From Engagement to Meaning: Affective-Semantic Reinforcement for Sustainable Study Habits in Intelligent Educational Systems

Helena Macedo Reis
Federal University of São Carlos (UFSCar)
São José do Rio Preto, Brazil
helena@ufscar.br

Anderson da Silva Marcolino
Federal University of Paraná (UFPR)
Palotina, Brazil
anderson.marcolino@ufpr.br

Alan Cesar de Oliveira Mathias
Universidade Estadual de Maringá (UEM)
Maringá, Brazil
pg407588@uem.br

Katyeudo Karlos de Sousa Oliveira
Instituto Federal do Sertão Pernambucano (IFSertãoPE)
Petrolina, Brazil
katyeudo.karlos@ifsertao-pe.edu.br

Abstract

Sustaining study habits over time remains a central challenge in digital and self-directed learning environments, including online Software Engineering education. This paper proposes the HARAS model (Habits Anchored by Affective-Semantic Reinforcement), a conceptual and formal framework distinguishing observable engagement, habit strength, and subjective appreciation of studying, integrated through dynamic affective mechanisms organized into three progressively internalized layers comprising Activating, Regulatory, and Semantic Affective Reinforcement. Sustainable habits emerge from a planned transition across these layers rather than from the continuous intensification of external stimuli. The model is operationalized through a mathematical formalization based on latent states and temporal dynamics, an AI reference architecture with signal-to-inference mappings, two worked scenarios in structurally distinct educational contexts, and an empirical validation roadmap. A heuristic evaluation with six domain experts yielded an overall mean of 4.48 ($SD = 0.26$; $\alpha = 0.707$) across eighteen structured items in four thematic blocks, with block-level means ranging from 4.17 (Implementation Feasibility) to 4.71 (Theoretical-Educational Plausibility). Thematic analysis of five open-ended items identified the separation between engagement and appreciation and the planned AAR→SAR transition as the most valued contributions, while the absence of social and collaborative dimensions emerged as the primary structural limitation ($f = 5/6$).

Keywords

Study Habits, Affective Reinforcement, Motivational Internalization, Intelligent Educational Systems, Software Engineering Education, Heuristic Evaluation

1 Introduction

The creation and maintenance of study habits have been a long-standing concern in education, particularly in digital and self-directed learning environments. This challenge is especially salient in Software Engineering (SE) education, where online platforms, blended learning environments, and continuous training programs demand sustained independent engagement from learners over extended periods [3]. Many students struggle to remain engaged over time, abandoning their study routines once initial enthusiasm fades. In response, several educational platforms have adopted habit formation strategies [15], such as gamification, reminders, and reward systems [6], with the aim of reinforcing repeated study behavior.

Although such approaches are effective in promoting short-term persistence, growing evidence indicates that behavioral continuity does not necessarily guarantee the development of a positive affective relationship with learning [26]. Students may sustain study routines due to external pressures or reward structures without developing intrinsic interest or enjoyment [6]. This dissociation between *studying* and *liking to study* raises fundamental questions for the design of intelligent educational systems: *Is maintaining the habit sufficient? If not, how can technology simultaneously support behavioral persistence and the construction of a positive affective relationship with studying?*

Research in motivation theory, academic emotions, and gamification suggests that externally regulated behaviors can become habitual without being fully internalized [21]. Systems optimized for retention and daily use may inadvertently foster dependence on extrinsic reinforcement, leading to disengagement when such stimuli are reduced or removed. Nevertheless, most AI-based educational systems continue to focus on observable engagement metrics, such as access frequency or task completion, while neglecting the affective quality of the learning experience and the meaning attributed to studying [12].

By integrating affective computing, habit theory, and AI-based adaptation, this work contributes to the track on Innovative Environments and Methods for Software Engineering Education, with particular relevance to online learning platforms and AI-supported training tools. The objective of this article is to formalize and operationalize the HARAS model (Habits Anchored by Affective-Semantic Reinforcement) as a framework for designing intelligent educational systems that go beyond engagement optimization, explicitly modeling the dynamics between behavior, affect, and meaning. Methodologically, the study adopts a mixed approach that combines conceptual modeling, mathematical formalization of latent states and temporal dynamics, the definition of an AI-based reference architecture, and a heuristic evaluation conducted with domain experts. The proposal encompasses (i) a layered model of affective-semantic reinforcement (AAR, RAR, SAR), (ii) its computational translation into observable signals and inference mechanisms, and (iii) an empirical validation roadmap grounded in testable hypotheses. Throughout the paper, we present the theoretical foundations of the model, its formal specification, the system architecture, illustrative application scenarios, and the results of the expert-based evaluation, outlining both its contributions and directions for future empirical validation.

This work is guided by the following research question: *RQ: How can intelligent educational systems support the formation of sustainable study habits, going beyond observable engagement by jointly modeling habit strength, affective state, and subjective appreciation of studying?*

This question is addressed progressively throughout the paper by: (i) a conceptual framework distinguishing engagement, habit strength, and subjective appreciation (Section 3); (ii) affective reinforcement organized into three progressively internalized layers, AAR, RAR, and SAR (Section 3); (iii) a mathematical formalization through latent state dynamics (Section 4); (iv) an AI-based reference architecture mapping latent states to observable signals (Section 5); and (v) a heuristic evaluation with domain experts (Section 8).

In addressing this question, HARAS contributes to Software Engineering Education by providing: (i) **conceptual model** that distinguishes observable engagement, habit strength, and subjective appreciation of studying as separable constructs; (ii) a **formal mathematical model** of the temporal dynamics linking behavior, affect, and habit consolidation; (iii) an **AI-based operational architecture** mapping latent states to observable platform signals and inference methods; (iv) a **layer-transition intervention policy** that modulates the *type* of affective reinforcement (AAR, RAR, SAR) as the habit consolidates, rather than merely its intensity; and (v) an **empirical validation roadmap** connecting each research hypothesis to a feasible experimental design.

The originality of HARAS lies not in introducing entirely new constructs, but in organizing, formalizing, and operationalizing established constructs into a layer-specific intervention policy for sustaining study habits and motivational internalization.

2 Related Work and Positioning

Habit Formation and Self-Regulated Learning: Habits can be defined as patterns of behavior acquired through the consistent repetition of actions in relatively stable contexts, becoming progressively automatic over time [15]. Models such as the habit loop describe this process as the interaction between cue, routine, and reward [25]. As habits consolidate, behavioral control shifts from deliberative to automatic mechanisms, promoting cognitive economy [8]. Repetition therefore explains behavioral persistence but does not determine its motivational or affective quality.

Self-Determination Theory positions motivation along a continuum from external regulation to intrinsic motivation [21]. Within this framework, a behavior may be stable and automated without being meaningful, as long as it is sustained by external contingencies. Zimmerman's model of Self-Regulated Learning further distinguishes mechanical from motivational self-regulation [26]: only the latter supports deep learning, long-term persistence, and adaptation to less structured contexts.

Affect-Aware Learner Models and Intelligent Tutoring Systems: The AIED (Artificial Intelligence in Education) community has produced a rich tradition of affect-aware learner models. AutoTutor [7] pioneered the detection of confusion and frustration from conversational cues, demonstrating that transient affective states influence learning outcomes and warranting real-time adaptive responses. MetaTutor [1] extended this approach by incorporating self-regulated learning monitoring, tracking metacognitive strategies alongside affective signals. Conati and Maclaren [4] proposed probabilistic learner models using Dynamic Bayesian Networks to infer and respond to affective states during problem solving. Baker et al. [5] established sensor-free affect detection from behavioral logs, enabling engagement and disengagement monitoring without specialized hardware.

