International Journal of Serious Games
ISSN: 2384-8766
https://journal.seriousgamessociety.org/

Article

Learning Mathematical Derivatives Through a Virtual Escape Room Application

Robertas Damaševičius[1] and Tatjana Sidekerskienė[2]

[1]*Department of Applied Informatics, Vytautas Magnus University, LT-44404 Kaunas, Lithuania;*
[2]*Department of Applied Mathematics, Kaunas University of Technology, LT-51368 Kaunas, Lithuania*

robertas.damasevicius@vdu.lt; tatjana.sidekerskiene@ktu.lt

Keywords:
virtual learning
scaffolding
digital escape room
gamification

Received: December 2025
Accepted: September 2026
Published: October 2026
DOI: 10.17083/0sv0r936

Abstract

Purpose: This study presents the design and pilot evaluation of a virtual learning environment (VLE) integrated with an in-person digital escape room (DER) for learning derivatives in a mathematics course. It addresses three research questions: RQ1: How can a scaffolded VLE help students connect the derivative-as-limit concept with rule-based differentiation fluency? RQ2: How can responsible AI-assisted reflection be embedded without replacing learner reasoning? RQ3: What feasibility, engagement, perception, and formative performance patterns emerge when the VLE is used as flipped preparation for an in-person DER? **Methodology:** The VLE combines concept priming, first-principles derivations, procedural differentiation tasks, logarithmic differentiation, tiered hints, worked-solution reveal logic, and AI-assisted reflection prompts. A pilot study with first-year undergraduate students ($N = 48$) collected data on completion, session duration, hint use, AI-reflection use, formative quiz results, item-level diagnostics, and post-DER perceptions. **Findings:** 94% of students completed the full VLE sequence, with a median duration of 56 minutes (IQR: 47–64); 62% used at least one hint, and 68% used AI-assisted reflection. Post-DER ratings were positive for Interest/Enjoyment ($M = 3.45$, $SD = 0.52$), Perceived Competence ($M = 3.18$, $SD = 0.58$), Relatedness/Teamwork ($M = 3.36$, $SD = 0.55$), and Perceived Educational Value ($M = 3.42$, $SD = 0.49$). The formative quiz average was 4.1/5 and revealed interpretable misconceptions in first-principles derivation, chain-rule use, and logarithmic differentiation. **Originality and Impact:** The VLE–DER blueprint links calculus learning objectives, scaffolded room design, validator-based feedback, hint tiers, AI-supported metacognition, telemetry, and item-level diagnostics. The model offers transferable feasibility and design evidence for serious-game-based mathematics learning.

1. Introduction

Calculus remains a foundational step to advanced study in science, engineering, and data-centric disciplines, yet persistent difficulties with the concept of the derivative continue to impede early progression for many first-year university students. Empirical classroom experience and prior literature indicate that novices often memorize rules without anchoring them to the underlying limit definition, leading to brittle performance when problems depart from familiar templates or require multistep reasoning [1, 2]. Despite extensive instructional effort, first-year university students routinely show misconceptions about the meaning of the derivative, particularly its interpretation as an instantaneous rate of change obtained via a limiting process [3, 4, 1]. When confronted with problems that require connecting the limit definition to algebraic manipulation or that combine multiple differentiation rules, university students demonstrate fragile performance and a tendency to rely on surface heuristics [5]. Conventional preparation, textbook reading, and unstructured problem sets often lack the scaffolding and timely feedback required for novices to build robust mental models. At the same time, the increasing availability of large language models (LLMs) introduces both opportunities for formative explanation and risks of overreliance without understanding [6].

This study addresses these difficulties by designing and piloting a virtual learning environment (VLE) that prepares university students for an in-person, problem-based session through a carefully scaffolded, pre-class sequence. The environment, created in Metasteps (https://metasteps.com/), integrates (i) concept priming and motivation, (ii) deliberate practice with the derivative from first principles, (iii) reinforcement of procedural fluency using standard rules and logarithmic differentiation, and (iv) a brief formative quiz and reflection activity. The design draws on constructivist principles and structured tutoring insights [7, 8], as well as self-determination theory to support autonomy, competence, and relatedness in learning [9]. Beyond content sequencing, the VLE operationalizes immediate, low-stakes feedback and optional worked solutions to enable self-correction, while prompting university students to use AI-based reflection responsibly for verification and explanation rather than solution outsourcing [10]. The flipped format repositions initial exposure and consolidation before class, reserving contact time for higher-order problem solving. In doing so, it seeks to narrow the gap between conceptual understanding (derivative-as-limit) and procedural fluency (application of rules) that often persists in traditional lecture-first arrangements [11, 12]. Prior studies have found that flipped classrooms can improve calculus performance and conceptual understanding relative to traditional formats [13]. Self-regulated and peer-teaching enhancements further support student engagement and self-efficacy in calculus [14].

This study presents the design and implementation of an interactive VLE created using Metasteps, aimed at supporting university students in the Mathematics 1 course as they prepare for an in-person digital escape room (DER) activity. The activity serves as a flipped classroom component, allowing university students to engage independently with foundational calculus concepts, while its core affordances (scaffolding, formative checks, worked-solution comparison, reflection prompts) are deliberately chosen to support the transition from derivative-as-limit to rule-based fluency. The approach draws upon principles of constructivist learning, gamification, and scaffolded problem-solving, aiming to enhance engagement, understanding, and confidence in differentiation [7, 8].

This study presents two design-oriented contributions. First, it documents a theoretically grounded VLE–DER framework for learning derivatives that links constructivist scaffolding and structured tutoring practices to a flipped delivery model. The design specifies how concept priming, first-principles derivations, procedural fluency tasks, tiered feedback, optional worked solutions, and collaborative escape-room activities are organized into a coherent learning sequence. Second, it proposes a practical protocol for responsible AI-assisted reflection in mathematics learning, positioning large language models as tools for verification and explanation.

2. Background and Related Work

Virtual Learning Environments (VLEs) in STEM have evolved from content repositories and discussion boards into rich, activity-centered spaces that integrate interactive tasks, immediate feedback, and analytics on learner progress [15]. Contemporary platforms emphasize structured practice, mastery-based progression, and multi-representational support (symbolic, graphical, narrative), which are especially salient in mathematics where conceptual and procedural knowledge must co-develop [16, 17]. Within this continuum, metaverse-inspired experiences (including virtual galeries [18], lightweight web-based 3D scenes [19], immersion [20, 21] and avatar-mediated collaboration [22]) promise heightened presence, social co-regulation, and embodied interaction [23].

Recent work on Educational Escape Rooms (EERs) in mathematics and STEM education shows a gradual shift from using escape rooms merely as motivational game-based activities toward treating them as structured instructional designs with explicit pedagogical, communicative, and assessment functions. Glavaš [24] emphasizes puzzle classification, collaboration, communication, problem solving, engagement, and the need to align game mechanics with learning objectives in mathematical EERs. Taraldsen [25] shows that escape rooms can stimulate curiosity and pedagogical reflection, although students may initially interpret them through traditional views of teaching and learning. Roig et al. [26] combine escape-room mechanics with capture-the-flag-style assessment in STEM, where puzzles re-

quire applied problem solving and hints are integrated into the scoring logic. Luque-Sánchez and Montejo-Gámez [27] demonstrate that virtual escape rooms can support mathematical communication by encouraging collaborative discussion, task leadership, and mathematical vocabulary during problem solving. These studies indicate that EERs are most educationally meaningful when narrative, puzzle structure, collaboration, feedback, and assessment are deliberately aligned with disciplinary learning goals; however, they also reveal the continuing need for more transparent design descriptions and empirically grounded models that connect virtual preparation, mathematical reasoning, and escape-room-based collaboration. Yet, the most durable learning gains reported in mathematics and introductory calculus typically trace to careful task sequencing, explicit scaffolding, and just-in-time formative feedback rather than immersion per se [28]. Consequently, the design challenge is not to maximize immersion, but to engineer activity flows that reduce extraneous cognitive load, surface productive struggle with timely hints, and align interactive elements to clear learning outcomes [29, 30].

Large language models (LLMs) and related AI tools have introduced credible new pathways for formative explanation, self-explanation prompts, and solution verification in undergraduate mathematics. When harnessed with explicit norms, AI can help university students interrogate intermediate steps, compare alternative derivations, and articulate why a step follows—activities known to strengthen metacognition and transfer [31, 32, 33]. At the same time, unstructured use can encourage answer-first behavior, hallucinated steps, and brittle over-reliance [34]. Recent discussions in the education literature underscore the need to channel AI toward reflective practices, e.g., requiring university students to request a hint rather than a final answer, to paraphrase the rationale in their own words, and to cross-check AI-generated steps against canonical definitions and course conventions [35]. Framed within self-determination theory, the approach aims to leverage AI to support autonomy (student control over when and how to seek explanations) and competence (accurate, step-aware feedback) while preserving relatedness through subsequent in-person collaboration [36].

Serious-game research emphasizes that educational effectiveness depends not only on content presentation but also on the alignment between learning objectives and gameplay mechanics. Frameworks such as Mechanics–Dynamics–Aesthetics (MDA) [37] argue that educational goals should be embedded within the rules and progression systems that govern player behavior rather than merely being attached to game-like interfaces. Educational Escape Rooms provide one mechanism for achieving this alignment by linking puzzle progression directly to domain-specific reasoning and collaborative problem solving [38].

Digital Escape Rooms (DERs) have emerged as engaging, game-based learning environments that support active, collaborative, and problem-oriented learning across diverse educational contexts. Prior studies have demonstrated that DERs can effectively promote student motivation, engagement, and deep learning by integrating narrative elements with domain-specific challenges and time-bound puzzles [39, 40, 41]. In STEM education, DERs have been particularly valuable for fostering procedural fluency and critical thinking through scaffolded tasks that simulate real-world problem solving [42]. Previously, DERs were employed as an in-person learning activity following virtual preparation in a VLE, offering an opportunity to apply learned principles in a socially interactive, immersive setting [43]. The integration of DERs complements the flipped instructional design by reinforcing conceptual understanding and encouraging peer collaboration, while also enabling formative assessment of learners' reasoning and engagement under authentic cognitive load [44].

Students' difficulties with derivatives extend beyond procedural errors and include challenges in coordinating intuitive interpretations with formal mathematical definitions. Tall and Vinner's distinction between concept image and concept definition suggests that learners may possess informal mental representations of derivatives that differ substantially from the formal limit-based definition taught in calculus courses [45]. Such discrepancies often manifest as rule-based manipulation without a corresponding conceptual understanding of derivative meaning. Understanding derivatives also requires covariational reasoning, namely the ability to coordinate changing quantities and interpret rates of change dynamically rather than merely execute symbolic procedures [46]. Effective calculus learning requires students to connect multiple representations, including symbolic expressions, graphical interpretations, numerical approximations, and verbal explanations. Difficulties frequently when learners can operate successfully within one representation but fail to establish meaningful connections between representations. These perspectives complement the constructivist and scaffolding-oriented framework adopted in this study. The derivative-as-limit activities were designed to support the development of more coherent concept images, while validator-based feedback, progressive hints, and worked-example gating were intended to facilitate movement between intuitive, symbolic, and increasingly formal forms of mathematical reasoning. This interpretation is also consistent with Tall's broader Three Worlds of Mathematics framework, which characterizes mathematical learning as progression through embodied, symbolic, and formal worlds of mathematical thinking [47].

