas medium and high achievers preferred real gifts or leaderboards as a gamification element to demonstrate their high-level cognitive abilities, virtual rewards (e.g., badges) were the primary choice of low achievers because they were easier to obtain. Hence, it is suggested that an adaptive gamification approach be considered to accommodate the diverse motivational needs and cognitive abilities of different students. In Marín et al. (2018), students who successfully completed a programming challenge by providing a correct answer were rewarded with the opportunity to access exclusive online reading materials and short video lessons. The results showed that the gamified group achieved a higher final exam score than the control group; however, this difference was not statistically significant.

3.4.6. *Observed mixed effects of the combination of reward contingency and reward stake on academic performance*

(1) The combination of “finishing or completing a task + low-stakes rewards” ($k = 2$) showed mixed findings. Maimaiti and Hew (2025) reported improved English language learning outcomes when students earned redeemable badges and extra materials for completing tasks across multiple phases of a flipped classroom. These findings suggest that low-stakes completion rewards may be less detrimental than high-stakes ones, particularly when paired with informational feedback, but not as effective as the performance-contingent rewards.

Another example is that Riedmann et al. (2024) integrated a social robot and gamification elements into an adult language learning setting (Introductory Spanish). The intervention lasted less than one session and included gamified activities where learners completed tasks to progress through the lesson. The tangible reward was a certificate of participation, which displayed the learner's name, score, and earned badges. This certificate qualifies as a low-stakes reward because it has symbolic value but no impact on course grades or formal assessment outcomes. The results showed a non-significant difference between the gamified and control groups.

(2) “Finishing or completing a task + high-stakes rewards” ($k = 3$) was the most common combination and generally showed mixed outcomes. For example, Metwally et al. (2021) implemented a homework-based system where students could redeem experience points from learning exercise completion for additional course final grades. This design associated with significantly lower performance in the gamified group compared to the control group (Hedges' $g = -1.00$). It is reported that students were likely to be intrinsically motivated when homework was attributed to factors under their control, when they considered that

Table 3
Moderator analysis of reward contingency, reward stakes, and reward combination.

Moderator variables		Effect size and 95 % CI						Heterogeneity	
Moderator	Studies	*k*	Hedges' *g*	*SE*	*p*-value	95 % *CI* (lower bound)	95 % *CI* (upper bound)	*Q* value	*p*-value
Reward contingency								3.435	0.488
Finishing or completing a task	Kwon & Özpolat (2021) Tsay et al. (2018) Metwally et al. (2021) Riedmann et al. (2024) Maimaiti & Hew (2025)	5	−0.918	0.831	0.269	−2.546	0.711		
Exceeding a norm	Murillo-Zamorano et al. (2021) Putz et al. (2020) (1) Putz et al. (2020) (2)	3	0.267	0.912	0.77	−1.52	2.054		
Surpassing a score	Marín et al. (2018) Cakıroglu & Guler (2021)	2	−0.249	0.997	0.803	−2.203	1.705		
Solving each unit	Ge (2018) (1) Ge (2018) (2)	2	0.659	0.701	0.348	−0.716	2.033		
Others[1]	Domínguez et al. (2013) (Rewards offered for doing a task) Zvarych et al. (2019) (Rewards offered for doing well) Schöbel et al. (2023) (Rewards offered for solving each unit + Rewards offered for finishing or completing a task) Sanchez et al. (2020) (Rewards offered for surpassing a score + Rewards offered for exceeding a norm)	4	0.017	0.863	0.985	−1.675	1.708		
Reward stakes[2]								2.592	0.274
Low-stakes reward	Putz et al. (2020) (1) Putz et al. (2020) (2) Sanchez et al. (2021) Marín et al. (2018) Zvarych et al. (2019) Schöbel et al. (2023) Cakıroglu & Guler (2021) Riedmann et al. (2024) Maimaiti & Hew (2025)	9	0.812	0.507	0.109	−0.181	1.806		
High-stakes reward	Kwon & Özpolat (2021) Tsay et al. (2018) Metwally et al. (2021) Domínguez et al. (2013) Ge (2018) (1) Ge (2018) (2)	6	−0.081	0.39	0.835	−0.846	0.683		
Mixed (Low- and high-stakes reward)	Murillo-Zamorano et al. (2021)	1	0.332	1.028	0.747	−1.684	2.347		
Reward contingency and stakes combination								4.274	0.511
Surpassing a score + Low-stakes reward	Marín et al. (2018) Cakıroglu & Guler (2021)	2	0.872	0.917	0.342	−0.926	2.669		
Finishing or completing a task + High-stakes reward	Kwon & Özpolat (2021) Tsay et al. (2018) Metwally et al. (2021)	3	−0.462	0.578	0.425	−1.595	0.671		
Finishing or completing a task + Low-stakes reward	Riedmann et al. (2024) Maimaiti & Hew (2025)	2	0.507	0.914	0.579	−1.285	2.298		
Exceeding a norm + Low-stakes reward	Putz et al. (2020) (1) Putz et al. (2020) (2)	2	1.746	0.926	0.059	−0.07	3.561		
Rewards offered for solving each unit + High-stakes reward	Ge (2018) (1) Ge (2018) (2)	2	1.12	0.911	0.219	−0.666	2.907		
Others[3]	Domínguez et al. (2013) (Rewards offered for doing a task + High-stakes reward) Zvarych et al. (2019) (Rewards offered for doing well + Low-stakes reward) Sanchez et al. (2020) (Rewards offered for surpassing a score & Rewards offered for exceeding a norm + Low-stakes reward)	5	1.05	0.733	0.152	−0.387	2.486		

(*continued on next page*)

Table 3 (*continued*)