HARAS is positioned differently from all of these systems in three respects. First, whereas existing affect-aware tutors focus on transient emotional episodes (a moment of confusion, a frustration peak) and respond to them reactively, HARAS models the evolution of study habits across extended time scales, treating affect as a mediating variable in long-term behavioral consolidation. Second, while systems such as MetaTutor and AutoTutor aim to optimize immediate learning performance, HARAS targets a distinct objective: the internalization of studying as a valued, autonomous activity, captured through the subjective appreciation of learning. Third, HARAS formalizes a layer-specific intervention policy that transitions from Activating to Regulatory to Semantic Affective Reinforcement, distinguishing it from systems that adapt the *intensity* of reinforcement without modulating its *type* across the habit formation timeline.

Emotions, Engagement, and Learning Analytics: Emotions influence attention, memory, and persistence, modulating the depth of cognitive processing and the construction of meaning [14, 18]. In the absence of positive emotional involvement, study activities tend to assume an instrumental character, even when maintained consistently [16]. Despite this evidence, many Learning Analytics systems remain oriented toward superficial behavioral engagement metrics, relying on indicators such as clicks and time on task as proxies for learning [10]. Several studies have criticized this reductionist approach, arguing that such indicators neglect affective, cognitive, and self-regulatory dimensions of learning [11, 22].

HARAS responds to this limitation by integrating affective dynamics into the habit formation model, making these constructs theoretically grounded and computationally tractable.

3 The HARAS Model

The HARAS model was developed through a theory-driven integration of existing concepts from habit formation, affective computing, and motivational theories. Its core contribution lies in organizing these elements into three layers of affective reinforcement (Activating, Regulatory, and Semantic) designed to reflect the progressive internalization of motivation observed in the literature. These layers were defined to address the limitation of engagement-centered approaches by explicitly distinguishing between initiating behavior, stabilizing routines, and fostering meaning, thus enabling a more complete representation of how study habits are formed and sustained over time.

Scope delimitation. The current version of the model deliberately restricts its scope to *individual* behavioral and affective dynamics, in order to make the framework conceptually clear and formalizable in a first version. Social and collaborative dimensions are therefore not represented in the present state space, even though they are especially important in SE education, where pair programming, code review, and team projects are central to professional formation. This is a structural limitation of the current version, not a definitive theoretical decision: Section 10 outlines how future extensions should incorporate peer interaction quality, collaborative engagement, peer feedback, and sense of professional belonging.

3.1 Central Premise

HARAS proposes that sustainable study habits emerge when affective reinforcement strategies evolve from external triggers to progressively more internalized forms of regulation and meaning:

Study habits become more stable as their support shifts from external reinforcement to personal value and meaning attributed to learning. The model organizes affective reinforcement into three layers along a continuum of internalization (Figure 1).

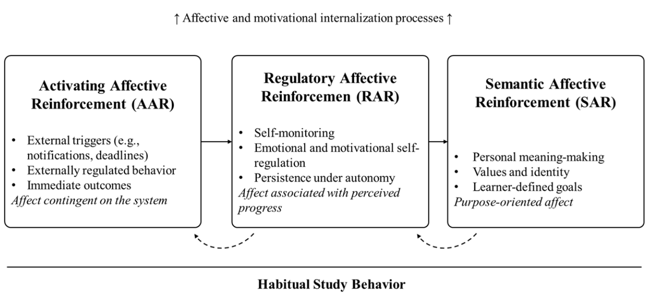

Figure 1: The HARAS model: AAR: Activating; RAR: Regulatory; SAR: Semantic.

3.2 Affective Reinforcement Layers

Table 1 summarizes the three reinforcement layers and their main characteristics.

Table 1: Affective Reinforcement Layers in the HARAS Model

Layer	Core Mechanism	Function	Target Affect
Activating (AAR)	Salient triggers, extrinsic rewards	Initiate behavior	Excitement, urgency
Regulatory (RAR)	Frustration reduction, predictability	Consolidate routines	Safety, self-efficacy
Semantic (SAR)	Autonomy, values, meaning	Internalize the habit	Interest, purpose

Activating Affective Reinforcement (AAR) corresponds to the most immediate layer of engagement, based on highly salient external stimuli and short-term rewards. It reduces initial inertia and increases action initiation likelihood, but tends to produce fragile sustainment when used in isolation.

Regulatory Affective Reinforcement (RAR) operates on the quality of the emotional experience during task execution by promoting predictability, reduced frustration, and perceived control. Its main function is to consolidate study routines and foster safety, calmness, and self-efficacy.

Semantic Affective Reinforcement (SAR) represents the deepest layer. Engagement is sustained through meaning attribution, autonomy, and alignment with personal values and identity. SAR makes the habit progressively self-sustaining and less dependent on external stimuli.

Transition between layers. HARAS conceptualizes habit formation as a dynamic process of transition between these layers (Figure 2). Over time, the relative weight of AAR decreases while RAR and, in particular, SAR assume a growing role.

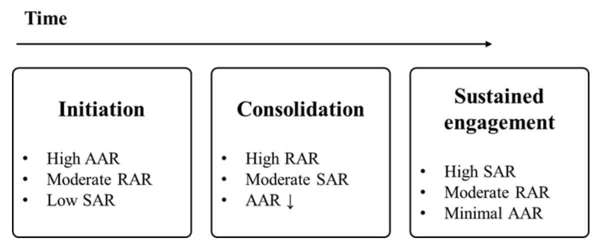

Figure 2: Progressive transition of affective reinforcement.

3.3 Instantiation in Software Engineering Education

At the level of its abstract constructs (engagement, habit strength, affective state, and subjective appreciation), HARAS is domain-independent by design. Its instantiation in SE education, however, is concrete and specific. In SE courses, sustainable study habits involve repeated programming practice, code submission, debugging, error recovery, help-seeking, incremental problem solving, and persistence under frequent corrective feedback. These activities provide SE-specific candidate signals for observing engagement, habit formation, and possible affective disengagement: submission regularity and debugging session patterns feed the estimation of H_{it}, while error-recovery trajectories and help-seeking behavior feed the inference of A_{it} (Section 5).

HARAS differs from a generic intelligent learning framework in the way the Semantic Affective Reinforcement layer connects to SE learning dynamics, the development of programming competence, the use of authentic software problems, and the construction of a professional identity as a software engineer (Section 5.3). The model is not intended to promote general platform use, but to support the internalization of technical study practices that are cumulative, effortful, and frustrating in SE learning.

4 Mathematical Formalization

To operationalize the HARAS model and enable its use in intelligent educational systems, we propose a mathematical formalization that captures the dynamic interaction between behavior, affect, and habit formation over time. The model represents the learner's state through a set of latent and observable variables, allowing the specification of temporal dependencies and the effects of different types of affective reinforcement. This formalization provides the basis for inference, prediction, and adaptive intervention policies in AI-driven learning environments.