Three gaps motivate the present study. First, reports of virtual or metaverse learning in STEM foreground immersion and engagement but under-specify the instructional mechanics that connect activities to conceptual change in calculus. Reproducible blueprints that detail the granularity of scaffolds (hint tiers, feedback triggers, worked-solution reveal logic) and their mapping to learning objectives remain sparse [48, 49]. Second, the role of AI in mathematics courses is often discussed at the level of policy or ethics, whereas course-level designs that operationalize responsible AI-assisted reflection, complete with prompts, guardrails, and evaluation rubrics, are relatively uncommon [50, 51].

Third, evaluation frequently relies on perception-only measures; fewer studies report aligned instruments that probe both conceptual understanding (derivative-as-limit) and procedural fluency (rule application, logarithmic differentiation) with item-level diagnostics and effect-size reporting [52, 53].

This work addresses these gaps by contributing: (i) a detailed, exportable VLE blueprint for derivatives that integrates constructivist scaffolding and structured tutoring principles; (ii) a practical protocol for responsible AI-assisted reflection embedded within the activity sequence, explicitly aligned with autonomy–competence–relatedness targets; and (iii) an evaluative framework spanning formative item blueprints, perception scales, and an analysis plan suitable for controlled comparisons and longitudinal tracking. The positioning is thus deliberately *design-centric*: rather than advocating immersion as an end in itself, we specify the scaffolds, feedback, and reflection practices that a VLE must instantiate for novices to link first-principles reasoning with rule-based fluency in differentiation.

3. System Overview

3.1 Pedagogical Rationale

The learning environment design is grounded in constructivist and socio-cultural learning theories. Drawing on Vygotsky's principles, the sequence of rooms represents a scaffolded learning pathway, where conceptual support gradually decreases as learner autonomy increases. Gamified and interactive design elements follow principles of situated learning and self-determination theory, aiming to enhance intrinsic motivation and engagement. Evaluation of the learning sequence is twofold:

1. Formative assessment within the virtual environment – implemented via the embedded quiz, allowing university students to test their conceptual and procedural understanding of derivatives.
2. Post-experience evaluation of the subsequent in-person DER – via a questionnaire assessing engagement, perceived competence, relatedness, and perceived educational value during the escape-room activity.

3.2 Platform: Metasteps

Metasteps is used as the creating and delivery layer for the VLE. Conceptually, the platform (Figure 1) is decomposed into four cooperating subsystems: (i) an *creating workspace* that defines activities, hints, and assessment checks; (ii) a *runtime interaction engine* that renders activities and executes feedback logic; (iii) a *state and telemetry bus* that persists learner progress and emits analytics events; and (iv) a *packaging/export toolchain* that bundles assets, logic, and schemas for deployment in Learning Management System (LMS) or standalone web delivery.

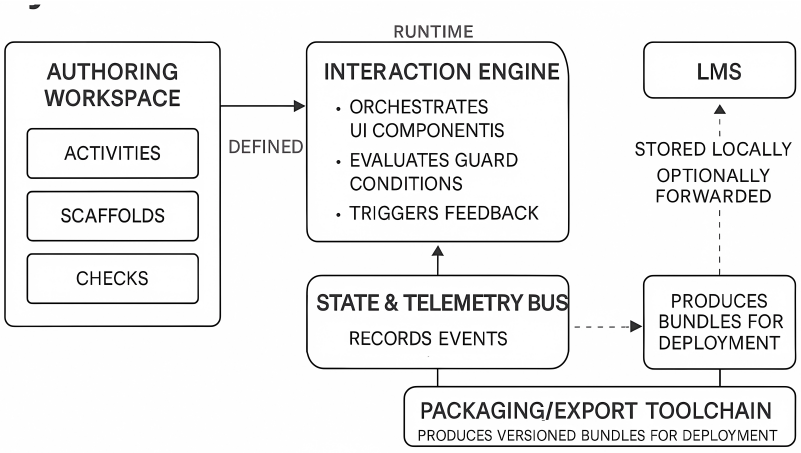


Figure 1. System architecture overview. Author defines activities, scaffolds, and checks in the creating workspace. At runtime, the interaction engine orchestrates UI components, evaluates guard conditions, and triggers feedback. The state & telemetry bus records fine-grained events (e.g., submissions, hint reveals, correctness) that are stored locally and optionally forwarded to the LMS. The packaging/export toolchain produces versioned bundles for deployment.

3.2.1 Creation Model and Interaction Types

Creation in Metasteps is *activity-centric*. Each activity is a typed component with: (a) `props` (prompt, stimuli, expected response schema), (b) `guards` (preconditions to unlock follow-on steps), (c) `checks` (validators for correctness, completeness, or structure), (d) `hints` (tiered scaffolds revealed on demand or via triggers), and (e) `effects` (feedback messages, progression, or logging). Activities are composed into a *flow* via directed edges with optional

branching (e.g., remediation arcs). To standardize implementation and analysis, we define a small set of interaction types (Table 1). Each type specifies its response format, validator class, feedback modality, and the analytics fields emitted on interaction:

Table 1. Interaction types and their runtime/analytics specifications.

Interaction Type	Description (Response Schema)	Validator/Check	Feedback Modality
Definition Recall	Short-text entry for the derivative-as-limit	Regex + semantic keyphrase matching (Levenshtein tolerance)	Inline correctness badges; expandable exemplar definition
First-Principles Derivation	Multi-step algebraic workflow; schema: ordered list of LaTeX fragments	Stepwise structural check (pattern graph) + final numeric/symbolic equivalence	Step-level hints; revealable worked solution; checkpointing
Rule Application (Power/ Product/ Quotient)	Single-expression differentiation	CAS-backed equivalence (normal form)	Granular messages on rule misuse; lightweight rule table tooltip
Chain Rule Composition	Nested compositions; schema: parse tree of input expression	Tree pattern match (outer/inner), derivative reconstruction check	Visual parse of $f(g(x))$; targeted nudge on inner/outer confusion
Logarithmic Differentiation	Product/power to log-derivative; sequence of transformation steps	Step normalization + identity checks (log laws)	Tiered hints; final consolidated solution comparison
Multiple-Choice Question (MCQ)/ Single-Select	Concept check	Immediate correctness; explanation for chosen distractor	
Reflection Prompt (AI-Assisted)	Free-text reflection	Length + keyword rubric (no correctness)	Metacognition checklist; responsible-use reminder

3.2.2 VLE Architecture and Flow

The virtual learning environment (VLE) is organized as a directed learning flow (Figure 2) composed of four scaffolded stages: (i) *Motivation & Concept Priming*, (ii) *Derivatives from First Principles*, (iii) *Application & Extension (Procedural Fluency)*, and (iv) *Assessment Space (Formative Quiz)*. Each stage is a subgraph of activities connected by guard conditions and instrumented with fine-grained telemetry (submit/step_submit/hint_show/reflect), enabling item-level difficulty estimation and hint-utilization analytics.

The architecture prioritizes (a) conceptual anchoring via the limit definition, (b) progressive fading of supports, and (c) immediate, actionable feedback. Activities are implemented using the canonical interaction types and validator pipelines described earlier, with stepwise normalization, structural checks, and CAS-backed equivalence.

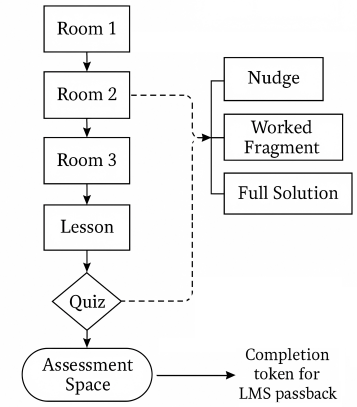


Figure 2. VLE architecture and learning flow. Nodes represent typed activities; edges represent guard-conditioned progression (solid = mastery progression, dashed = remediation arc). Hint tiers (nudge/worked fragment/full solution) are available in Rooms 2–3 and are rate-limited by an impasse detector (time-on-step, repeated structural mismatch). The Assessment Space aggregates item-level evidence and emits a completion token for LMS passback (optional).

Table 2 enumerates the mapping from learning objectives to room-level activity clusters and validator emphasis; Table 3 details the hint policy and auto-trigger criteria.

3.3 Serious Game Design Perspective

From a serious-game perspective, the proposed VLE–DER framework is defined by the integration of educational objectives with gameplay mechanics. To make this relationship explicit, the design can be interpreted through the Mechanics–Dynamics–Aesthetics (MDA) [37] framework as follows.

Mechanics. The core mechanics include puzzle solving, code discovery, room progression, time constraints, progressive hints, validation checkpoints, and collaborative decision making. Mathematical reasoning serves as the

Table 2. Learning objectives mapped to rooms, activity archetypes, and validator emphasis.

Learning Objective	Room	Activity Archetypes
Explain derivative as limit	Definition recall; concept MCQ; graphical rate-of-change probes	Regex/keyword match; distractor-specific feedback
Compute from first principles	Multi-step derivation ($\Delta y/\Delta x \to 0$); limit algebra	Stepwise structural checks; limit law invariants
Apply rules (power/product/quotient)	Single-expression derivative; error diagnosis on rule misuse	CAS-backed normal form equivalence; rule-flagging
Apply chain rule for compositions	Parse-tree identification of $f(g(x))$; derivative reconstruction	Tree pattern match; depth-sensitive feedback
Logarithmic differentiation	Transform-then-differentiate sequences	Identity checks for log laws; step normalization
Formative self-evaluation	Mixed blueprint (concept + procedure + trig/log-diff)	Keyed/structural/Equiv. checks; confidence tagging

Table 3. Hint policy, tiers, and auto-trigger thresholds (impasse detection).

Tier	Content	User Trigger	Auto-Trigger (any)
Nudge	Law/definition cue (e.g., "apply quotient rule"; "factor h before limit")	Hint button	$\geq$ 90 s dwell on step; 2 structural mismatches
Worked fragment	One correct next step with rationale	Reveal after nudge	3 incorrect attempts; oscillation between two errors
Full solution	Complete, annotated derivation	Confirmed request + reflection prompt	Not auto-triggered; gated behind reflection

primary interaction mechanic because advancement depends on successfully solving derivative-related challenges rather than on exploration alone. Puzzle dependency structures ensure that later challenges cannot be completed without knowledge generated in earlier stages.

