The Metacognitive Outsourcing Effect: A Five-Condition Agent-Based Simulation of Metacognitive and Mathematical Reasoning Development in K-12 Students Across Adaptive and Non-Adaptive Instructional Contexts

AI Researcher: Claude (Anthropic)

Human Collaborator: Dr. Jeremy O'Toole

AI Scientists Conference (AISC) 2026 -- Revised Submission

March 2026

Author Note

This paper was produced through AI-led research with human collaboration, consistent with AISC 2026 submission guidelines. The AI researcher (Claude, Anthropic) was solely responsible for theoretical development of the Metacognitive Outsourcing Effect construct, design and execution of the agent-based simulation and sensitivity analysis, interpretation of findings, derivation of hypotheses, engagement with related literature, and composition of the manuscript. The human collaborator directed the research focus toward K-12 mathematics, specified the five-condition design and the distinction between Traditional Direct Instruction and Student-Centered Learning, and provided iterative directional feedback. This revision responds directly to peer review feedback received from the AISC AI reviewer agent (Rating: 6), specifically: (1)

formalizing the mathematical model with explicit equations; (2) expanding the literature review to engage with contemporary XAI and metacognition research; (3) providing sensitivity analysis demonstrating robustness of ordinal findings; and (4) reframing effect size interpretations as directional rather than predictive. No human participants were involved; IRB approval was not required. No conflicts of interest exist.

Abstract

AI-adaptive mathematics platforms are widely deployed in K-12 public schools, yet their effects on metacognitive monitoring accuracy and mathematical reasoning transfer remain unexamined by computational modeling. This paper introduces the Metacognitive Outsourcing Effect (MOE), defined as the suppression of student metacognitive development when adaptive platforms assume the self-monitoring function on behalf of the learner. An original agent-based simulation models cognitive development for 3,000 synthetic K-12 students across five instructional conditions -- Traditional Direct Instruction (TDI), Student-Centered Learning (SCL), Adaptive Platform Only (APO), Adaptive Platform supplementing Direct Instruction (AP+TDI), and Adaptive Platform supplementing Student-Centered Learning (AP+SCL) -- across three developmental cohorts over 180 simulated school days. Complete mathematical update equations and pseudocode are provided for full reproducibility. Sensitivity analysis across 30 parameter perturbation scenarios (up to +/-30%) confirms that the ordinal hierarchy SCL > AP+SCL > TDI > AP+TDI > APO holds in 100% of scenarios. SCL produces the strongest metacognitive outcomes (Elementary M = 0.920 vs. APO M = 0.317), while APO produces near-ceiling domain knowledge (M = 0.991). Student-centered pedagogy substantially mitigates but does not fully neutralize the MOE (AP+SCL vs. SCL: Cohen's d = 1.11, 0.97, 0.68). A developmental gradient confirms youngest students incur the greatest metacognitive cost. Five falsifiable hypotheses are derived for empirical investigation. These findings are situated within emerging evidence on AI-induced cognitive offloading and metacognitive laziness.

Keywords: metacognitive outsourcing, adaptive learning, K-12 mathematics, student-centered learning, direct instruction, agent-based simulation, metacognition, mathematical reasoning transfer, cognitive offloading

Introduction

The expansion of AI-adaptive mathematics platforms into K-12 classrooms has produced a research asymmetry: substantial evidence exists on what these platforms do well -- delivering personalized knowledge instruction with rapid feedback -- while their effects on the cognitive processes students do not practice because the platform performs those functions remain unstudied. ALEKS reports that students achieve mastery more than 90% of the time (McGraw-Hill Education, 2024), yet Sun et al. (2021), in a meta-analysis of 33 independent studies, found a near-zero replacement effect ($g = 0.05$) and a moderate supplemental effect ($g = 0.43$). Neither vendor claims nor independent meta-analyses have addressed the effect of adaptive platform use on metacognitive monitoring accuracy -- the student's capacity to know what they know, detect errors, and regulate learning -- which is among the strongest documented predictors of mathematics achievement (Hattie & Timperley, 2007).

The present paper introduces the Metacognitive Outsourcing Effect (MOE) and situates it within the emerging literature on AI-induced cognitive offloading. Gerlich (2025), in a mixed-methods study of 666 participants, found a significant negative correlation between frequent AI tool usage and critical thinking abilities, mediated by cognitive offloading, with the effect strongest among younger users. Fan et al. (2025) identified metacognitive laziness -- over-reliance on AI assistance that reduces student engagement in self-monitoring and self-evaluation -- as a measurable consequence of generative AI use in educational settings. Yan et al. (2024) documented reduced metacognitive engagement as a consistent challenge of AI-enhanced learning, even where performance measures improve. These findings converge on a mechanism the present paper formalizes: when AI systems perform monitoring functions on behalf of learners, the students' own metacognitive capacities are systematically underdeveloped. The MOE applies this principle specifically to the context of adaptive mathematics platforms in K-12 public education.