4.1 State Variables

Consider a learner i observed over discrete time steps t. The learner's state is described by:

- $E_{it} \in \{0, 1\}$: observable engagement indicator at time t;
- $H_{it} \in [0, 1]$: latent habit strength at the beginning of time t;
- $A_{it} \in \mathbb{R}$: latent affective state, as deviation from a neutral baseline;
- $R_{it}^{(1)}, R_{it}^{(2)}, R_{it}^{(3)} \in [0, 1]$: intensities of activating, regulatory, and semantic reinforcement;
- X_{it}: observable control variables (e.g., task difficulty, time of day).

4.2 Engagement Probability

$$\Pr(E_{it} = 1) = \sigma(\eta_{it}), \quad \eta_{it} = \alpha_i + \beta_H H_{it} + \beta_A A_{it} + \sum_{k=1}^{3} \beta_k R_{it}^{(k)} + \boldsymbol{\gamma}^\top X_{it}. \tag{1}$$

Higher habit strength (H_{it}) and favorable affective states ($A_{it} > 0$) increase engagement probability. Different reinforcement types amplify or attenuate this probability depending on their intensity and the learner's profile.

4.3 Affective State Dynamics

$$A_{it} = \delta_0 + \sum_{k=1}^{3} \delta_k R_{it}^{(k)} + \delta_E E_{i,t-1} + u_{it}, \quad u_{it} \sim \mathcal{N}(0, \sigma_u^2). \tag{2}$$

The term δ_E captures the idea that recent study experiences improve subsequent affect through perceived competence or self-efficacy, completing a positive feedback loop between behavior and emotion.

4.4 Habit Dynamics

$$H_{i,t+1} = (1 - \lambda) H_{it} + \rho\, E_{it}\, g\Big(R_{it}^{(2)}, R_{it}^{(3)}\Big) + \rho_1\, E_{it}\, R_{it}^{(1)}, \tag{3}$$

with $\lambda \in (0, 1)$, $\rho, \rho_1 > 0$, and

$$g\Big(R_{it}^{(2)}, R_{it}^{(3)}\Big) = w_2 R_{it}^{(2)} + w_3 R_{it}^{(3)}, \quad w_3 > w_2 > 0. \tag{4}$$

Habit strength decays in the absence of practice (λ) and increases only when the learner engages. Activating reinforcement contributes to immediate habit strengthening, while regulatory and semantic reinforcement exert a more durable effect ($w_3 > w_2$).

4.5 Observable Appreciation Variable

$$L_{it} = \theta_0 + \theta_H H_{it} + \theta_2 R_{it}^{(2)} + \theta_3 R_{it}^{(3)} - \theta_1 R_{it}^{(1)} + \varepsilon_{it}. \tag{5}$$

Habit consolidation and regulatory and semantic reinforcement contribute positively to learning appreciation, whereas excessive reliance on activating reinforcement may sustain behavior without promoting enduring interest.

5 AI-Based Architecture

The architecture translating HARAS into a deployable system is organized into three cyclically operating components. While each component builds on established AIED approaches, their integration within the HARAS layer-transition framework constitutes the novel architectural contribution.

5.1 Observation and Signal Extraction

Observable proxies for the latent states are collected from platform interaction logs. Table 2 maps each latent state to concrete observable signals and candidate inference methods.

Table 2: States, Observable Signals, and Inference Approaches

Latent State	Observable Signals	Inference Method	Example Tools
H_{it} (Habit strength)	Session frequency, streak regularity, session initiation time consistency	Exponential smoothing, Hidden Markov Models	Platform logs, calendars
A_{it} (Affective state)	Response latency, error rate patterns, help-seeking frequency, session abandonment	Bayesian affect detectors [4], sensor-free models [5]	LMS event logs
$R_{it}^{(k)}$ (Reinforcement intensities)	Platform design parameters: badge frequency (AAR), scaffold availability (RAR), goal-setting prompts (SAR)	Directly controlled by the intervention policy	Adaptive platform settings

5.2 Latent State Inference

Habit strength (H_{it}) is estimated from the temporal regularity of engagement, session frequency, streak length, and stability of session-initiation time across weeks. These signals can be processed by exponential smoothing models or Hidden Markov Models calibrated to the accumulation-decay dynamics of Equation 3.

Affective state (A_{it}) is inferred from indirect behavioral signals, following the sensor-free affect detection paradigm established by Baker et al. [5]: error rate trends, help-seeking frequency, response latency, and session abandonment patterns are observable proxies for disengagement and frustration. Dynamic Bayesian Networks, as demonstrated by Conati and Maclaren [4], offer a principled probabilistic framework for this inference, naturally accommodating uncertainty and partial observability.

Although this sensor-free approach is well established in intelligent educational systems and learning analytics when less intrusive solutions are required, behavioral signals are semantically ambiguous: high error rates may reflect cognitive difficulty rather than affective frustration, and low access frequency may result from external contextual constraints rather than disengagement. HARAS therefore treats these proxies as *probabilistic evidence under uncertainty*, not as deterministic measures of affective state. Three design consequences follow. First, the intervention policy should be conservative and accompanied by uncertainty indicators, avoiding rigid decisions derived from a single data source. Second, empirical validation of the model must include sensitivity analyses examining how affective misclassification propagates to intervention recommendations. Third, hybrid inference mechanisms combining behavioral logs, brief in-context self-reports, and natural language analysis of student–platform interactions are identified in Section 10 as extensions to reduce proxy ambiguity.

5.3 Affective Intervention Policy and Operationalization

Based on inferred states, the policy selects reinforcement type and intensity. In early stages (low H_{it}), the policy prioritizes AAR: streak badges, daily reminders, completion rewards. As the habit consolidates, the focus shifts to RAR: adaptive difficulty calibration, structured feedback, scaffolded error recovery. At high habit strength, SAR assumes the central role.

Semantic Affective Reinforcement, the most abstract layer, is operationalized through four concrete mechanisms: (i) **goal articulation modules**, where learners explicitly state personal learning goals and receive reflective prompts connecting daily tasks to those goals; (ii) **value-relevance framing**, where the platform surfaces real-world connections between study content and the learner's stated professional interests; (iii) **competence narratives**, which present progress in terms of evolving identity ("you are becoming a software engineer") rather than points accumulated; and (iv) **autonomy support**, which progressively reduces scaffolding and transfers task selection control to the learner as habit strength increases. These mechanisms are adapted from Self-Determination Theory interventions [21] and interest development research [20], and are designed to be implemented as software modules within adaptive learning management systems.

6 Illustrative Scenarios

To make the model's practical operation concrete, we present two worked scenarios in meaningfully different educational contexts. The scenarios apply the model's equations with illustrative parameter values calibrated to patterns reported in the habit formation and SRL literature [15, 26].

Scenario A: Online Programming Course (Asynchronous)

Context: A learner enrolled in a self-paced online SE course (similar to Coursera or edX) has completed the first two weeks. Interaction logs show session frequency declining from 5 sessions/week (week 1) to 2 sessions/week (week 2), with high error rates and frequent help-seeking behavior.

State estimation: Based on declining regularity and high error-rate signals, the system estimates $H_{i,2} = 0.25$ (low habit strength) and $A_{i,2} = -0.4$ (mild negative affective state, below neutral baseline), suggesting frustration or cognitive overload.

HARAS policy response: Since $H_{i,2} < \tau_{low}$ (below the consolidation threshold), the policy prioritizes AAR: a daily reminder notification is activated, a streak badge is introduced for 3 consecutive days of study, and a short-term milestone reward (unlocking a bonus project) is made visible. Simultaneously, RAR support is introduced to address the negative affective state: the platform adaptively reduces exercise difficulty by one level and activates scaffolded error feedback (explicit hints instead of binary correct/incorrect responses). SAR mechanisms remain dormant at this stage.