Dynamics. The interaction of these mechanics produces several intended dynamics. Time pressure encourages prioritization and communication, puzzle dependencies promote collaboration and information sharing, and progressive hints balance challenge with support. Validation checkpoints provide immediate feedback and prevent teams from progressing using incorrect solutions. These dynamics create cycles of hypothesis generation, verification, correction, and collaborative problem solving.

Aesthetics. The intended player experiences include challenge, accomplishment, curiosity, cooperation, and immersion. The narrative context of restoring a malfunctioning "Derivative Engine" provides purpose and urgency, while room progression and code unlocking provide a sense of advancement and achievement. Rather than emphasizing entertainment alone, the aesthetic experience is designed to support engagement with mathematical reasoning.

The resulting gameplay loop can be summarized as: explore → interpret clues → solve mathematical challenge → validate solution → unlock progression → collaborate on the next puzzle. Educational objectives are embedded directly within progression mechanics because advancement depends on demonstrating understanding of derivative concepts. Mathematical reasoning functions simultaneously as a learning activity and as the primary game mechanic. To clarify the serious-game structure of the proposed VLE–DER activity, Figure 3 represents the design using a Machinations-inspired interpretation of the Mechanics–Dynamics–Aesthetics (MDA) framework. The figure shows how designed mechanics, such as puzzles, hints, time pressure, progression, feedback, rewards, collaboration, and narrative, generate emergent player dynamics and intended learner experiences in support of the educational goals.

3.4 Design of the In-Person Educational Escape Room

The in-person digital escape room was designed as an Educational Escape Room (EER) that complemented the scaffolded VLE preparation sequence. Its purpose was not to introduce new derivative concepts, but to provide a collaborative transfer activity in which students could apply concepts previously practised in the VLE. The design followed a backward-design logic: learning objectives were defined first, then translated into puzzle mechanics, clues, and progression rules. The main learning objectives were to reinforce students' ability to: (i) interpret the derivative as a limit, (ii) apply standard differentiation rules, (iii) identify outer and inner functions in chain-rule problems, and (iv) use logarithmic differentiation in a multi-step task. The narrative framed students as members of a mathematical rescue team attempting to restart a malfunctioning "Derivative Engine" before a system shutdown (Figure 4). To restore the engine, teams had to solve a sequence of derivative-based puzzles. Each correct solution generated a clue, code, or mathematical artifact needed for the next stage. The narrative was intentionally lightweight so that it supported immersion and urgency without distracting from mathematical reasoning. The puzzle sequence was organized into four stages of increasing complexity. In the first stage, students reconstructed a fragmented derivative-as-limit expression and identified the missing algebraic steps required to move from the difference quotient to the final derivative. In the second stage, teams matched standard differentiation rules to partially completed solutions and corrected intentionally

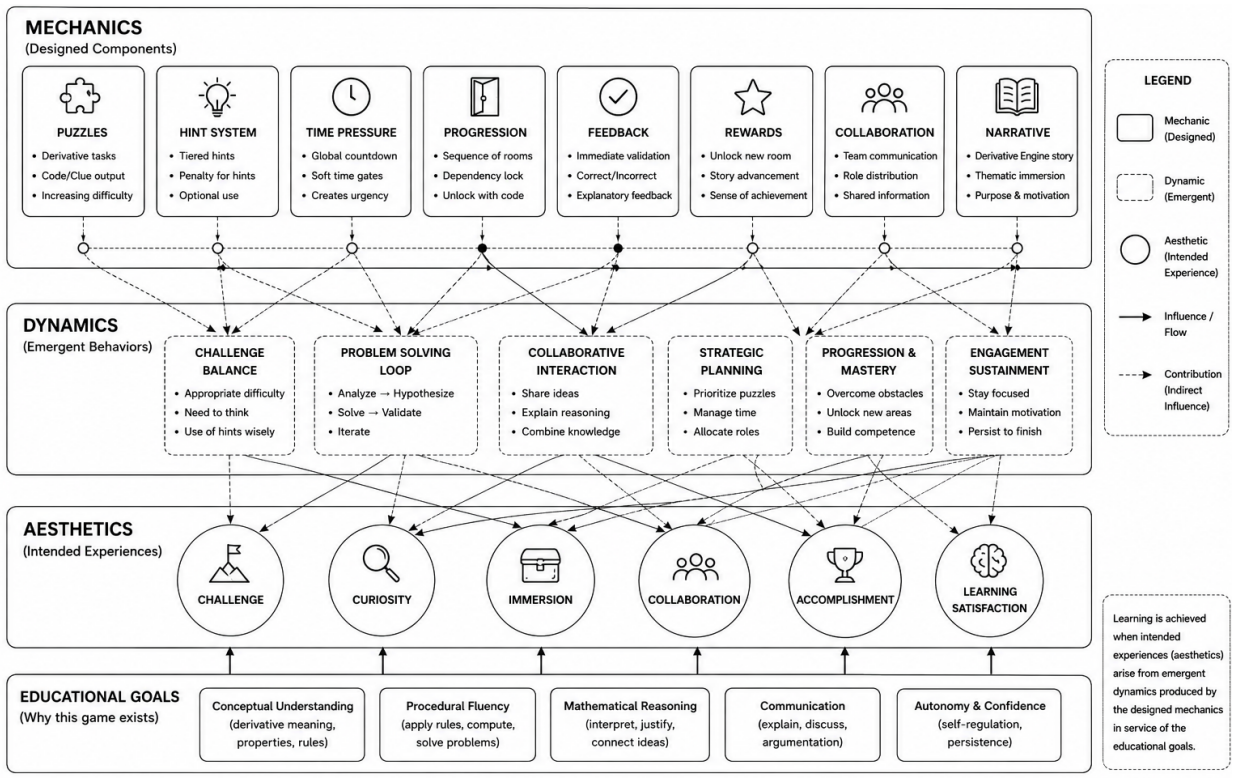


Figure 3. MDA representation of the escape room design for learning derivatives. The designed mechanics include puzzles, hint system, time pressure, progression, validation feedback, rewards, collaboration, and narrative. The associated aesthetics include challenge, curiosity, immersion, collaboration, accomplishment, and learning satisfaction, which are aligned with educational goals of conceptual understanding, procedural fluency, mathematical reasoning, communication, and autonomy.

introduced errors. In the third stage, students solved chain-rule puzzles by identifying the outer and inner functions of composite expressions and using the resulting derivatives to decode the next clue. In the final stage, teams completed a logarithmic differentiation problem whose solution produced the final escape code. The puzzles were designed according to three principles: alignment with learning objectives, progressive difficulty, and interdependence between tasks. Each puzzle required mathematical reasoning rather than simple code searching, and later puzzles depended on artifacts obtained from earlier ones. This structure encouraged students to connect conceptual and procedural knowledge while maintaining the collaborative and time-constrained nature of an escape-room activity.

The scaffolding strategy was informed not only by common procedural errors but also by theoretical perspectives on mathematical cognition. Activities based on first-principles derivatives were intended to strengthen connections between concept images and formal derivative definitions, whereas later procedural tasks supported transitions from conceptual understanding to symbolic fluency. The combination of explanations, hints, validation rules, and worked examples was designed to help learners coordinate multiple representations and gradually move from intuitive reasoning toward more formal mathematical thinking.

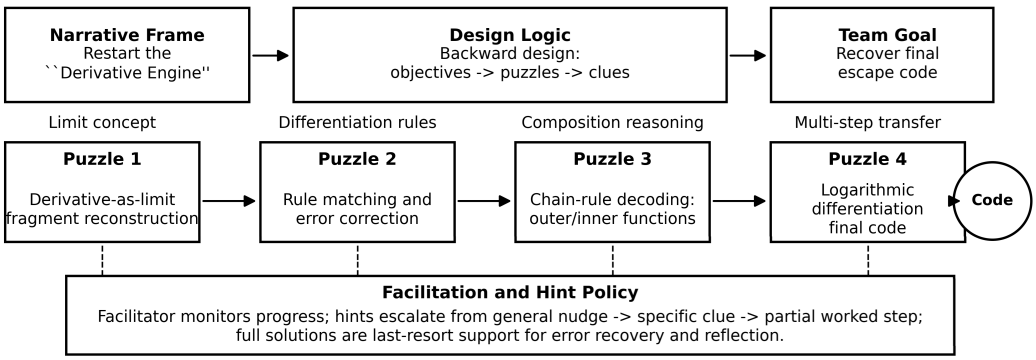


Figure 4. Design flow of Educational Escape Room (EER) for learning derivatives. The activity uses a lightweight narrative in which teams restore a malfunctioning "Derivative Engine" by solving four progressively connected derivative-based puzzles.

3.5 VLE Design and Development

The virtual environment consists of three interconnected spaces (Figure 5); each designed with a distinct pedagogical function:

1. Learning Environment (Rooms 1–3) – three interactive rooms focusing on motivation, conceptual reinforcement, and procedural application.
2. Assessment Space – a structured quiz for knowledge consolidation and self-evaluation.
3. Post-Experience Evaluation Space – a reflective survey collecting student feedback on learning experience and perceived effectiveness.

All rooms are visually linked through a consistent 3D environment and thematic cues (colors, characters, and visual prompts). The design promotes progressive learning, starting from informal recall to increasingly complex applications.



Figure 5. Overview of the virtual interactive learning environment developed in Metasteps. The space consists of three thematic learning rooms (motivation, conceptual reinforcement, and application), followed by the Assessment Space and Post-Experience Evaluation Space. Each room supports scaffolded learning and integrates visual, interactive, and reflective elements.

3.5.1 Room 1 – Motivation & Concept Priming Room

The first room (Figure 6) introduces the activity and reminds university students of their semester-long creative assignment: to design a pedagogical meme related to a topic in Mathematics 1. The room displays humorous meme examples on the topic of limits, featuring concepts such as indeterminate forms and the L'Hôpital's rule. This space aims to foster emotional engagement, creativity, and concept recall, encouraging university students to connect personal expression with mathematical reasoning. Humor acts as a cognitive anchor, reducing anxiety and establishing a low-stakes, positive entry into the virtual experience.



Figure 6. Room 1–Introduction and Motivation: Mathematical memes on the concept of limits. Virtual room shows humorous mathematics-related memes illustrating the concept of limits and L'Hôpital's rule. The space serves as an introductory environment designed to motivate university students, foster creativity, and promote emotional engagement with mathematical ideas.

3.5.2 Room 2 – Concept Reinforcement: Derivatives from First Principles

The second room (Figure 7) reinforces the conceptual foundation of differentiation by focusing on the definition of the derivative as a limit. The room features visual aids displaying the derivative formula and presents three tasks requiring university students to calculate derivatives using this definition. To support autonomy, university students can

watch an embedded video tutorial where the instructor solves the same examples step by step. A separate wall features a fully worked-out solution, allowing learners to compare their reasoning and outcomes. This space operationalizes Vygotskian scaffolding principles by combining guided instruction with self-directed exploration [7]. It supports zone of proximal development learning, where university students gradually internalize expert problem-solving strategies through mediated interaction. This room operationalizes the limit definition via guided derivations (Table 4). Each derivation is decomposed into checkpoints that enforce algebraic invariants (difference quotient formation, rationalization/factoring, correct limit transition). Learners can request tiered hints; a final revealable solution is gated behind a reflection prompt documenting the key idea they were missing.