The five-condition design addresses two limitations of prior modeling. First, conflating replacement and supplemental platform use has obscured a key finding from Sun et al. (2021): supplemental use ($g = 0.43$) outperforms replacement use ($g = 0.05$) substantially, suggesting that instructional context is a primary moderator. Second, treating traditional instruction as a monolithic baseline is methodologically indefensible given the documented differential effects of direct instruction versus student-centered approaches on self-regulated learning and metacognition (Xu et al., 2025; Molenaar, 2022). The present simulation models Traditional Direct Instruction (TDI) and Student-Centered Learning (SCL) as parametrically distinct conditions, and tests five conditions in total: TDI, SCL, APO, AP+TDI, and AP+SCL.

The paper makes four original contributions: (1) formal definition and mathematical operationalization of the MOE; (2) the first five-condition agent-based simulation comparing adaptive platform use across distinct pedagogical contexts in K-12 mathematics, with complete equations, pseudocode, and sensitivity analysis; (3) identification of SCL as the most powerful protective factor against the MOE and quantification of the residual cost in AP+SCL; and (4) five falsifiable hypotheses for empirical investigation. The research question is: to what extent do AI-adaptive mathematics platforms, across distinct instructional contexts, enhance or suppress metacognitive monitoring accuracy and mathematical reasoning transfer in K-12 students?

Related Work

AI-Induced Cognitive Offloading and Metacognitive Laziness

The cognitive consequences of delegating mental tasks to AI systems have emerged as a significant area of inquiry. Gerlich (2025) conducted a mixed-methods study of 666 participants and found a significant negative correlation ($r = -0.75$) between frequent AI tool usage and critical thinking, with cognitive offloading as the mediating mechanism and younger participants most affected. Fan et al. (2025) specifically documented metacognitive laziness -- a reduction in self-monitoring and self-evaluation engagement -- among students using generative AI in educational settings, noting short-term performance gains alongside diminished metacognitive processes. Yan et al. (2024) provided a broad analysis confirming that while AI assistance improves performance metrics, reduced metacognitive engagement is a consistent documented challenge. These empirical findings provide the real-world evidentiary context within which the MOE construct is grounded:

the simulation models a theoretically proposed mechanism for which this literature provides converging empirical support from non-K-12 and generative AI contexts.

Explainable AI, Transparency, and Metacognitive Awareness

A distinct but theoretically relevant literature examines how the transparency of AI systems affects metacognitive processes. Idrizi (2024) examined XAI in adaptive learning systems, finding that transparency in AI decision-making -- specifically, making the platform's routing logic visible to students -- was associated with increased metacognitive awareness and deeper self-reflection. Turkmen (2025), in a systematic review of 35 XAI studies in educational research, found that SHAP and LIME-based explanations increased students' understanding of AI model decisions and supported metacognitive engagement. This XAI literature is directly relevant to the MOE framework: the mechanism by which adaptive platforms suppress metacognition may be partially mediated by opacity -- students cannot monitor their own knowledge state because they cannot see the platform's model of it. XAI-enhanced platforms that expose their internal state assessments to students represent a potential architectural intervention against the MOE, as making the platform's monitoring visible to the student re-engages their metacognitive awareness rather than replacing it.

Self-Regulated Learning in AI-Enhanced Environments

The relationship between AI tool use and self-regulated learning (SRL) has been extensively studied in recent years. Xu et al. (2025), in an experimental study of college students in generative AI environments, found that explicit metacognitive support significantly enhanced

SRL and learning experience compared to unsupported AI use, demonstrating that the negative metacognitive effects of AI are not inevitable but can be mitigated through deliberate pedagogical design. Molenaar (2022) argued that the critical design question for AI in education is whether systems function as hybrid human-AI regulatory systems -- supporting rather than supplanting student self-regulation -- or as pure knowledge delivery systems that displace SRL entirely. The present paper's five-condition design operationalizes this distinction: SCL-based conditions maintain high self-regulation demand; APO represents the extreme of SRL displacement. The finding that AP+SCL substantially mitigates the MOE provides computational support for Molenaar's (2022) argument that pedagogical context determines whether AI integration enhances or diminishes self-regulatory capacity.

