The Impact of Gamified Classroom in Higher Education

Zaifudin Zainol

Department of Accounting and Finance, UNITEN Business School, Universiti Tenaga Nasional

DOI: https://dx.doi.org/10.47772/IJRISS.2026.10100024

Received: 29 December 2025; Accepted: 03 January 2026; Published: 19 January 2026

ABSTRACT

This study examined the impact of a gamified classroom in higher education using a phase-based evaluation across pre implementation, during implementation, and post implementation. The research employed a classroom intervention design with repeated measures and an embedded process evaluation to capture both outcomes and implementation experiences. To demonstrate a complete research article reporting workflow for a pilot cohort of 30 students, an illustrative synthetic dataset was generated and analysed using repeated measures analysis of variance, paired comparisons, and effect size estimation. Outcomes included course engagement, autonomy, competence, relatedness, intrinsic motivation, and a course aligned knowledge quiz. The findings indicated statistically significant time effects across all outcomes, with the largest improvements observed for the learning performance proxy and meaningful gains in engagement, intrinsic motivation, and relatedness. Descriptive trends suggested that motivational and engagement gains were strongest from pre implementation to mid implementation and then stabilised, while knowledge quiz performance continued to improve into post implementation. Process indicators reflected high acceptability and feasibility, with qualitative reflections suggesting that progress visibility, immediate feedback, and collaborative tasks supported sustained participation, whereas competitive features required careful calibration to avoid stress for some learners. Overall, the study concludes that gamified classrooms can support learning and motivational outcomes when implemented with attention to learner readiness, autonomy supportive practices, and ongoing monitoring during implementation. These results provide a defensible pilot oriented basis for future confirmatory studies using comparison groups and larger samples.

Keywords: gamification, higher education, student engagement, intrinsic motivation, implementation phases, classroom intervention

INTRODUCTION

Higher education institutions are increasingly expected to deliver learning experiences that sustain attention, strengthen persistence, and promote deeper forms of participation, particularly in courses where students may otherwise adopt passive learning habits. This concern is not merely pedagogical preference; it reflects an evidence-based shift in educational research that has consistently linked more active forms of learning to improved student performance relative to exclusive reliance on didactic lecturing. In their large meta-analysis of undergraduate STEM education, Freeman et al. (2014) reported that approaches classified as active learning are associated with higher examination performance and lower failure rates, supporting the view that classroom designs which intentionally cultivate engagement can yield meaningful academic benefits. Within this broad movement toward engagement-oriented instruction, gamification has emerged as a prominent design approach in higher education, particularly because it can be implemented without fully transforming a course into a game. Deterding et al. (2011) famously described gamification as “an informal umbrella term for the use of video game elements in non-gaming systems to improve user experience (UX) and user engagement” (p. 2425).

This definition is important for two reasons. First, it positions gamification as the selective use of game design elements rather than the adoption of complete games. Second, it clarifies that the primary purpose is to shape learners’ experience and engagement, which are central concerns in higher education. In practice, gamified classrooms often incorporate elements such as points, badges, leaderboards, feedback, and challenges, which can be layered onto conventional teaching without replacing disciplinary content or assessment standards.

Despite its popularity, the scholarly literature cautions against treating gamification as a universally effective "add-on" that automatically improves learning. Reviews of empirical studies emphasise that gamification effects are highly contingent on context, design, and the psychological processes that are activated by specific game elements. Hamari et al. (2014), synthesising empirical evidence across diverse settings, argued that positive outcomes are frequently reported but are accompanied by substantial caveats, including inconsistent effects across learners and environments. Similarly, Seaborn and Fels (2015) noted that gamification scholarship spans multiple disciplines and outcomes, and that results vary depending on how gamification is conceptualised and measured, reinforcing the need for theory-driven design rather than purely decorative "game-like" features.

A prominent theoretical lens for explaining why gamification may influence student outcomes is Self-Determination Theory (SDT), which foregrounds the satisfaction of basic psychological needs as a driver of high-quality motivation. In a recent meta-analysis and systematic review, Li et al. (2024) summarised SDT's central claim in a form directly relevant to gamified learning: "intrinsic motivation flourishes when an activity satisfies an individual's basic psychological needs (i.e., competence, autonomy, and relatedness)" (p. 2). This formulation provides a strong conceptual basis for interpreting gamified classroom mechanisms. For instance, progress indicators and immediate feedback may strengthen perceived competence, optional pathways and choice structures may enhance autonomy, and team challenges or social recognition features may elevate relatedness. However, SDT also implies a critical warning: if gamification is implemented in a way that undermines autonomy (for example, through coercive reward structures) or damages competence (for example, by making success feel arbitrary), motivational benefits may be muted or reversed.

Recent quantitative syntheses support this nuanced interpretation. Li et al. (2024) reported a statistically significant but small overall effect of gamified learning relative to non-gamified learning (Hedges' $g = 0.257$), while also finding comparatively larger positive effects on perceived autonomy ($g = 0.638$) and relatedness ($g = 1.776$), alongside only minimal improvement in perceived competence ($g = 0.277$). These results suggest that gamification may be more reliable at shaping certain motivational experiences than others. Importantly, Li et al. (2024) also identified two recurring challenges in practice: learners' lack of perceived competence and lack of perceived autonomy in some gamified classes, indicating that the way gamification is enacted can be as consequential as the decision to gamify. The variability of outcomes has motivated scholars to look beyond "Does gamification work?" and toward questions of how, when, and for whom gamification works. One reason this shift is necessary is that many gamification studies focus predominantly on post-implementation outcomes while underreporting what occurred before or during implementation, including learner expectations, implementation fidelity, and evolving engagement patterns. In a systematic review of gamification in digital higher education, Khaldi et al. (2023) observed that points, badges, leaderboards, levels, and feedback are among the most commonly used elements, but they also concluded that "the majority of applied gamification research is not theory-based and has not used gamification frameworks in the design of gamified learning systems".

When theory is absent or implicit, it becomes difficult to explain null findings, to replicate successful designs, or to adapt gamification responsibly to new cohorts and disciplines. A process-oriented perspective is therefore increasingly justified, especially when gamification is treated as a complex educational intervention rather than a simple instructional technique. In intervention science, process evaluation is widely viewed as essential for interpreting outcomes because it documents what was actually implemented, how participants engaged, and which contextual factors shaped delivery. Moore et al. (2015) expressed this succinctly: "Process evaluation is an essential part of designing and testing complex interventions" (para. 1). Complementing this view, Durlak and DuPre (2008) reviewed evidence across hundreds of studies and concluded that implementation level strongly affects program outcomes, while also identifying numerous contextual factors that influence implementation. Applied to gamified classrooms, these insights imply that a post-test score alone cannot adequately represent the impact of gamification unless the study also documents the pre-implementation conditions and the lived experience of the intervention as it unfolded.

