

Evolution of Educational Gamification Design: From Points-Based Systems to AI-Enhanced Narrative Integration

Alisha Janowsky

Department of Psychology, University of Central Florida
Orlando, Florida, USA
e-mail: alisha.janowsky@ucf.edu

Martha Hubertz

Department of Psychology, University of Central Florida
Orlando, Florida, USA
e-mail: martha.hubertz@ucf.edu

Abstract— Educational gamification has evolved through three distinct phases: points-badges-leaderboards (PBL) dominance (pre-2017), narrative integration (2015-2024), and AI-enhanced design (2023-present). This paper asks: does the shift toward narrative and AI-enhanced design actually improve learning outcomes, and do those effects hold across integration methods? Drawing on a systematic review of 39 studies and empirical data from four course implementations (N = 635), we find that narrative gamification correlates with significant learning gains (+4-5%, $d = 0.27-0.32$) and 65-89% student approval. Critically, AI-generated narrative content proves most effective when invisibly embedded in course design (66-73% undetectable) rather than surfaced as explicit AI communication (33% undetectable). This evolution represents an evidence-based progression reducing technical barriers for educators while introducing new ethical obligations around transparency and oversight.

Keywords- Gamification; Educational design; Artificial Intelligence (AI); Narrative learning; Generative AI; Self-Determination Theory (SDT); Constructivism.

I. INTRODUCTION

Educational gamification, applying game design elements in learning contexts, has undergone substantial transformation since mainstream adoption. Early implementations (pre-2017) relied predominantly on Points, Badges, and Leaderboards (PBL) [1]. Research revealed limitations, prompting shifts toward narrative and storytelling [2][3], and generative AI tools (2023-2024) have since created new possibilities for narrative creation [4][5]. This paper traces this evolution through three phases, grounded in Constructivism [9], Self-Determination Theory (SDT) [10], and Engagement Theory [11].

The rest of this paper is organized as follows. Section II presents the three-phase evolution overview. Section III describes Phase I, Section IV addresses Phase II, and Section V, Phase III. In Section VI, we describe the implications for practice. We conclude the article in Section VII.

II. THREE PHASE EVOLUTION OVERVIEW

Table I summarizes the three phases, distinguishing literature-derived evidence from the authors' own implementation findings.

TABLE I. THREE-PHASE GAMIFICATION EVOLUTION

Phase	Characteristics & Limitation	Evidence Source
Phase I: PBL (pre-2017)	Structural overlay; technical simplicity. Limitation: inconsistent outcomes; "pointsification".	Literature [1][2][6]
Phase II: Narrative (2015-2024)	Narrative transforms content; SDT-aligned. Limitation: high implementation burden.	Lit. [1][7] Authors: N=448
Phase III: AI-Enhanced (2023-present)	AI creates narrative assets; reduces barriers. Limitation: transparency, ethics, oversight.	Lit. [4][5] Authors: N=635

III. PHASE I: PBL DOMINANCE (PRE-2017)

A. Implementation Advantages

PBL elements offered technical simplicity, direct parallels to traditional assessment, and minimal redesign requirements, enabling rapid adoption [1]. From an SDT perspective [10], PBL addresses competence through achievement signals but neglects autonomy and relatedness — the needs most associated with sustained intrinsic motivation.

B. Research Limitations

Research revealed inconsistent learning outcomes. Balci et al. [6] conducted controlled experiments ($n = 102$; $n = 88$) finding badges and leaderboards did not improve academic performance despite increased reported motivation — consistent with constructivist critiques of passive structural overlays [9]. By 2015-2017, critique of "pointsification" identified core problems: points losing meaning,

leaderboards creating harmful competition, and rewards disconnected from learning [2].

IV. PHASE II: NARRATIVE INTEGRATION (2015-2024)

A. Theoretical Distinction and Literature Evidence

Research distinguishes structural gamification (PBL overlay on unchanged content) from content gamification (narrative transforms content itself), grounded in Constructivist Theory [9]. Khaldi et al. [1] reviewed 39 studies, finding 20 (51%) incorporated narrative elements. Jarrah et al. [7] examined 500 students and found narrative variables the strongest predictor of skill acquisition, followed by gamified engagement. Narrative design aligns with SDT [10] by providing meaningful context (relatedness) and learner choice (autonomy), and with Engagement Theory [11] through collaborative, project-based frameworks.

