European Journal of Education Studies
ISSN: 2501 - 1111
ISSN-L: 2501 - 1111
Available online at: www.oapub.org/edu

04DOI: 10.46827/ejes.v13i9.6958
Volume 13 | Issue 9 | 2026

FALSIFIABILITY AND BEHAVIORAL MODELING: PROPOSAL FOR A SCIENTIFIC FRAMEWORK FOR GAMIFICATION DESIGN

Cláudio Eduardo Marques Lacerda[1],
Silvio Luiz Rutz da Silva[2i]
[1]PPG em Ensino de Ciências e Educação Matemática – UEPG,
Ponta Grossa, PR, Brasil
orcid.org/0009-0003-4557-207X
[2]Departamento de Física,
PPG em Ensino de Ciências e Educação Matemática – UEPG,
Ponta Grossa, PR, Brasil
orcid.org/0000-0003-1859-9018

Abstract:
This literature review examines the intersection of falsifiability principles and behavioral modeling approaches in designing gamified processes, synthesizing findings from 271 peer-reviewed publications. This study addresses a critical gap in contemporary research: the need for scientifically rigorous methodologies that go beyond superficial Pointsification to develop effective behavior-change interventions. The proposed framework integrates Karl Popper's epistemology with B. F. Skinner's operant conditioning and is structured across four levels of integration: epistemological, theoretical, methodological, and practical. Theoretically, this study articulates Self-Determination Theory, Flow Theory, and Behavioral Economics under the lens of empirical testability. Methodologically, intellectual rigor is operationalized through a three-tiered scoring system (0, 5, and 10 points) that prioritizes "resistance to refutation" over impulsive speed, thereby transforming the assessment into a continuous scientific validation process. The practical application is consolidated into a replicable four-stage didactic sequence: conceptualization, structural mapping, mechanics design, and behavioral calibration. The main conclusions indicate that the success of gamified design requires (1) explicit theoretical grounding that allows the generation of falsifiable hypotheses, (2) behavioral models that consider individual variations and temporal dynamics, and (3) multidimensional metrics that integrate behavioral, psychological, and system data. The study concludes that integrating these principles consolidates gamification as a rigorous design engineering discipline, capable of generating predictable results and enabling deep learning. This study offers an essential methodological roadmap for the field's maturation and for moving beyond intuition-based practices.

[2] Correspondence: email rutz@uepg.br

Keywords: gamification; falsifiability; behavioral modeling; educational psychology; instructional design

1. Introduction

Gamification, defined as applying game design elements and principles in contexts unrelated to games, has emerged as a prominent strategy for increasing engagement, motivation, and behavior change in education, health, organizational management, and civic participation (Deterding *et al.*, 2013). Since its popularization in the early 2010s, gamification has evolved from a predominantly industrial phenomenon marked by the superficial implementation of points, badges, and leaderboards to a more theoretically grounded and empirically validated field of research and practice.

However, significant challenges remain regarding the scientific rigor, theoretical coherence, and predictability of gamified interventions. Many implementations fail to achieve intended outcomes, and some studies report null or negative effects on motivation and performance (Dichev & Dicheva, 2017). This inconsistency highlights the lack of falsifiable theoretical frameworks and validated behavioral models to guide the design decisions and predict outcomes.

Integrating the principles of falsifiability (Popper, 1959) with behavioral modeling offers a promising path toward more rigorous and effective design. Falsifiability requires that theoretical propositions be formulated to allow for empirical testing and potential refutation. Applied to gamification, falsifiability demands that design decisions be based on explicit, testable hypotheses about user behavior (Lorig *et al.*, 2017) and that outcomes be measured objectively, with evaluation criteria that distinguish successful interventions from unsuccessful ones (Levy *et al.*, 2018; Mora *et al.*, 2017). Complementarily, behavioral modeling provides the necessary mechanisms to understand, predict, and influence user behavior in gamified systems through motivation, ability, and stimulus variables (Grunburg, 2025; Skinner, 1953).

Given this scenario, this systematic review synthesizes evidence at the intersection of the epistemology of falsifiability and behavioral modeling to propose a transition from intuitive gamification to rigorous design engineering. Based on an analysis of 271 publications, this study clarifies the theoretical frameworks underpinning behavioral prediction, assesses the practical application of Popperian rigor in outcome measurement, and ultimately establishes a research agenda that prioritizes hypothesis testability and construct validity as foundational pillars for the discipline's maturation. Integrating falsifiability with behavioral modeling has profound implications.

(I) Theoretical: Addresses the current proliferation of descriptive frameworks that lack predictive power and empirical testability (Mora *et al.*, 2017), requiring theoretical commitments that enable the cumulative development of knowledge (Lorig *et al.*, 2017).

(II) Methodological: establishes standards that differentiate the science of gamification from mere advocacy of the practice (Dichev & Dicheva, 2017), requiring that theory specify the conditions that would constitute evidence against itself (Popper, 1959), and that the measures used possess strict scientific validity (Levy *et al.*, 2018). Behavioral modeling provides tools to operationalize theoretical constructs and account for individual and contextual variation through unified models of behavior (Grunburg, 2025) and personalization algorithms (Khoshkangini *et al.*, 2021).

(III) Practice: Designers equipped with the theory of falsifiability can make evidence-based decisions regarding which elements to include and how to configure them for specific populations (Schöbel *et al.*, 2018; Challco *et al.*, 2023). This scientific approach contrasts with the 'best practices' approach that dominates the industry, which often relies on anecdotal accounts and lacks rigor in outcome evaluation (Mora *et al.*, 2017; Dichev & Dicheva, 2017).

A compelling, practical, and educational application, for example (Lacerda, 2025a, 2025b), transforms the traditional game 'Stop' into a scoring system based on falsifiability (Popper, 1959). The metric distinguishes: (a) falsified answers (0 points), generic or based on common sense; (b) partial answers (5 points), vague or incomplete; and (c) unfalsified answers (10 points), which demonstrate conceptual quality and resistance to refutation. This system operationalizes scientific epistemology by directly influencing student behavior through differential reinforcement contingencies (Skinner, 1953), serving as a promising methodological hypothesis for operationalizing construct validity (Levy *et al.*, 2018) by rewarding intellectual rigor rather than impulsive speed, and remains open to future empirical validation.