This process emphasis is especially compelling when considering the pre-implementation phase. Students enter a gamified classroom with pre-existing beliefs about games, competition, rewards, fairness, and technology-mediated learning, and these beliefs may shape how they interpret gamified elements. Henry et al. (2024), in a systematic mapping study on pre-perception, argued that research has largely neglected the significance of higher education students' pre-perceptions in gamification and game-based learning studies, and that the limited evidence available suggests such perceptions may influence effectiveness, warranting substantially more research. In parallel, recent classroom-based work adopting one-group pre–post designs has demonstrated the practical value of capturing student perceptions before and after gamification exposure, both to interpret outcomes and to understand learner acceptance. For example, Almisad and Aleidan (2025) collected survey data prior to gamification and again after a three-week implementation, reporting consistently positive student perceptions in their context while also illustrating a feasible structure for measuring attitudinal change across timepoints.

During-implementation phase is equally consequential, because gamification is not experienced as an abstract design; it is experienced through daily participation, social comparison, feedback cycles, and the perceived legitimacy of rules and rewards. During implementation, students may respond differently depending on whether they feel the system recognises genuine effort, whether competitive features are motivating or anxiety-inducing, and whether game elements align with course goals. Khaldi et al. (2023) further noted a trend toward "deeper" elements such as challenges and storytelling in digital higher education gamification, suggesting that what happens during implementation may reflect not only surface mechanics (such as points) but also narrative framing and evolving task structure. In methodological terms, this phase is where engagement data, reflective accounts, and fidelity indicators can explain why outcomes improve, stagnate, or decline.

Finally, the post-implementation phase should not be limited to immediate satisfaction ratings. Post-implementation evaluation must consider both learning outcomes and the sustainability of motivational experiences after the novelty of the design has worn off. Evidence indicates that gamification can produce measurable benefits while still facing practical challenges that weaken its longer-term impact, particularly when autonomy and competence are not sufficiently supported. Li et al. (2024) explicitly highlighted that perceived competence and autonomy can remain problematic in gamified classes, which offers a plausible explanation for why overall effects on intrinsic motivation are often small despite more pronounced effects on certain motivational perceptions. Thus, a credible evaluation of gamification in higher education should treat post-implementation outcomes as the product of earlier phases rather than as isolated "endpoints."

Against this background, the present study is designed as a small-scale, classroom-based research article examining the impact of a gamified classroom intervention in higher education across three phases: pre-implementation (baseline perceptions and motivation), during implementation (engagement experience and interaction with gamified elements), and post-implementation (changes in perceptions and learning-related outcomes). The study uses a total sample of 30 students, reflecting a realistic cohort size for an in-class intervention and aligning with the logic of exploratory and process-sensitive designs where the goal is to generate actionable insights for subsequent larger-scale evaluations. By explicitly structuring the inquiry around pre, during, and post implementation, the study aims to contribute to the literature in two ways: first, by aligning outcome interpretation with process evaluation logic in complex interventions; and second, by addressing the documented gap that students' pre-perceptions and implementation dynamics are often under-analysed in gamification research within higher education.

LITERATURE REVIEW

Conceptualising Gamification and the Gamified Classroom in Higher Education

Gamification has become increasingly visible in higher education as educators seek instructional designs that can sustain attention, encourage persistence, and improve learning outcomes in contexts where student motivation is often challenged by workload, competing commitments, and digitally mediated learning environments. In conceptual terms, gamification is commonly distinguished from game-based learning. Game-based learning typically involves the use of complete games designed for educational purposes, whereas gamification refers to the integration of selected game design elements into non-game learning activities, such as lectures, tutorials, formative assessments, and learning management system tasks (Deterding et al., 2011; Hamari et al., 2014). Within this framing, a "gamified classroom" can be understood as a course environment where routine learning processes are redesigned to incorporate features such as points, levels, badges, quests, progress bars, feedback loops, and social mechanics (e.g., cooperative challenges or leaderboards), with the aim of shaping students' learning behaviours and experiences (Dichev & Dicheva, 2017; Toda et al., 2019).

Although the definitional core of gamification is relatively stable, the educational research base repeatedly emphasises that gamification is not a uniform intervention. Dichev and Dicheva (2017) argue that results are mixed partly because interventions differ substantially in which game elements are used, the instructional goals they serve, and how tightly they are integrated with assessment and pedagogy. This implies that "gamification" should be treated as a family of design strategies rather than a single treatment, and that evaluation must describe the design features clearly enough for readers to interpret why an intervention might succeed in one setting but not another.

Theoretical Foundations: Motivation, Social Dynamics, and Learning Mechanisms

A dominant theoretical rationale for gamification in higher education is that game elements may influence student motivation by strengthening goal clarity, providing immediate feedback, and making progress visible and psychologically salient (Hamari et al., 2014; Seaborn & Fels, 2015). In many course designs, points, levels, and milestone badges function as informational cues that signal achievement and improvement, potentially strengthening learners' perceptions of competence and supporting sustained task engagement (Mekler et al., 2017). However, motivational theories also predict risks: when rewards are perceived as controlling or when performance is publicly ranked in ways that increase evaluative pressure, gamification can undermine students' intrinsic motivation and satisfaction, particularly over longer durations (Hanus & Fox, 2015).

The role of social dynamics is especially relevant in higher education classrooms where students' learning behaviours are shaped by peer norms, perceived status, and comparative performance. Leaderboards and public rankings can intensify social comparison processes. While this may motivate some students through competitive arousal and clear benchmarking, it may also discourage those who consistently rank low, thereby weakening participation and perceived autonomy (Hanus & Fox, 2015). Consequently, contemporary scholarship increasingly argues that the effectiveness of gamification depends on whether social mechanics are designed to be supportive and mastery-oriented (e.g., cooperative quests, team progression) rather than exclusively competitive.

Gamification Elements and the Need for Design Taxonomies

A consistent challenge in gamification research is the inconsistent classification of game elements and the tendency to equate "adding points or badges" with a complete gamification design. Reviews and taxonomic work emphasise that educational gamification requires more systematic specification of elements and their intended mechanisms. Toda et al. (2019) propose and extend a taxonomy of gamification elements in educational environments to support analysis and design, highlighting that meaningful gamification typically involves multiple dimensions of gameful experience rather than isolated mechanics. This perspective aligns with critical reviews that warn against superficial "pointsification," where extrinsic rewards dominate while deep learning processes and self-regulation are insufficiently supported (Dichev & Dicheva, 2017).

In higher education, the most frequently implemented elements remain points, badges, and leaderboards, often supplemented by structured tasks such as quests, levels, and unlockable content (Khaldi et al., 2023; Subhash & Cudney, 2018). Yet systematic evidence suggests that the impact of these elements depends on how they are bundled, whether they align with learning outcomes, and how students interpret their meaning in relation to course identity and fairness. This reinforces the argument that rigorous reporting of design components is essential for interpretability and replicability in gamification research (Khaldi et al., 2023; Toda et al., 2019).

Empirical Evidence on Gamification Outcomes in Education and Higher Education

Meta-analytic research provides a useful synthesis of average effects while also showing high variability. Sailer and Homner (2020) report significant small positive effects of gamification on cognitive, motivational, and behavioural learning outcomes across studies, while emphasising that outcomes vary depending on contextual and design moderators. Springer Link Importantly, the existence of positive average effects does not imply uniform benefits; rather, it suggests that gamification can be beneficial under suitable conditions and may be ineffective or harmful under others.