B. Authors' Implementation Data

The following findings derive from the authors' implementations, not the literature review. Narrative was operationalized through thematic framing: all course materials, assessments, and assignments were reframed within a coherent storyline (Mario Party-themed), with structured narrative progression across the semester and role-based participation elements. A quasi-experimental comparison of Fall 2023 versus Fall 2024 sections of the same undergraduate psychology course (same instructor, institution, and assessments; $N = 448$ across three sections) introduced this narrative gamification while holding all other variables constant. Results showed significant improvements across three standardized assessment domains: math performance +4.3% ($t(378) = 2.87, p = .004, d = 0.29$), theory understanding +3.8% ($t(378) = 2.63, p = .009, d = 0.27$), and lab skills +4.2% ($t(378) = 2.75, p = .006, d = 0.28$). Each domain was assessed by a standardized instrument tied to course content delivered under the Mario Party narrative: math performance via weekly problem sets framed as mini-game challenges, theory understanding via scenario-based quizzes embedded in the storyline, and lab skills via practical exercises presented as in-game tasks. These gains are promising but should be interpreted cautiously: as a quasi-experimental design, novelty effects, instructor enthusiasm, or unmeasured cohort differences cannot be fully ruled out. Visual gamification elements received 65-89% student approval; 76% reported reduced anxiety and 75% found assignments more enjoyable, consistent with SDT's relatedness and competence dimensions [10].

V. PHASE III: AI-ENHANCED DESIGN (2023-PRESENT)

A. Addressing Technical Barriers — Literature Evidence

Narrative implementation remained technically demanding. Wei et al. [4] conducted a 20-week experiment ($n = 60$) comparing AI-assisted digital storytelling (ChatGPT, Midjourney, Runway) versus conventional approaches; the AI group showed significantly enhanced collaborative problem-solving and creativity. De Vicente-

Yagüe-Jara et al. [5] found ChatGPT improved creative writing fluency, flexibility, and originality in 193 students, while emphasizing AI functions best as collaborative assistant, not replacement for human judgment.

B. Integration Method Matters — Authors' Implementation Data

The following findings derive from the authors' four course implementations ($N = 635$). A direct comparison examined two AI integration strategies: (1) embedded AI, in which AI-generated content formed invisible narrative infrastructure (Mario Party/Pokemon themes woven into course materials), and (2) explicit AI, in which AI-generated text appeared as direct course communications. The results are summarized in Table II.

TABLE II. AI INTEGRATION METHOD COMPARISON (Authors' Data, $N = 635$)

Outcome Measure	Embedded AI (n=448)	Explicit AI (n=187)
Undetectable as AI	66-73%	33%
Added educational value	49-72%	32-44%
Natural and engaging	61-72%	Not assessed

Note. N/A = not assessed in explicit AI condition.

VI. IMPLICATIONS FOR PRACTICE

A. Evidence-Based Design

Educators should prioritize narrative-integrated gamification over PBL-only approaches. Visual gamification elements receive 65-89% approval with learning gains of $d = 0.27-0.32$, small-to-medium effect sizes representing meaningful educational improvement.

B. Strategic AI Integration

Employ AI as invisible narrative infrastructure rather than an explicit course "voice." Embedded AI significantly outperforms explicit AI on detectability (66-73% vs. 33%) and perceived value (49-72% vs. 32-44%). Theme selection should map deliberately to learning objectives: for example, Mario Party's mini-game structure maps naturally to distinct skill domains (math, theory, lab), enabling themed assessments that reinforce specific competencies rather than applying a generic overlay.