2. Literature Review

The construction of a scientifically rigorous gamification framework requires a foundation that connects the philosophy of science with human behavior psychology (Mora *et al.*, 2017). This section presents the conceptual pillars that support the transition from intuitive design to evidence-based engineering, focusing on Karl Popper's criterion of falsifiability as the demarcation between descriptive practice and scientific investigation (Popper, 1959). Additionally, it explores leading motivation theories and behavioral models, which provide the necessary apparatus for predicting and shaping outcomes (Skinner, 1953; Grunburg, 2025). Integrating these domains establishes the basis for gamified design to move beyond an ad hoc intervention and become a process guided by testable, validatable hypotheses (Lorig *et al.*, 2017).

The pillar of active experience, grounded in Dewey (1979), serves as a basis for ensuring that playful engagement is not merely passive, but an immersion that generates knowledge through experimental practice. This educational perspective complements scientific rigor, ensuring that gamification not only controls behavior but also fosters meaningful learning through action and reflection.

2.1 Falsifiability in Scientific Design

Falsifiability, a concept introduced by the philosopher Karl Popper (1959), serves as the fundamental criterion that distinguishes science from pseudoscience. In gamification design, this principle requires that an intervention not be merely 'attractive' or '' engaging', thus overcoming the limitations of purely prescriptive frameworks (Mora *et al.*, 2017). Instead, the design must be grounded in theories that generate predictions falsifiable through empirical observation (Popper, 1959), using a rigorous, hypothesis-driven approach that clearly specifies the expected causal mechanisms (Lorig *et al.*, 2017).

Scientific design based on falsifiability transforms gamification from an "art based on intuition and creativity" into behavioral engineering. This means that every design decision (e.g., choosing a leveling or reward system) must be justified by an explicit theoretical commitment. If the expected behavioral outcome does not occur, the theory is refuted or requires revision, thereby promoting the cumulative development of knowledge (Lorig *et al.*, 2017).

One of the most practical contributions to scientific design is transforming traditional assessment into a process of scientific validation, shifting the focus from the 'correct answer' to 'error resistance' (Popper, 1959). This logic is operationalized through a falsifiability table based on contingencies of differential reinforcement (Skinner, 1953), which defines the game mechanics on three levels: (i) Falsified (0 points), penalizing logical failure; (ii) Partial/Ambiguous (5 points), identifying vulnerable knowledge; and (iii) Not Falsified (10 points), rewarding conceptual quality that resists refutation. Table 1 synthesizes these criteria and their epistemological foundations.

Table 1: Evaluation and Scoring Matrix Based on Falsifiability

Response Level	Score	Epistemological Description (Popper)	Behavioral Characteristics (Skinner)	Validation Criterion
Counterfeit	**0**	Generic answer based on common sense or with an evident logical flaw.	**Impulsive Behavior:** Guesswork or lack of a technical foundation.	**Refuted:** Does not withstand any validity tests.
Partial or Ambiguous	**5**	Meets the basic criteria but remains vague, incomplete, or imprecise.	**Vulnerable Knowledge:** Unsystematic use of terminology or concepts.	**Inconclusive:** Partial resistance to refutation requires refinement**.**
Does-Not-Falsify	**10**	Judicious use of knowledge with a high conceptual quality.	**Rigorous Thinking:** Reflective Analysis and Technical Terminological Precision.	**Validated:** Withstands rigorous attempts at refutation (domain).

Note: Authors' elaboration is based on Popper (1959) and Skinner (1953).

While this structure provides a theoretical basis for operationalizing construct validity in scientific measurement (Levy *et al.*, 2018), it should be regarded as a promising methodological hypothesis rather than an already validated assessment instrument. From a scientific design perspective, the game shifts from a motivational resource to a measurement instrument. For a gamified system to be scientifically valid, it must meet the same standards as laboratory instruments (Levy *et al.*, 2018): (i) Construct Validity: measuring exactly what it is intended to measure (e.g., intellectual rigor rather than typing speed). (ii) Reliability: Producing consistent results across different applications. (iii) Sensitivity: The ability to detect significant differences in performance among users.

When applying falsifiability, success in the game becomes inseparable from the unquestionable mastery of the content. Knowledge is validated only when it withstands attempts at refutation (Popper, 1959). This eliminates 'getting it right by chance' or by 'intuition,' transforming design into a science grounded in testable hypotheses (Lorig *et al.*, 2017). Consequently, reinforcement contingencies force users to slow down impulsive processes and engage in reflective analysis, which Popperian epistemology defines as operating directly on behavior (Skinner, 1953).

2.2 Main Theoretical Approaches in Gamification

Contemporary research has identified three main theoretical frameworks that offer complementary perspectives on how gamification influences human behavior and decision-making (Mora *et al.*, 2017):

1) **Self-Determination Theory (SDT):** Formulated by Deci and Ryan (1985), SDT posits that human motivation exists on a continuum and identifies three basic psychological needs: autonomy (sense of choice), competence (sense of mastery), and relatedness (sense of connection), which, when satisfied, sustain intrinsic motivation. SDT generates falsifiable predictions about which game elements enhance or undermine motivation, thereby enabling predictions of intervention success (Schöbel *et al.*, 2018).
2) **Flow Theory,** proposed by Csikszentmihalyi (1990), describes optimal experience as a state of complete absorption in an activity. Flow occurs when challenges and the user's skills are balanced. Design based on this theory seeks to keep the user in a constant zone of engagement, avoiding both anxiety and boredom (Challco *et al.*, 2023).
3) **Behavioral Economics:** This perspective examines how game mechanics exploit documented cognitive biases and heuristics, such as loss aversion, temporal discounting, social proof, and choice architecture (Butler, 2014). While it offers powerful explanations for the psychological power of certain design patterns, this approach raises crucial ethical questions regarding behavioral manipulation (Butler, 2014; McGloin *et al.*, 2018).

Integrating these three theoretical frameworks (Figure 1) enables gamification design to move beyond intuition. While SDT underpins long-term motivation (Deci & Ryan, 1985), Flow Theory optimizes the intensity of present engagement (Csikszentmihalyi, 1990), and Behavioral Economics explains immediate reactions to choice stimuli (Butler, 2014). As shown in the matrix (Figure 1), each framework not only describes phenomena but also generates specific, empirically testable predictions, meeting the falsifiability criteria proposed by Popper (1959).