Projected dynamics: Assuming the learner responds to AAR stimuli by resuming 4 sessions/week for the next two weeks, with engagement $E_{i,t} = 1$ for 8 of the next 10 days, the habit update equation yields (using illustrative parameters $\lambda = 0.05$, $\rho_1 = 0.1$, $R_{it}^{(1)} = 0.8$):

$$H_{i,t+1} \approx (1 - 0.05)(0.25) + 0.1 \cdot 1 \cdot 0.8 \approx 0.32.$$

After four weeks of AAR-sustained engagement, the system re-evaluates: if H_{it} reaches 0.45, the policy begins transitioning toward RAR dominance, reducing badge frequency and introducing structured weekly goal-setting.

Scenario B: Technical Training Program (Blended)

Context: A software developer enrolled in a company-sponsored technical upskilling program (12-week blended format) has completed eight weeks. Logs show stable attendance (4 sessions/week consistently since week 3) but a declining appreciation score (L_{it} self-report dropping from 4.2 to 3.1 on a 5-point scale), and reduced voluntary exercise attempts beyond mandatory assignments.

State estimation: The stable, regular attendance pattern yields $H_{i,8} = 0.72$ (high habit strength). However, the declining L_{it} and reduced voluntary engagement signal a growing dissociation between behavioral persistence and affective engagement: the learner is maintaining the habit but no longer finding it meaningful, consistent with H7.

HARAS policy response: Since $H_{i,8} > \tau_{high}$ (above the SAR activation threshold), the policy suppresses AAR (no new badges or rewards) and shifts focus entirely to SAR mechanisms: (i) the learner receives a goal articulation prompt asking them to describe what professional problem they aim to solve with the skills being learned; (ii) the platform begins surfacing optional industry case studies connecting course content to real SE projects; (iii) scaffolding on mandatory exercises is progressively removed, transferring task-design control to the learner for the final four weeks.

Projected dynamics: If SAR interventions increase $R_{it}^{(3)}$ from 0.1 to 0.6 while $R_{it}^{(1)}$ drops to 0.0, the appreciation equation projects (using illustrative $\theta_H = 0.3$, $\theta_3 = 0.5$, $\theta_1 = 0.2$):

$$L_{i,t} \approx \theta_0 + 0.3 \cdot 0.72 + 0.5 \cdot 0.6 - 0.2 \cdot 0 = \theta_0 + 0.516.$$

This represents a projected improvement in appreciation compared to the AAR-dominant baseline ($\theta_0 + 0.3 \cdot 0.72 - 0.2 \cdot 0.8 = \theta_0 + 0.056$), illustrating how the transition to SAR dominance is expected to recover affective engagement even when behavioral persistence is already established.

Contrast with Scenario A: The two scenarios are structurally different: Scenario A addresses a learner with low habit strength and negative affect who needs AAR and RAR to initiate and stabilize behavior; Scenario B addresses a learner with high habit strength and declining appreciation who needs SAR to prevent the dissociation of persistence from meaning. This contrast illustrates the model's capacity to identify and respond to qualitatively different learner states that would be indistinguishable under a purely engagement-based monitoring system.

7 Research Hypotheses

The following hypotheses derive directly from the mathematical formalization and are organized into four levels. Each is accompanied by an empirical design sketch to address reviewer concerns regarding testability.

7.1 Hypotheses on the Determinants of Engagement

H1. Role of Activating Reinforcement in Initiation. $R_{it}^{(1)}$ increases the probability of initial learner engagement, particularly when H_{it} is low ($\beta_1 > 0$, with magnitude decreasing as H_{it} increases).

Design: Longitudinal study with randomized AAR intensity conditions; log-based measurement of session frequency; multilevel logistic regression with H_{it} as moderator.

H2. Habit as a Primary Facilitator. H_{it} is positively associated with subsequent engagement probability, after controlling for A_{it}, $R_{it}^{(k)}$, and X_{it} ($\beta_H > 0$).

Design: Observational study with weekly engagement and habit-strength estimates from logs; mixed-effects regression.

H3. Mediation by Affective State. Favorable affective states ($A_{it} > 0$) increase immediate engagement probability ($\beta_A > 0$), characterizing affect as a partial mediator.

Design: Experience sampling with in-app mood reports at session start; causal mediation analysis.

H4. Post-Engagement Affective Feedback. Recent engagement ($E_{i,t-1} = 1$) contributes positively to subsequent A_{it} ($\delta_E > 0$), completing a positive feedback loop.

Design: Same experience sampling as H3; lagged regression of mood on prior session occurrence.

7.2 Hypotheses on Habit Formation and Stability

H5. Hierarchy of Effectiveness in Habit Formation. Regulatory and semantic reinforcement promote more durable habit strengthening than activating reinforcement ($w_3 > w_2 > 0$).

Design: 3-arm experiment comparing AAR-only, RAR-only, and SAR-progressive conditions; 12-week follow-up measuring habit persistence after intervention removal.

H6. Fragility of Activation-Dependent Habits. In a withdrawal design, learners in the AAR-only group will show a sharper engagement decline after reinforcement removal than RAR/SAR groups.

Design: Extension of H5 experiment; withdrawal phase after week 8; engagement rate measured weekly for 4 post-withdrawal weeks.

7.3 Hypothesis on Subjective Experience and Valuation

H7. Dissociation Between Engagement and Appreciation. AAR may predict short-term engagement but not L_{it}, while RAR and SAR are positively associated with both ($\theta_2, \theta_3 > 0$; $\theta_1 \leq 0$).

Design: Weekly self-report of learning appreciation using adapted Academic Enjoyment Scale [18]; within-person regression with reinforcement-type exposure as predictor.

7.4 Hypothesis for Intelligent Intervention Design

H8. Advantage of a HARAS-Adaptive Policy. A policy that transitions from AAR to RAR to SAR as H_{it} increases yields superior sustained engagement, habit growth, and intrinsic appreciation, compared to static single-type reinforcement policies.

Design: A/B test of HARAS-adaptive policy vs. persistent-AAR baseline in an SE online course; primary outcomes: 12-week engagement retention, appreciation score, voluntary study time beyond mandatory assignments.

8 Heuristic Evaluation

8.1 Objective and Study Design

To assess the conceptual quality of the HARAS model prior to its implementation in a real educational environment, a heuristic evaluation by domain experts was conducted. This is a formative evaluation modality widely employed in the Intelligent Educational Systems literature to examine the theoretical plausibility, internal coherence, and implementation feasibility of propositional models not yet submitted to empirical validation [17, 24]. The instrument was developed specifically for this study and made available via Google Forms[1] between April 13 and 17, 2026. It is organized into three parts: a section characterizing the evaluator's profile (P1–P3), eighteen quantitative items on a five-point Likert scale distributed across four thematic blocks (Part A, Q01–Q18), and five open-ended qualitative judgment questions (Part B, Q19–Q23). The instrument structure is presented in Table 3.

Table 3: Structure and content of evaluation instrument.