C. Transparency, Ethics, and Human Oversight

The effectiveness of invisible AI integration creates an inherent ethical tension: students consistently express concerns about accuracy and over-reliance [8], and agency requires some awareness of the tools shaping their learning environment. We propose a *transparency with boundaries* framework that preserves engagement benefits while respecting student agency: (1) **disclose at the syllabus level** that AI tools support course design, without foregrounding specific implementations; (2) **maintain explicit human**

control over all pedagogical objectives, assessment criteria, and quality assurance — educators review and curate all AI-generated content before deployment; (3) **model critical AI evaluation** as a transferable skill by periodically inviting students to assess AI-generated content for accuracy and bias; and (4) **set clear boundaries** distinguishing instructor-authored content from AI-assisted infrastructure. This framework acknowledges AI's role without obscuring it entirely, reducing the risk of over-reliance while preserving the engagement advantages of embedded narrative.

D. Limitations

Several limitations should be noted. First, all implementations were quasi-experimental rather than randomized; novelty effects, instructor enthusiasm, and unmeasured cohort differences cannot be excluded. Second, all data derive from a single institution and discipline (undergraduate psychology, UCF), limiting generalizability. Third, student approval ratings are self-reported and subject to social desirability bias. Fourth, the URM classification in student samples aggregates groups with potentially different responses to gamification elements. Future work should employ randomized designs, validated engagement instruments, and multi-institution samples.

VII. CONCLUSION

Educational gamification has evolved from PBL systems through narrative integration to AI-enhanced design. Narrative gamification produces measurable learning gains (4-5%, $p < .01$, $d = 0.27-0.32$) and strong student approval (65-89%), grounded in Constructivism [9], SDT [10], and Engagement Theory [11]. AI tools enable sophisticated narrative creation at scale, but effectiveness depends critically on method: embedded narratives substantially outperform explicit AI communications. Provided educators maintain transparent human oversight, this evolution creates unprecedented opportunity to implement research-validated narrative approaches without technical barriers.

REFERENCES

- [1] A. Khaldi, R. Bouzidi, and F. Nader, "Gamification of e-learning in higher education: a systematic literature review," *Smart Learning Environments*, vol. 10, Article 10, 2023.
- [2] C. Dichev and D. Dicheva, "Gamifying education: what is known, what is believed and what remains uncertain," *International Journal of Educational Technology in Higher Education*, vol. 14, Article 9, 2017.
- [3] S. Deterding, D. Dixon, R. Khaled, and L. Nacke, "From game design elements to gamefulness: defining gamification," *Proc. 15th Int. Academic MindTrek Conf.*, pp. 9-15, 2011.
- [4] X. Wei, L. Wang, L. K. Lee, and R. Liu, "The effects of generative AI on collaborative problem-solving and team creativity performance in digital story creation," *International Journal of Educational Technology in Higher Education*, vol. 22, Article 23, 2025.
- [5] M. I. de Vicente-Yagüe-Jara, O. López-Martínez, V. Navarro-Navarro, and F. Cuéllar-Santiago, "Writing, creativity, and artificial intelligence: ChatGPT in the university context," *Comunicar*, vol. 31, no. 77, pp. 45-54, 2023.
- [6] S. Balci, J. M. Secaur, and B. J. Morris, "Comparing the effectiveness of badges and leaderboards on academic performance," *Education and Information Technologies*, vol. 27, no. 6, pp. 8669-8704, 2022.
- [7] H. Y. Jarrah et al., "The impact of storytelling and narrative variables on skill acquisition in gamified learning," *International Journal of Data and Network Science*, vol. 8, no. 2, pp. 1161-1168, 2024.
- [8] C. K. Y. Chan and W. Hu, "Students' voices on generative AI: perceptions, benefits, and challenges in higher education," *International Journal of Educational Technology in Higher Education*, vol. 20, Article 43, 2023.
- [9] J. Piaget, *The Psychology of Intelligence*. London: Routledge, 1950.
- [10] R. M. Ryan and E. L. Deci, "Self-Determination Theory and the facilitation of intrinsic motivation, social development, and well-being," *American Psychologist*, vol. 55, no. 1, pp. 68-78, 2000.
- [11] G. Kearsley and B. Shneiderman, "Engagement theory: a framework for technology-based teaching and learning," *Educational Technology*, vol. 38, no. 5, pp. 20-23, 1998.