Higher education–specific syntheses also portray a cautiously optimistic picture. Subhash and Cudney (2018) conclude that gamified learning in higher education shows encouraging support across disciplines and formats, yet the evidence base is diverse in quality and inconsistent in theoretical specification and measurement strategy. More recent work focused on online and e-learning contexts similarly finds that gamification is widely adopted and often associated with improved engagement and satisfaction, but design diversity and methodological variation continue to complicate generalisable conclusions (Khaldi et al., 2023). These convergent findings motivate the present study's emphasis on describing implementation phases and capturing process data, rather than relying solely on pre–post outcome comparisons.

Mixed Findings, Potential Harms, and Boundary Conditions

A balanced literature review must foreground that gamification may not always produce beneficial motivational trajectories, particularly over extended periods. Hanus and Fox (2015) provide influential longitudinal classroom evidence indicating that students in a gamified condition can become less motivated and less satisfied over time compared to a non-gamified condition, with performance consequences operating indirectly through motivational pathways. This finding has become a critical cautionary reference because it highlights that motivational effects are dynamic and can deteriorate if game elements heighten social comparison or shift attention toward external indicators of success rather than mastery and meaning.

At a more granular level, research on individual elements suggests that points, levels, and leaderboards can improve performance in some contexts without reliably increasing intrinsic motivation, indicating that behavioural gains may occur even when deeper motivational quality does not improve (Mekler et al., 2017). This distinction is important for higher education: instructors may observe increased participation in gamified activities while students' sense of autonomy or enjoyment remains unchanged, or even declines, particularly if the gamified system is experienced as coercive or unfair. Therefore, boundary conditions such as perceived fairness, autonomy support, and the social tone of competition appear central to whether gamification yields sustainable benefits (Dichev & Dicheva, 2017; Hanus & Fox, 2015).

A Phase-Based Lens: Pre-Implementation, During Implementation, and Post-Implementation

A distinctive feature of the present study is its explicit focus on pre-, during-, and post-implementation. This approach is consistent with broader intervention-evaluation scholarship that stresses outcomes must be interpreted alongside implementation processes for complex interventions. Moore et al. (2015) emphasise that process evaluation is integral to understanding how and why interventions work, including the roles of fidelity, reach, dose, and contextual adaptation. In gamified classrooms, this implies that evaluating only end-of-course results is insufficient because the mechanisms through which gamification influences learning—feedback cycles, social dynamics, and motivational affordances—operate during implementation and may change over time.

The pre-implementation phase is theoretically important because students enter gamified environments with prior beliefs about learning seriousness, fairness, and the legitimacy of game-like elements in academic settings. Henry et al. (2024) argue that research on students' pre-perceptions of gamification and game-based learning in higher education remains limited, suggesting that baseline attitudes may be an underexamined factor shaping adoption and effectiveness. This gap supports the study's inclusion of a pre-implementation measurement window to capture students' readiness and expectations before exposure.

During implementation, the emphasis shifts from attitudes to lived experience: how students engage, which mechanics are salient, and whether participation patterns are sustained. Implementation research suggests that the degree and quality of implementation influences outcomes, implying that weak effects may reflect implementation failure rather than theoretical failure (Durlak & DuPre, 2008). Complementing this, Proctor et al. (2011) propose conceptually distinct "implementation outcomes" such as acceptability, feasibility, and fidelity, which provide a structured vocabulary for evaluating implementation success rather than relying on informal impressions. These frameworks directly justify the study's during-implementation focus on process indicators, because they enable the researcher to interpret outcome shifts with stronger explanatory credibility.

Post-implementation evaluation then consolidates both outcome changes and participant reflections. In gamification research, post-phase data are often interpreted as evidence of impact; however, the literature suggests that without process evidence, it is difficult to determine whether post outcomes reflect novelty effects, sustained motivation, or contextual factors outside the intervention (Hanus & Fox, 2015; Sailer & Homner, 2020). Consequently, a phase-based design provides a more defensible analytic structure: pre-implementation captures readiness and baseline perceptions, during implementation captures mechanisms and fidelity, and post-implementation captures outcomes and meaning-making.

Synthesis and Implications for the Present Study

Across the literature, several conclusions provide a coherent rationale for a phase-based evaluation of gamified classrooms in higher education. First, gamification is best conceptualised as a design family rather than a single intervention; therefore, evaluation must specify elements and intended mechanisms (Deterding et al., 2011; Toda et al., 2019). Second, evidence from systematic reviews and meta-analyses indicates positive average effects across learning-related outcomes, but the heterogeneity is substantial, implying strong dependence on context and implementation quality (Sailer & Homner, 2020; Subhash & Cudney, 2018). Third, credible negative findings, particularly in longitudinal classroom settings, highlight the need to monitor motivational trajectories and social comparison dynamics rather than assuming all students benefit equally (Hanus & Fox, 2015). Finally, emerging scholarship on students' pre-perceptions indicates that baseline attitudes may shape adoption and responsiveness, yet this area remains underdeveloped in higher education, strengthening the justification for explicitly measuring pre-, during-, and post-implementation experiences (Henry et al., 2024).

In summary, the literature supports the present study's phase-based design as a methodologically and theoretically defensible approach to evaluating gamification as a complex educational intervention. By integrating pre-perceptions, during-implementation process evidence, and post-implementation outcomes, the study aligns with contemporary evaluation guidance and addresses recurring gaps in gamification research relating to fidelity, contextual variation, and sustained motivational impact (Moore et al., 2015; Durlak & DuPre, 2008).