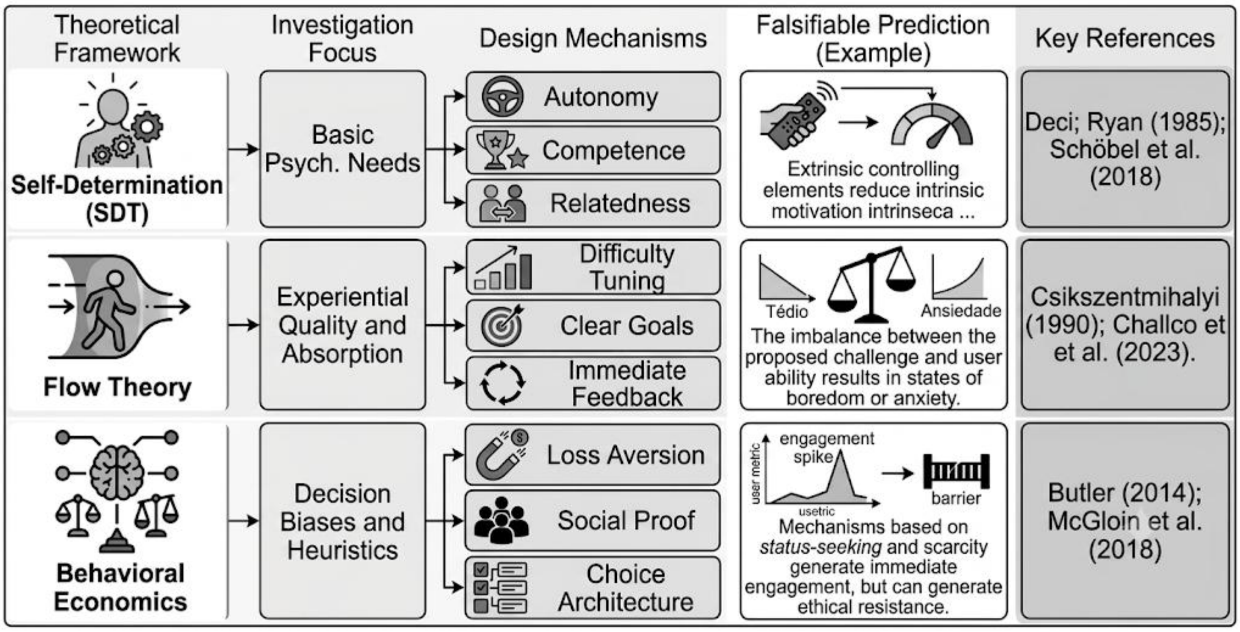

Figure 1: Theoretical Integration Matrix in Scientific Design

These theoretical frameworks make it possible to transform gamification from an intuitive art into an engineering discipline that is based on testable hypotheses. Each framework provides a conceptual apparatus for formulating precise predictions of user behavior in different system configurations, enabling the scientific validation of interventions (Lorig *et al.*, 2017; Mora *et al.*, 2017).

2.3 Behavioral Theories and Models Applied to Gamification

Beyond motivational theories, gamification design draws on behavioral models that explain and predict user actions through contingencies and cognitive processes.

Operant Conditioning (Skinner, 1953) provides principles for understanding how consequences shape behavior. In scientific gamification, differential reinforcement is used to carry out "shaping," reinforcing successive approximations to the target behavior. For example, educational gamification is cited, with scoring systems recalibrated to reward technical quality rather than speed, creating contingencies in which rigorous thinking is more advantageous than impulsive, superficial responses (Skinner, 1953).

Social Cognitive Theory (Bandura, 1986) highlights reciprocal determinism among personal factors, environment, and behavior, positioning self-efficacy, the belief in one's own ability to perform tasks, as the central mediator of engagement. Gamification elements, such as progressive challenges and mastery experiences, are designed to increase self-efficacy, which generates falsifiable predictions: interventions that elevate this belief must necessarily result in improved performance and sustained persistence (Bandura, 1986).

These models explain how gamification can support long-term behavior change. An effective strategy should shift the user from extrinsic motivation (behavior driven by goals and outcome expectations) to automatic habits triggered by contextual cues. This requires the careful design of "cue-behavior-reward" cycles and the gradual fading of explicit rewards as the behavior becomes more intrinsic and habitual.

Synthesizing Popperian falsifiability (Popper, 1959) and Skinnerian modeling (Skinner, 1953) creates a powerful framework for behavior change. In the scoring system described (Table 1), falsifiability provides the quality criterion (resistance to refutation), while operant conditioning ensures the game mechanics reinforce only rigorous behaviors.

By rewarding only "Non-Falsified" responses with the highest score, the system induces a "reflective slowdown," making technical rigor the sole path to victory. This integration turns assessment into an irrefutable test of mastery, ensuring that success in the game stems directly from internalizing high intellectual standards (Popper, 1959; Skinner, 1953).

2.4 Integration of Falsifiability and Behavioral Modeling

Integrating falsifiability principles into behavioral modeling creates a comprehensive framework that transforms gamification design from an "intuitive art" into "systematic engineering" (Mora *et al.*, 2017). This integration operates synergistically at four fundamental levels (Popper, 1959; Skinner, 1953):

1) **Epistemological Level:** establishes the criteria for scientific knowledge claims, requiring gamification theories to generate testable predictions subject to empirical falsification (Popper, 1959)
2) **Theoretical Level:** This requires the explicit specification of causal mechanisms and mediating pathways, describing how the elements of the game influence psychological processes and, consequently, behavioral outcomes (Mora *et al.*, 2017; Lorig *et al.*, 2017).
3) **Methodological Level:** guides research design and evaluation, prescribing the use of pre-specified hypotheses and objective metrics that allow effective interventions to be distinguished from design failures (Lorig *et al.*, 2017; Levy *et al.*, 2018).
4) **Practical Level:** informs design decisions, enabling professionals to select and configure game elements based on evidence and their ability to shape the desired behavior (Schöbel *et al.*, 2018).