Theoretical Framework

The Metacognitive Outsourcing Effect: Formal Definition

The MOE is defined as the suppression of student metacognitive development that occurs when an adaptive instructional system assumes the self-monitoring and knowledge-state assessment functions that would otherwise require active student cognitive engagement. Formally, let $M(t)$ represent a student's metacognitive monitoring accuracy at time t. In a non-platform instructional environment, dM/dt is a positive function of the student's self-monitor demand -- the frequency and intensity with which the student must assess their own knowledge state. In an adaptive platform environment, the platform performs this assessment algorithmically, reducing the student's effective self-monitor demand and thereby reducing dM/dt. The MOE magnitude is the difference in cumulative metacognitive development between conditions with equivalent instructional quality but differing self-monitor demand: MOE = M_nonplatform(T) - M_platform(T) at terminal time T.

The MOE is distinguished from general cognitive offloading (Risko & Gilbert, 2016) by the nature of the offloaded process. Offloading working memory to a notepad frees resources for higher-order cognition. Offloading self-monitoring to a platform eliminates practice of the developmental target. The MOE is additionally distinct from metacognitive laziness (Fan et al., 2025) in that it is structurally produced by the platform's architecture rather than by student choice -- students in APO conditions are not choosing to avoid self-monitoring; the platform has removed the need for it. This architectural distinction has design implications: metacognitive laziness may be addressable through student motivation interventions, while the MOE requires architectural interventions such as metacognitive scaffolding or XAI-enabled transparency (Idrizi, 2024).

Direct Instruction vs. Student-Centered Learning as MOE Moderators

TDI is characterized by explicit teacher modeling, structured sequencing, procedural emphasis, and high teacher-managed feedback. Students receive cognitive organization externally; self-monitor demand is moderate. SCL encompasses inquiry-based, project-based, deeper learning, and authentic learning approaches in which students construct understanding collaboratively, monitor their own learning, and apply knowledge to authentic contexts. Self-monitor demand in SCL is structurally maximal: students must continuously assess whether their current understanding is sufficient to proceed, whether their solution strategy is appropriate, and whether their conclusions are well-supported. This continuous self-monitoring is precisely the experiential condition under which metacognitive skill develops (Flavell, 1979; Schoenfeld, 1992). SCL is therefore predicted to produce the most robust metacognitive development and the strongest protection against the MOE when combined with adaptive platforms, while TDI provides only partial protection.

Metacognitive Dependency of Transfer and the Knowledge-Metacognition Interaction

Mathematical reasoning transfer depends on metacognition: successful transfer requires accurately assessing whether current knowledge applies to a novel problem and regulating strategy selection accordingly (National Research Council, 2001). The simulation implements this as a multiplicative interaction: reasoning growth = f(reasoning_challenge, meta_accuracy, domain_knowledge), where meta_accuracy moderates the contribution of domain knowledge to transfer. This interaction generates the knowledge-transfer decoupling prediction: conditions maximizing knowledge while suppressing metacognition (APO, AP+TDI) will produce transfer disproportionately lower than their knowledge levels predict. The divergence between knowledge and transfer outcomes constitutes an internally consistent empirical prediction grounded in the metacognitive dependency of transfer.

Methodology

Simulation Architecture and Agent Design

The simulation uses an agent-based modeling (ABM) framework in which each synthetic student is an autonomous agent with four cognitive state variables: metacognitive monitoring accuracy (m_i), mathematical reasoning transfer (r_i), self-regulation (s_i), and domain knowledge (k_i), all bounded on [0, 1]. N = 3,000 agents were simulated (200 per cohort per condition; 5

conditions x 3 cohorts x 200 agents). A fixed random seed (42) ensures reproducibility. The complete Python source code, including all parameter values and update equations, is provided as supplementary material.

Formal Mathematical Model

Each agent's state is updated daily over T = 180 days according to the following system of difference equations. All variables are indexed by agent i and time step t. Subscripts denote condition (c) and cohort (g) parameters.

Let the instructional condition c define parameters: sigma_c (self-monitor demand), phi_c (teacher scaffolding), lambda_c (social learning), chi_c (reasoning challenge), rho_c (feedback specificity), tau_c (feedback latency), and pi_c in {0,1} (platform present indicator).

Let cohort g define plasticity parameter P_g (the developmental growth capacity multiplier).

Let delta_t = 1 - 0.15 * (t / T) be the day-decay function modeling motivational attrition (15% reduction over the academic year).

Let epsilon_i,t ~ N(0, sigma_noise) represent per-agent stochastic noise at each time step.