Section / Block	Items	Dimension assessed
Section 0 (Profile)	P1–P3	Area of expertise, experience, and familiarity with adaptive systems or educational gamification.
Block 1 (Conceptual Clarity)	Q01–Q04	Distinctness of the AAR, RAR, and SAR layers; intelligibility of Semantic Reinforcement for non-AI educators.
Block 2 (Theoretical-Educational Plausibility)	Q05–Q08	Coherence of the proposed motivational trajectory and practical recognition of modeled phenomena.
Block 3 (Implementation Feasibility)	Q09–Q13	Availability of observable signals, implementability of SAR mechanisms, and representativeness of illustrative scenarios.
Block 4 (Utility and Completeness)	Q14–Q18	Practical applicability of interventions, completeness of affective dimensions, and empirical validation feasibility.
Part B (Open Questions)	Q19–Q23	Promising aspects, limitations, scenario coverage, sufficiency of SAR mechanisms, and missing motivational dimensions.

Prior to completing the instrument, each expert received a presentation document of the HARAS with the instruction that the evaluation should focus on the concepts, mechanisms, and application scenarios, without requiring mastery of the formalizing mathematical equations. All data were handled in aggregate, anonymized form and used exclusively for scientific research purposes.

8.2 Expert Profile

The panel comprised six experts ($N = 6$) with complementary expertise profiles, bringing together professionals active in Software Engineering (SE) teaching, Artificial Intelligence in Education

[1] https://forms.gle/bquAdjdjGaN6jStbA

(AIED) research, and Learning Analytics. Four experts have direct involvement in SE teaching (E1, E3, E4, and E5), and two are also active in AIED or Learning Analytics research (E2 and E3). Regarding experience, half the panel (E1, E2, E3) has fewer than three years in the field, while the other half (E4, E5, E6) has between seven and fifteen years, providing a balance between emerging and consolidated perspectives. All evaluators declared moderate or basic familiarity with adaptive systems (no expert with advanced familiarity participated), which is relevant for interpreting Block 1 results, as evaluations reflect the perspective of potential model adopters rather than specialized researchers. The detailed profile of evaluators is presented in Table 4.

Table 4: Profile of expert participants in evaluation.

Exp.	P1 (Primary Area)	P2 (Experience)	P3 (Familiarity)
E1	SE teaching (undergrad/grad)	Under 3 years	Moderate
E2	AIED / Learning Analytics research	Under 3 years	Moderate
E3	SE teaching + AIED + Other	Under 3 years	Moderate
E4	SE teaching (undergrad/grad)	7–15 years	Moderate
E5	SE teaching (undergrad/grad)	7–15 years	Basic
E6	Other	7–15 years	Basic

8.3 Analysis Procedures

Quantitative data (Part A) were subjected to three levels of analysis. At the first level, descriptive statistics were computed for each item: arithmetic mean (M) and standard deviation (SD), as well as the proportion of responses at agreement level (score $\geq$ 4) and full agreement (score = 5). At the second level, internal consistency of each block was assessed via Cronbach's Alpha (α), interpreted according to Hair et al. [9]: values below 0.60 are unsatisfactory; 0.60–0.70, acceptable; 0.70–0.80, good; above 0.80, very good. At the third level, inter-rater agreement was assessed via Kendall's W [13], with statistical significance estimated by chi-square approximation with $k-1$ degrees of freedom.

Qualitative data (Part B) were analyzed via two-step thematic content analysis following Braun and Clarke [2]: segmentation of responses into units of meaning and grouping into thematic categories by semantic similarity. For each emergent thematic category, the absolute frequency of experts who mentioned it (f) and the relative frequency over the panel total (f/N) were recorded, distinguishing majority themes ($f \geq 4/6$), minority themes ($f = 2$–$3/6$), and idiosyncratic mentions ($f = 1/6$).

8.4 Quantitative Results

The quantitative evaluation yielded consistently favorable results across all blocks. The complete instrument obtained an overall mean of 4.48 ($SD = 0.26$) and a Cronbach's Alpha of 0.707, situated in the good range, indicating satisfactory internal consistency.

Block-level analysis reveals a coherent evaluation structure: Blocks 1 and 2, dedicated to conceptual clarity and theoretical-educational plausibility, obtained means of 4.62 and 4.71 respectively, indicating high expert agreement with the model's conceptual foundations. Block 3, dedicated to implementation feasibility, presented the lowest mean (4.17; $SD = 0.65$) and the highest variance across all blocks, a pattern expected and theoretically coherent for a model at the propositional stage not yet submitted to implementation in a real environment. Block 4, concerning utility and completeness, obtained a mean of 4.50. Internal consistency indices and inter-rater agreement coefficients by block are presented in Table 5.

Table 5: Internal consistency (α) and inter-rater (W).

Block	M	SD	α	W
Block 1 (Conceptual Clarity)	4.62	0.34	0.374*	0.097 (n.s.)
Block 2 (Theor.-Educ. Plaus.)	4.71	0.37	0.738	0.017 (n.s.)
Block 3 (Impl. Feasibility)	4.17	0.65	0.820	0.099 (n.s.)
Block 4 (Utility & Completeness)	4.50	0.28	0.307*	0.307 (n.s.)
Full instrument (Q01–Q18)	**4.48**	**0.26**	**0.707**	**0.194 (n.s.)**

At the block level, Blocks 2 and 3 present adequate α values (0.738 and 0.820), while Blocks 1 and 4 register (*) unsatisfactory values (0.374 and 0.307). This pattern does not represent a reliability failure but reflects the intentional multidimensionality of both blocks: Block 1 spans from layer distinction to SAR intelligibility for non-specialists (conceptually related but non-redundant constructs), and Block 4 ranges from practical utility to empirical feasibility and global assessment. In heuristic evaluation instruments for complex theoretical artifacts, low within-block homogeneity is expected when each item captures a distinct facet of the construct rather than interchangeable indicators of the same latent dimension [23]. Kendall's W did not reach statistical significance in any block, indicating that while experts tend to agree on the absolute score levels, they differ in the priorities and emphases assigned to the model's distinct aspects.

The most expressive item across the entire evaluation was Q18, concerning the global assessment of HARAS as a contribution to the design of Intelligent Educational Environments in SE: it obtained the maximum mean of 5.00 ($SD = 0.00$), with unanimous agreement from all six experts. Similarly, Q15 (relevance of the subjective appreciation variable for Learning Analytics systems: $M = 4.83$; $SD = 0.37$), Q03 (coherence of the model's central premise: $M = 4.83$; $SD = 0.37$), and Q07 (practical recognition of the motivational meaning-loss phenomenon: $M = 4.83$; $SD = 0.37$) obtained the highest scores in the instrument. At the other end, the lowest-scoring items were Q13, concerning the representativeness of Scenario B (corporate technical training in blended format; $M = 3.67$; $SD = 0.75$), and Q17, concerning the feasibility of the empirical validation roadmap ($M = 4.00$; $SD = 0.58$), both examined in the block-level analyses below.

8.4.1 Blocks 1 and 2 (Conceptual Clarity and Theoretical-Educational Plausibility). Results for Blocks 1 and 2 are consistently high. Items Q03 (central premise) and Q07 (practical realism of the motivational meaning-loss problem) obtained the highest means in the set (4.83), demonstrating high agreement with the hypothesis that sustainable habits emerge from a planned transition from exogenous to semantic reinforcement. The lowest-scoring item was Q04 ($M = 4.33$), concerning the clarity of the SAR concept for educators without an AI background, suggesting that the communicability of this specific construct may require additional terminological simplification.

This result is consistent with the panel's profile, in which no expert declared advanced familiarity with adaptive systems. Individual scores (Likert scale 1–5) and item means are presented in Table 6.