This integration can be operationalized through a four-stage didactic sequence designed for the teaching and implementation of intellectually rigorous gamification (Popper, 1959; Skinner, 1953):

1) **Stage 1 (Conceptualization and Intentionality):** Focuses on the distinction between products (games) and processes (gamification), which requires the definition of clear goals and key performance indicators (KPIs).
2) **Stage 2 (Structural Mapping and Theoretical Pillars):** Analyzes traditional mechanics in relation to the pillars of active experience (Dewey, 1979; Kapp, 2012) and Popperian falsifiability, shifting the focus from the "correct answer" to "resistance to error";

3) **Stage 3 (Mechanics, Design, and Adaptability):** Applies reflection-in-action to create new rules that demonstrate applicability in different domains.
4) **Stage 4 (Rules, Reinforcement, and Synthesis):** Uses Skinnerian modeling to carry out behavioral calibration through the construction of falsifiability tables that establish the reinforcement contingencies necessary for the internalization of standards of intellectual rigor (Skinner, 1953).

This methodology exemplifies "theory-to-design translation," ensuring that each implemented game element directly follows from testable theoretical commitments (Schöbel *et al.*, 2018). Table 2 synthesizes this process and the objectives of each stage.

Table 2: Didactic Sequence for Scientific Gamification Design

Internship	Denomination	Focus	Theoretical Framework	Deliverables (KPIs/Outputs)
1	**Conceptualization**	Differentiate games (products) from gamification (processes).	Design Intentionality	Declaração de Missão e Definição de KPIs.
2	**Mapping**	Analysis of failures in traditional mechanics.	Dewey, Popper and Vygotsky	Identification of learning and rigor "Gaps
3	**Mechanics Design**	Creation of new rules and adaptability.	Schön (Reflection-in-action)	Prototype of rules applicable to new domains.
4	**Rules and Reinforcement**	Behavioral calibration and technical accuracy.	Skinner (Differential Reinforcement)	Falsifiability Tables and Playbooks.

Note: Authors' elaboration is based on Schöbel *et al.* (2018).

3. Methodological Procedures

This theoretical review uses a systematic approach to map and synthesize knowledge at the intersection of falsifiability and behavioral modeling in gamification design. The selection and analysis processes were structured to ensure replicability and rigor in constructing the proposed framework. The document corpus comprised 271 peer-reviewed publications identified through systematic search strategies across multiple academic databases (including Scopus, Web of Science, IEEE Xplore, and Google Scholar). The search encompassed a combination of Boolean terms related to "gamification," "behavioral modeling," and "falsifiability/popper."

The selected documents included journal articles, proceedings from high-impact international conferences, and foundational theoretical treatises in psychology, philosophy of science, and human-computer interaction.

The selection focused on works that explicitly addressed (i) fundamental theoretical frameworks, (ii) models of human behavior, (iii) design methodologies, (iv) verification and validation approaches, or (v) empirical evaluations of gamified systems.

We prioritized educational and organizational contexts for application domains. This choice is justified by the marked tension observed in these environments between user engagement (the ludic dimension) and effective learning results or performance (the technical dimension). Nonetheless, we included relevant findings from health, civic participation, and sustainability, provided these studies contributed significantly to the theoretical understanding or methodological innovation of the framework proposed here.

We conducted a categorical content analysis of data extracted from the 271 sources to identify effectiveness patterns, methodological gaps, and behavior change mechanisms. This synthesis enabled the transition from the state of the art to the practical implementation roadmap proposed in the subsequent sections of this article.

4. Results and Discussion

4.1 Behavioral Modeling Approaches in Gamification

Behavioral modeling in gamification provides the technical and conceptual apparatus needed to operationalize theoretical constructs as measurable, predictable variables (Grunburg, 2025). While theoretical foundations provide an explanatory basis for interventions, modeling focuses on specific mechanisms to understand, predict, and influence user behavior through defined causal pathways (Lorig *et al.*, 2017; Mora *et al.*, 2017).

This section explores modeling approaches in the field, ranging from qualitative frameworks based on psychological needs to adaptive computational systems grounded in machine learning and multi-agent systems (Khoshkangini *et al.*, 2021; Belliato *et al.*, 2025). By integrating principles from operant conditioning and social cognitive theory, modeling enables designers to specify testable behavioral trajectories, turning the gamified experience into a dynamic, scientifically validated system (Skinner, 1953; Bandura, 1986).

Self-Determination Theory (SDT), originally formulated by Deci and Ryan (1985), is the most widely adopted theoretical framework for modeling motivation in gamified systems. The theory posits that human motivation exists on a continuum ranging from amotivation through extrinsic motivation to intrinsic motivation, which involves engaging in activities for the inherent satisfaction they provide.

SDT makes falsifiable and testable predictions by identifying three basic psychological needs that, when satisfied, foster intrinsic motivation: (i) Autonomy: the sense of volition, choice, and self-direction. (ii) Competence: A sense of efficacy, mastery, and appropriate challenge. (iii) Relatedness: A sense of connection and belonging with other users.

In design, autonomy is supported by options for choice and personalization; competence by progressive challenges and constructive feedback; and relatedness by social resources and collaborative mechanics (Schöbel *et al.*, 2018). Interventions that address these needs tend to increase sustained engagement (Silic & Lowry, 2020).

A critical design flaw occurs when gamification shifts the focus from the activity itself to isolated external rewards, generating the so-called overjustification effect. In these cases, extrinsic incentives can undermine pre-existing intrinsic motivation, especially in inherently interesting tasks such as this one. Thus, SDT-based predictions allow designers to avoid mechanics that reduce perceived competence or feel manipulative (Khoshkangini *et al.*, 2021).

Flow Theory, introduced by Csikszentmihalyi (1990), describes the ideal experience as a state of total absorption in an activity characterized by intense focus, loss of self-consciousness, and intrinsic enjoyment. In the context of gamification, this theory is fundamental for understanding the quality of the engagement experience and predicting user behavior when challenged.

The theory predicts that flow occurs when perceived challenges and user skills are both high and perfectly balanced. The disruption of this homeostasis produces predictable and testable outcomes: (i) Anxiety arises when challenges exceed an individual's competencies. (ii) Boredom: This occurs when the user's skills surpass the proposed challenges.