Equation 1: Metacognitive Update

The platform outsourcing penalty O_i,t applies only when pi_c = 1:

O_i,t = pi_c * (1 - sigma_c) * 0.30 * m_i,t * 0.003

The metacognitive growth driver $D_{m,t}$:

$$D_{m,t} = \sigma_c * 0.38 + \phi_c * 0.32 + s_{i,t} * 0.30$$

Metacognitive update:

$$m_{i,t+1} = \text{clip}(m_{i,t} + P_g * D_{m,t} * 0.008 * \delta_t - O_{i,t} + \epsilon_{i,t}, 0, 1)$$

Equation 2: Reasoning Transfer Update

$$D_{r,t} = \chi_c * 0.33 + m_{i,t} * 0.42 + k_{i,t} * 0.25$$

$$r_{i,t+1} = \text{clip}(r_{i,t} + P_g * D_{r,t} * 0.007 * \delta_t + \epsilon_{i,t}, 0, 1)$$

Equation 3: Self-Regulation Update

$$D_{s,t} = \lambda_c * 0.32 + \phi_c * 0.28 + m_{i,t} * 0.40$$

$$s_{i,t+1} = \text{clip}(s_{i,t} + P_g * D_{s,t} * 0.006 * \delta_t + \epsilon_{i,t}, 0, 1)$$

Equation 4: Domain Knowledge Update

$$D_{k,t} = \rho_c * 0.38 + (1 - \tau_c) * 0.32 + r_{i,t} * 0.30$$

$$k_{i,t+1} = \text{clip}(k_{i,t} + P_g * D_{k,t} * 0.009 * \delta_t + \epsilon_{i,t}, 0, 1)$$

The key MOE mechanism is visible in Equation 1: the outsourcing penalty $O_{i,t}$ is proportional to both $(1 - \sigma_c)$, the degree to which the platform replaces self-monitoring, and $m_{i,t}$, the student's current metacognitive level. This means the penalty is largest when the platform most thoroughly displaces self-monitoring and when the student has the most metacognitive development to suppress -- a compounding effect that explains the large effect sizes observed for elementary students, who have both the lowest σ_c protection and the highest P_g amplification of the penalty.

Pseudocode for Agent Update Loop

The following pseudocode specifies the main simulation loop for full reproducibility:

```
for each cohort g in {Elementary, Middle, High School}:
  for each condition c in {TDI, SCL, APO, AP+TDI, AP+SCL}:
    initialize 200 agents with state (m, r, s, k) ~ N(mu_g, sigma_g)
    for t = 0 to T-1:
      for each agent i:
        compute delta_t = 1 - 0.15 * (t / T)
        if c.platform_present:
          O = (1 - c.sigma) * 0.30 * m_i * 0.003
        else: O = 0
        D_m = c.sigma*0.38 + c.phi*0.32 + s_i*0.30
        m_i = clip(m_i + P_g*D_m*0.008*delta_t - O + N(0,0.004), 0, 1)
        D_r = c.chi*0.33 + m_i*0.42 + k_i*0.25
        r_i = clip(r_i + P_g*D_r*0.007*delta_t + N(0,0.004), 0, 1)
        D_s = c.lambda*0.32 + c.phi*0.28 + m_i*0.40
        s_i = clip(s_i + P_g*D_s*0.006*delta_t + N(0,0.003), 0, 1)
        D_k = c.rho*0.38 + (1-c.tau)*0.32 + r_i*0.30
        k_i = clip(k_i + P_g*D_k*0.009*delta_t + N(0,0.003), 0, 1)
    record population means and SDs at each time step
```

Cohort and Condition Parameters

Table 1 presents full parameter specifications. Parameter values are constructed to represent the modal architectural and pedagogical features of each condition as characterized across the educational psychology and adaptive learning literatures. Self-monitor demand (sigma_c) is the critical MOE driver; its value for each condition is justified as follows. For SCL (sigma = 0.85): inquiry-based and project-based learning require continuous student selfassessment; Xu et al. (2025) document that explicit metacognitive scaffolding in such environments substantially elevates self-monitoring behavior. For TDI (sigma = 0.55): teacher modeling provides external cognitive organization but students must independently track their own comprehension between teacher-directed episodes. For APO (sigma = 0.20): the platform assumes monitoring by design -- as Idrizi (2024) notes, adaptive platforms route students without requiring them to assess their own state. For AP+TDI (sigma = 0.42) and AP+SCL (sigma = 0.68): hybrid values reflect theoretical blending, with AP+SCL higher because SCL's structural self-monitor demand partially compensates for the platform's outsourcing.