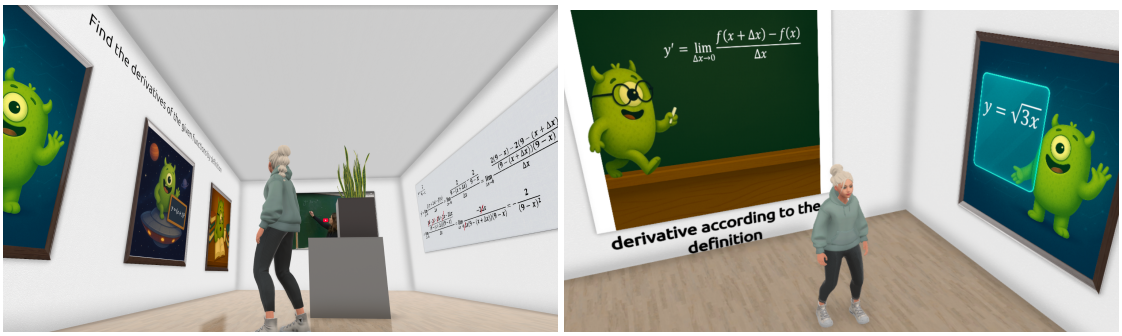


Figure 7. Room 2, Derivatives from First Principles. The interface guides university students to compute a derivative using the limit definition, with optional scaffolds (video hint, step-by-step derivation, and revealable worked solution). Inline checks provide immediate feedback on algebraic steps and limit evaluation.

For a target f, the activity enforces the sequence:

$$\textbf{(S1)}\ \frac{f(x+h)-f(x)}{h} \rightarrow \textbf{(S2)}\ \text{algebraic simplification} \rightarrow \textbf{(S3)}\ \lim_{h\to 0}\ \text{evaluation}. \tag{1}$$

Validators check stepwise structure before allowing the next step.

Table 4. First-principles derivations: item catalog and validator invariants.

ID	Target $f(x)$	Critical Invariants Checked	Common Error Flags
F1	$(x+1)^2$	(i) Correct difference quotient; (ii) expand $(x+h+1)^2$; (iii) factor h; (iv) evaluate limit	Skips factoring; cancels non-common terms
F2	$\sqrt{3x}$	(i) Multiply by conjugate; (ii) factor h; (iii) domain note $x>0$	Rationalize wrong side; evaluate before simplification
F3	$3x-x^2$	(i) Correct difference quotient; (ii) correct substitution $3(x+h)-(x+h)^2$; (iii) expand $(x+h)^2$; (iv) factor h; (v) evaluate limit	Students incorrectly write $f(x+h)$ as $3x-x^2+h$ instead of substituting $x+h$ everywhere; skipping factoring; applying derivative rules instead of using the definition

3.5.3 Room 3 – Application & Extension Room (Procedural Fluency)

Room 3 (Figure 8) extends university students' procedural fluency by engaging them with derivative rules and properties. A comprehensive table of derivatives (including power, exponential, logarithmic, and trigonometric functions) is presented alongside the sum, product, quotient, and chain rules. Students solve a sequence of tasks of increasing difficulty, beginning with straightforward rule applications and culminating in a logarithmic differentiation problem introduced by a playful robot character. The challenge highlights that "the robot cannot compute it, but you can," emphasizing human reasoning beyond algorithmic calculation. The tasks are accompanied by visual cues and friendly character illustrations to sustain motivation. At the end, university students are encouraged to verify their answers using ChatGPT, integrating AI-assisted reflection [10]. This fosters metacognitive awareness and bridges traditional calculus with digital-era learning tools.

This room consolidates rule-based differentiation, emphasizes error diagnosis, and extends to compositions and logarithmic differentiation (Table 5). Activities adopt single-step submissions with immediate CAS-backed equivalence checks; rule misuse is flagged and paired with concise corrective messaging. Parse-tree visualization supports the chain rule by rendering $f(g(x))$ with outer/inner identification.

3.5.4 Assessment Space (Formative Quiz) – Knowledge Consolidation and Self-Evaluation

The Assessment Space (Figure 9) implements a short quiz that samples both conceptual understanding and procedural fluency. Items are auto-graded with typed validators; confidence ratings are optionally collected to enable calibration

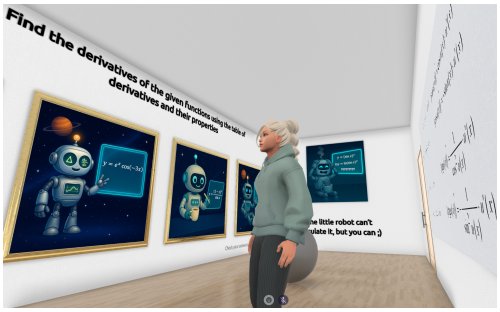


Figure 8. Application and Extension Room, reinforcing procedural fluency in differentiation. Tasks progress from base rules (power, product, quotient) to chain-rule compositions and a culminating logarithmic differentiation challenge. A compact rule table and context-sensitive hints support just-in-time recall without displacing reasoning.

Table 5. Procedural fluency task set with rule emphasis and feedback signals.

ID	Prompt (Abbrev.)	Primary Rule(s)	Feedback Signals (Examples)
R1	$\frac{d}{dx}\left(e^{\cos(2x)}\right)$	Chain	"Outer = $e^{(\cdot)}$, inner = $\cos(2x)$; multiply by derivative of $\cos(2x)$"
R2	$\frac{d}{dx}\left(e^{x}\cos(-3x)\right)$	Product (+ chain)	"Apply product rule; differentiate e^x and $\cos(-3x)$ separately; include derivative of inner $-3x$"
R3	$\frac{d}{dx}\left(\frac{(1-x)^3}{\sin x}\right)$	Quotient (+ chain)	"Use quotient rule; keep denominator squared; apply chain rule to $(1-x)^3$"
R4	$\frac{d}{dx}(\sin^2 x)$	Chain	"Outer = square, inner = $\sin x$; multiply by $\cos x$"
R5	$\frac{d}{dx}\left((\sin x)^x\right)$ (log-diff)	Logarithmic diff.	"Take ln both sides; use $\ln a^b = b \ln a$ before differentiating"

analysis. Immediate correctness indicators are provided, while full worked solutions remain behind an explicit reveal to encourage self-attempt first.

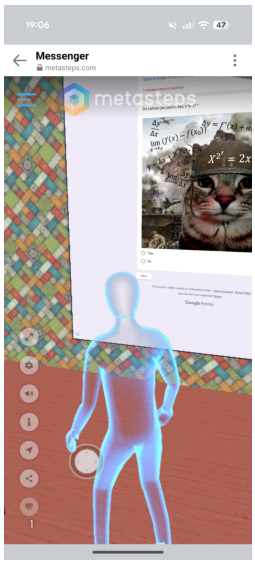


Figure 9. Assessment Space, a short formative quiz designed to evaluate conceptual understanding and procedural fluency in differentiation. Items span recognition of the derivative-as-limit, recall of elementary derivatives, and rule application (including chain rule). Immediate correctness indicators and deferred full solutions enable self-checking and reflection.

The assessment covers five cognitive dimensions:

1. Concept recognition – identifying the derivative definition formula.
2. Recall of derivative tables – verifying knowledge of elementary derivatives.
3. Basic computation – applying differentiation to simple functions.
4. Rule integration – applying the chain rule to composite expressions (such as $y = e^{2\tan(x)}$, $y = \cos^2 x$).
5. Extension – log-differentiation variants such as $y = (x^2 + 1)^x$.

This formative evaluation provides immediate feedback, reinforcing knowledge and supporting self-assessment before the in-person session. The quiz follows a fixed blueprint (Table 6) to ensure coverage and comparability across course cohorts . Item telemetry (latency, attempt count, hint use) is exported with an anonymized session identifier to support item diagnostics (difficulty, discrimination proxies) and to inform iterative improvement.

Table 6. Formative quiz blueprint (concept + procedure + extension).

Q#	Construct	Prompt (Abbrev.)	Type / Validator	Scoring/Telemetry
Q1	Concept (definition)	State derivative-as-limit at x_0	Short-text / regex+keyphrase	Binary; latency, lev_dist
Q2	Recall (base rules)	$\frac{d}{dx}(\cos(x))$, $\frac{d}{dx}(\ln x)$	MCQ (multi) / keyed	Partial credit; choice vector
Q3	First principles	Derive $f'(x)$ for $f(x) = x^2$	Multi-step / structural+equiv.	Step scores; hint_tier
Q4	Rule application	$\frac{d}{dx}(e^{2\tan(x)})$, $\frac{d}{dx}(\cos^2(x))$	Single-step / CAS equivalence	Binary; rule_flag
Q5	Extension (log-diff)	$\frac{d}{dx}((x^2+1)^x)$	Single-step / CAS/identity	Binary; attempts

Each item optionally collects a confidence rating on a 3-point scale (low/medium/high). A post-quiz panel visualizes personal calibration (e.g., over/under-confidence) and recommends targeted review. Worked solutions are presented with *why-this-step* notes linking back to the relevant rules or limit laws used.

For example, for the First Principles construct, with $f(x) = x^2$,

$$f'(x) = \lim_{h\to 0} \frac{(x+h)^2 - x^2}{h} = \lim_{h\to 0} \frac{x^2 + 2xh + h^2 - x^2}{h} = \lim_{h\to 0} \frac{2xh + h^2}{h} = \lim_{h\to 0} (2x + h) = 2x. \tag{2}$$

For example, for the Rule Application construct, with $f(x) = e^{2\tan(x)}$,

$$\frac{d}{dx}\left(e^{2\tan(x)}\right) = e^{2\tan(x)} \cdot 2 \cdot \frac{1}{\cos^2 x} \quad \text{(chain rule, outer} = e^{(\cdot)}\text{, inner} = 2\tan(x)). \tag{3}$$

For a log-differentiation variant, with $y = (x^2+1)^x$,

$$\ln y = x\ln(x^2+1) \;\Rightarrow\; \frac{y'}{y} = \ln(x^2+1) + x \cdot \frac{2x}{x^2+1}, \quad y' = (x^2+1)^x \left[\ln(x^2+1) + \frac{2x^2}{x^2+1}\right]. \tag{4}$$

3.6 Content Design for Calculus: Limits and Derivatives

The content sequence is engineered to fuse conceptual anchoring (derivative-as-limit) with procedural fluency (rules, compositions, logarithmic differentiation), while minimizing extraneous cognitive load (Table 7). Each activity specifies (i) a learning objective, (ii) an expected representation (symbolic, graphical, or mixed), (iii) validator invariants (structural and equivalence checks), and (iv) a tiered-hint policy with explicit reveal conditions. Conceptual tasks favor multiple representations, e.g., connecting the secant-to-tangent limiting process with symbolic limit evaluation, whereas procedural tasks emphasize rule selection diagnostics, parse-tree reasoning for compositions, and identity management for logarithmic differentiation. Worked solutions are embedded but gated behind reflection prompts to preserve desirable difficulty; when revealed, they highlight *why* each step is valid and reference the governing law (limit law, product rule, chain rule, log identities).