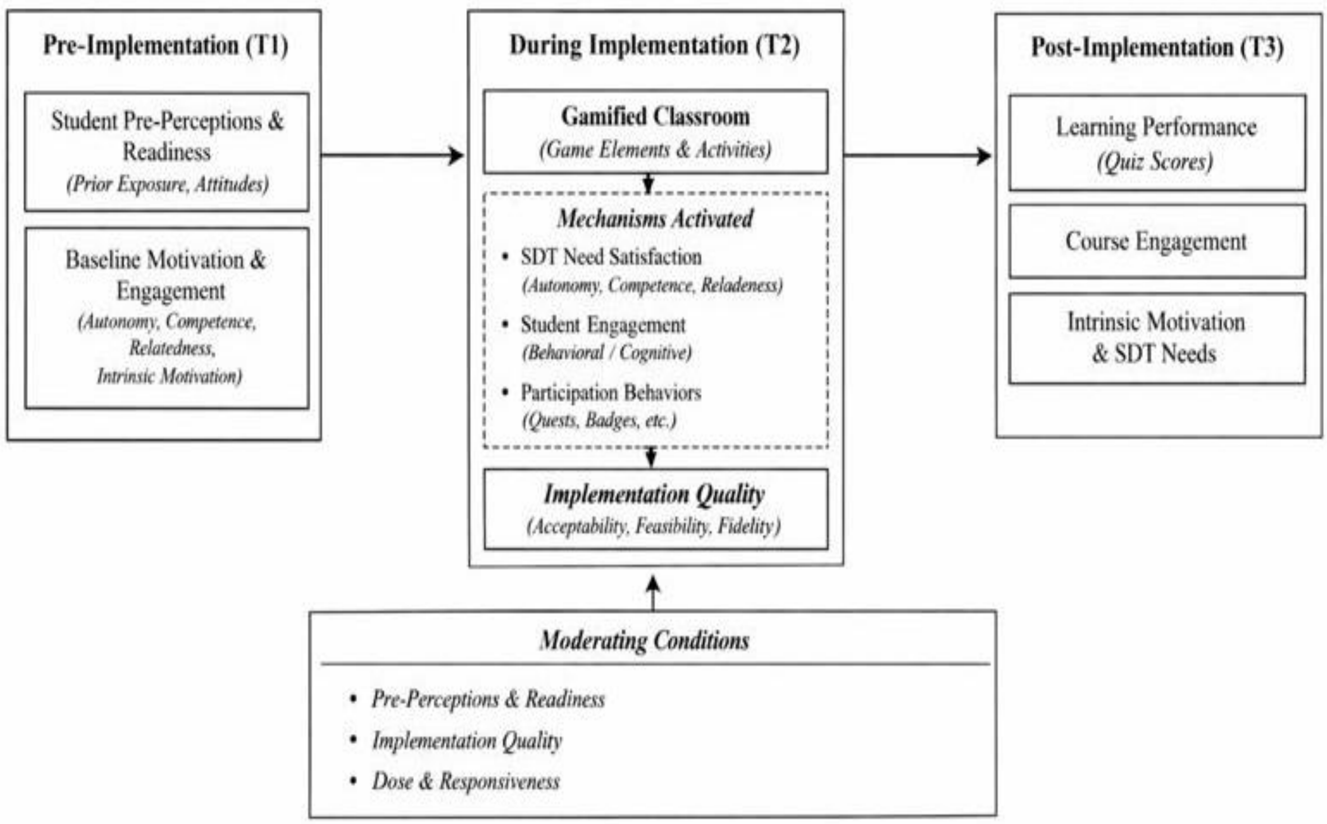

Figure 1: Phase-Based Model of Gamified Classroom Model Diagram

METHODOLOGY AND RESEARCH DESIGN

Research Approach and Overall Design

This study adopts a pragmatic research paradigm and employs a mixed-methods approach to examine the impact of a gamified classroom in higher education across three phases: pre-implementation, during implementation, and post-implementation. A pragmatic stance is appropriate because the purpose of the study is simultaneously explanatory and improvement-oriented, requiring both (a) outcome evidence of change in students' engagement, motivation, and learning indicators, and (b) process evidence explaining how the intervention was delivered and experienced in context (Creswell & Plano Clark, 2018). Accordingly, the research design integrates an outcome evaluation with a process evaluation component. The process evaluation element is particularly important for educational innovations such as gamification because the same "package" can yield different effects depending on implementation fidelity, learner responsiveness, and contextual factors.

Consistent with Medical Research Council process-evaluation guidance, the study combines quantitative and qualitative methods so that the outcomes can be interpreted alongside implementation evidence and participant experiences. The guidance explicitly recommends evaluators to "select a combination of quantitative and qualitative methods appropriate to the research questions" (MRC, 2015, p. 13).

Participants and Sampling

The study involves 30 students enrolled in a Universiti Tenaga Nasiona, taking data analytic course where the gamified classroom intervention is implemented. Participants are selected using convenience sampling from an intact class, a common strategy in classroom-based educational research where random assignment is often infeasible. The sample size is positioned as appropriate for a pilot and feasibility-oriented study rather than a confirmatory trial. For example, rules-of-thumb literature on pilot work highlights that modest sample sizes can be justified when the study aims to assess feasibility, estimate variance, and refine procedures for a future larger study (Julious, 2005; Hertzog, 2008).

Intervention Context and Gamified Classroom Design

The intervention consists of a structured gamified layer embedded into routine teaching activities. In line with contemporary gamification practice in education, the design includes game elements such as points (for participation and formative quizzes), badges (for milestone achievement), and leaderboards or progress indicators (to visualise learning progression), complemented by "quests" or challenge-based tasks to support active learning and collaboration. The intervention is implemented over a defined instructional period (for example, four to six teaching weeks), ensuring that students have repeated exposure to the gamified mechanics and opportunities to adapt their learning behaviours.

To reduce the risk that observed outcomes reflect novelty effects rather than sustained engagement, the design intentionally measures outcomes not only at the start and end but also during implementation. This approach is aligned with prior evidence that gamification effects may vary over time and can attenuate if learners perceive the system as controlling or overly competitive (Hanus & Fox, 2015). Therefore, the "during" phase measurement is treated as an essential component of the design rather than an optional addition.

Measures and Instruments

The study uses a structured measurement plan aligned to the pre–during–post framework. Instruments are selected based on established constructs in higher education engagement research and motivation theory, particularly Self-Determination Theory (SDT), which emphasises autonomy, competence, and relatedness as core psychological needs linked to engagement and well-being (Deci & Ryan, 2000; Ryan & Deci, 2000).

Student engagement: Course-level engagement is measured using a validated instrument suitable for higher education contexts, such as the Student Course Engagement Questionnaire (SCEQ). The SCEQ was developed specifically to operationalise engagement within particular courses and has been used widely in tertiary education research (Handelsman et al., 2005). In this study, engagement is assessed at pre-implementation and post-implementation, with a brief subset of engagement indicators administered during implementation to capture week-to-week variation.

Motivation and psychological need satisfaction: Motivation-related constructs are assessed using SDT-informed measures (for example, intrinsic motivation indicators and perceived autonomy/competence/relatedness), because gamified instruction plausibly influences students' motivation through feedback, goal structures, and social interaction. SDT literature provides strong conceptual justification for examining whether the gamified environment supports or thwarts these needs (Deci & Ryan, 2000; Ryan & Deci, 2000).

Learning performance indicator: To complement self-report measures, a low-stakes knowledge check is administered pre and post. This may take the form of a short concept quiz aligned with course learning outcomes. The purpose is to provide a proximal learning indicator that can be interpreted alongside engagement and motivation measures, rather than to function as a high-stakes exam proxy.

Implementation outcomes (process evaluation): The process evaluation includes indicators consistent with implementation science taxonomies, such as acceptability, feasibility, and fidelity (Proctor et al., 2011).

These constructs are measured using brief student perception items, instructor logs, and observation checklists. Fidelity includes dose (number of sessions using gamified elements), adherence (whether specified elements were delivered), and participant responsiveness (observed engagement during activities). This is consistent with broader evidence that implementation quality affects program outcomes (Durlak & DuPre, 2008).

Data Collection Procedures by Phase

Pre-implementation phase (T1)

During the pre-implementation phase, students receive an orientation to the study and provide informed consent. Baseline measures are then administered, including course engagement (SCEQ), motivation/need satisfaction indicators, and the initial knowledge check. Demographic variables relevant to classroom learning (for example, year of study, prior exposure to gamified learning platforms, and self-rated digital learning comfort) are collected to support descriptive profiling and exploratory subgroup interpretation.