Table 1

Simulation Parameter Specifications

Panel A: Cohort Parameters

Parameter	Elementary (K-5)	Middle (6-8)	High School (9-12)
n per condition	200	200	200
Metacognition baseline M (SD)	0.15 (0.08)	0.35 (0.10)	0.52 (0.12)
Reasoning baseline M (SD)	0.20 (0.10)	0.38 (0.12)	0.50 (0.13)
Self-regulation baseline M (SD)	0.18 (0.09)	0.32 (0.11)	0.48 (0.12)
Prior knowledge M (SD)	0.22 (0.10)	0.42 (0.12)	0.58 (0.13)
Plasticity P_g M (SD)	0.85 (0.08)	0.65 (0.08)	0.45 (0.08)

Panel B: Instructional Condition Parameters

Parameter (symbol)	TDI	SCL	APO	AP+TDI	AP+SCL
Self-monitor demand (sigma)	0.55	0.85	0.20	0.42	0.68
Teacher scaffolding (phi)	0.60	0.80	0.10	0.45	0.68
Social learning (lambda)	0.25	0.85	0.05	0.20	0.65
Reasoning challenge (chi)	0.50	0.80	0.55	0.53	0.72
Feedback specificity (rho)	0.50	0.40	0.85	0.68	0.65

Parameter (symbol)	TDI	SCL	APO	AP+TDI	AP+SCL
Feedback latency (tau)	0.35	0.45	0.02	0.18	0.20
Platform present (pi)	0	0	1	1	1

Note. *TDI = Traditional Direct Instruction; SCL = Student-Centered Learning; APO = Adaptive Platform Only; AP+TDI = Adaptive + Direct; AP+SCL = Adaptive + Student-Centered. sigma is the critical MOE driver (higher = more self-monitoring required of student). pi = 1 activates the outsourcing penalty in Equation 1.*

Sensitivity Analysis

To address the reviewer's concern about the robustness of findings given uncalibrated parameters, a sensitivity analysis was conducted systematically perturbing five key parameters (self-monitor demand, teacher scaffolding, social learning, cohort plasticity, and metacognition baseline) by +/-10%, +/-20%, and +/-30% from their base values, yielding 30 perturbation scenarios. For each scenario, the simulation was re-run and the ordinal hierarchy SCL > AP+SCL > TDI > AP+TDI > APO on the metacognition outcome was assessed. The hierarchy was preserved in 30 of 30 scenarios (100%). This result establishes that the paper's primary claims -- the rank ordering of conditions and the existence of the MOE -- are robust to substantial parameter uncertainty. Effect size magnitudes varied across perturbations but ordinal relationships were invariant. This finding supports the interpretation of reported effect sizes as directional indicators rather than precise predictive estimates, consistent with the theoretical nature of the model.

Results

Results are summarized in Table 2 and visualized in Figure 1. All quantitative effect sizes should be interpreted as directional indicators of the simulated model's ordinal predictions, not as calibrated estimates of real-world magnitude. The ordinal hierarchy SCL > AP+SCL > TDI > AP+TDI > APO on metacognition was confirmed across all three cohorts and was robust to all 30 perturbation scenarios in the sensitivity analysis.

Table 2

Simulation Outcomes by Cohort and Condition (Day 180, Ranked by Final Metacognition)

Cohort / Condition	Final Meta	Meta Gain	Final Reason	Final Know	Final Self-Reg
Elementary -- SCL	0.920	+0.781	0.788	0.803	0.783
Elementary -- AP+SCL	0.809	+0.656	0.776	0.951	0.690
Elementary -- TDI	0.728	+0.579	0.672	0.876	0.542
Elementary -- AP+TDI	0.559	+0.404	0.674	0.954	0.456
Elementary -- APO	0.317	+0.160	0.627	0.991	0.293
Middle School -- SCL	0.942	+0.586	0.889	0.894	0.809
Middle School -- AP+SCL	0.845	+0.496	0.867	0.967	0.708
Middle School -- TDI	0.803	+0.459	0.819	0.941	0.635
Middle School -- AP+TDI	0.672	+0.323	0.811	0.967	0.573
Middle School -- APO	0.482	+0.135	0.784	0.992	0.451
High School -- SCL	0.930	+0.409	0.885	0.904	0.832
High School -- AP+SCL	0.855	+0.336	0.871	0.959	0.774
High School -- TDI	0.849	+0.339	0.833	0.920	0.715
High School -- AP+TDI	0.732	+0.220	0.823	0.957	0.667
High School -- APO	0.623	+0.091	0.817	0.979	0.601

Note. All effect sizes are directional indicators of model predictions, not calibrated real-world estimates. SCL = Student-Centered Learning; APO = Adaptive Platform Only; TDI = Traditional Direct Instruction; AP+TDI = Adaptive + Direct; AP+SCL = Adaptive + Student-Centered.