Moderator variables		Effect size and 95 % CI						Heterogeneity	
Moderator	Studies	*k*	Hedges' *g*	*SE*	*p*-value	95 % *CI* (lower bound)	95 % *CI* (upper bound)	*Q* value	*p*-value
	Murillo-Zamorano et al. (2021) (Rewards offered for exceeding a norm + High- & low-stakes reward) Schöbel et al. (2023) (Rewards offered for solving each unit & Rewards offered for finishing or completing a task + Low-stakes reward)								

Note. k = number of interventions; *SE* = standard error; *n* = total sample size; CI = confidence interval.
Note [1]*:* The category 'Others' under reward contingency includes studies using reward conditions that do not fall into the previously specified categories (e.g., rewards offered simply for doing a task Domínguez et al. (2013)) or studies that implemented multiple reward contingencies within the same intervention (e.g., Schöbel et al. (2023) and Sanchez et al. (2020)). These studies were grouped together because each individual type contained fewer than two studies.
Note [2]*:* There is one study involving both high- and low-stakes rewards, which could not be classified exclusively into either category, as it encompasses elements of both.
Note [3]*:* The category 'Others' under reward combination includes studies using reward conditions that do not fall into the previously specified categories. These studies were grouped together because each individual type contained fewer than two studies.

> they had the expertise needed to be successful learners, and when they were interested in doing homework to learn, not just to obtain high grades.

Similarly, Kwon and Özpolat (2021) tied rewards to the submission of additional reports and offered redeemable "power-ups" that influenced exam scores; yet, students in the gamified group underperformed (Hedges' $g = -0.80$) compared to the non-gamified group. The authors stated that gamification can harm content knowledge and student perceptions when applied only to coursework assessments, and latent social competition for higher grades may lead to lower intrinsic motivation, thereby affecting students' content knowledge. The authors suggested that instead of applying simplistic game mechanics to narrow areas of the coursework, the gamified system should consider encompassing a multitude of mechanics and elements that cater to the diverse needs of students. Yet, which mechanics and elements should be included were not discussed in the report. According to self-determination theory, aside from supporting students' competency, as this study did, we suggested that mechanisms to fulfil students' needs of autonomy and relatedness are also important and worth considering.

3.4.7. No observed association between the combination of reward contingency and reward stake and academic performance

The combination of reward contingency and stakes only occurred once in the primary studies; hence, no conclusive association can

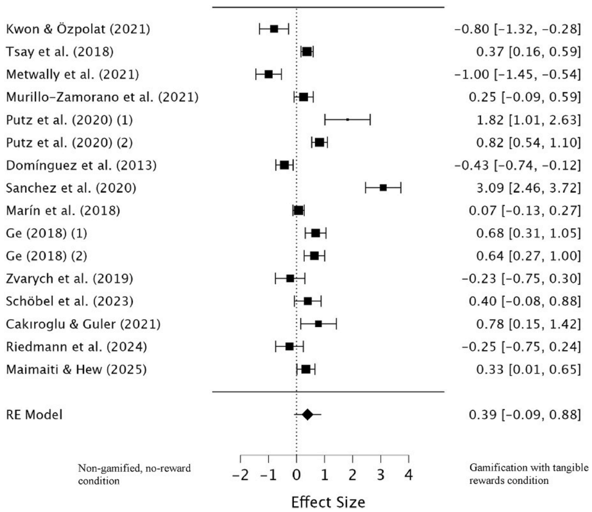

Fig. 2. The Forest Plot in Favour of Gamified Learning with Tangible Rewards Versus the Non-Gamified, No-Reward Group
Note. Points and lines represent effect sizes (Hedges' *g*) and confidence intervals, respectively. The diamond at the bottom of the figure shows the overall effect size, with its width reflecting the corresponding confidence interval, RE Model = Random-effects model.

be found.

(1) “Doing a task + high-stakes rewards” ($k = 1$): In the study of Domínguez et al. (2013), students could earn up to 5 % of their final score based on their participation in activities in the classroom and on the e-learning platform. It was found that the students in the gamified group scored significantly lower in the final exam than those in the control group (Hedge's g = −0.43).
(2) “Doing well + low-stakes rewards” ($k = 1$): In the design of Zvarych et al. (2019), students would obtain a certain number of Lego pieces as they completed quality works, but the quality was not pre-specified. No significant difference was found between the gamified and control groups.
(3) “Surpassing a score and exceeding a norm + low-stakes rewards” ($k = 1$): In Sanchez et al. (2020), each student received real badges for correctly answering all questions (surpassing a score), and they competed to earn the greatest number of badges during a reading comprehension program in primary education. The top-performing student received a trophy as a tangible reward (exceeding a norm), which qualifies as a low-stakes incentive because it did not affect course grades. The results found significantly higher reading comprehension scores and a higher level of interest in learning in the gamified group compared to the control group.
(4) “Exceeding a norm + low- and high-stakes rewards” ($k = 1$): Murillo-Zamorano et al. (2021) applied both low- and high-stakes rewards for the winning team. For instance, students accumulated points by outperforming other teams on tests (faster and more accurate). The team receiving more points (exceeding a norm) for performing better on advanced-level tests would receive extra points for the final exams (high-stakes rewards). Additionally, the team with the most points in Kahoot quizzes during the grand finale was awarded badges at the award ceremony (low-stakes rewards). The results found no significant difference between the gamified and control groups.
(5) “Solving each unit and finishing a task + low-stakes rewards” ($k = 1$): In Schöbel et al. (2023), trophy badges were awarded to students upon completion of a unit of online training exercises. Regular badges were awarded to students when they had viewed supplementary learning materials that was not necessary to finish the online training but suggested for successful completion. These low-stakes badges were not tied to the course grades but only used to encourage further self-learning outside of the class. The results found no significant difference in problem-solving performance.