Table 7. Content blocks and intended learning outcomes with representation and validator focus.

Content Block	Learning Outcomes	Primary Representation	Validator Focus
Derivative as a Limit	Define $f'(x_0)$ via difference quotient; interpret instantaneous rate of change	Symbolic + graphical	Regex+keyphrase; distractor-specific MCQ feedback
First-Principles Derivations	Execute $f'(x)$ by limit definition (e.g., x^2, $\sqrt{3x}$)	Symbolic, stepwise	Structural invariants, limit-law gating; final equivalence
Rule Application	Apply power, product, quotient rules to single expressions	Symbolic	CAS-backed equivalence; rule misuse flags
Chain Rule (Compositions)	Identify outer/inner; reconstruct derivative of $f(g(x))$	Symbolic + parse tree	Tree pattern checks; depth-sensitive feedback
Logarithmic Differentiation	Transform products/powers to sums; differentiate & back-transform	Symbolic, multi-step	Identity checks (log laws); step normalization

3.6.1 Task Design: Progression, Difficulty, and Hints

Tasks progress along a calibrated path that increases structural complexity while keeping step granularity stable (Table 8). We parameterize difficulty by (a) *algebraic load* (number of valid transformations before the limit resolves), (b)

structural depth (nesting for compositions), and (c) *identity load* (number/type of identities required, e.g., $\ln(a^b) = b\ln a$). Hint tiers (*nudge* → *worked fragment* → *full solution*) are rate-limited by impasse detectors (dwell time, repeated structural mismatches). The target is to sustain *productive struggle* without tipping into frustration.

Table 8. Progression and difficulty calibration for core tasks with hint policy.

Task Family (Example)	Algebraic Load	Structural Depth	Identity Load	Hint Policy (Triggers)
First Principles: $f(x) = x^2$	Low	Shallow	None	Nudge @ 90 s; worked fragment after 2 wrong structures; full solution gated by reflection
First Principles: $f(x) = \sqrt{3x}$	Medium	Shallow	Conjugate	Nudge on missing conjugate; worked fragment shows rationalization step
First Principles: $3x - x^2$	Medium	Shallow	None	Nudge on incorrect $f(x+h)$ substitution (e.g., $3x - x^2 + h$); fragment shows $3(x+h) - (x+h)^2$ and factoring h
Rules+chain: $e^x \cos(-3x)$	Low	Shallow	None	Rule table tooltip on product misuse; fragment shows $u'v+uv'$
Rules+chain: $\frac{(1-x)^3}{\sin x}$	Medium	Shallow	None	Nudge on quotient rule structure; misuse flag if only numerator diff'd
Chain: $\sin^2(x)$	Medium	Moderate	None	Parse tree overlay; nudge to mark outer/inner; fragment shows $2\sin x \cos x$
Log-Diff: $(\sin x)^x$	High	Moderate	Log laws	Nudge: "take ln"; fragment: $\ln y = x\ln(\sin x)$; identity-check on $\ln a^b$

Early tasks minimize identity load to foreground the limit structure (difference quotient, factoring/rationalization). Next tasks add composition depth. Supports fade in two ways: (i) hints escalate more slowly and require a reflection note before full-solution reveal; (ii) validator messaging transitions from prescriptive (*"factor h before taking the limit"*) to diagnostic (*"denominator not squared, check quotient rule"*).

3.6.2 Worked Solutions, and Common Misconceptions

Worked examples are provided for transparency and error recovery, but are revealed only after a reflection prompt that asks the learner to articulate the missing idea. Each solution marks the governing rule/identity at the right margin and highlights invariants checked by the validator at each step.

Worked examples are aligned 1:1 with checkpoint tasks and escalate in identity complexity. Each includes (i) a *template* (e.g., product rule schema), (ii) a *counterexample* displaying a common error, and (iii) a *minimal fix* highlighting the precise repair. This trio supports near transfer during the Assessment Space while avoiding over-scaffolding.

Each misconception in Table 9 maps to at least one quiz item with distractor logic keyed to that misconception. Post-quiz feedback links back to the specific example panel, completing the loop between instruction, diagnosis, and targeted review.

Table 9. Common misconceptions, validator flags, and corrective feedback.

Misconception	Corrective Feedback (Hint Tier)
Cancelling non-common terms in $\frac{(x+h)^2-x^2}{h}$	"Expand, then factor the common h before the limit." (nudge)
Taking $h \to 0$ before removing indeterminate form	"Simplify to remove 0/0 using factoring or conjugates." (nudge)
Quotient rule without squaring denominator	"Check quotient rule template; denominator is squared." (nudge)
Misusing $\ln(a^b)$ as $\ln a^b = b\ln a$ with wrong grouping	"Apply $\ln(a^b) = b\ln a$; keep a grouped before differentiating." (worked fragment)
Confusing average vs instantaneous rate	MCQ rationale explains secant-to-tangent limit and local linearity (explanation card)

4. Study Design and Methodology

4.1 Participants and Context

The study was conducted with first-year undergraduate university students enrolled in a compulsory calculus module (Mathematics 1) at Kaunas University of Technology (Lithuania) (cohort: N=48). The virtual learning environment (VLE) implemented in Metasteps was integrated as a required flipped preparation activity, while the subsequent in-person session employed a digital escape-room (DER) format focused on collaborative problem solving with derivatives. Recruitment occurred during the first month of the semester; informed consent was obtained prior to data collection. To preserve anonymity, all VLE telemetry, quiz responses, and questionnaire data were linked to a pseudonymous session identifier. Completion of the VLE contributed a small participation mark but did not affect

summative course grades; similarly, the DER and questionnaire were treated as low-stakes learning activities. The study timeline is summarized in Figure 10.

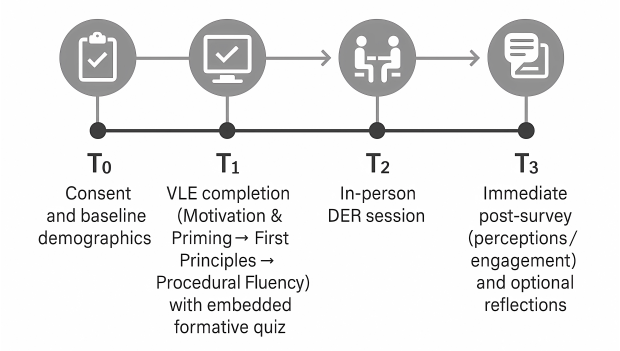


Figure 10. Study timeline. $\mathbf{T_0}$**:** Consent and baseline demographics; $\mathbf{T_1}$**:** VLE completion (Motivation & Priming → First Principles → Procedural Fluency) with embedded formative quiz; $\mathbf{T_2}$**:** In-person DER session; $\mathbf{T_3}$**:** Immediate post-survey (perceptions/engagement) and optional reflections.

4.2 Study Protocol (Flipped Preparation + In-Person Session)

The protocol (Table 10) comprised two phases: (i) asynchronous flipped preparation in the VLE (Metasteps) and (ii) a subsequent on-campus session emphasizing collaborative transfer. Learners complete a scaffolded sequence culminating in a short formative quiz. Telemetry logs include event type (`submit`, `step_submit`, `hint_show`, `reflect`), correctness, latency, and hint tier used. Worked solutions are gated behind a mandatory reflection prompt to discourage answer-first behavior. Students solve multi-step derivative challenges. Artifacts (scratch work, intermediate answers) are not graded but are discussed in a closing debrief to reinforce conceptual anchors (derivative-as-limit) and rule-based fluency. Immediately after the DER, participants complete a perception/engagement questionnaire (four-point Likert scale). A short free-text reflection prompts university students to summarize a key misconception they corrected.

Table 10. Protocol logistics and data capture by phase.

Phase	Activity	Data Captured	Duration
T_0	Consent & demographics	Enrollment, prior calculus exposure	10–15 min
T_1	VLE (flipped)	Item-level telemetry (correctness, latency, hints), formative quiz responses	45–70 min
T_2	DER (in-person)	Team artifacts (optional), facilitator notes (non-identifying)	60–75 min
T_3	Post-survey & reflection	Likert items, free-text reflection	10–12 min

4.3 Instruments

Outcomes are measured using (i) a blueprint-driven formative quiz aligned to learning objectives and (ii) a perception/engagement survey with theory-informed subscales. Scoring and equivalence checks are automated; reliability and validity checks are planned as part of the analysis.

4.3.1 Formative Quiz

The quiz samples (Table 11) both conceptual understanding (derivative-as-limit) and procedural fluency (rules, compositions, logarithmic differentiation). Items are auto-graded using a validator pipeline: `normalize` → `structural_check` → `equivalence_check`. For algebraic equivalence, expressions are reduced to a normal form via a computer-algebra back-end; stepwise derivations must satisfy structural invariants (e.g., factoring of x prior to limit evaluation).

Internal consistency for procedural items is estimated via a congeneric model (McDonald's ω) in addition to Cronbach's α; item difficulty is approximated by p-values and time-on-task. We conducted an exploratory two-parameter IRT fit (discrimination, difficulty) where sample size permits. Face/content validity is supported by blueprint mapping to learning outcomes (Table 11).

4.3.2 Questionnaire

The post-DER questionnaire was designed as a short, context-specific perception instrument rather than a full standardized motivation scale. Although established instruments such as the Intrinsic Motivation Inventory (IMI) provide validated subscales for constructs including interest/enjoyment and perceived competence, the present pilot study fo-

Table 11. Formative quiz: constructs, exemplar prompts, scoring, and validator focus.

Construct (LO)	Exemplar Prompt (Abbrev.)	Scoring	Validator Focus
Concept: derivative-as-limit	State $f'(x_0)$ using a limit	Binary; partial for keyphrases	Regex + keyphrase set; Levenshtein tolerance
Recall: base derivatives	$\frac{d}{dx}(\cos x)$, $\frac{d}{dx}(\ln x)$	Multi-select partial credit	Keyed options; distractor rationales
First principles (worked)	Derive $f'(x)$ for $f(x) = x^2$	Step-level points	Structural invariants (difference quotient, factor h), final equivalence
Rule application	$\frac{d}{dx}(e^{2\tan(x)})$, $\frac{d}{dx}(\cos^2(x))$	Binary	CAS normal form; rule-misuse flag (product vs chain)
Extension: trig/log-diff	$\frac{d}{dx}((x^2+1)^x)$	Binary	Identity checks; transform-then-differentiate sequence

cused primarily on feasibility, usability, and immediate perceptions of a specific derivative-focused VLE–DER activity. Therefore, we used a concise set of self-devised items adapted to the learning context, the collaborative DER format, and the limited post-session response time. The items were intended to capture immediate perceptions of interest/enjoyment, perceived competence, relatedness/teamwork, perceived educational value, and recognition/extrinsic orientation after the escape-room activity. A survey was conducted with participants after the DER experience, using a four-point Likert scale and questions presented in Table 12.