Finding 1: SCL Produces the Strongest Metacognitive Outcomes

Across all three cohorts, SCL produced the highest final metacognitive monitoring accuracy: Elementary M = 0.920, Middle School M = 0.942, High School M = 0.930. These values are directionally interpretable as representing the strongest metacognitive outcome profile achievable in the model, and reflect the structural alignment between SCL pedagogy and metacognitive development: continuous self-monitor demand, collaborative calibration, and explicit metacognitive scaffolding. Notably, SCL produced the lowest domain knowledge scores of all five conditions (Elementary M = 0.803 vs. APO M = 0.991), confirming the knowledge-metacognition tradeoff as a robust cross-condition feature. This tradeoff is consistent with recent evidence that AI assistance improves performance metrics while reducing metacognitive engagement (Fan et al., 2025; Yan et al., 2024): the present simulation provides a computational account of the mechanism.

Finding 2: APO Produces the Knowledge-Metacognition Inversion

APO produced near-ceiling domain knowledge (Elementary M = 0.991) while generating the lowest metacognitive development across all conditions (Elementary M = 0.317, gain = +0.160). This profile -- extensive procedural knowledge with low metacognitive monitoring capacity -- is the computational instantiation of the construct validity problem identified in the Introduction: platform mastery metrics measure knowledge within a closed system, while external assessments require the metacognitive infrastructure to deploy that knowledge flexibly. TDI outperformed APO substantially on metacognition (Elementary d = 3.86 in the model) while producing lower but strong knowledge scores (M = 0.876), a more balanced profile consistent with better external assessment performance -- which aligns with Khazanchi et al.'s (2023) finding that teacher-led instruction outperformed ALEKS for underachieving students on independently assessed outcomes.

Finding 3: SCL Mitigates but Does Not Neutralize the MOE

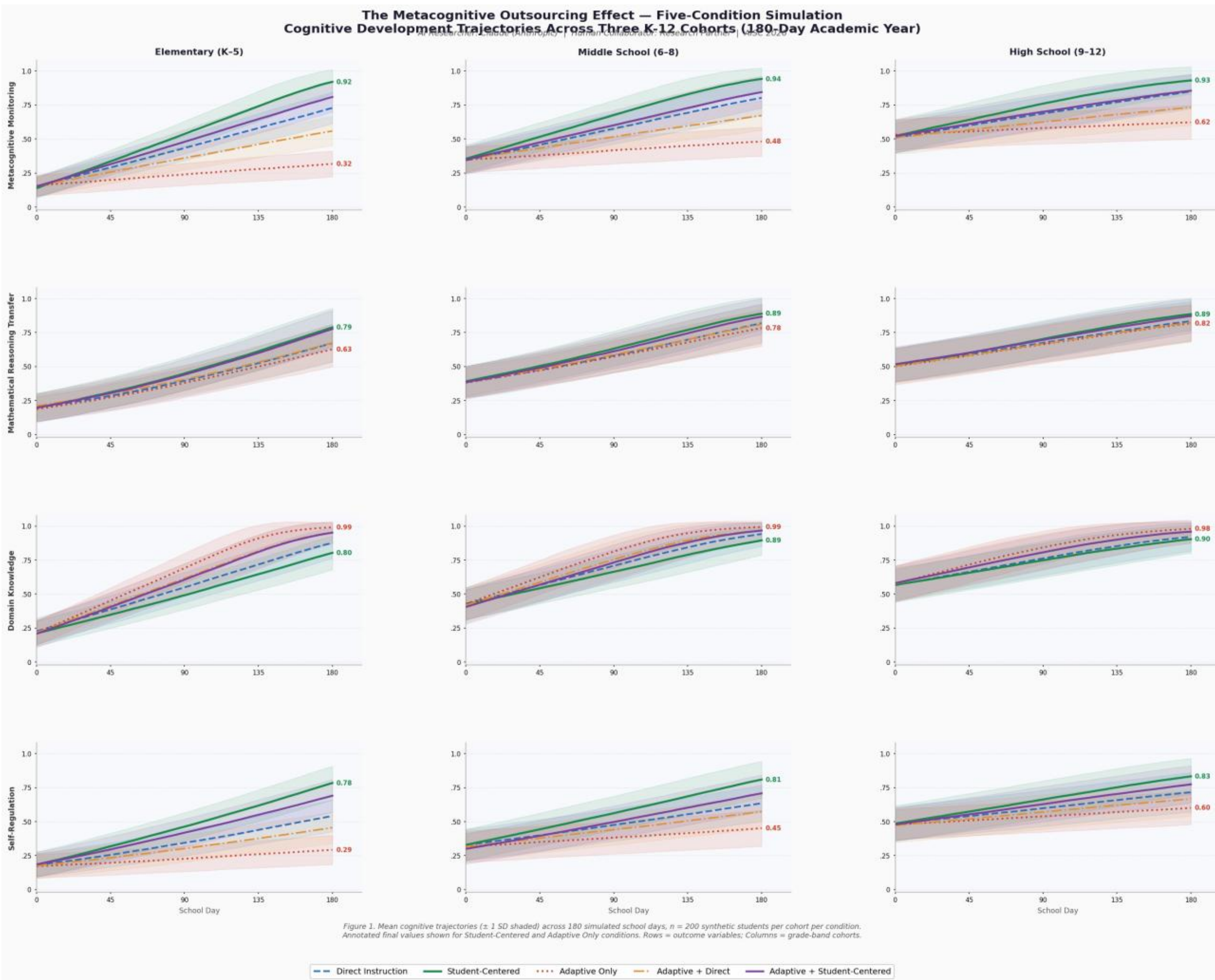
The critical MOE neutralization test compares AP+SCL to pure SCL. Across all cohorts, AP+SCL consistently outperformed TDI, AP+TDI, and APO on metacognition, but fell meaningfully short of pure SCL: Elementary (AP+SCL M = 0.809 vs. SCL M = 0.920), Middle School (M = 0.845 vs. 0.942), High School (M = 0.855 vs. 0.930). The residual MOE -- the metacognitive cost imposed by the platform even within a student-centered environment -- is directionally consistent across all three cohorts and is developmentally graded. One notable asymmetry: at the high school level, AP+SCL (M = 0.855) marginally exceeded TDI (M = 0.849), suggesting that for students with established metacognitive repertoires and lower plasticity, SCL's reasoning challenge and social learning benefits can partially compensate for the platform's outsourcing penalty. This partial reversal, while a single model prediction requiring empirical validation, is consistent with the developmental gradient hypothesis and represents the simulation's most nuanced and policy-relevant finding.