Flow-based gamification that maintains this balance produces greater engagement and performance. However, apply the theory with caution, as flow can occur in superficial activities that do not necessarily lead to effective learning, raising questions about when the experience is truly beneficial versus merely enjoyable.

The GamiFlow framework proposed by Challco *et al.* (2023) exemplifies the practical application of these principles. This structure operationalizes flow in learning scenarios by means of (i) adaptive adjustment of difficulty: the use of computational models to estimate, in real time, the learner's skill level; and (ii) clear goals and immediate feedback: essential elements to maintain focus and immersion.

The falsifiability of Flow Theory depends on the precise operationalization of these states. While questionnaires (such as the Flow Short Scale) measure subjective experience, objective indicators such as time spent on the task, performance metrics, and physiological measurements provide evidence to validate or refute the design's effectiveness (Challco *et al.*, 2023; Levy *et al.*, 2018).

Computational approaches complement theoretical frameworks by employing machine learning, data mining, and statistical modeling to predict user behavior, personalize experiences, and optimize system parameters in real time. These techniques enable data-driven discoveries of behavioral patterns and relationships that existing qualitative theories may not capture.

Khoshkangini *et al.* (2021) proposed a system for automatically generating and recommending personalized challenges. This approach uses machine learning to model individual characteristics, such as performance history, learning styles, and motivational profiles, to predict which challenges will be most effective for each user. By dynamically balancing task difficulty with demonstrated skill level, the system promotes the maintenance of the flow state and the satisfaction of the need for competence (SDT), adapting to temporal changes in users' states (Khoshkangini *et al.*, 2021).

In social and organizational contexts, Belliato *et al.* (2025) presented a formal, multi-agent systems approach to align individual motivation with collective objectives. The framework uses modeling techniques to represent motivational states, group goals, and game mechanics that mediate this relationship. This formal specification generates falsifiable predictions about how individual behaviors aggregate into collective outcomes, enabling design optimization to achieve organizational goals without sacrificing participants' personal engagement (Belliato *et al.*, 2025).

These innovations demonstrate how computational modeling transforms gamification into an adaptive system that responds in real time to user actions, thereby significantly increasing the theoretical accuracy and effectiveness of interventions (Khoshkangini *et al.*, 2021; Belliato *et al.*, 2025).

Personalization is based on the premise that individuals substantially differ in their responses to game elements. Two main approaches to system adaptation are distinguished (Khoshkangini *et al.*, 2021):

1) **Explicit Personalization (User Modeling):** This uses computational models and machine learning to identify dimensions of individual differences (such as performance history and motivational profile) and generate challenges adapted to the user's current state (Khoshkangini *et al.*, 2021).
2) **Implicit Personalization (Behavioral Contingencies):** exemplified by the integration of falsifiability and behavioral modeling. The system naturally adapts to everyone's starting point, creating contingencies that reinforce their rigorous thinking. Impulsive students receive consistent feedback (0 points), which directs their behavior toward reflective approaches, while those who already demonstrate rigor receive reinforcement that maintains this pattern. This allows for the testing of general theoretical principles while accommodating individual variability (Schöbel *et al.*, 2018).

Understanding the causal pathways through which gamification influences users is essential for scientific design, with four main mechanisms:

1) **Immediate Reinforcement:** Based on operant conditioning (Skinner, 1953), this reinforcement uses points, feedback, and progression mechanics to provide immediate consequences that shape users' responses. Effectiveness depends on contingency (consistency), immediacy, and magnitude of the consequence.
2) **Cognitive Mechanisms:** Gamification directs attention and effort through clear goals (missions), self-monitoring (progress bars), and information processing. Model Compatibility Theory (McGloin *et al.*, 2018) explains how the alignment between game mechanics and users' mental models influences engagement and learning.
3) **Social Mechanisms:** Social influence is employed through rankings, team challenges, and social recognition. Formal models seek to align individual motivation with collective goals to avoid the risks of excessive competition or exclusion (Belliato *et al.*, 2025).

4) **Mechanisms of Identity and Self-Concept:** In the long term, repeated and socially validated engagement can lead the user to incorporate behaviors into their self-concept. The scoring system based on falsifiability demonstrates this by orchestrating reinforcements that shape the student's identity as a "rigorous thinker," internalizing standards of intellectual quality that transcend the game's context.

4.2 Principles of Falsifiability in Gamification Design

Applying falsifiability transforms gamification from an intuitive practice into a data-driven scientific endeavor. This process requires rigor from hypothesis formulation to final results analysis.

Scientific design begins with explicit hypotheses that specify the game elements, psychological mechanisms, and expected outcomes under defined boundary conditions (Lorig *et al.*, 2017).

Following the guidelines of Lorig *et al.* (2017), the formulation of a hypothesis must be sufficiently precise to allow for falsification. For example, consider the following proposition: progressive adjustment of difficulty will increase the average session duration by at least 20% compared with fixed difficulty for users with moderate skill. This hypothesis is scientifically valid because it clearly defines the intervention, the outcome metric, the effect size, and the target population, allowing immediate empirical refutation if the data do not confirm the prediction and thus meeting the Popperian criterion.

Falsifiability requires the translation of abstract concepts into observable and measurable variables through three types of metrics (Levy *et al.*, 2018; Mora *et al.*, 2017): (i) Behavioral Metrics: observable actions such as test performance (accuracy), time spent on the task, and completion rates. (ii) Psychological Metrics: measures of internal states, such as motivation (Intrinsic Motivation Inventory) and flow experiences (Flow Short Scale). (iii) System Metrics: detailed interaction data, such as click logs, response times, and navigation patterns.

The scoring system based on falsifiability (Table 1) operationalizes the construct of "intellectual rigor" into concrete categories (0, 5, 10 points), creating an objective metric that directly reflects the student's technical proficiency.

To ensure a comprehensive analysis, the GATUGU Framework (Swacha *et al.*, 2023) proposes six assessment perspectives: Objective (goals), Audience (users), Technology (implementation), User Experience (subjectivity), Game (mechanics), and Utility (practicality). Effective gamification should score highly across all perspectives, and shortcomings in any one can undermine the system's effectiveness.