3.4.8. Supplementary effect sizes of the combination of reward contingency and reward stake on academic performance

The pooled effect for this category of “exceeding a norm + low-stakes rewards” was largest but not statistically significant (Hedges' $g = 1.746$, 95 % *CI* [−0.07, 3.561], $p = .059$). Similarly, the pooled effect for this “solving each unit + high-stakes rewards” was positive but not statistically significant (Hedges' $g = 1.12$, 95 % *CI* [−0.666, 2.907], $p = .219$). The pooled effect for this “surpassing a score + low-stakes rewards” was positive but not statistically significant (Hedges’ $g = 0.872$, 95 % *CI* [−0.962, 2.669], $p = .342$).

The pooled effect for this category of “finishing or completing a task + low-stakes rewards” was positive but not statistically significant (Hedges' $g = 0.507$, 95 % *CI* [−1.285, 2.298], $p = .579$). The pooled effect for this category of “finishing or completing a task + high-stakes rewards” was negative but not statistically significant (Hedges’ $g = -0.462$, 95 % *CI* [−1.595, 0.671], $p = .425$).

Overall, these patterns suggest that low-stakes, performance-contingent combinations (e.g., surpassing a score or exceeding a norm) are more likely to coincide with positive outcomes, whereas high-stakes, completion-contingent combinations often align with negative or mixed results. However, given the small sample sizes and heterogeneity, these observations remain exploratory in nature.

3.4.9. Heterogeneity analysis and publication bias assessment

For reward-related moderators, we did not find significant differences in heterogeneity for reward contingency ($Q = 3.435$, $p = .488$), reward stakes ($Q = 2.592$, $p = .274$) and reward combination ($Q = 4.274$, $p = .511$; see Table 3). For non-reward-related moderators, no significant variation in heterogeneity was found for educational context ($Q = 4.447$, $p = .108$), course subject ($Q = 0.682$, $p = .877$), and intervention duration ($Q = 2.827$, $p = .587$; see Table 4).

The funnel plot of all effect sizes in terms of standardised errors (see Fig. 3) is generally symmetrical, indicating a lack of publication bias (Egger et al., 1997). This result was confirmed by the nonsignificant *p*-value ($p = .114$) of Egger's regression test. The Begg and Mazumdar rank correlation test (Kendall's tau with continuity correction) also yielded a nonsignificant tau value of 0.15 ($p = .45$, two-tailed), indicating that publication bias was nonsignificant.

4. Discussion

To better understand how tangible rewards influence academic performance in gamified learning environments, this section explores emerging patterns in reward strategies and their psychological implications. While our meta-analysis suggests that rewards offered for “exceeding a norm + low-stakes rewards” may be associated with positive outcomes, this observation is based on a limited number of interventions and considerable variation in implementation. As such, it should be interpreted as a preliminary trend rather than a definitive conclusion.

This is a systematic review with supplementary effect sizes to discuss the observed association between the use of tangible rewards in gamified learning and students’ academic achievement, the small number of studies and uneven distribution across reward contingency categories and stake levels limit the generalizability of the findings. Instead of treating the results as conclusive, we propose that the primary contribution of this review lies in its descriptive synthesis of how tangible rewards have been operationalised in practice.

4.1. Performance-contingency rewards are suggestive but inconclusive

Our findings indicate mixed evidence regarding the effects of tangible rewards on students' academic performance when comparing gamified learning with and without such rewards. One possible explanation is that Bai et al. (2021, pp. 899–904) employed a performance-contingent scheme of rewards for "exceeding a norm," in which only the top five students on the leaderboard and the five students who improved the most each week could receive a tangible reward. This limited eligibility may have reduced other participants' expectations of receiving the reward, which could partly explain the lack of significant performance differences. In contrast, Xiao and Hew (2023) used a reward scheme that included "surpassing a score" and "completing a task" with no limit on the number of tangible reward winners each week. This broader eligibility might have increased students' expectations of receiving a reward and coincided with a significant improvement in academic performance in that specific context. However, these observations are based on only two studies, and more research is needed to confirm whether these differences are attributable to reward design or other contextual factors.

In comparing gamified learning environments with tangible rewards to non-gamified, no-reward groups, performance-contingent rewards, specifically those offered for "exceeding a norm", were generally associated with a small positive effect on academic outcomes (Hedges' $g = 0.267$, 95 % *CI* [–1.52, 2.054]), although this effect was not statistically significant ($p = .77$). These rewards, which reflect achievement to a relative standard, can intensify competition among students due to increased social comparison, thereby driving greater effort and improved performance (Dissanayake et al., 2018). However, motivation tends to rise only when participants perceive a genuine chance of success. The educational impact of such competitive strategies depends greatly on how competition is structured; while constructive competition may foster cooperation and mutual support (Sailer & Homner, 2020), excessive competitiveness could reduce motivation or heighten anxiety (Hanus & Fox, 2015). Given the limited evidence base, it remains unclear whether these mechanisms translate reliably to gamified contexts. Future research might explore collaborative reward structures, such as team-based incentives for joint performance (Putz et al., 2020) or cumulative team points based on individual contributions (Murillo-Zamorano et al., 2021), which could offer a more balanced approach, promote shared engagement and mitigate the risks associated with unhealthy competition.

4.2. Extra course grades as high-stakes rewards warrant further examination

The impact of high-stakes rewards was negative (Hedges' $g = -0.081$, $p = .835$), indicating that these rewards were associated with lower academic performance. However, this effect was not statistically significant, indicating that there is no reliable evidence to suggest that high-stakes rewards influence outcomes in this context.