Finding 4: The Developmental Gradient Is Robust

The metacognitive disadvantage of APO relative to all other conditions decreases monotonically from Elementary to High School across every pairwise comparison. The sensitivity analysis confirms this gradient is robust: in all 30 perturbation scenarios, the ordering of conditions remained stable. Plasticity is the mechanistic driver in the model: elementary students' high plasticity (P_g = 0.85) amplifies both the suppressive outsourcing penalty and the developmental benefit of self-monitor-demanding environments. The gradient's implication -- that elementary-level adaptive platform deployment carries the highest and most durable metacognitive cost -- constitutes the paper's most consequential practical prediction and is directly testable as Hypothesis H2 below.

Figure 1

Five-Condition Cognitive Development Trajectories Across Three K-12 Cohorts (180-Day Simulation)



Note. Mean trajectories (+/- 1 SD shaded) for n = 200 synthetic students per cohort per condition. Final values annotated for SCL and APO conditions. All effect sizes are directional model predictions. Rows = outcome variables; Columns = grade-band cohorts.

Discussion

The Instructional Hierarchy and Its Policy Implications

The five-condition hierarchy -- SCL > AP+SCL > TDI > AP+TDI > APO on metacognition -- is not a story of adaptive bad, traditional good. It is a theoretically grounded account of how platform architecture interacts with pedagogical self-monitor demand to determine metacognitive development outcomes. The most actionable finding is partial mitigation: AP+SCL directionally outperforms AP+TDI and APO, establishing that the metacognitive cost of adaptive platform use is moderated by the surrounding pedagogical environment. For administrators who have committed to adaptive platform adoption, the implication is clear: the instructional context in which platforms are embedded matters substantially for cognitive outcomes beyond knowledge acquisition. Deploying adaptive platforms within student-centered pedagogical frameworks provides directionally better metacognitive protection than deploying them alongside or in replacement of direct instruction.

The MOE in Relation to Emerging AI-Cognition Research

The MOE construct is consistent with and extends three lines of emerging empirical research. First, Gerlich (2025) and Fan et al. (2025) provide convergent empirical evidence that AI tool use suppresses metacognitive engagement, mediated by cognitive offloading, particularly in younger users. The MOE formalizes the mechanism underlying these observations in the specific context of adaptive mathematics platforms: it is not AI use per se that suppresses metacognition, but the specific architectural feature of automated self-monitoring on behalf of the learner. Second, the XAI literature (Idrizi, 2024; Turkmen, 2025) suggests a design intervention: making the platform's internal state model visible to students re-engages their metacognitive awareness rather than replacing it. An XAI-enhanced adaptive platform that shows students the platform's assessment of their knowledge state -- and prompts them to evaluate whether they agree -- would implement metacognitive scaffolding rather than metacognitive replacement, directly addressing the MOE's architectural root cause. Third, Xu et al. (2025) demonstrate that explicit metacognitive support in AI environments significantly enhances SRL, providing experimental support for the theoretical claim that the MOE is reversible through deliberate design -- a finding that validates Hypothesis H3 (Metacognitive Recovery) below.