Table 12. Questionnaire about the DER session in the VLE

No.	Question
Q1	I would describe this escape room as very interesting.
Q2	Playing the escape room was an interesting way to learn.
Q3	Playing the escape room was a boring activity.*
Q4	Playing the escape room was fun.
Q5	Playing the escape room was quite enjoyable.
Q6	I play well in the escape room compared to other university students.
Q7	I was quite good at playing the escape room.
Q8	I was unable to play the escape room well.*
Q9	I would be willing to play the escape room again because it was valuable to me.
Q10	I believed I learnt about the topic while I was playing the escape room.
Q11	I believed playing the escape room was beneficial to me.
Q12	I felt close with my peers while playing the escape room.
Q13	I did not feel like I could really trust my peers in playing the escape room.*
Q14	I like to play the escape room with my peers because we could achieve something together.
Q15	Playing the escape room provided more chances to interact with my peers.
Q16	Playing the escape room helped me to improve my grade.
Q17	I want others to know that I am a good student if I play the escape room well.
Q18	I want to play this escape room well so that my effort will be recognized by my classmates.

Items were grouped *a priori* into five subscales (Table 13) informed by theory and content review: *Interest/Enjoyment*, *Perceived Competence*, *Relatedness/Teamwork*, *Perceived Educational Value*, and *Recognition/Extrinsic Orientation*. The post-experience instrument uses a four-point Likert scale (1=*Strongly disagree*, 4=*Strongly agree*) and includes negatively worded items to mitigate acquiescence bias. Negatively worded items (Q3, Q8, Q13) are reverse coded as $x' = 5 - x$ for $x \in \{1, 2, 3, 4\}$. For each subscale, we compute the mean (higher scores indicate greater alignment with the target construct), standard deviation, and Cronbach's α; $\alpha \geq .70$ is treated as acceptable for exploratory use. Where appropriate, we report McDonald's ω to corroborate internal consistency. Calibration is examined by binning self-rated competence (low/medium/high) and comparing mean quiz scores across bins with 95% confidence intervals.

Table 13. Subscale composition and scoring rules for the DER questionnaire.

Subscale	Items	Reverse-Coded	Score (Range)
Interest / Enjoyment	Q1, Q2, Q3*, Q4, Q5	Q3	Mean of items
Perceived Competence	Q6, Q7, Q8*	Q8	Mean of items
Relatedness / Teamwork	Q12, Q13*, Q14, Q15	Q13	Mean of items
Perceived Educational Value	Q9, Q10, Q11, Q16	None	Mean of items
Recognition / Extrinsic Orientation	Q17, Q18	None	Mean of items

The instrument was administered immediately after the in-person DER experience via a web form embedded in the VLE. Item order was fixed for comparability across cohorts. Responses were anonymized at collection; telemetry linkage to task performance used a pseudonymous session identifier to enable calibration analyses without storing personally identifying information. Missing responses were rare; if present, subscale means were computed with

pairwise deletion provided at least half the subscale items were answered.

Primary outcomes include subscale means and 95% confidence intervals (CIs), internal consistency (α, ω), and correlations between Perceived Competence and objective indicators (e.g., quiz score, number of hints, time-on-task). We further examine calibration by binning self-rated competence (low/medium/high) and plotting average quiz performance per bin.

5. Results

We report outcomes for both the flipped virtual learning environment (VLE) and the subsequent in-person digital escape room (DER), focusing on (i) completion and engagement in the VLE, (ii) perceived understanding, competence, and social experience in the DER, and (iii) formative quiz performance at both aggregate and item levels. Unless stated otherwise, descriptive values are reported as mean (SD) with 95% confidence intervals (CIs) computed via normal approximation; proportions include Wilson CIs.

5.1 Completion and Engagement Metrics

Figure 11 presents a diverging stacked Likert chart summarizing participant responses to items Q1 through Q18 on a 4-point agreement scale. To aid interpretability, responses to reverse-coded items (Q3, Q8, Q13) were recoded so that all rightward bars indicate more favorable perceptions. The chart centers responses at the boundary between "Disagree" and "Agree," visually emphasizing attitudinal polarity. Notably, items such as Q5 and Q12 show strong agreement, suggesting a high level of endorsement for the corresponding statements. In contrast, items like Q3 and Q16 reveal a relatively higher proportion of disagreement, indicating areas of potential concern or misunderstanding. This visualization provides a clear, comparative view of perception distributions across the instrument, highlighting both strengths and possible targets for instructional refinement.

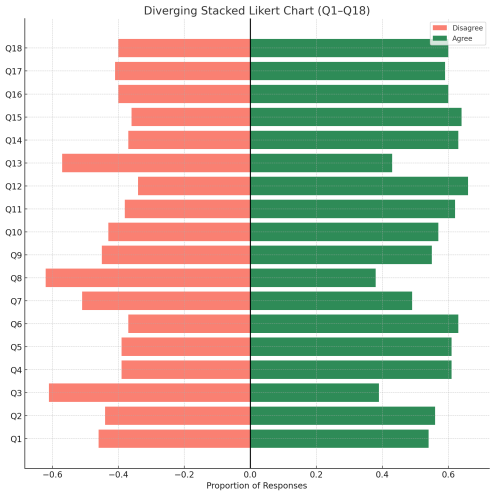


Figure 11. Visualization of item responses (Q1–Q18) to the DER questionnaire. Bars show the percentage distribution across the four categories (1–4), centered at the neutral boundary between *Disagree* and *Agree*. Reverse-coded items (Q3, Q8, Q13) are plotted after recoding so that rightward bars always indicate more favorable perceptions.

Overall completion of the flipped VLE was high (Table 14, Figure 12): 94% (45/48; 95% CI [83, 98]%). The median time-on-VLE was 56 min (IQR: 47–64), with 62% of university students using at least one hint (median 3 hints among users; IQR: 2–5). AI-assisted reflection was used by 68% of university students (at least once); among users, the median number of reflection prompts was 2. The proportion of learners who accessed at least one full solution and those who completed the VLE without requesting any full solution provides insight into the extent to which learners engaged in independent problem solving and productive struggle, thereby enabling a more accurate interpretation of the educational value of the scaffolded learning experience.

5.2 Perceived Understanding and Confidence

Post-experience ratings on a four-point Likert scale indicated positive subjective evaluations of the VLE across all considered subscales(Table 15, Figure 13). Interest/Enjoyment and Perceived Educational Value exhibited the highest means, while Perceived Competence and Relatedness/Teamwork were also above indifference levels on average. Internal consistency was acceptable to good ($\alpha \geq .72$). As the questionnaire used context-specific items rather than a full established instrument such as the Intrinsic Motivation Inventory [54], these results should be interpreted as

Table 14. Completion and engagement metrics for the flipped VLE phase.

Metric	Value with 95% CI / IQR
Completion rate (completed all rooms)	94% (45/48) [83%, 98%]
Session duration (minutes), median	56 IQR: 47–64
Activities completed per learner	18.9 (SD = 3.1) [18.0, 19.8]
Any hint usage (at least one hint)	62% (30/48) [47%, 75%]
Hints per hint-user, median	3 IQR: 2–5
Any AI reflection usage	68% (33/48) [53%, 80%]
Reflections per AI-user, median	2 IQR: 1–3
Submission attempts per item	1.6 (SD = 0.5) [1.5, 1.8]
Full-solution reveal usage (at least once)	29% (14/48) [18%, 43%]
Completed without any full-solution reveal	71% (34/48) [57%, 82%]

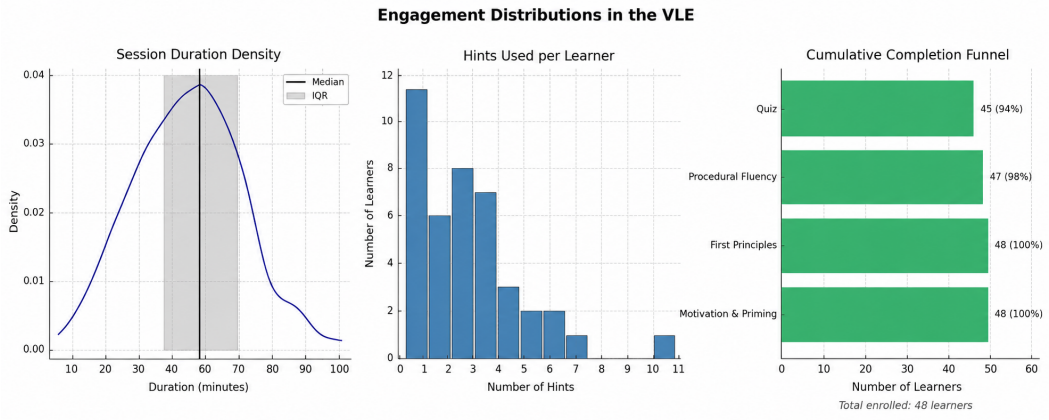


Figure 12. Engagement distributions. **Left:** session duration density with median (solid) and IQR (shaded). **Center:** per-learner histogram of total hints used. **Right:** cumulative completion funnel across rooms (Motivation & Priming → First Principles → Procedural Fluency → Quiz).

immediate learner perceptions of the DER experience rather than validated measures of intrinsic motivation.

Table 15. Perception subscales: descriptive statistics and exploratory internal consistency estimates (4-point Likert).

Subscale	Mean	SD	α	ω
Interest / Enjoyment	3.45	0.52	0.83	0.84
Perceived Competence	3.18	0.58	0.78	0.79
Relatedness / Teamwork	3.36	0.55	0.80	0.81
Perceived Educational Value	3.42	0.49	0.82	0.83
Recognition / Extrinsic Orientation	2.88	0.70	0.72	0.73

Calibration analysis showed that self-reported competence bins (low/medium/high) tracked average quiz scores monotonically; however, a modest overconfidence bias was observed in the medium bin (mean self-rating exceeded performance by ≈0.2 points on the 4-point scale).

5.3 Quiz Performance (Item-Level and Aggregate)

Aggregate quiz performance was 4.1 ± 0.8 out of 5. Item-level statistics (Table 16, there was a satisfactory difficulty spread (p-values 0.58–0.83) and positive discrimination (point-biserial correlations 0.31–0.52). Latency was longest for the first-principles derivation (Q3), consistent with its multi-step structure and validator gating.

Table 16. Item-level statistics for the formative quiz (N=48). *pbis* = point-biserial discrimination; latency in seconds.