Ryan and Deci (2020) argue that employing grades as high-stakes rewards rarely offers meaningful feedback about actual learning and instead tends to emphasise students' standing in relation to their peers. Although grading is widely used in educational settings, there is little robust evidence to suggest that such strategies genuinely enhance motivation or learning. In fact, research points to a range of negative outcomes, with grades most frequently experienced as controlling rather than supportive of students' autonomy and intrinsic motivation.

Table 4
Moderator analysis of educational context, subject, and intervention duration.

Moderator variables	Effect size and 95 % *CI*						Heterogeneity	
Moderator	k^1	Hedges' *g*	*SE*	*p*-value	95 % *CI* (lower bound)	95 % *CI* (upper bound)	*Q* value	*p*-value
Educational context							4.447	0.108
K–12	3	0.921	0.535	0.086	−0.129	1.97		
Tertiary	9	−0.947	0.614	0.123	−2.15	0.256		
Others[2]	4	0.037	0.704	0.958	−1.342	1.416		
Subject							0.682	0.877
Business, management, and economics	6	0.46	0.44	0.297	−0.403	1.323		
Computer science/Information and communication technologies	2	−0.639	0.869	0.462	−2.342	1.064		
Language	6	0.014	0.623	0.982	−1.206	1.235		
Others[3]	2	0.091	0.882	0.918	−1.639	1.821		
Intervention duration[4]							2.827	0.587
<1 week	4	−0.403	0.727	0.579	−1.828	1.021		
1 week–3 months	4	0.773	0.517	0.135	−0.239	1.786		
3 months–1 semester	3	−1.096	0.783	0.162	−2.63	0.439		
≥ 1 semester	4	−0.04	0.727	0.957	−1.464	1.385		

Note [1]: k = number of interventions.
Note [2]: The category 'Others' under educational contexts includes studies that do not fall into the previously specified categories. These studies were grouped together because each individual type contained fewer than two studies.
Note [3]: The category 'Others' under subjects includes studies that do not fall into the previously specified categories. These studies were grouped together because each individual type contained fewer than two studies.
Note [4]: One study did not report the duration of its intervention.

Another possible explanation is that students may view their educational journey primarily as a means to achieve a higher grade rather than as an opportunity to understand and genuinely enjoy the learning material. They may become trapped in a reward loop and stop exhibiting the learning behaviours associated with high-stakes rewards, such as grades, once those rewards are no longer offered (Zichermann & Cunningham, 2011). Consequently, the controlling aspect of high-stakes rewards becomes much more salient than their informational aspect, which undermines intrinsic motivation (Ryan et al., 1983). This lack of intrinsic motivation has been linked in prior research to poorer performance (Cerasoli et al., 2014), as individuals who dislike their tasks tend to perform worse than those who enjoy them.

4.3. Evidence for low-stakes rewards is suggestive but inconclusive

One pattern that appears more consistent is that low-stakes rewards tended to show a positive direction of effect (Hedges' $g = 0.812$, $p = .109$), but this effect was also not statistically significant. For instance, Sánchez et al. (2021) awarded real badges to students who answered all questions correctly. Marín et al. (2018) awarded a medal to students who answered a question correctly. Such rewards may emphasise personal engagement and a sense of connection with the learning process. These rewards are typically small rewards, offered as incentives that are too weak to influence people's behaviour (Garaus et al., 2015).

Approaches such as providing supplementary learning resources (e.g., Marín et al., 2018) could strengthen the personal connection between students and their studies. Similarly, group activities that culminate in cooperative learning badges (e.g., Murillo-Zamorano et al., 2021) might foster a sense of community and shared engagement, promoting a more authentic bond with the learning environment. These small rewards may help shift students' attention from their final grades to the learning process, leading to longer knowledge retention (Ge, 2018).

This low-stakes reward design aligns with Cognitive Evaluation Theory and Self-Determination Theory, which suggest that high-stakes rewards can be perceived as controlling and may undermine intrinsic motivation, whereas low-stakes rewards are less likely to threaten autonomy and may support informational feedback (Deci et al., 2001; Ryan & Deci, 2020). However, given the lack of statistical significance and the limited number of studies, these interpretations should be viewed as preliminary and hypothesis-generating rather than conclusive.

4.4. Three nuanced patterns of reward strategies

First, competitive rewards, such as those for "exceeding a norm," can effectively motivate students in short-term, collaborative settings like group competitions or leaderboards, though their impact varies and may not generalize across all contexts or sustain long-term engagement. Second, Progress-based rewards that recognize ongoing achievement or task completion help maintain engagement, but their influence on academic performance may depend on how well the gamification mechanism aligns with learning outcomes. Third, low-stakes, informational rewards, such as badges or certificates for recording task completion, may help mitigate negative effects associated with high-stakes or controlling rewards by providing feedback and recognition that supports autonomy and intrinsic motivation without exerting strong external control.

4.5. Summary of qualitative evidence

The qualitative evidence obtained in this study also provides some tentative suggestions for applying tangible rewards in gamified learning. First, students may benefit from having autonomy to choose their preferred tangible rewards. For instance, Çakıroğlu and

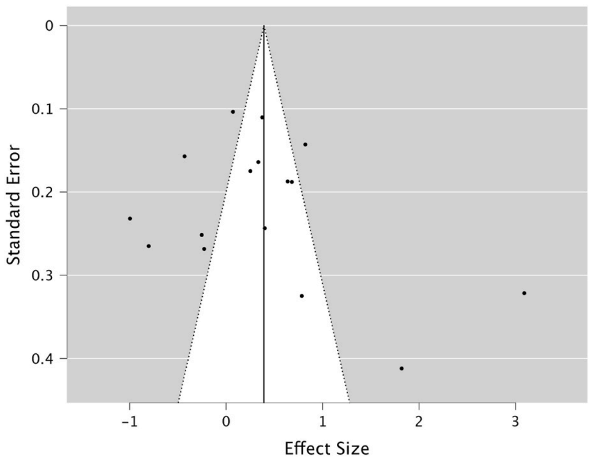

Fig. 3. Funnel Plot Assessing Publication Bias
Note. The middle line lies in the overall effect value. The triangular shapes the 95 % confidence interval. The points are individual studies.