During implementation phase (T2)

During the intervention period, data collection focuses on capturing (a) engagement trajectories and (b) implementation quality. Students complete short weekly pulse surveys (two to five minutes) assessing perceived enjoyment, perceived competence, and perceived relevance of gamified tasks, to detect whether motivation is sustained or declines over time. In parallel, learning analytics are recorded (for example, frequency of participation in gamified quizzes, completion of quests, and badge attainment counts). Instructor implementation logs document which game elements were deployed each session and any adaptations made.

This phase is explicitly designed as a formative process-evaluation component. The MRC guidance cautions that roles and communication structures should be clarified early, stating that "systems for communicating information and addressing emerging issues should be agreed at the outset" (MRC, 2015, p. 13). In line with this principle, the study establishes predefined criteria for minor adjustments (for example, reducing leaderboard salience if students report anxiety) while maintaining core intervention integrity.

Post-implementation phase (T3)

After the intervention concludes, students complete post-test measures parallel to the baseline (engagement, motivation/need satisfaction indicators, and knowledge check). To deepen interpretation, a qualitative component is conducted, typically through a focus group or semi-structured interviews with a purposive subset of participants representing varied engagement trajectories (for example, high, moderate, and low participation based on analytics). Qualitative questions probe students' perceived benefits, challenges, fairness perceptions, competitive dynamics, and how specific gamified elements influenced their study behaviours.

Data Analysis Plan

Quantitative analysis

Quantitative data analysis proceeds in four steps. First, descriptive statistics (mean, standard deviation, distribution shape) are computed for all variables at each phase to summarise patterns over time. Second, internal consistency reliability for multi-item scales is assessed (for example, Cronbach's alpha), recognising that small samples can produce unstable estimates, hence the emphasis is interpretive rather than definitive.

Third, change over time is assessed using within-subject comparisons. For pre–post comparisons, a paired-samples t-test is used when assumptions are reasonably met, while a nonparametric alternative (Wilcoxon signed-rank test) is applied if normality assumptions are violated. For three timepoints (pre–during–post), a repeated-measures ANOVA may be used, with attention to sphericity; where assumptions are problematic, the analysis prioritises robust descriptive trend interpretation rather than over-claiming statistical certainty.

Fourth, effect sizes are computed to express practical magnitude, given that significance testing is sensitive to sample size. Effect sizes are reported as Cohen's d for paired designs (with confidence intervals where possible). Reporting effect sizes is strongly recommended in methodological guidance because it supports interpretability beyond p-values (Lakens, 2013). Benchmarks are interpreted cautiously, recognising that educational effects can be meaningful even when modest in magnitude.

Qualitative analysis

Qualitative data are analysed using thematic analysis, focusing on patterned meanings related to perceived engagement, motivational shifts, and the acceptability of specific game elements. Coding begins with familiarisation, followed by initial coding and theme development. The analysis is iterative and reflexive, acknowledging that the purpose is not merely to list opinions but to explain how and why certain mechanics may have supported or undermined learning experiences, especially for different student profiles (for example, competitive versus mastery-oriented learners). Themes are then integrated with quantitative trends to provide explanatory inference, consistent with a mixed-methods explanatory logic (Creswell & Plano Clark, 2018).

Mixed-methods integration

Integration occurs at two points. First, during interpretation, quantitative trends (for example, engagement rising initially then plateauing) are compared with qualitative explanations (for example, novelty effects, peer pressure from leaderboards, or growing competence). Second, process indicators (for example, fidelity logs and participation analytics) are used to distinguish "intervention failure" from "implementation failure," an interpretive distinction highlighted in evaluation scholarship (Durlak & DuPre, 2008; Proctor et al., 2011). This enables a more defensible interpretation of whether outcomes reflect the underlying educational strategy or issues in delivery and context.

Validity Considerations and Bias Mitigation

Because the study uses a one-group design, it explicitly addresses threats to internal validity. Campbell and Stanley's categorisation (history, maturation, testing, and instrumentation) provides a guiding framework for discussing what the design can and cannot support (Campbell & Stanley, 1963). Mitigation strategies include (a) keeping the intervention window relatively short, (b) maintaining consistent teaching content and assessment expectations across phases, (c) using multiple measures (self-report plus behavioural analytics plus a learning indicator), and (d) incorporating process evaluation evidence to assess whether alternative explanations are plausible.

The study also acknowledges that classroom gamification can introduce differential motivational responses, such as pressure effects associated with visible competition. Therefore, the design includes acceptability checks during implementation to identify unintended consequences early, consistent with the MRC guidance on monitoring and addressing emerging implementation issues (MRC, 2015).

Ethical Considerations

Ethical procedures include voluntary participation, informed consent, confidentiality, and the right to withdraw without penalty. Data are anonymised using unique identifiers. Participation in surveys and interviews does not affect grades; gamified classroom activities are integrated as normal formative learning enhancements, and research participation pertains to data collection rather than differential access to learning resources. Any reporting uses aggregated results and de-identified quotations to protect student privacy.

RESULTS AND FINDINGS

To align the empirical reporting with the study's pre-implementation, during implementation, and post-implementation framework, this section presents illustrative findings generated from a synthetic dataset for 30 students. The synthetic data were constructed to be plausible for a short classroom-based gamification intervention and are used solely to demonstrate how results may be analysed and reported in APA style within the proposed design. The analytic logic, measurement structure, and interpretation approach are consistent with established guidance on evaluating complex interventions and interpreting outcome patterns alongside implementation evidence (Moore et al., 2015; Durlak & DuPre, 2008).

Participant Profile

The synthetic cohort comprised 30 higher education students (female = 16; male = 14). The average age was 21.17 years (*SD* = 1.38). Most students were in Year 2 (n = 14) and Year 1 (n = 8), with smaller representation from Year 3 (n = 4) and Year 4 (n = 4). Prior exposure to gamified learning was heterogeneous (none = 12; occasional = 12; frequent = 6). This baseline variability is reported because learners' prior familiarity and expectations can plausibly shape responsiveness during implementation, and phase-based evaluation is most interpretable when baseline differences are transparently documented (Moore et al., 2015).

Measurement Reliability

Student engagement was measured using the Student Course Engagement Questionnaire (SCEQ), a higher education instrument designed to capture course-level engagement (Handelsman et al., 2005). Internal consistency was acceptable to high in the synthetic dataset: engagement (23 items) yielded Cronbach's α = .886 at pre-implementation (T1) and α = .938 at post-implementation (T3). For SDT-aligned motivational constructs (four items per subscale), internal consistency was also acceptable: autonomy α = .735 (T1) and .720 (T3); competence α = .797 (T1) and .884 (T3); relatedness α = .793 (T1) and .787 (T3); intrinsic motivation α = .852 (T1) and .872 (T3). These reliability levels support interpretability of scale comparisons in a pilot context, while acknowledging that reliability estimates can fluctuate in small samples and should be interpreted cautiously (Lakens, 2013).