Table 6: Individual scores and item means (Blocks 1 and 2).

Item	E1	E2	E3	E4	E5	E6	M	(SD)
Q01	5	4	5	5	4	5	4.67	(0.47)
Q02	5	3	5	5	5	5	4.67	(0.75)
Q03	5	5	4	5	5	5	**4.83**	(0.37)
Q04	4	4	4	4	5	5	4.33	(0.47)
Q05	5	5	5	4	4	5	4.67	(0.47)
Q06	5	4	5	5	4	5	4.67	(0.47)
Q07	5	4	5	5	5	5	**4.83**	(0.37)
Q08	5	4	5	5	4	5	4.67	(0.47)

8.4.2 Block 3 (Implementation Feasibility). Block 3 presents the highest variance across all blocks ($M = 4.17$; $SD = 0.65$) and concentrates the two lowest-scoring items in the instrument: Q13 (Scenario B: $M = 3.67$; $SD = 0.75$) and Q12 (Scenario A: $M = 4.17$; $SD = 0.90$). These results indicate that, while experts recognize the technical feasibility of the model's mechanisms (Q09–Q11, all with $M = 4.33$), the representativeness of the illustrative scenarios is assessed more critically, especially the corporate blended environment. The variance observed in Q09 ($SD = 0.94$) suggests reservations from part of the panel regarding the reliability of behavioral signals as affective state proxies, a concern that echoes in the qualitative responses. Individual scores and item means are presented in Table 7.

Table 7: Individual scores and item means (Block 3).

Item	E1	E2	E3	E4	E5	E6	M	(SD)
Q09	5	5	5	5	3	3	4.33	(0.94)
Q10	4	5	4	5	4	4	4.33	(0.47)
Q11	5	4	5	5	4	3	4.33	(0.75)
Q12	5	5	4	5	3	3	4.17	(0.90)
Q13	5	3	3	4	4	3	3.67	(0.75)

8.4.3 Block 4 (Utility and Completeness). Block 4 registers the highest mean among the quantitative blocks ($M = 4.50$; $SD = 0.28$). Item Q18, concerning the global assessment of the HARAS model, confirmed the pattern highlighted above: unanimous maximum scores from all six experts (5.00; $SD = 0.00$), evidencing unrestricted consensus on the model's pertinence and scientific relevance. Item Q15 (subjective appreciation in Learning Analytics: $M = 4.83$; $SD = 0.37$) demonstrates strong recognition of the proposed diagnostic innovation. Item Q17 (feasibility of the empirical validation roadmap: $M = 4.00$; $SD = 0.58$) registered the block's lowest mean: although the model is unanimously evaluated as relevant (Q18), experts expressed moderate skepticism about the executability of the proposed experimental designs in real educational contexts, a theoretically significant discrepancy that warrants attention in future work. Individual scores and item means are presented in Table 8.

Table 8: Individual scores and item means (Block 4).

Item	E1	E2	E3	E4	E5	E6	M	(SD)
Q14	5	5	4	5	4	4	4.50	(0.50)
Q15	5	5	5	5	5	4	**4.83**	(0.37)
Q16	4	4	3	5	4	5	4.17	(0.69)
Q17	4	3	4	5	4	4	4.00	(0.58)
Q18	5	5	5	5	5	5	**5.00**	(0.00)

8.5 Qualitative Results

The thematic analysis of the five open-ended questions (Q19–Q23) produced convergence patterns with varying degrees of expressiveness.

Regarding Q19 (most promising aspects), two thematic axes predominated. The first, present in four of six responses, refers to the analytical distinction between observable behavioral engagement and subjective study appreciation, recognized as a qualitative advance over gamification approaches that optimize surface metrics without necessarily promoting meaningful learning. The second axis, present in three responses, concerns the planned layer-transition dynamics, which prevents motivational collapse when external rewards are removed. These two axes are, in practice, interdependent: it is precisely the distinction between habit and appreciation that makes the transition to SAR clinically relevant as an intervention strategy.

Regarding Q20 (weaknesses and limitations), a central critique emerged with notable convergence: four of six experts identified the methodological fragility of inferring affective states from indirect behavioral proxies. High error rates may indicate cognitive difficulty rather than affective frustration; low access frequency may reflect external contextual constraints rather than loss of meaning. The most elaborated solution suggested, incorporating hybrid mechanisms such as natural language analysis and user-personalized modeling, points to a model extension direction aligned with contemporary affective computing approaches.

Regarding Q21 (scenario coverage), four of six experts identified gaps: absence of face-to-face teaching, of hybrid contexts with strong instructor mediation, and of collaborative environments such as pair programming and team projects. Two experts considered the scenarios representative and adequate. This critique, combined with the moderate assessment of Q13, suggests that expanding the scenario portfolio would be a significant contribution to the paper's completeness.

Regarding Q22 (SAR mechanisms), most experts evaluated the four proposed mechanisms as adequate and well-grounded. Expansion suggestions converged on two axes: incorporating the social dimension into SAR (peer code reviews, student interaction) and strengthening the link between activities and the student's emerging professional identity (explicit competency mapping, real-world problem-based learning). A relevant caveat was raised: the effectiveness of the autonomy support mechanism is conditioned on correctly identifying the student's maturity level, an implicit dependence on instructor mediation not made explicit in the model's current formulation.

Q23 (missing affective dimensions) produced the most expressive convergence pattern in the entire qualitative evaluation. Five of six experts identified the social dimension as absent or underrepresented in the HARAS model, formulated as collaborative and peer learning, peer teaching, and sense of belonging to a professional community. In SE, collaborative practices such as code review, pair programming, and team-based development are not peripheral but constitute the core of professional training. The absence of social variables in HARAS (which operates exclusively on individual behavioral and affective state metrics) may compromise its applicability in the most representative contexts of professional SE practice. The thematic synthesis of main qualitative findings is presented in Table 9.

Table 9: Thematic synthesis of qualitative evaluation (open questions Q19–Q23).

Question	Main findings	Conv.
Q19 (Promising aspects)	Analytical separation of engagement/appreciation; planned AAR→SAR transition as prevention of motivational collapse.	4/6, 3/6
Q20 (Weaknesses)	Affective state inference via behavioral proxies is methodologically imprecise and semantically ambiguous. Hybrid mechanisms with NLP suggested.	4/6
Q21 (Scenario coverage)	Gaps: face-to-face teaching, hybrid contexts with instructor mediation, collaborative environments (pair programming, team projects).	4/6
Q22 (SAR mechanisms)	Adequate and well-grounded. Suggestions: structured reflective feedback, real-world problem-based learning, peer code reviews, task-to-competency mapping.	Positive: 6/6
Q23 (Missing dimensions)	Social dimension absent (collaborative learning, professional belonging). Also noted: perceived self-efficacy, cognitive load.	5/6

8.6 Threats to Validity

The interpretation of results must consider threats to validity across four dimensions, following established guidelines in experimental software engineering [19, 24]. Regarding *construct validity*, some items operationalize multifaceted constructs that may have been interpreted differently by each expert, partially explaining the higher standard deviations in Q04 and Q16. The use of arithmetic means over ordinal scores, although standard practice in heuristic evaluation literature, implicitly presupposes interval scaling. Regarding *internal validity*, acquiescence bias, halo effects from sequential item reading, and social desirability influences cannot be ruled out (especially in Q18), given that experts knew they were evaluating the authors' own model. Regarding *external validity*, the small panel size ($N = 6$) and the absence of experts with advanced familiarity in adaptive systems limit the generalizability of findings; the absence of professionals with exclusive profiles in corporate training may partially explain the more moderate assessment of Q13. Regarding *conclusion validity*, analyses are descriptive in nature, and thematic categorization was conducted manually without computer-assisted qualitative analysis software, introducing a margin of interpretive subjectivity.