Güler (2021) found that medium- and high-performing individuals preferred real gifts as rewards, whereas low-performing individuals preferred badges because they were easier to obtain. Second, rewards might be more effective when given to recognize progress in learning and involve playful elements (e.g., Lego bricks in Zvarych et al., 2019, and Marvel superhero profiles in Marín et al., 2018). Otherwise, students could perceive tangible rewards as manipulative when the rewards are only driven by prescribed behaviours (Domínguez et al., 2013). Third, a high level of anxiety was experienced by the subjects assigned to the forfeit-or-prize reward scheme (Ge, 2018), suggesting that penalty-based mechanisms may not be advisable in gamified learning in the long run. These insights should be interpreted as exploratory and require further empirical validation.

5. Limitations and future research

Our review has some limitations that open up avenues for future research. First, as only two studies (i.e., Bai et al., 2021, pp. 899–904; Xiao & Hew, 2023) directly compared gamification with tangible rewards, our analysis encountered difficulties in achieving a comprehensive meta-analytic synthesis. In addition, the small number of studies comparing the effects of gamified learning with tangible rewards with those of non-gamified, no-reward groups on students' academic performance may have contributed to the large heterogeneity of the effects observed. This heterogeneity may also be linked to the various implementations of gamification. Despite these limitations, our study synthesised the results of studies comparing non-gamified, no-reward learning scenarios with gamified learning with tangible rewards.

Second, while our coding scheme was guided by the revised reward contingency categories proposed by Cameron et al. (2001), which define seven distinct types of reward contingencies, we also incorporated a binary classification of reward stakes (high vs. low). This theoretically yields a 7×2 matrix of 14 possible combinations. However, the empirical evidence is more limited. Among the 16 studies (18 interventions) included in our review, only 10 of the 14 theoretical combinations were represented. The remaining four combinations were not observed, likely due to underreporting of reward design details or the novelty of certain gamification strategies. Rather than artificially populating all cells, we chose to present only empirically grounded combinations to maintain analytical rigour and transparency. This approach also highlights important gaps in the literature, such as the absence of studies involving "doing well + high-stakes reward", which we now explicitly identify as an area for future research.

Third, while our systematic review synthesizes evidence on the effects of tangible rewards in gamified learning, it is important to acknowledge the considerable variation observed across studies. Notably, research such as Çakıroğlu and Güler (2021) demonstrates that even with an overall positive gamification effect, substantial differences between low and high-achieved students can emerge, where low-achieved students did not benefit much compared to the high-achieved counterparts. These findings suggest that focusing solely on average or pooled outcomes may overlook critical individual or contextual factors influencing effectiveness. This variation underscores the need not only for more studies in this area but also for more in-depth investigations that can unpack the mechanisms underlying these differences. Future research should employ designs that allow for the exploration of within-group variability and the identification of moderating factors.

Fourth, the descriptive effect sizes consistently showed clear trends towards positive or negative effects, thus corroborating the results of our systematic review. However, it is important to acknowledge that our analysis may not fully elucidate the precise impact of tangible rewards due to the small number of primary studies and many nonsignificant results. Nevertheless, our findings offer valuable insights into the potential associations between various reward administration schemes in gamified learning and students' academic performance. When interpreting the effect sizes of reward administration schemes that did not yield significant *p*-values, caution should be exercised.

6. Conclusion

This systematic review aims to understand how tangible reward administration schemes influence students' academic performance in gamified learning. The results of 18 independent interventions revealed that, at the descriptive level, offering rewards for "exceeding a norm" and "solving each unit" are associated with positive academic performance in gamified classes. Specifically, small rewards offered for "exceeding a norm" are more likely to improve students' academic performance when no extra course grades are directly awarded. Teachers are suggested to use a combination of performance-contingent and low-stakes rewards to promote students' academic performance. This study highlights the nuanced interplay between reward mechanisms and educational outcomes, offering insights to guide effective reward strategies in gamified classes. However, the evidence base remains limited and heterogeneous, warranting caution in making broad claims about the effectiveness of gamified reward strategies. Future research should critically examine these interventions and explore individual differences to better understand when and for whom such strategies are most effective.

Author statement

Shurui Bai: Conceptualization, Methodology, Data curation, Writing- Original draft preparation, Funding acquisition.
Yingxue Liu: Data curation, Writing- Original draft preparation.
Khe Foon Hew: Investigation, Validation, Writing - Review & Editing.
Michael Sailer: Investigation, Validation, Writing - Review & Editing.
Ya Xiao: Investigation, Validation.

Acknowledgment

The research is supported by a grant from the Research Grants Council of Hong Kong (Project reference no: ECS 28600024). Large language Models (e.g., Copilot) was used to improve the clarity of some writings in this manuscript. After using this tool, the authors reviewed and edited the specific contents and take full responsibility for the accuracy of the contents. The manuscript was later sent for professional English proofreading for final editing.