Item	Construct	p	pbis	Latency	Hint Use
Q1	Concept: derivative-as-limit	0.83	0.38	38 (SD = 19)	8%
Q2	Recall: base derivatives	0.79	0.31	44 (SD = 21)	10%
Q3	First principles (x^2)	0.71	0.52	126 (SD = 48)	35%
Q4	Chain rule $\cos^2(x)$	0.65	0.45	72 (SD = 29)	24%
Q5	Extension (trig/ log-diff)	0.58	0.41	81 (SD = 34)	29%

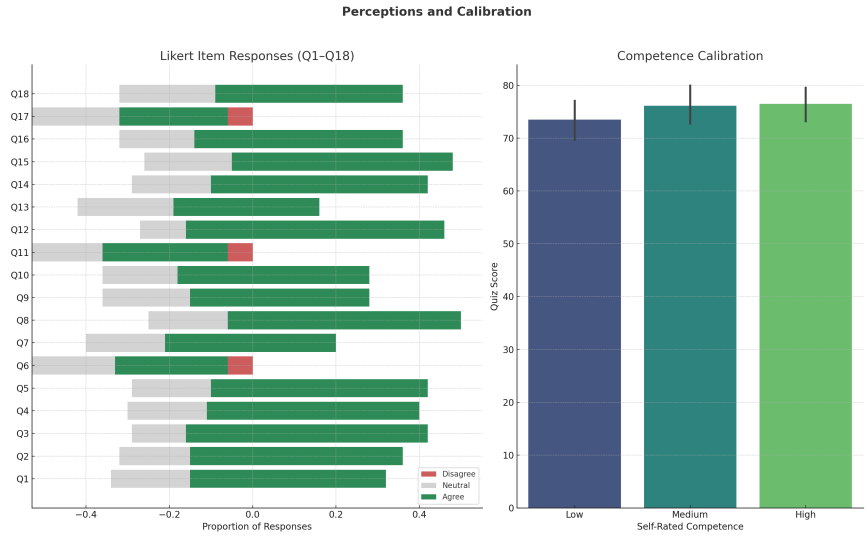


Figure 13. Perceptions and calibration. **Left:** diverging stacked Likert chart (items Q1–Q18, reverse-coded items recoded so rightward indicates more favorable). **Right:** competence calibration: quiz score by self-rated competence bin (low/medium/high) with error bars.

Across learners, total hints used exhibited a small negative correlation with total quiz score ($r = -0.26$, 95% CI $[-0.51, -0.01]$), consistent with hints being concentrated where difficulty was higher. Time-on-VLE showed a weak positive correlation with score ($r = 0.19$, 95% CI $[-0.10, 0.45]$). Learners who used AI-assisted reflections (at least once) had marginally higher scores ($d = 0.36$), though this difference was not statistically conclusive at $\alpha = 0.05$.

Validator flags corroborated the misconception catalog: the most frequent structural error in Q3 was failure to factor the common h before limit evaluation; in Q4, a subset applied the product rule to a composition; in Q5 (log-differentiation variant), misuse of $\ln(a^b)$ appeared in 21% of incorrect attempts. Post-quiz feedback linked each error to a specific worked-example panel, enabling targeted review.

Internal consistency of the procedural subset (Q2, Q4, Q5) yielded $\omega = 0.76$ (95% CI $[0.62, 0.86]$). Omitting Q3 (multi-step) from the aggregate score did not alter the qualitative pattern of associations with engagement measures. Sensitivity analyses using nonparametric correlations (Spearman's ρ) produced similar magnitudes and inferences.

6. Discussion and Conclusions

study was guided by three research questions. First, we asked how a scaffolded virtual learning environment can be designed to help first-year undergraduate students connect the derivative-as-limit concept with rule-based differentiation fluency. Second, we asked how responsible AI-assisted reflection can be embedded in mathematics learning in a way that supports metacognition rather than answer substitution. Third, we asked what feasibility, engagement, perception, and formative performance patterns emerge when such a VLE is used as flipped preparation for an in-person digital escape room activity. The findings provide initial answers to these questions by showing that a staged sequence moving from conceptual priming to first-principles derivations, procedural rule application, logarithmic differentiation, formative quiz tasks, and collaborative DER problem solving can be implemented with high completion, sustained engagement, positive perceptions, and interpretable diagnostic error patterns. The results also indicate that the proposed approach is feasible to implement and is associated with positive learner engagement, perceived competence, and perceived educational value.

The main advancement over the state of the art is not merely the use of a virtual environment, gamification, or an escape room, but the explicit instructional orchestration of these elements into a reproducible calculus-learning model. Previous VLE and DER studies often emphasize immersion, engagement, or general motivation, while providing limited detail on the micro-level instructional mechanics that connect interaction design to conceptual change. In contrast, the present study specifies how learning objectives are mapped to rooms, task types, validator checks, hint tiers, worked-solution reveal conditions, telemetry events, misconception flags, and post-task reflection. This makes the proposed VLE–DER model transferable and inspectable, especially for instructors seeking to teach derivatives through a structured combination of conceptual anchoring and procedural fluency.

The second advancement concerns the integration of AI-assisted reflection into a mathematics learning sequence. Rather than treating large language models as unrestricted answer generators, the revised design positions AI as a reflective support tool for explanation, verification, and metacognitive monitoring. This is operationalized through explicit responsible-use prompts, reflection checkpoints, and worked-solution gates that require learners to articulate

what they did not understand before viewing complete solutions. In this way, AI support is aligned with autonomy and competence while preserving the learner's own mathematical effort.

The flipped VLE achieved high completion with sustained engagement at conceptual checkpoints. Students reported strong interest, perceived educational value, and solid perceived competence; calibration analyses aligned self-ratings and performance. The formative quiz exhibited reasonable item difficulty spread and discriminative capacity, with error patterns that matched anticipated misconceptions and were addressed via targeted feedback loops.

A limitation of this pilot study is that the post-DER questionnaire used short, context-specific items rather than a validated instrument such as the Intrinsic Motivation Inventory. Thus, the subscale scores can be interpreted as exploratory perception indicators, and future studies should use validated motivation scales. The pilot evidence can be interpreted as feasibility and design evidence rather than causal proof of learning gains. The high completion rate, frequent use of hints, use of AI-assisted reflection, positive DER perception scores, and item-level quiz patterns suggest that the environment was usable, engaging, and diagnostically informative. However, because the study did not include a control group or pre–post performance measures, it cannot establish that the VLE or DER caused improvements in conceptual understanding or procedural fluency. Future work could compare the proposed VLE–DER model with traditional flipped instruction, non-gamified VLE preparation, and DER-only formats using controlled pre–post designs, delayed retention tests, and larger multi-cohort samples.

References

[1] F. Feudel and R. Biehler, "Students' understanding of the derivative concept in the context of mathematics for economics," *Journal für Mathematik-Didaktik*, vol. 42, pp. 273–305, 2020. doi: 10.1007/s13138-020-00174-z .

[2] S. Jones and K. Watson, "Recommendations for a "target understanding" of the derivative concept for first-semester calculus teaching and learning," *International Journal of Research in Undergraduate Mathematics Education*, vol. 4, pp. 199–227, 2017. doi: 10.1007/s40753-017-0057-2 .

[3] M. Kertil and H. G. Dede, "Promoting prospective mathematics teachers' understanding of derivative across different real-life contexts," *International Journal for Mathematics Teaching and Learning*, 2022. doi: 10.4256/ijmtl.v23i1.361 .

[4] N. A. Talib, S. S. Abdul Rahim, and H. Zulnaidi, "Slope conception as rate of change starter kit: Malaysian pre-service secondary mathematics teachers' subject matter knowledge in rate of change," *Eurasia Journal of Mathematics, Science and Technology Education*, 2023. doi: 10.29333/ejmste/13862 .

[5] F. Yu, "Promoting productive understandings of rate of change in calculus courses," *PRIMUS*, vol. 33, pp. 965–980, 2023. doi: 10.1080/10511970.2023.2214891 .

[6] X. Lan and Y. Zhou, "Teaching derivative concept using 6 questions cognitive model," *Journal of Didactic Mathematics*, vol. 1, no. 3, pp. 127–137, 2020. doi: 10.34007/JDM.V1I3.371 .

[7] L. S. Vygotsky, *Mind in Society: The Development of Higher Psychological Processes*. Cambridge, MA: Harvard University Press, 1978.

[8] D. Wood, J. S. Bruner, and G. Ross, "The role of tutoring in problem solving," *Journal of Child Psychology and Psychiatry*, vol. 17, no. 2, pp. 89–100, 1976. doi: 10.1111/j.1469-7610.1976.tb00381.x .

[9] E. L. Deci and R. M. Ryan, *Intrinsic Motivation and Self-Determination in Human Behavior*. New York, NY: Springer, 1985.

[10] E. Kasneci, K. Sessler, S. Küchemann, M. Bannert, D. Dementieva, F. Fischer, U. Gasser, G. Kasneci, S. Krusche, and U. Schmid, "Chatgpt for good? on opportunities and challenges of large language models for education," *Learning and Individual Differences*, vol. 103, p. 102274, 2023. doi: 10.1016/j.lindif.2023.102274 .

[11] Y. R. S. Leuyacc and J. L. C. Parejas, "Developing student understanding of calculus concepts through flipped classroom," in *2021 XVI Latin American Conference on Learning Technologies (LACLO)*, 2021, pp. 494–497. doi: 10.1109/LACLO54177.2021.00087 .

[12] A. Istiandaru, F. Setyawan, A. Hidayat, and V. Istihapsari, "Learning differential calculus using self-regulated flipped classroom approach," *Journal of Physics: Conference Series*, vol. 1321, 2019. doi: 10.1088/1742-6596/1321/2/022134 .

[13] N. H. Wasserman, C. Quint, S. A. Norris, and T. Carr, "Exploring flipped classroom instruction in calculus iii," *International Journal of Science and Mathematics Education*, vol. 15, pp. 545–568, 2015. doi: 10.1007/s10763-015-9704-8 .

[14] Ramadoni and Mustofa, "Enhancing flipped classroom with peer teaching to promote students' conceptual understanding and self-efficacy in calculus courses," *Pegem Journal of Education and Instruction*, 2022. doi: 10.47750/pegegog.12.03.17 .

[15] Y. Zhang, M. A. Feijoo-Garcia, Y. Gu, V. Popescu, B. Benes, and A. J. Magana, "Virtual and augmented reality in science, technology, engineering, and mathematics (stem) education: An umbrella review," *Information*, vol. 15, no. 9, p. 515, Aug. 2024. doi: 10.3390/info15090515 .

[16] K. Huang, C.-H. Chen, W.-S. Wu, and W.-Y. Chen, "Interactivity of question prompts and feedback on secondary students' science knowledge acquisition and cognitive load," *Journal of Educational Technology & Society*, vol. 18, no. 4, pp. 159–171, 2015.