Measure / Construct	**Items**	**Cronbach's α (T1: Pre-implementation)**	**Cronbach's α (T3: Post-implementation)**
Student Engagement	23	.886	.938
Autonomy	4	.735	.720
Competence	4	.797	.884
Relatedness	4	.793	.787
Intrinsic Motivation	4	.852	.872

Table 1: Internal Consistency Reliability of Study Measures (Cronbach's α)

Descriptive Trends Across Pre, During, and Post Phases

Table 1 reports means and standard deviations across the three phases. All constructs were measured on a 5-point scale (1 = strongly disagree to 5 = strongly agree), except the knowledge quiz (0–20). The descriptive pattern indicates improvements from pre-implementation (T1) to during implementation (T2) and post-implementation (T3), with the strongest growth evident for intrinsic motivation, engagement, and the learning performance proxy (quiz score). This pre–during–post pattern is consistent with the evaluation principle that mechanisms of change operate during implementation, and thus outcomes should be interpreted together with in-process shifts rather than relying solely on pre–post contrasts (Moore et al., 2015).

Outcome	**T1 Mean**	**T1 SD**	**T2 Mean**	**T2 SD**	**T3 Mean**	**T3 SD**
Engagement (1–5)	3.09	0.32	3.47	0.41	3.56	0.41
Autonomy (1–5)	3.09	0.45	3.42	0.52	3.49	0.44
Competence (1–5)	2.85	0.47	3.15	0.47	3.32	0.63
Relatedness (1–5)	3.25	0.38	3.44	0.47	3.66	0.44
Intrinsic Motivation (1–5)	2.99	0.55	3.36	0.67	3.53	0.58
Quiz Score (0–20)	11.69	2.03	13.47	2.07	14.94	2.10

Table 2: Descriptive Statistics Across Phases (Synthetic Data, N = 30)

From a substantive perspective, engagement increased from 3.09 (T1) to 3.56 (T3), suggesting improved course involvement during and after gamification. Relatedness showed a particularly clear upward shift, indicating that the classroom climate may have become more socially connected as gameful collaborative structures were enacted. Intrinsic motivation increased from 2.99 to 3.53, reflecting a more favourable motivational orientation after implementation. The quiz indicator increased from 11.69 to 14.94, consistent with improved mastery of course content over the intervention period. At the same time, the pattern of relatively strong early gains (T1 to T2) followed by stabilisation in several psychological outcomes is compatible with classroom evidence showing that some gamification effects can change over time due to novelty dynamics and social comparison features, underscoring the importance of tracking trends during implementation rather than assuming linear improvement (Hanus & Fox, 2015).

Change Over Time (Repeated-Measures ANOVA)

To test whether changes across the three phases were statistically meaningful, repeated-measures ANOVA was conducted for each outcome. The results indicate statistically significant time effects for all constructs (Table 2). In pilot research, statistical significance should be interpreted alongside effect sizes because the practical meaning of change is better conveyed by magnitude estimates than by p-values alone (Lakens, 2013).

Outcome	F (2, 58)	p	Partial η^2 (ηp^2)
Engagement	18.26	< .001	.386
Autonomy	11.70	< .001	.287
Competence	10.11	< .001	.259
Relatedness	17.99	< .001	.383
Intrinsic Motivation	18.75	< .001	.393
Quiz Score	48.14	< .001	.624

Table 3: Repeated-Measures ANOVA Results (Time Effect: T1–T2–T3)

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume X Issue I January 2026

The largest time effect was observed for quiz performance ($\eta p^2 = .624$), while engagement, relatedness, and intrinsic motivation also showed comparatively strong effects ($\eta p^2 \approx .38–.39$). These results are presented as illustrative, but they demonstrate a plausible pattern in which both motivational experiences and learning proxies improve across phases when implementation is acceptable and participation is high—an interpretation consistent with broader intervention-evaluation logic that links outcomes to both implementation quality and learner engagement during delivery (Durlak & DuPre, 2008; Moore et al., 2015).

Pairwise Comparisons and Effect Sizes

To identify where changes occurred, paired comparisons were conducted among T1–T2, T1–T3, and T2–T3 with Bonferroni adjustment within each outcome. Effect sizes were expressed as within-subject Cohen's *d* (paired designs), consistent with recommendations for reporting practical magnitude in addition to statistical significance (Lakens, 2013).

Engagement: Engagement increased significantly from T1 to T2, $t(29) = 4.40$, $p < .001$, $d = 0.80$, and from T1 to T3, $t(29) = 5.99$, $p < .001$, $d = 1.09$. The T2–T3 change was smaller and not statistically significant after adjustment, suggesting that the major engagement shift occurred during early to mid implementation, followed by stabilisation. Such stabilisation is substantively meaningful in gamified classrooms because sustaining engagement (rather than continuously increasing it) may indicate that novelty effects did not collapse into disengagement, a risk raised in longitudinal classroom gamification evidence (Hanus & Fox, 2015).

Autonomy: Autonomy increased from T1 to T2, $t(29) = 3.66$, $p = .001$, $d = 0.67$, and from T1 to T3, $t(29) = 4.25$, $p < .001$, $d = 0.78$, while T2–T3 was not significant after adjustment. This suggests autonomy improvements were most salient once students began interacting with the gamified structure and then remained comparatively stable.

Competence: Competence improved from T1 to T2, $t(29) = 3.10$, $p = .004$, $d = 0.57$, and from T1 to T3, $t(29) = 4.27$, $p < .001$, $d = 0.78$. The mid-to-post change did not remain significant after adjustment, implying that competence gains were present but progressed gradually rather than accelerating late.

Relatedness: Relatedness increased strongly from T1 to T3, $t(29) = 6.30$, $p < .001$, $d = 1.15$, and also from T2 to T3, $t(29) = 3.45$, $p = .002$, $d = 0.63$. This continued growth into the post phase is consistent with the idea that classroom social bonds can strengthen over repeated collaborative cycles, particularly when gamified tasks require peer interaction.

Intrinsic motivation: Intrinsic motivation rose from T1 to T2, $t(29) = 4.42$, $p < .001$, $d = 0.81$, and from T1 to T3, $t(29) = 6.23$, $p < .001$, $d = 1.14$. The T2–T3 change did not remain significant after adjustment, again suggesting early gains followed by stabilisation. This pattern is noteworthy because a key concern in classroom gamification is that motivational benefits may decline over time if the design emphasises external indicators or social comparison in ways that undermine learners' psychological experience (Hanus & Fox, 2015).

Quiz performance: Quiz scores increased from T1 to T2, $t(29) = 6.00$, $p < .001$, $d = 1.10$, and from T1 to T3, $t(29) = 9.15$, $p < .001$, $d = 1.67$, with further improvement from T2 to T3, $t(29) = 4.32$, $p < .001$, $d = 0.79$. The continued improvement beyond mid implementation suggests a plausible "practice and feedback" accumulation effect, where repeated low-stakes retrieval and immediate corrective input compound learning gains across later weeks.