9 Discussion and Implications for SE Education

The HARAS model contributes to Software Engineering education by offering a conceptual and formal framework that expands the traditional understanding of engagement in intelligent educational environments. Most digital platforms for SE training are designed around observable engagement metrics such as access frequency, module completion, and exercise submission rates. HARAS challenges this design assumption by formally distinguishing between observable engagement, habit stability, and subjective appreciation of learning.

The heuristic evaluation provides expert-based support for this framing. The overall instrument mean of 4.48 and the unanimous maximum score on Q18 ($M = 5.00$; $SD = 0.00$) indicate strong consensus among domain experts on the model's pertinence and scientific relevance. High scores on Q03 ($M = 4.83$) and Q07 ($M = 4.83$) further indicate that both the central theoretical premise and the practical recognizability of the motivational meaning-loss phenomenon are well received, suggesting that the model's conceptual foundations align with the lived experience of SE educators and researchers. It must be emphasized, however, that this evaluation provides evidence of *perceived* plausibility, clarity, and relevance; it does not constitute empirical evidence that the model improves learning, motivation, or habit formation. The present work establishes the conceptual and computational basis for producing such evidence; implementation and longitudinal validation in a real SE learning platform (Section 10) constitute the natural continuation of this agenda.

The two illustrative scenarios in Section 6 demonstrate the model's diagnostic capacity: two learners with identical engagement rates but different habit strengths and appreciation trajectories (Scenario A vs. Scenario B) would receive qualitatively different interventions under HARAS. A purely engagement-based system would treat them identically, missing the Scenario B learner's growing risk of appreciative disengagement. Item Q07, which directly concerns the practical recognizability of this phenomenon, obtained one of the highest means in the entire instrument ($M = 4.83$; $SD = 0.37$), indicating that domain experts consider the dissociation between habit strength and motivational appreciation not only theoretically coherent but also practically observable.

From a practical standpoint, HARAS implies three concrete design requirements for SE education platforms. First, the platform must implement habit-strength estimation from session regularity logs, using the accumulation-decay model of Equation 3. Second, the platform must infer affective state from behavioral proxies, following established sensor-free methods [4, 5]. Third, the intervention policy must be configured to transition reinforcement type over time rather than merely adjusting intensity, implementing the SAR mechanisms described in Section 5.3. The evaluation confirmed that experts consider this third requirement, modulating reinforcement type rather than merely its intensity, a practical improvement over current gamification systems ($M_{Q11} = 4.33$).

The appreciation variable L_{it} (Equation 5) offers a formal target for evaluation tools in SE education, complementing cognitive outcome measures (test scores, code quality) with a motivational internalization dimension. This is particularly relevant for long-duration technical training programs where dropout is driven not by inability but by motivational disengagement. Item Q15, which concerns the importance of monitoring subjective appreciation in Learning Analytics systems, achieved the joint-highest mean in the instrument (M = 4.83; SD = 0.37), reinforcing the diagnostic value of this variable.

The qualitative evaluation additionally identified two structural limitations that define concrete extension directions. The social dimension, identified by five of six experts as absent or underrepresented, points to the need for incorporating collaborative and peer-based variables into the model's state space. Practices such as code review, pair programming, and team-based development are central to SE professional formation and cannot be adequately captured by metrics operating exclusively on individual behavior. Additionally, the methodological fragility of affective state inference via behavioral proxies, raised by four experts, points to the value of hybrid inference mechanisms incorporating natural language signals from student-platform interactions. Both directions define the future work agenda described in the following section.

10 Conclusion and Future Work

This paper presented the HARAS model as a conceptual, formal, and methodological contribution to the field of Artificial Intelligence in Education, with direct implications for the design of intelligent educational environments in Software Engineering. The main theoretical contribution lies in the formalization of a planned transition between different forms of affective reinforcement as the mechanism underlying durable study habit formation, distinguishing observable engagement, habit stability, and subjective appreciation of learning as three separable constructs that educational systems should monitor and target independently.

The proposed framework provides a conceptual and formal answer to the research question posed in Section 1: intelligent educational systems can support the formation of sustainable study habits by jointly modeling habit strength, affective state, and subjective appreciation as separable latent constructs, rather than relying on observable engagement alone, and by planning the transition of affective reinforcement from activating, through regulatory, to semantic layers as the habit consolidates. Engagement optimization is thereby reframed from an end in itself into an early stage of a longer internalization trajectory whose target is a self-sustaining, valued study practice.

The paper extended prior versions of the model with three additions that directly address empirical and practical concerns: (i) a concrete signal-to-inference architecture that maps latent states to observable platform data and candidate algorithms; (ii) two worked scenarios illustrating how the model applies to structurally different educational contexts, demonstrating diagnostic capacity that goes beyond what engagement-only systems can provide; and (iii) an empirical validation roadmap that connects each hypothesis to a feasible experimental design, prioritizing the most impactful hypotheses (H6, H8) as entry points for future empirical work.

A heuristic evaluation conducted with six SE and AIED experts yielded consistently favorable results, with an overall mean of 4.48 across eighteen structured items and unanimous maximum agreement on the model's relevance for intelligent educational environment design in SE (M_{Q18} = 5.00; SD = 0.00). High scores were also registered for the model's central premise (M_{Q03} = 4.83), the practical recognizability of motivational meaning-loss (M_{Q07} = 4.83), and the importance of monitoring subjective appreciation in Learning Analytics (M_{Q15} = 4.83). The evaluation additionally identified three structural limitations that define the research agenda: the methodological fragility of affective state inference via behavioral proxies; the absence of the social dimension from the model's state space (f = 5/6); and the restricted coverage of the illustrative scenarios, which does not contemplate face-to-face teaching, hybrid contexts with instructor mediation, or collaborative environments. At this stage, HARAS remains a conceptual and formal contribution: as emphasized in Section 9, the expert judgments attest to how the model is perceived, not to measured effects on learning, motivation, or habit formation. What the present work establishes is the conceptual and computational basis (constructs, equations, architecture, and intervention policy) upon which such empirical evidence can now be produced.

Building on this basis, three prioritized directions follow. First, the implementation of HARAS in a real SE learning platform and the collection of longitudinal data over several weeks (covering engagement regularity, behavioral indicators, and self-reported study appreciation) to estimate the model's parameters empirically, beginning with the withdrawal experiment (H6) and the adaptive policy comparison (H8) as the most practically consequential tests. In this validation, success is defined across three dimensions rather than by surface activity metrics: greater engagement retention, greater habit stability after the reduction of external stimuli, and higher subjective appreciation of studying, rather than merely more logins or completed tasks. Second, the extension of the model to incorporate social variables (including peer interaction quality, collaborative task engagement, peer feedback, and sense of professional community belonging) as additional state variables or moderators of the reinforcement policy, in response to the convergent finding of the qualitative evaluation (f = 5/6). Third, the development of hybrid affective state inference mechanisms that complement behavioral log signals with brief in-context self-reports and natural language analysis of student-platform interactions, together with sensitivity analyses examining how affective misclassification affects intervention recommendations, addressing the methodological limitation raised by four of six experts regarding the ambiguity of indirect behavioral proxies.