Appendix A. Basic Information on Gamification Research with Tangible Rewards

Number	Research source: Author/s (year)	Geographic location	Educational context	Subject/ Discipline	Intervention duration	Sample size (CG, GG)	Gamified component (s)	Theoretical perspectives for gamification design
1	Kwon & Özpolat (2021)	US	Undergraduate	Business: Operations and Supply Chain Management	1 semester (16 weeks)	CG = 29, GG = 33	Activity level: Assessment (in-class and out-of-class activities)	Motivational theory: Self-determination theory
2	Metwally et al. (2021)	Egypt	Primary school (6th grade)	English	1 month	CG = 44, GG = 40	Activity level: Assignment/ homework (out-of-class activities)	Gamification theory: The gamified framework consists of dynamics, mechanics, and components
3	Cakıroglu & Guler (2021)	Turkey	Secondary school (7th grade)	Statistics	5 weeks	CG = 20, GG = 21	Course level (in-class and out-of-class activities)	Statistical literacy framework, and gamification theory
4	Sánchez et al. (2021)	Spain	Primary school (4th grade)	English: Reading	2 months	CG = 43, GG = 43	Course level (in-class activities)	Motivational and gamification theories, and L1 reading comprehension
5	Marín et al. (2018)	Chile	Undergraduate (first year)	Programming	2 semesters	CG 2014 = 407, CG 2015 = 143 GG 2015 = 267	Course level (in-class and out-of-class activities)	Gamification theory: The gamified framework consists of dynamics, mechanics, and components.
6	Putz et al. (2020)	Austria	Secondary and tertiary	Workshop for Sustainable Supply Chain Management	June 2015–May 2017	Workshop 1: CG = 18, GG = 16 Workshop 2: CG = 73, GG = 261	Course/ workshop level (in-class activities)	Gamification theory
7	Tsay et al. (2018)	UK	Undergraduate	Personal and Professional Development Course	24 weeks	CG = 175, GG = 160	Course level (out-of-class/ preclass activities)	User-centred design and student-centred learning
8	Murillo-Zamorano et al. (2021)	Spain	Undergraduate	Macroeconomics	15 weeks	CG = 65, GG = 67	Course level (in-class and out-of-class activities)	Gamification and active learning theories (flipped and cooperative learning approaches)
9	Zvarych et al. (2019)	Ukraine	Undergraduate	English for Specific Purposes	Not reported	CG = 27, GG = 29	Course level (in-class and out-of-class activities)	Gamification theory
10	Domínguez et al. (2013)	Spain	Undergraduate	Qualification for Users of ICT	1 semester (15 weeks)	CG = 68, GG = 106	Course level (in-class and	Motivational and gamification theories

(continued on next page)

(*continued*)

Number	Research source: Author/s (year)	Geographic location	Educational context	Subject/ Discipline	Intervention duration	Sample size (CG, GG)	Gamified component (s)	Theoretical perspectives for gamification design
							out-of-class activities)	
11	Ge, Z.-G. (2018)	China	Continuing education courses	English (fully online)	1 session (3-h online classes)	CG = 60, GG1 (forfeit-or-prize) = 60, GG2 (prize-only) = 60	Activity level: Quizzes	Gamification theories and reward strategies in gamification
12	Schöbel et al. (2023)	A western European country	Undergraduate students (first year)	Management Education: Business and Information Systems Engineering (fully online)	The training lasts between 90 and 100 min	CG = 32, GG = 36	Course/ training programme level (in-class and after-class activities)	Gamification theories
13	Bai et al. (2021)	Hong Kong, China	Postgraduate students	E-Learning Strategies and Management	10 weeks	CG' = 26, GG = 26	Course level (in-class and out-of-class activities)	Gamification theories
14	Xiao & Hew (2023)	China	Undergraduate students	International Business	8 sessions	CG' = 29, GG = 28	Course level (in-class and out-of-class activities)	Gamification and flipped learning theories grounded in the First Principles of Instruction
15	Riedmann et al. (2024)	Germany	Not reported	(Introductory) Spanish	Less than 1 session	CG = 34, GG = 30	Course-level (in- and out-of-class activities)	Motivational theories (e.g., self-determination theory) and gamification theories
16	Maimaiti & Hew (2025)	China	Undergraduate (first year)	Software development	10 weeks	CG = 75, GG = 76	Course-level (in- and out-of-class activities)	Gamified framework (Goal-Access-Feedback-Challenge-Collaboration), flipped learning and self-regulated learning

Note. CG = non-gamified, no-reward group; CG' = gamified group without tangible rewards; GG = gamified group with tangible rewards.

Data availability

Data will be made available on request.

References

References marked with an asterisk * indicate studies included in the review.

Amrein, A. L., & Berliner, D. C. (2002). High-stakes testing, uncertainty, and student learning. *Education Policy Analysis Archives, 10*, 18. https://doi.org/10.14507/epaa.v10n18.2002

* Bai, S., Gonda, D. E., & Hew, K. F. (2021). Effects of tangible rewards on student learning performance, knowledge construction, and perception in fully online gamified learning. *2021 IEEE international conference on engineering, technology & education (TALE)*. https://doi.org/10.1109/TALE52509.2021.9678741.

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

Becker, B. J. (1994). Combining significance levels. In H. Cooper, & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 215–230). Russell Sage Foundation.

Borenstein, M. (2005). Software for publication bias. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis* (pp. 193–220). Wiley. https://doi.org/10.1002/0470870168.ch11.

* Çakıroğlu, Ü., & Güler, M. (2021). Enhancing statistical literacy skills through real life activities enriched with gamification elements: An experimental study. *E-learning and Digital Media, 18*(5), 441–459. https://doi.org/10.1177/2042753020987016.

Cameron, J., Banko, K. M., & Pierce, W. D. (2001). Pervasive negative effects of rewards on intrinsic motivation: The myth continues. *The Behavior Analyst, 24*(1), 1–44. https://doi.org/10.1007/BF03392017

Cerasoli, C. P., Nicklin, J. M., & Ford, M. T. (2014). Intrinsic motivation and extrinsic incentives jointly predict performance: A 40-year meta-analysis. *Psychological Bulletin, 140*(4), 980–1008. https://api.semanticscholar.org/CorpusID:2371919.