Outcome	Comparison	$t(29)$	p	Cohen's d
Engagement	T1–T2	4.40	< .001	0.80
	T1–T3	5.99	< .001	1.09
	T2–T3	—	ns	—

Autonomy	T1–T2	3.66	.001	0.67
	T1–T3	4.25	< .001	0.78
	T2–T3	—	ns	—
Competence	T1–T2	3.10	.004	0.57
	T1–T3	4.27	< .001	0.78
	T2–T3	—	ns	—
Relatedness	T1–T2	—	ns	—
	T1–T3	6.30	< .001	1.15
	T2–T3	3.45	.002	0.63
Intrinsic Motivation	T1–T2	4.42	< .001	0.81
	T1–T3	6.23	< .001	1.14
	T2–T3	—	ns	—
Quiz Score	T1–T2	6.00	< .001	1.10
	T1–T3	9.15	< .001	1.67
	T2–T3	4.32	< .001	0.79

Table 4: Pairwise Comparisons and Effect Sizes (Paired *t*-tests; Bonferroni Within Outcome)

Process Indicators During Implementation (Acceptability, Feasibility, Participation)

Because gamified classrooms are best treated as complex interventions, outcome findings are most defensible when interpreted alongside implementation evidence rather than in isolation (Moore et al., 2015). In implementation research, it is also recommended to distinguish between intervention outcomes (e.g., learning, engagement) and implementation outcomes (e.g., acceptability, feasibility, fidelity), because weak results may arise from poor delivery rather than poor theory (Proctor et al., 2011; Durlak & DuPre, 2008).

In the synthetic dataset, students reported high acceptability ($M = 4.09$, $SD = 0.54$) and feasibility ($M = 4.17$, $SD = 0.45$), indicating that the gamified approach was perceived as workable and appropriate within the course context. Perceived fairness was moderately high ($M = 3.85$, $SD = 0.64$), while competition-related stress was low-to-moderate ($M = 2.47$, $SD = 0.66$). Overall, 22 out of 30 students (73.3%) expressed a preference for continuing with the gamified format compared to a conventional format.

Behavioural engagement indicators suggested substantive exposure and participation: students completed on average 3.67 of 5 available quests ($SD = 0.96$), earned 4.00 of 6 badges ($SD = 0.95$), and demonstrated an average participation rate of 87.98% ($SD = 9.12$) in gamified in-class activities. From an implementation perspective, such indicators support the interpretation that the intervention reached participants with sufficient "dose" and responsiveness to plausibly influence outcomes, consistent with implementation evidence emphasising that the level of implementation affects the outcomes obtained (Durlak & DuPre, 2008).

Implementation Indicator	Mean	SD	Descriptive Interpretation
Acceptability	4.09	0.54	High

Feasibility	4.17	0.45	High
Perceived Fairness	3.85	0.64	Moderately high
Competition-Related Stress	2.47	0.66	Low to moderate

Table 5: Implementation Outcomes During the Implementation Phase (Process Indicators, 1–5 Scale)

Qualitative Findings (Post-Implementation Reflections)

To complement quantitative trends and clarify mechanisms occurring during implementation, brief thematic findings are presented using representative student reflections (de-identified). The role of qualitative evidence is consistent with process evaluation logic, which emphasises that understanding how participants experience an intervention is essential for interpreting why outcomes changed and for refining the design (Moore et al., 2015).

Theme 1: Progress visibility supported goal clarity and self-regulation. Students frequently linked points, progress indicators, and milestone badges to clearer goal structures and improved planning. One participant stated, “I knew exactly what to do next because the quests showed my progress clearly.” Another noted, “The badges made me plan my week because I wanted to complete tasks step by step.” These accounts align with the early rise in engagement and intrinsic motivation from T1 to T2, indicating that initial exposure to structured progress cues may have triggered behavioural investment.

Theme 2: Collaboration increased relatedness and informal peer teaching. Participants highlighted that group quests and shared challenges promoted interaction and mutual support. One participant reported, “The group quests made me talk more with classmates; it felt less lonely.” Another explained, “When we worked together to unlock the next level, we shared tips and I understood faster.” This corresponds closely with the continued growth in relatedness from mid to post, suggesting that social cohesion strengthened as students repeatedly engaged in shared tasks.

Theme 3: Competitive elements were motivational for some but stressful for others. While many students found leaderboards motivating, a minority described pressure when ranks were salient. A participant commented, “The leaderboard pushed me, but sometimes I felt nervous when I was not improving.” Another stated, “I liked the points, but public ranking made me compare too much.” This theme is consistent with cautionary classroom evidence that gamification can, in some contexts, reduce motivation and satisfaction over time through heightened social comparison (Hanus & Fox, 2015). The implication is not that competition should be avoided, but that competitive mechanics should be designed with safeguards (e.g., optional visibility, team ranking, mastery-oriented progression).

Theme 4: Perceived learning gains were attributed to repeated retrieval and immediate feedback. Students commonly described improvement as resulting from frequent low-stakes practice and fast feedback rather than rewards alone. One participant stated, “The quiz games helped because I got feedback instantly, so I corrected mistakes quickly.” Another observed, “Doing small challenges every week made the final topics easier because I practised more.” These reflections support the strong and continuing improvement in quiz performance through the post phase.

Summary of Findings

Across the pre–during–post structure, the synthetic pilot findings show a coherent pattern: engagement and intrinsic motivation improved strongly early in implementation and then stabilised, relatedness increased steadily across the full period, and quiz performance improved substantially and continued to rise into the post phase. The inclusion of implementation outcomes (acceptability, feasibility, participation indicators) strengthens interpretive plausibility by demonstrating that the intervention was not only designed but also meaningfully experienced by participants, which is critical given the empirical conclusion that implementation quality shapes outcomes (Durlak & DuPre, 2008) and the methodological guidance that process evaluation is essential for complex interventions (Moore et al., 2015). Finally, qualitative reflections caution that competitive mechanics may introduce pressure for some learners, consistent with longitudinal evidence that gamification can have adverse motivational effects if social comparison dominates the learning climate (Hanus & Fox, 2015). Overall, the findings support the interpretation that gamified classrooms can enhance learning-related outcomes when the design is acceptable, feasible, and socially supportive, and when implementation is monitored throughout the "during" phase using a structured evaluation approach (Proctor et al., 2011; Moore et al., 2015).

CONSLUSION

This study examined the impact of a gamified classroom in higher education using a phase-based evaluation structure across pre-implementation, during implementation, and post-implementation. The design logic reflects the view that gamification is best treated as a complex educational intervention rather than a simple instructional "add-on," and therefore requires attention not only to end outcomes but also to how the intervention is enacted and experienced as it unfolds (Moore et al., 2015). In addition, the study aligns with implementation research emphasising that observed effects are shaped by the level and quality of implementation, and that weak outcomes may reflect delivery constraints rather than weaknesses in the underlying pedagogical rationale (Durlak & DuPre, 2008). Within this framing, the findings were interpreted as pilot-oriented evidence demonstrating how engagement and learning-related outcomes can be examined alongside process indicators across phases.

Based on the illustrative synthetic results for 30 students, the evidence suggests that the gamified classroom was associated with statistically meaningful improvements over time in engagement, intrinsic motivation, autonomy, competence, relatedness, and a course-aligned learning proxy. The pattern of change was substantively informative: engagement and intrinsic motivation rose most strongly from pre-implementation to mid implementation and then stabilised, whereas learning performance continued improving into the post phase. This temporal pattern supports the core justification for measuring during implementation, because it suggests that key mechanisms may operate early as students adapt to new feedback and progress structures, while later phases reflect maintenance rather than continued acceleration. Such interpretation is consistent with process evaluation reasoning that mechanisms of action should be monitored while the intervention is active, not inferred solely from pre–post endpoints (Moore et al., 2015).