A central challenge in gamification is the disconnect between engagement during gameplay and learning outcomes assessed afterward. Integrating assessment into game mechanics through a scoring system based on falsifiability addresses this problem by making assessment continuous and inseparable from the experience. This ensures immediate feedback for the student and provides a structured, testable mechanism to align in-game performance with genuine conceptual learning, subject to empirical verification (Levy *et al.*, 2018; Mora *et al.*, 2017.

4.3 Empirical and Ethical Evidence

Applying the principles of falsifiability and behavioral modeling in real-world contexts reveals a complex landscape of empirical evidence across domains such as education, organizational management, and health (Deterding *et al.*, 2013). An analysis of 271 publications shows that although gamification has strong potential to increase engagement and motivation, results are often heterogeneous, with studies reporting outcomes ranging from significant improvements to null or negative effects (Dichev; Dicheva, 2017; Mora *et al.*, 2017).

This section examines the evidence on the effectiveness of theory-based interventions, contrasting them with superficial 'pontification' approaches, and discusses the ethical implications of deliberately influencing behavior. By scrutinizing case studies and impact analyses, this study seeks to show that methodological transparency and respect for user autonomy are inseparable pillars of scientifically sound and socially responsible gamification (Silic; Lowry, 2020; Butler, 2014; Popper, 1959).

4.3.1 Educational Gamification — The Case of the Game “Stop”

As an example, a detailed case study shows how scientific gamification transforms the traditional game “Stop” (Lacerda, 2025a, 2025b). The intervention uses a scoring system based on falsifiability (Popper, 1959) that rewards intellectual rigor (i.e., resistance to refutation) rather than response speed.

This approach allows (i) overcoming the conceptual confusion between games as entertainment and gamification as design, (ii) establishing learning objectives with measurable KPIs, (iii) eliminating responses based on common sense, and (iv) adapting mechanics to different domains. Integrating operant conditioning (Skinner, 1953) ensures that differential reinforcement shapes student behavior, promoting a "reflective deceleration" essential for deep learning (Popper, 1959; Skinner, 1953).

4.3.2 Organizational and Health Contexts

In corporate environments, gamification is applied to security and compliance training. Silic and Lowry (2020) demonstrated that scientifically designed interventions increase engagement and retention, although the effects vary depending on employees’ prior experience with games and organizational culture.

In health, the focus is on sustained behavior change (e.g., medication adherence). Verschueren *et al.* (2020) proposed a framework for serious games that requires explicit theoretical grounding and rigorous clinical trials. Applying falsifiability in this domain demands that failures to produce clinical outcomes lead to refinement or abandonment of the intervention, rather than retrospective justification (Silic & Lowry, 2020; Verschueren *et al.*, 2020).

4.3.3 Manipulation vs. Autonomy

Behavioral economics (Butler, 2014) reveals that game mechanics can exploit cognitive biases (such as loss aversion), raising ethical dilemmas. This document argues that scientific gamification must respect user autonomy. The scoring system based on falsifiability exemplifies this ethic by making evaluation criteria transparent: students understand the rules and deliberately choose their intellectual effort, unlike manipulative systems that bypass rational deliberation (Butler, 2014; Popper, 1959).

4.3.4 Equity, Privacy and Long-Term Effects

Individual differences in personality traits and abilities can widen performance gaps if gamification is not tailored to the user. In the context of privacy, the collection of detailed data for personalization must be transparent and limited to educational or clinical purposes. Finally, longitudinal studies are needed, since initial effectiveness (novelty effect) should give way to value internalization and habit formation, ensuring that gamification promotes intrinsic motivation rather than ongoing dependence on external rewards (Dichev & Dicheva, 2017; Verschueren *et al.*, 2020).

5. Recommendations

5.1 Proposed Framework for Implementation and Validation

Based on the synthesis of the 271 publications analyzed and the theoretical connection between Popperian epistemology and behaviorist modeling, this section proposes an original framework for designing scientific gamification. The primary objective is to provide researchers and practitioners with a systematic methodology that goes beyond the intuitive nature of ad hoc interventions and the superficiality of points (Mora *et al.*, 2017). By integrating the rigor of hypothesis testability with the precision of behavioral prediction, the framework advocated here is organized into operational axes designed to ensure construct validity and enable causal inference.

Therefore, the aim is to consolidate the transition from gamification toward evidence-driven design engineering, in which each game mechanic is treated as a scientific proposition subject to falsification (Lorig *et al.*, 2017; Popper, 1959). From this perspective, the proposed methodological framework rests on a paradigmatic shift: abandoning purely intuitive practice in favor of a scientifically rigorous foundation. In this regard, randomized controlled trials (RCTs) and factorial designs are justified by the need to establish causal inferences and isolate the effects of multiple elements independently (Lorig *et al.*, 2017).

This methodological rigor requires samples with sufficient statistical power to detect complex interactions between mechanisms, overcoming the intrinsic limitations of purely exploratory case studies (Dichev; Dicheva, 2017; Mora *et al.*, 2017). To validate this approach, the selected application domains, education, organizational management, and health, corroborate the evidence regarding the heterogeneity of intervention effects (Silic

& Lowry, 2020) and align with contemporary guidelines for game development (Verschueren *et al.*, 2020).

At the core of this model, the experimental condition uses a three-level scoring system (0, 5, 10), operationalizing Popperian falsifiability (Popper, 1959) through Skinnerian contingency (Skinner, 1953). This model contrasts with ad hoc gamification (pointsification) because it rewards resistance to refutation. To this end, hypotheses must be falsifiable, specifying magnitude and boundary conditions, following the model of Lorig *et al.* (2017).

Complementarily, the mediation analysis used validated constructs from the dominant theories to explain the results: (i) self-efficacy (Bandura, 1986), (ii) basic needs satisfaction (Deci & Ryan, 1985), and (iii) flow states (Csikszentmihalyi, 1990). Measurement will integrate behavioral, psychological, and system metrics to provide convergent evidence (Levy *et al.*, 2018), and the preregistration of analysis plans will serve as a safeguard against p-hacking and post-hoc rationalization (Lorig *et al.*, 2017).