[17] J. G. Cromley, R. Chen, and L. Lawrence, "Meta-analysis of stem learning using virtual reality: Benefits across the board," *Journal of Science Education and Technology*, vol. 32, no. 3, p. 355–364, Mar. 2023. doi: 10.1007/s10956-023-10032-5 .

[18] C. Papasarantou, R. Alimisi, and D. Alimisis, "Virtual galleries as learning scaffolds for promoting problem-based learning," *Education Sciences*, vol. 13, no. 12, p. 1168, Nov. 2023. doi: 10.3390/educsci13121168 .

[19] J. S. Liang, "Modeling an immersive vr driving learning platform in a web-based collaborative design environment," *Computer Applications in Engineering Education*, vol. 20, no. 3, p. 553–567, Aug. 2012. doi: 10.1002/cae.20424 .

[20] D. Parmar, L. Lin, N. DSouza, S. Jörg, A. E. Leonard, S. B. Daily, and S. V. Babu, "How immersion and self-avatars in vr affect learning programming and computational thinking in middle school education," *IEEE Transactions on Visualization and Computer Graphics*, vol. 29, no. 8, p. 3698–3713, Aug. 2023. doi: 10.1109/tvcg.2022.3169426 .

[21] W. Huang, R. D. Roscoe, M. C. Johnson-Glenberg, and S. D. Craig, "Motivation, engagement, and performance across multiple virtual reality sessions and levels of immersion," *Journal of Computer Assisted Learning*, vol. 37, no. 3, p. 745–758, Dec. 2020. doi: 10.1111/jcal.12520 .

[22] R. Kaźmierczak, G. Grunwald, R. Skowroński, L. Kaźmierczak, and C. Kowalczyk, "Augmented reality tools for mathematics and geoscience education," *Scientific Reports*, vol. 15, no. 1, May 2025. doi: 10.1038/s41598-025-02090-z .

[23] A. Pregowska, M. Osial, and A. Gajda, "What will the education of the future look like? how have metaverse and extended reality affected the higher education systems?" *Metaverse Basic and Applied Research*, 2023. doi: 10.56294/mr202457 .

[24] A. Glavaš, "A socio-constructivist conceptual and design framework for educational escape room games," *Education Sciences*, 2026. doi: 10.3390/educsci16030375 ., all Open Access; Gold Open Access; Green Open Access. [Online]. Available: https://www.scopus.com/pages/publications/105033958112?origin=resultslist

[25] L. H. Taraldsen, "Introducing teacher education students to escape room as a didactic tool in mathematics." *Mathematics Teaching-Research Journal*, 2025. [Online]. Available: https://www.scopus.com/pages/publications/105012042441?origin=resultslist

[26] P. J. Roig, S. Alcaraz, K. Gilly, C. Bernad, and C. Juiz, "A ctf-like escape room for stem assessment," *Education Sciences*, 2026. doi: 10.3390/educsci16040616 ., all Open Access; Gold Open Access; Green Open Access. [Online]. Available: https://www.scopus.com/pages/publications/105036823023?origin=resultslist

[27] F. Luque-Sánchez and J. Montejo-Gámez, "A virtual escape room for the enhancement of mathematical communication in secondary education," *Technology, Knowledge and Learning*, 2026. doi: 10.1007/s10758-023-09706-1 . [Online]. Available: https://www.scopus.com/pages/publications/85178210104?origin=resultslist

[28] C.-C. Chang and S.-T. Yang, "Interactive effects of scaffolding digital game-based learning and cognitive style on adult learners' emotion, cognitive load and learning performance," *International Journal of Educational Technology in Higher Education*, vol. 20, no. 1, Mar. 2023. doi: 10.1186/s41239-023-00385-7 . [Online]. Available: http://dx.doi.org/10.1186/s41239-023-00385-7

[29] W.-S. Wang, C.-J. Lin, H.-Y. Lee, T.-T. Wu, and Y.-M. Huang, "Feedback mechanism in immersive virtual reality influences physical hands-on task performance and cognitive load," *International Journal of Human–Computer Interaction*, vol. 40, no. 15, p. 4103–4115, May 2023. doi: 10.1080/10447318.2023.2209837 .

[30] G. Cosentino, J. Anton, K. Sharma, M. Gelsomini, M. Giannakos, and D. Abrahamson, "Generative ai and multimodal data for educational feedback: Insights from embodied math learning," *British Journal of Educational Technology*, vol. 56, no. 5, pp. 1686–1709, 2025. doi: 10.1111/bjet.13587 .

[31] S. Steinert, K. E. Avila, S. Ruzika, J. Kuhn, and S. Kuchemann, "Harnessing large language models to develop research-based learning assistants for formative feedback," *Smart Learning Environments*, vol. 11, no. 1, p. 62, 2024. doi: 10.1186/s40561-024-00354-1 .

[32] H. Kumar, R. Xiao, B. Lawson, I. Musabirov, J. Shi, X. Wang, H. Luo, J. J. Williams, A. N. Rafferty, J. Stamper, and M. Liut, "Supporting self-reflection at scale with large language models: Insights from randomized field experiments in classrooms," in *Proceedings of the Eleventh ACM Conference on Learning @ Scale*, 2024, pp. 86–97. doi: 10.1145/3657604.3662042 .

[33] Y. Wang and Y. Zhao, "Metacognitive prompting improves understanding in large language models," *ArXiv*, vol. abs/2308.05342, 2023. doi: 10.48550/arXiv.2308.05342 .

[34] F. Scholten, T. Rebholz, and M. Hütter, "Metacognitive myopia in large language models," *ArXiv*, vol. abs/2408.05568, 2024. doi: 10.48550/arXiv.2408.05568 .

[35] L. Zhang and W. Zhang, "Integrating large language models into project-based learning based on self-determination theory," *Interactive Learning Environments*, 2024. doi: 10.1080/10494820.2024.2446535 .

[36] H. Li, J. Yu, X. Cong, Y. Dang, Y. Zhan, H. Liu, and Z. Liu, "Exploring llm-based student simulation for metacognitive cultivation," *arXiv preprint arXiv:2502.11678*, 2025. doi: 10.48550/arXiv.2502.11678 .

[37] R. Hunicke, M. LeBlanc, and R. Zubek, "Mda: A formal approach to game design and game research," in *Proceedings of the AAAI Workshop on Challenges in Game AI*, 2004, pp. 1–5.

[38] J. Bakke and B. E. Lewis, "Escaping the ordinary: Mechanics, dynamics, and aesthetic-guided design and evaluation of an escape room for high yield review in basic medical sciences," *Anatomical Sciences Education*, Aug. 2025. doi: 10.1002/ase.70100 . [Online]. Available: http://dx.doi.org/10.1002/ase.70100

[39] F. Yllana-Prieto, J.-S. Jeong, and D. González-Gómez, "An online-based edu-escape room: A comparison study of a multidimensional domain of psts with flipped sustainability-stem contents," *Sustainability*, 2021. doi: 10.3390/SU13031032 .

[40] A. Mathieson and E. Duca, "Stem escape rooms for public engagement," *Research for All*, 2021. doi: 10.14324/rfa.05.2.10 .

[41] S. Pais, A. Sousa, and A. Pires, "Using an escape room activity to enhance the motivation of undergraduate life science students in mathematics classes - a case study," in *European Conference on Games Based Learning*, 2023. doi: 10.34190/ecgbl.17.1.1431 .

[42] T. Sidekerskiene and R. Damaševičius, "Out-of-the-box learning: Digital escape rooms as a metaphor for breaking down barriers in stem education," *Sustainability*, 2023. doi: 10.3390/su15097393 .

[43] S. Pennucci, "Flip it and break it: Using flipped lessons and breakouts to energize learning," *Pennsylvania Libraries: Research & Practice*, 2020. doi: 10.5195/palrap.2020.220 .

[44] S. P. Dimeo, C. Astemborksi, J. Smart, and E. L. Jones, "A virtual escape room versus lecture on infectious disease content: Effect on resident knowledge and motivation," *Western Journal of Emergency Medicine*, vol. 23, no. 1, pp. 9–14, 2022. doi: 10.5811/westjem.2021.12.54010 .

[45] D. Tall and S. Vinner, "Concept image and concept definition in mathematics with particular reference to limits and continuity," *Educational Studies in Mathematics*, vol. 12, no. 2, pp. 151–169, 1981. doi: 10.1007/BF00305619 .

[46] M. Carlson, S. Jacobs, E. Coe, S. Larsen, and E. Hsu, "Applying covariational reasoning while modeling dynamic events," *Journal for Research in Mathematics Education*, vol. 33, no. 5, pp. 352–378, 2002. doi: 10.2307/4149958 .

[47] D. Tall, *How Humans Learn to Think Mathematically: Exploring the Three Worlds of Mathematics*. Cambridge University Press, 2013. doi: 10.1017/CBO9781139565202 .

[48] T. Kim, J. Planey, and R. Lindgren, "Theory-driven design in metaverse virtual reality learning environments: Two illustrative cases," *IEEE Transactions on Learning Technologies*, vol. 16, pp. 1141–1153, 2023. doi: 10.1109/TLT.2023.3307211 .

[49] D. Beck, L. Morgado, and P. M. O'Shea, "Educational practices and strategies with immersive learning environments: Mapping of reviews for using the metaverse," *IEEE Transactions on Learning Technologies*, vol. 17, pp. 319–341, 2024. doi: 10.1109/TLT.2023.3243946 .

[50] K. Almeman, F. EL Ayeb, M. Berrima, B. Issaoui, and H. Morsy, "The integration of ai and metaverse in education: A systematic literature review," *Applied Sciences*, 2025. doi: 10.3390/app15020863 .

[51] M. Fadhel, A. M. Duhaim, A. Albahri, Z. Al-qaysi, M. A. Ahmed, M. Chyad, W. Abd-Alaziz, O. Albahri, A. Alamoodi, L. Alzubaidi, A. Gupta, and Y. Gu, "Navigating the metaverse: Unraveling the impact of artificial intelligence - a comprehensive review and gap analysis," *Artificial Intelligence Review*, vol. 57, p. 264, 2024. doi: 10.1007/s10462-024-10881-5 .

[52] J. Solanes, S. Montava-Jorda, E. Golf-Laville, V. Colomer-Romero, L. Gracia, and A. Muñoz, "Enhancing stem education through interactive metaverses: A case study and methodological framework," *Applied Sciences*, 2023. doi: 10.3390/app131910785 .

[53] X. Chen, D. Zou, H. Xie, and F. L. Wang, "Metaverse in education: Contributors, cooperations, and research themes," *IEEE Transactions on Learning Technologies*, vol. 16, pp. 1111–1129, 2023. doi: 10.1109/TLT.2023.3277952 .

[54] R. M. Ryan, "Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory," *Journal of Personality and Social Psychology*, vol. 43, no. 3, pp. 450–461, 1982. doi: 10.1037/0022-3514.43.3.450 .