Cochran, W. G. (1954). The combination of estimates from different experiments. *Biometrics, 10*(1), 101. https://doi.org/10.2307/3001666

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Deci, E. L. (1975). *Intrinsic motivation*. New York: Plenum.

Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin, 125*(6), 627–668. https://doi.org/10.1037/0033-2909.125.6.627

Deci, E. L., Koestner, R., & Ryan, R. M. (2001). Extrinsic rewards and intrinsic motivation in education: Reconsidered once again. *Review of Educational Research, 71*(1), 1–27. https://doi.org/10.3102/00346543071001001

Dicheva, D., Dichev, C., Agre, G., & Angelova, G. (2015). Gamification in education: A systematic mapping study. *Journal of Educational Technology & Society, 18*(3), 75–88. http://www.jstor.org/stable/jeductechsoci.18.3.75.

Dissanayake, I., Zhang, J., Yasar, M., & Nerur, S. P. (2018). Strategic effort allocation in online innovation tournaments. *Information & Management, 55*(3), 396–406. https://doi.org/10.1016/j.im.2017.09.006

* Domínguez, A., Saenz-De-Navarrete, J., De-Marcos, L., Fernández-Sanz, L., Pagés, C., & Martínez-Herráiz, J. J. (2013). Gamifying learning experiences: Practical implications and outcomes. *Computers & Education, 63*, 380–392. https://doi.org/10.1016/j.compedu.2012.12.020.

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

Eisenberger, R., & Cameron, J. (1996). Detrimental effects of reward: Reality or myth? *American Psychologist, 51*(11), 1153–1166. https://doi.org/10.1037/0003-066X.51.11.1153

Fu, R., Gartlehner, G., Grant, M., Shamliyan, T., Sedrakyan, A., Wilt, T. J., Griffith, L., Oremus, M., Raina, P., Ismaila, A., Santaguida, P., Lau, J., & Trikalinos, T. A. (2011). Conducting quantitative synthesis when comparing medical interventions: AHRQ and the effective Health Care Program. *Journal of Clinical Epidemiology, 64*(11), 1187–1197. https://doi.org/10.1016/j.jclinepi.2010.08.010

Garaus, C., Furtmuller, G., & Güttel, W. H. (2015). The hidden power of small rewards: The effects of insufficient external rewards on autonomous motivation to learn. *The Academy of Management Learning and Education, 15*(1), 45–59. https://doi.org/10.5465/AMLE.2012.0284

* Ge, Z.-G. (2018). The impact of a forfeit-or-prize gamified teaching on e-learners' learning performance. *Computers & Education, 126*, 143–152. https://doi.org/10.1016/j.compedu.2018.07.009.

Hanus, M. D., & Fox, J. (2015). Assessing the effects of gamification in the classroom: A longitudinal study on intrinsic motivation, social comparison, satisfaction, effort, and academic performance. *Computers & Education, 80*, 152–161. https://doi.org/10.1016/j.compedu.2014.08.019

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine, 21*(11), 1539–1558. https://doi.org/10.1002/sim.1186

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

Kalogiannakis, M., Papadakis, S., & Zourmpakis, A. I. (2021). Gamification in science education. A systematic review of the literature. *Education Sciences, 11*(1), 1–36. https://doi.org/10.3390/educsci11010022

Kim, J., & Castelli, D. M. (2021). Effects of gamification on behavioral change in education: A meta-analysis. *International Journal of Environmental Research and Public Health, 18*(7), Article 3550. https://doi.org/10.3390/ijerph18073550

* Kwon, H. Y., & Özpolat, K. (2021). The dark side of narrow gamification: Negative impact of assessment gamification on student perceptions and content knowledge. *INFORMS Transactions on Education, 21*(2), 67–81. https://doi.org/10.1287/ITED.2019.0227.

Levitt, S. D., List, J. A., Neckermann, S., & Sadoff, S. (2016). The behavioralist goes to school: Leveraging behavioral economics to improve educational performance. *American Economic Journal: Economic Policy, 8*(4), 183–219. https://doi.org/10.1257/pol.20130358

Li, L., Hew, K. F., & Du, J. (2024). Gamification enhances student intrinsic motivation, perceptions of autonomy and relatedness, but minimal impact on competency: A meta-analysis and systematic review. *Educational Technology Research & Development, 72*, 765–796. https://doi.org/10.1007/s11423-023-10337-7

Lipsey, M. W., & Wilson, D. B. (2001). Practical meta-analysis. In *Practical meta-analysis*. Sage Publications, Inc.

Liu, Y., Hau, K., Liu, H., Wu, J., Wang, X., & Zheng, X. (2020). Multiplicative effect of intrinsic and extrinsic motivation on academic performance: A longitudinal study of Chinese students. *Journal of Personality, 88*(3), 584–595. https://doi.org/10.1111/jopy.12512

* Maimaiti, G., & Hew, K. F. (2025). Gamification bolsters self-regulated learning, learning performance and reduces strategy decline in flipped classrooms: A longitudinal quasi-experiment. *Computers & Education, 230*, Article 105278. https://doi.org/10.1016/j.compedu.2025.105278.

* Marín, B., Frez, J., Cruz-Lemus, J., & Genero, M. (2018). An empirical investigation on the benefits of gamification in programming courses. *ACM Transactions on Computing Education, 19*(1), Article 4. https://doi.org/10.1145/3231709.