Maintaining this scientific framework requires ethical procedures that prioritize transparency in mechanisms of influence and protect user autonomy, mitigating the risks of psychological manipulation (Verschueren *et al.*, 2020). Monitoring adverse effects, such as a decline in intrinsic motivation (the overjustification effect), is essential to the intervention's ethical integrity (Deci & Ryan, 1985; Verschueren *et al.*, 2020).

In summary, this study consolidates a systematized roadmap for future interventions to move beyond an intuitive approach and adopt the rigor of scientific design. The final framework is structured around four fundamental pillars:

1) **Experimental Design Protocol:** Future research should move from isolated case studies to experimental designs that allow for causal inference. In the Design, prioritize RCTs or factorial designs (e.g., 2×2) to isolate the effect of falsifiability-based scoring from other elements, such as leaderboards or narratives (Lorig *et al.*, 2017). In the Contextualization, interventions should be tested in domains where the tension between engagement and performance is critical, specifically in technical education, organizational safety, and behavioral health (Silic & Lowry, 2020).
2) **Operationalization of Hypotheses:** The protocol requires that a highly precise and falsifiable hypothesis precede all gamification. Example of a Standard: "Intervention X will increase behavior Y by Z% in context W, compared to control V." For mechanisms, researchers must specify the theoretical mediators (self-efficacy, SDT needs, or flow states) and pre-register analysis plans to mitigate publication bias and p-hacking (Lorig *et al.*, 2017; Levy *et al.*, 2018).
3) **Integrated Measurement Framework**: Effectiveness assessment must be multidimensional, integrating the data discussed in Section 4.2. It includes Outcome Metrics (technical performance and error resilience in the 0-5-10 system), Process Metrics (system logs, clicks, response times to detect the "reflective slowdown" predicted by Popper-Skinner integration), and Psychological Metrics

(validated inventories to measure the transition from extrinsic motivation to habit formation).

4) **Ethical Governance and Sustainability:** The research agenda must prioritize user autonomy. Transparency and scoring criteria must be explicit, transforming the rules of the game into a commitment to intellectual mastery (Verschueren *et al.*, 2020). In Longitudinal Studies, it is urgent to conduct research extending beyond eight weeks to validate whether gamification based on falsifiability promotes sustained intrinsic motivation or merely temporary habituation.

6. Conclusion

Integrating the principles of falsifiability (Popper, 1959) with behavioral modeling (Skinner, 1953) represents a significant step toward establishing gamification as a mature scientific discipline. Beyond the theoretical systematization of 271 publications, this study contributes a replicable, operational method synthesized in the framework. By structuring the design into stages ranging from intentional conceptualization to behavioral calibration through the 0-5-10 scoring system, the study provides a roadmap that can transform intuitive 'pointsification' into rigorous engineering.

This approach ensures that user success is not a byproduct of superficial engagement but irrefutable evidence of technical mastery and error resistance, enabling cumulative knowledge development based on testable and validated hypotheses (Lorig *et al.*, 2017; Mora *et al.*, 2017). The guidelines for future research are as follows:

1) **Development of Unified Frameworks:** Prioritize the integration of Falsifiability with dominant structures, such as SDT, Flow Theory, and behavioral economics, specifying not only what increases motivation but also which observations would falsify such theories (Mora *et al.*, 2017).
2) **Specification of Causal Pathways:** Theories must detail complete trajectories (game elements → psychological mechanisms → behavioral outcomes), identifying specific mediators and moderators (Lorig *et al.*, 2017; Mora *et al.*, 2017).
3) **Temporal Dynamics:** Overcoming the limitations of short-term studies to investigate sustained behavior change, habit formation, and the internalization of values (Verschueren *et al.*, 2020).
4) **Methodological Innovations:** Directly testing the proposed 0–5–10 scoring framework and comparing it with alternative assessment systems through randomized controlled trials (RCTs), factorial designs, computational behavioral analyses, and replication studies to empirically validate its construct validity and long-term behavioral benefits (Lorig *et al.*, 2017; Levy *et al.*, 2018).
5) **Ethical Responsibility:** Maintain continuous attention to issues of autonomy, manipulation, and equity, ensuring that interventions promote human flourishing and respect data privacy (Verschueren *et al.*, 2020).

Ultimately, the transition to a design grounded in scientific rigor and methodological transparency is imperative and will allow gamification to fulfill its original promise: to transcend superficial engagement to effectively enhance human motivation and deep learning across the various domains of social activity.

Supplementary Materials
Not applicable.

Authors' Contribution
All authors contributed equally to the conceptualization, methodology, literature review, writing, and critical revision of the manuscript. All authors reviewed and agreed to the published version of the article.

Funding Statement
This research received no external funding.

Institutional Review Board Statement
Not applicable.

Data Availability Statement
Not applicable.

Acknowledgments
To CAPES, CNPQ, the Graduate Program in Science Teaching and Mathematics Education, and the State University of Ponta Grossa.

Conflicts of Interest Statement
The authors declare no conflict of interest.

About the Author(s)
Cláudio Eduardo Marques Lacerda holds a degree in Biological Sciences from Universidade Estadual de Ponta Grossa (2018), a degree in Pedagogy from Faculdade Mantenense dos Vales Gerais (2021), and a postgraduate specialization—EAD in Science Teaching—Final Years of Elementary Education from Universidade Estadual de Londrina (2021). Master's student in the Graduate Program in Science Teaching and Mathematics Education at the State University of Ponta Grossa.
ORCID: https://orcid.org/0009-0003-4557-207X
Silvio Luiz Rutz da Silva graduated in Chemical Sciences from the State University of Ponta Grossa, Brazil. Completed a Master's in Materials Engineering at the Federal University of São Carlos and a PhD in Materials Science at the Federal University of Rio Grande do Sul. Professor in the Department of Physics at the State University of Ponta Grossa, Brazil. I am a permanent faculty member in the Graduate Program in Physics Teaching and in the Graduate Program in Science Teaching and Mathematics Education at UEPG.