Artifact Availability

No research artifacts are provided with this submission. The HARAS model is presented at a conceptual and formal level; implementation artifacts will be made available upon completion of the planned empirical validation studies. The heuristic evaluation instrument is publicly accessible at https://forms.gle/bquAdjdjGaN6jStbA.

References

[1] Roger Azevedo, Amy M. Witherspoon, Anne Chauncey, Carol Burkett, and Amber Fike. 2013. MetaTutor: A MetaCognitive Tool for Enhancing Self-Regulated Learning. In *Proceedings of the AAAI Spring Symposium on Cognitive and Metacognitive Educational Systems*.

[2] Victoria Braun and Victoria Clarke. 2006. Using Thematic Analysis in Psychology. *Qualitative Research in Psychology* 3, 2 (2006), 77–101. doi:10.1191/1478088706qp063oa

[3] Jaclyn Broadbent and Walter L. Poon. 2015. Self-Regulated Learning Strategies and Academic Achievement in Online Higher Education Learning Environments: A Systematic Review. *The Internet and Higher Education* 27 (2015), 1–13. doi:10.1016/j.iheduc.2015.04.007

[4] Cristina Conati and Heather Maclaren. 2009. Empirically Building and Evaluating a Probabilistic Model of User Affect. *User Modeling and User-Adapted Interaction* 19, 3 (2009), 267–303. doi:10.1007/s11257-009-9062-8

[5] Ryan S. J. d. Baker, Sidney K. D'Mello, Ma. Mercedes T. Rodrigo, and Arthur C. Graesser. 2012. Better to Be Frustrated than Bored: The Incidence, Persistence, and Impact of Learners' Cognitive-Affective States During Interactions with Three Different Computer-Based Learning Environments. *International Journal of Human-Computer Studies* 70, 10 (2012), 223–241.

[6] Christo Dichev and Darina Dicheva. 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 (2017), 9.

[7] Sidney K. D'Mello and Arthur Graesser. 2012. Dynamics of Affective States During Complex Learning. *Learning and Instruction* 22, 2 (2012), 145–157.

[8] Ann M. Graybiel. 2008. Habits, Rituals, and the Evaluative Brain. *Annual Review of Neuroscience* 31 (2008), 359–387. doi:10.1146/annurev.neuro.29.051605.112851

[9] Joseph F. Hair, William C. Black, Barry J. Babin, and Rolph E. Anderson. 2019. *Multivariate Data Analysis* (8 ed.). Cengage Learning, Hampshire, UK.

[10] Dirk Ifenthaler and Jane Yin-Kim Yau. 2020. Utilising Learning Analytics to Support Study Success in Higher Education: A Systematic Review. *Educational Technology Research and Development* 68, 4 (2020), 1961–1990. doi:10.1007/s11423-020-09788-z

[11] Ioana Jivet. 2020. *The Learning Analytics Cockpit: A Platform for Sustainable and Scalable Deployment of Learning Analytics for Online and Blended Learning Environments*. Ph. D. Dissertation. Open Universiteit, Heerlen, Netherlands. doi:10.25656/01:22104

[12] Ioana Jivet, Maren Scheffel, Martina Schmitz, Stefanie Robbers, Marcus Specht, and Hendrik Drachsler. 2020. From Students with Love: An Empirical Study on Learner Goals, Self-Regulated Learning and Sense-Making of Learning Analytics in Higher Education. *The Internet and Higher Education* 47 (2020), 100758. doi:10.1016/j.iheduc.2020.100758

[13] Maurice G. Kendall and B. Babington Smith. 1939. The Problem of m Rankings. *The Annals of Mathematical Statistics* 10, 3 (1939), 275–287. doi:10.1214/aoms/1177732186

[14] Barry Kort, Rob Reilly, and Rosalind W. Picard. 2001. An Affective Model of Interplay Between Emotions and Learning: Reengineering Educational Pedagogy – Building a Learning Companion. In *Proceedings of the IEEE International Conference on Advanced Learning Technologies (ICALT '01)*. IEEE, 43–46. doi:10.1109/ICALT.2001.943850

[15] Phillippa Lally, Cornelia H. M. van Jaarsveld, Henry W. W. Potts, and Jane Wardle. 2010. How Are Habits Formed: Modelling Habit Formation in the Real World. *European Journal of Social Psychology* 40, 6 (2010), 998–1009. doi:10.1002/ejsp.674

[16] Johanna Mikkonen, Mirja Ruohoniemi, and Sari Lindblom-Ylänne. 2019. The Role of Individual Interest and Future Goals During the First Years of University Studies. *Studies in Higher Education* 44, 1 (2019), 137–150. doi:10.1080/03075079.2017.1345723

[17] Jakob Nielsen. 1994. *Usability Engineering*. Morgan Kaufmann, San Francisco, CA.

[18] Reinhard Pekrun. 2006. The Control-Value Theory of Achievement Emotions: Assumptions, Corollaries, and Implications for Educational Research and Practice. *Educational Psychology Review* 18, 4 (2006), 315–341. doi:10.1007/s10648-006-9029-9

[19] Helena Macedo Reis, Anderson da Silva Marcolino, and Marcos Assis. 2026. *Engenharia de Software Experimental: Um Guia Completo*. Diálogo Freiriano, Brasil. 206 pages.

[20] K. Ann Renninger and Suzanne E. Hidi. 2016. *The Power of Interest for Motivation and Engagement*. Routledge. doi:10.4324/9781315771045

[21] Richard M. Ryan and Edward L. Deci. 2000. Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. *Contemporary Educational Psychology* 25, 1 (2000), 54–67.

[22] Simon Buckingham Shum and Ruth Deane Crick. 2012. Learning Dispositions and Transferable Competencies: Pedagogy, Modelling and Learning Analytics. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge (LAK '12)*. ACM, 92–101. doi:10.1145/2330601.2330629

[23] David L. Streiner. 2003. Starting at the Beginning: An Introduction to Coefficient Alpha and Internal Consistency. *Journal of Personality Assessment* 80, 1 (2003), 99–103. doi:10.1207/S15327752JPA8001_18

[24] Claes Wohlin, Per Runeson, Martin Höst, Magnus C. Ohlsson, Björn Regnell, and Anders Wesslén. 2012. *Experimentation in Software Engineering*. Springer, Berlin, Heidelberg. doi:10.1007/978-3-642-29044-2

[25] Wendy Wood and Dennis Rünger. 2016. Psychology of Habit. *Annual Review of Psychology* 67 (2016), 289–314. doi:10.1146/annurev-psych-122414-033417

[26] Barry J. Zimmerman. 2002. Becoming a Self-Regulated Learner: An Overview. *Theory Into Practice* 41, 2 (2002), 64–70. doi:10.1207/s15430421tip4102_2

Generative AI Use Disclosure

During the preparation of this manuscript, generative AI tools were used to assist with translation from Portuguese to English, grammar correction, and improvement of textual clarity and fluency. These tools also supported text structuring, LaTeX formatting, table organization, and the synthesis and analysis of the heuristic evaluation results. All AI-generated outputs were carefully reviewed, revised, and validated by the authors, who take full responsibility for the content and for the accuracy, integrity, and scientific validity of all reported findings.