Meder, M., Plumbaum, T., Raczkowski, A., Jain, B., & Albayrak, S. (2018). Gamification in e-commerce: Tangible vs. intangible rewards. *ACM International Conference Proceeding Series*, 11–19. https://doi.org/10.1145/3275116.3275126

* Metwally, A. H. S., Chang, M., Wang, Y., & Yousef, A. M. F. (2021). Does gamifying homework influence performance and perceived gameful experience?. *Sustainability, 13*(9), 1–18. https://doi.org/10.3390/su13094829.

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Journal of Clinical Epidemiology, 62*(10), 1006–1012. https://doi.org/10.1016/j.jclinepi.2009.06.005

* Murillo-Zamorano, L. R., López Sánchez, J.Á., Godoy-Caballero, A. L., & Bueno Muñoz, C. (2021). Gamification and active learning in higher education: Is it possible to match digital society, academia and students' interests?. *International Journal of Educational Technology in Higher Education, 18*(1), 1–27. https://doi.org/10.1186/s41239-021-00249-y.

* Putz, L.-M., Hofbauer, F., & Treiblmaier, H. (2020). Can gamification help to improve education? Findings from a longitudinal study. *Computers in Human Behavior, 110*, Article 106392. https://doi.org/10.1016/j.chb.2020.106392.

* Riedmann, A., Schaper, P., & Lugrin, B. (2024). Integration of a social robot and gamification in adult learning and effects on motivation, engagement and performance. *AI & Society, 39*(1), 369–388. https://doi.org/10.1007/s00146-022-01514-y.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin, 86*(3), 638–641. https://doi.org/10.1037/0033-2909.86.3.638

Ryan, R. M., & Deci, E. L. (2020). Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions. *Contemporary Educational Psychology, 61*, Article 101860. https://doi.org/10.1016/j.cedpsych.2020.101860

Ryan, R. M., Mims, V., & Koestner, R. (1983). Relation of reward contingency and interpersonal context to intrinsic motivation: A review and test using cognitive evaluation theory. *Journal of Personality and Social Psychology, 45*(4), 736–750. https://doi.org/10.1037/0022-3514.45.4.736

Ryan, R. M., & Weinstein, N. (2009). Undermining quality teaching and learning: A self-determination theory perspective on high-stakes testing. *Theory and Research in Education, 7*(2), 224–233. https://doi.org/10.1177/1477878509104327

Sailer, M., & Homner, L. (2020). The gamification of learning: A meta-analysis. *Educational Psychology Review, 32*(1), 77–112. https://doi.org/10.1007/S10648-019-09498-W/TABLES/7

* Sánchez, G. P., Cózar-Gutiérrez, R., del Olmo-Muñoz, J., & González-Calero, J. A. (2021). Impact of a gamified platform in the promotion of reading comprehension and attitudes towards reading in primary education. *Computer Assisted Language Learning, 36*(4), 669–693. https://doi.org/10.1080/09588221.2021.1939388.

Scammacca, N., Roberts, G., & Stuebing, K. K. (2014). Meta-analysis with complex research designs. *Review of Educational Research, 84*(3), 328–364. https://doi.org/10.3102/0034654313500826

* Schöbel, S. M., Janson, A., & Leimeister, J. M. (2023). Gamifying online training in management education to support emotional engagement and problem-solving skills. *Journal of Management Education, 47*(2), 166–203. https://doi.org/10.1177/10525629221123287.

Stenlund, T., Eklöf, H., & Lyrén, P.-E. (2017). Group differences in test-taking behaviour: An example from a high-stakes testing program. *Assessment in Education: Principles, Policy & Practice, 24*(1), 4–20. https://doi.org/10.1080/0969594X.2016.1142935

Subhash, S., & Cudney, E. A. (2018). Gamified learning in higher education: A systematic review of the literature. *Computers in Human Behavior, 87*, 192–206. https://doi.org/10.1016/j.chb.2018.05.028

Taylor, G., Jungert, T., Mageau, G. A., Schattke, K., Dedic, H., Rosenfield, S., & Koestner, R. (2014). A self-determination theory approach to predicting school achievement over time: The unique role of intrinsic motivation. *Contemporary Educational Psychology, 39*(4), 342–358. https://doi.org/10.1016/j.cedpsych.2014.08.002

* Tsay, C. H.-H., Kofinas, A., & Luo, J. (2018). Enhancing student learning experience with technology-mediated gamification: An empirical study. *Computers & Education, 121*, 1–17. https://doi.org/10.1016/j.compedu.2018.01.009.

* Xiao, Y., & Hew, K. F. T. (2023). Intangible rewards versus tangible rewards in gamified online learning: Which promotes student intrinsic motivation, behavioural engagement, cognitive engagement and learning performance?. *British Journal of Educational Technology, 55*(1), 297–317. https://doi.org/10.1111/bjet.13361.

Yıldırım, İ., & Şen, S. (2021). The effects of gamification on students' academic achievement: A meta-analysis study. *Interactive Learning Environments, 29*(8), 1301–1318. https://doi.org/10.1080/10494820.2019.1636089

Zainuddin, Z., Chu, S. K. W., Shujahat, M., & Perera, C. J. (2020). The impact of gamification on learning and instruction: A systematic review of empirical evidence. *Educational Research Review, 30*, 1–24. https://doi.org/10.1016/j.edurev.2020.100326

Zichermann, G., & Cunningham, C. (2011). *Gamification by design: Implementing game mechanics in web and mobile apps* (1st ed.). O'Reilly Media.

* Zvarych, I., Kalaur, S. M., Prymachenko, N. M., Romashchenko, I. V., & Romanyshyna, O. I. (2019). Gamification as a tool for stimulating the educational activity of students of higher educational institutions of Ukraine and the United States. *European Journal of Educational Research, 8*(3), 875–891. https://doi.org/10.12973/eu-jer.8.3.875.