ORCID: https://orcid.org/0000-0003-1859-9018

References

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Prentice-Hall. Retrieved from https://books.google.ro/books/about/Social_Foundations_of_Thought_and_Action.html?id=DKK3AAAAIAAJ&redir_esc=y

Belliato, M., *et al.* (2025). Aligning individual motivation with collective goals: A formal approach to gamification. In *Proceedings of the IEEE International Conference on Distributed Computing Systems Workshops* (ICDCSW) (pp. 738–745). IEEE. https://doi.org/10.1109/DCOSS-IOT65416.2025.00123

Butler, C. (2014). *Game design through the lens of behavioral economics* (Doctoral dissertation, New York University).

Challco, G. C., *et al.* (2023). GamiFlow: Towards a Flow theory-based gamification framework for learning scenarios. In *Proceedings of the International Conference on Artificial Intelligence in Education* (pp. 418–424). Springer. https://doi.org/10.1007/978-3-031-36336-8_65

Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. Harper & Row. Retrieved from https://www.researchgate.net/publication/224927532_Flow_The_Psychology_of_Optimal_Experience

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. Plenum Press. Retrieved from https://link.springer.com/book/10.1007/978-1-4899-2271-7

Deterding, S., *et al.* (2013). Designing gamification: Creating gameful and playful experiences. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems* (pp. 3263–3266). ACM. https://doi.org/10.1145/2468356.2479662

Dewey, J. (1979). *Experiência e educação* (R. Gasparian, Trad.). Companhia Editora Nacional. (Obra original publicada em 1938). Retrieved from https://books.google.com.br/books/about/Experi%C3%AAncia_e_educa%C3%A7%C3%A3o.html?id=UiC2EAAAQBAJ&redir_esc=y

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(1), 1–36. https://doi.org/10.1186/S41239-017-0042-5

Grunburg, M. (2025). *The Unified Behavior Model (UBM): An elemental, falsifiable framework for behavior change*. Zenodo. https://doi.org/10.5281/zenodo.15844153

Kapp, K. M. (2012). *The gamification of learning and instruction: Game-based methods and strategies for training and education*. Wiley. Retrieved from https://books.google.ro/books/about/The_Gamification_of_Learning_and_Instruc.html?hl=es&id=GLr81qqtELcC&redir_esc=y

Khoshkangini, R., *et al.* (2021). Automatic generation and recommendation of personalized challenges for gamification. *User Modeling and User-Adapted Interaction*, 31(1), 1–34. https://doi.org/10.1007/S11257-019-09255-2

Klemke, R., Antonaci, A., & Limbu, B. (2020). Designing and implementing gamification: GaDeP, Gamifire, and applied case studies. *International Journal of Serious Games*, 7(3), 91–108. https://doi.org/10.17083/IJSG.V7I3.357

Lacerda, C. E. M. (2025a). *Sequência didática: Do lúdico ao sério, como transformar um jogo em um processo de ensino gamificado* [Apresentação de slides].

Lacerda, C. E. M. (2025b). *Sequência didática (4 aulas): Rigor lógico e estrutura*l [Manuscrito não publicado].

Levy, K., *et al.* (2018). Method in the madness: The design of games as valid and reliable scientific tools. In *Proceedings of the 13th International Conference on the Foundations of Digital Games*. ACM. https://doi.org/10.1145/3235765.3235793

Lorig, F., *et al.* (2017). Hypothesis-driven experiment design in computer simulation studies. In *Proceedings of the Winter Simulation Conference* (pp. 1360–1371). IEEE Press. https://doi.org/10.1109/WSC.2017.8247880

McGloin, R., Farrar, K. M., & Fishlock, J. (2018). Model matching theory: A framework for examining the alignment between game mechanics and mental models. *Media and Communication*, 6(2), 126–136. https://doi.org/10.17645/MAC.V6I2.1326

Mora, A., *et al.* (2017). Gamification: A systematic review of design frameworks. *Journal of Computing in Higher Education*, 29(3), 516–548. https://doi.org/10.1007/S12528-017-9150-4

Popper, K. R. (1959). *The logic of scientific discovery*. Hutchinson. Retrieved from https://www.routledge.com/The-Logic-of-Scientific-Discovery/Popper/p/book/9780415278447

Schöbel, S., Söllner, M., & Leimeister, J. M. (2018). Gamifying information systems: Improving motivation and effective use through a theory-motivated method. In *Proceedings of the European Conference on Information Systems (ECIS*). Association for Information Systems. Retrieved from https://pubs.wi-kassel.de/wp-content/uploads/2019/01/JML_717.pdf?

Schön, D. A. (2000). *Educando o profissional reflexivo: Um novo design para o ensino e a aprendizagem* (R. C. Costa, Trad.). Artmed. Retrieved from https://books.google.ro/books/about/Educando_o_profissional_reflexivo.html?id=mXKBDeUKc2kC&redir_esc=y

Silic, M., & Lowry, P. B. (2020). Using design-science based gamification to improve organizational security training and compliance. *Journal of Management Information Systems*, 37(1), 129–161. https://doi.org/10.1080/07421222.2019.1705512

Skinner, B. F. (1953). *Science and human behavior*. Macmillan. Retrieved from https://books.google.ro/books/about/Science_And_Human_Behavior.html?id=Pjjknd1HREIC&redir_esc=y

Swacha, J., Queirós, R., & Paiva, J. C. (2023). GATUGU: Six perspectives of evaluation of gamified systems. *Information*, 14(2), Article 136. https://doi.org/10.3390/info14020136

Verschueren, S., *et al.* (2020). Developing theory-driven, evidence-based serious games for health: Framework based on research community insights. *JMIR Serious Games*, 8(2), Article e18515. https://doi.org/10.2196/18515

Vygotsky, L. S. (1991). *A formação social da mente: O desenvolvimento dos processos psicológicos superiores* (J. C. Neto, L. S. M. Barreto, & S. C. Afeche, Trads., 4ª ed.). Martins Fontes. Retrieved from https://books.google.ro/books/about/A_forma%C3%A7%C3%A3o_social_da_mente.html?id=BDm-AQAACAAJ&redir_esc=y

Yessad, A., Labat, J.-M., & Kermorvant, F. (2014). Have you found the error? A formal framework for learning game verification. In C. Rensing *et al.* (Eds.), Open learning and teaching in educational communities (pp. 578–583). Springer. https://doi.org/10.1007/978-3-319-11200-8_45