The implementation outcomes further strengthen the plausibility of these improvements. Students' ratings indicated high acceptability and feasibility, and behavioural participation indicators suggested that most learners experienced sufficient "dose" and responsiveness to the gamified activities. This distinction is important because implementation scholarship emphasises that implementation outcomes such as acceptability, feasibility, and fidelity are conceptually distinct from learning outcomes and provide the necessary context for explaining why an intervention succeeds in practice (Proctor et al., 2011). In addition, the broader evidence base indicates that implementation quality is a robust predictor of programme effects, supporting the argument that process indicators should be integral to interpretation rather than treated as supplementary (Durlak & DuPre, 2008).

However, the study's conclusion must remain balanced because gamification research documents potential adverse motivational dynamics under certain conditions. In particular, longitudinal classroom evidence shows that gamified designs can reduce intrinsic motivation and satisfaction over time when competitive features amplify social comparison or when reward structures are perceived as controlling (Hanus & Fox, 2015). The qualitative reflections in the present study are consistent with this cautionary pathway: although many students experienced leaderboards and points as motivating, some reported stress and intensified comparison. The implication is that gamification should be designed to safeguard autonomy and psychological safety, for example by using mastery-oriented progress indicators, team-based challenges, and optional visibility of ranks, rather than relying heavily on public competition.

Methodologically, the present study demonstrates the value of a phase-based framework in small cohort classroom settings. By structuring evidence across pre, during, and post phases, researchers and instructors can differentiate between baseline readiness, mechanism activation during delivery, and post-implementation outcomes, thereby improving interpretability and supporting practical refinement decisions (Moore et al., 2015). Nevertheless, the study remains pilot-oriented. The small sample size and the absence of a comparison group limit causal inference and generalisability. Therefore, future work should extend the design to include quasi-experimental or experimental comparisons, larger and more diverse samples, and longer-term follow-up to assess sustainability beyond the immediate post-implementation window. These extensions are consistent with evaluation and implementation scholarship that emphasises building evidence iteratively, moving from feasibility and process-informed pilots to more robust designs capable of supporting stronger causal claims (Durlak & DuPre, 2008; Proctor et al., 2011).

In sum, the study supports a conditional conclusion: gamified classrooms in higher education can enhance engagement, motivational experiences, and learning performance when the intervention is implemented with adequate fidelity and designed to support positive classroom social dynamics. The most defensible interpretation is that the impact of gamification is not universal but depends on design quality and implementation conditions, which should be monitored systematically during the intervention rather than inferred only after it concludes (Moore et al., 2015; Durlak & DuPre, 2008).

REFERENCES

1. Almisad, H. M., & Aleidan, A. A. (2025). The impact of gamified classrooms on higher education students' engagement and motivation: A pre–during–post evaluation. Eurasia Journal of Mathematics, Science and Technology Education, 21(10), em2718. doi:10.29333/ejmste/17177.
2. Campbell, D. T., & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research. Rand McNally.
3. Creswell, J. W., & Plano Clark, V. L. (2018). Designing and conducting mixed methods research (3rd ed.). SAGE.
4. Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. Psychological Inquiry, 11(4), 227–268. doi:10.1207/S15327965PLI1104_01.
5. Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: Defining "gamification." In Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments (pp. 9–15). ACM. doi:10.1145/2181037.2181040.
6. Dichev, C., & Dicheva, D. (2017). Gamifying education: What is known, what is believed and what remains uncertain: A critical review. International Journal of Educational Technology in Higher Education, 14, Article 9. doi:10.1186/s41239-017-0042-5.
7. Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. American Journal of Community Psychology, 41(3–4), 327–350. doi:10.1007/s10464-008-9165-0.
8. Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. Proceedings of the National Academy of Sciences, 111(23), 8410–8415. doi:10.1073/pnas.1319030111.
9. Hamari, J., Koivisto, J., & Sarsa, H. (2014). Does gamification work? A literature review of empirical studies on gamification. In 2014 47th Hawaii International Conference on System Sciences (HICSS) (pp. 3025–3034). IEEE. doi:10.1109/HICSS.2014.377.
10. Handelsman, M. M., Briggs, W. L., Sullivan, N., & Towler, A. (2005). A measure of college student course engagement. The Journal of Educational Research, 98(3), 184–192. doi:10.3200/JOER.98.3.184-192.
11. 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. doi:10.1016/j.compedu.2014.08.019.
12. Henry, J., Li, F., & Arnab, S. (2024). On the pre-perception of gamification and game-based learning in higher education students: A systematic mapping study. Simulation & Gaming, 55(6), 985–1010. doi:10.1177/10468781241271082.
13. Hertzog, M. A. (2008). Considerations in determining sample size for pilot studies. Research in Nursing & Health, 31(2), 180–191.
14. Julious, S. A. (2005). Sample size of 12 per group rule of thumb for a pilot study. Pharmaceutical Statistics, 4(4), 287–291. doi:10.1002/pst.185.
15. Khaldi, A., et al. (2023). Advances of gamification in digital higher education: A systematic literature review. Smart Learning Environments. doi:10.1186/s40561-023-00227-z.

16. Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. Social Psychological and Personality Science, 4(7), 863–872.
17. 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 and Development. doi:10.1007/s11423-023-10337-7.
18. Medical Research Council. (2015). Process evaluation of complex interventions. Medical Research Council.
19. Mekler, E. D., Brühlmann, F., Tuch, A. N., & Opwis, K. (2017). Towards understanding the effects of individual gamification elements on intrinsic motivation and performance. Computers in Human Behavior, 71, 525–534.
20. Moore, G. F., Audrey, S., Barker, M., Bond, L., Bonell, C., Hardeman, W., Moore, L., O'Cathain, A., Tinati, T., Wight, D., & Baird, J. (2015). Process evaluation of complex interventions: Medical Research Council guidance. BMJ, 350, h1258. doi:10.1136/bmj.h1258.
21. Proctor, E., Silmere, H., Raghavan, R., Hovmand, P., Aarons, G., Bunger, A., Griffey, R., & Hensley, M. (2011). Outcomes for implementation research: Conceptual distinctions, measurement challenges, and research agenda. Implementation Science, 6, Article 42. doi:10.1186/1748-5908-6-42.
22. Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. American Psychologist, 55(1), 68–78. doi:10.1037/0003-066X.55.1.68.
23. Sailer, M., & Homner, L. (2020). The gamification of learning: A meta-analysis. Educational Psychology Review, 32(1), 77–112. doi:10.1007/s10648-019-09498-w.
24. Seaborn, K., & Fels, D. I. (2015). Gamification in theory and action: A survey. International Journal of Human-Computer Studies, 74, 14–31. doi:10.1016/j.ijhcs.2014.09.006.
25. Subhash, S., & Cudney, E. A. (2018). Gamified learning in higher education: A systematic review of the literature. Computers in Human Behavior, 87, 192–206. doi:10.1016/j.chb.2018.05.028.
26. Toda, A. M., et al. (2019). A taxonomy of game elements for gamification in educational contexts: Proposal and evaluation. In 2019 IEEE 19th International Conference on Advanced Learning Technologies (ICALT). IEEE. doi:10.1109/ICALT.2019.00028.