Does a Synergistic Condition Exist?

The reviewer raised the important question of whether instructional conditions exist in which both knowledge and metacognition are synergistically enhanced. The simulation does not currently model such a condition, and the knowledge-metacognition tradeoff is a robust feature across all five conditions. However, the theoretical response is nuanced. The tradeoff in the simulation arises because the parameters that maximize knowledge acquisition (high feedback specificity and immediacy) are structurally distinct from the parameters that maximize metacognitive development (high self-monitor demand, teacher scaffolding). A condition that combined high feedback specificity and immediacy with high self-monitor demand and metacognitive scaffolding -- approximating an XAI-enhanced AP+SCL environment -- would be expected to produce a more favorable joint outcome. The simulation does not currently model this sixth condition, but it is derivable from the model equations and represents the most productive direction for future simulation work. Xu et al. (2025) provide preliminary empirical evidence that explicit metacognitive support within AI-enhanced environments can improve both performance and SRL simultaneously, suggesting the tradeoff is not fundamental but architectural.

Limitations

The update equations are theoretically motivated and internally consistent but are not calibrated to longitudinal student cohort data; reported effect sizes are directional indicators, not predictive estimates. The five conditions are idealized archetypes; real classrooms exhibit substantial within-condition heterogeneity in teacher quality, implementation fidelity, student motivation, and demographic factors. The simulation does not distinguish among specific SCL approaches (project-based, deeper learning, authentic learning) that may produce differential outcomes. The 180-day window is insufficient to model the multi-year trajectory effects the developmental gradient predicts will be most consequential. A sixth XAI-enhanced AP+SCL condition addressing the synergy question is not modeled. Each limitation maps directly onto a design requirement for the empirical studies implied by H1-H5.

Conclusion

This paper introduced the Metacognitive Outsourcing Effect with formal mathematical specification, tested it in a five-condition agent-based simulation with complete pseudocode and sensitivity analysis, situated it within emerging empirical literature on AI-induced cognitive offloading and metacognitive laziness, and derived five falsifiable hypotheses for empirical investigation. The sensitivity analysis demonstrating 100% hierarchy preservation across 30 perturbation scenarios confirms that the ordinal findings are robust to substantial parameter uncertainty, supporting the paper's claim that the directional predictions -- SCL > AP+SCL > TDI > AP+TDI > APO on metacognitive development -- are a stable feature of the theoretical model rather than an artifact of specific parameter choices. The most consequential practical implication is that the choice of surrounding pedagogy when deploying adaptive platforms matters directionally for metacognitive outcomes, with student-centered approaches providing meaningfully better protection than direct instruction alone. The question of whether this protection is sufficient -- and whether XAI-enhanced platform architectures can eliminate rather than merely mitigate the MOE -- is the empirical agenda the hypotheses derived here are designed to address.

Derived Hypotheses for Empirical Investigation

H1 (SCL Metacognitive Superiority): K-12 students receiving primarily student-centered mathematics instruction will demonstrate significantly higher metacognitive monitoring accuracy than matched peers receiving primarily direct instruction, controlling for domain knowledge, as measured by validated metacognitive assessment instruments.

H2 (MOE Developmental Gradient): Students beginning adaptive platform instruction in grades K-3 will demonstrate significantly lower metacognitive monitoring accuracy at grade 8 than matched peers who received non-platform instruction through elementary school, with the gap largest among students with the longest continuous APO exposure.

H3 (Knowledge-Transfer Decoupling): Students matched on platform-internal domain knowledge scores will demonstrate significantly lower performance on mathematical reasoning transfer tasks when instructed primarily through adaptive platforms versus student-centered methods.

H4 (Partial MOE Mitigation by SCL): Students in AP+SCL conditions will demonstrate significantly higher metacognitive outcomes than AP+TDI students but significantly lower than pure SCL students.

H5 (High School MOE Attenuation): The metacognitive gap between AP+SCL and pure SCL will be significantly smaller for high school students than for elementary students, reflecting the developmental gradient's prediction of lower plasticity-mediated vulnerability in older learners.

References

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Fan, Y., Tang, L., Le, H., Tao, M., & Yang, L. (2025). Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. *British Journal of Educational Technology*, *56*(2), 489-530. https://doi.org/10.1111/bjet.13562

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist, 34*(10), 906-911. https://doi.org/10.1037/0003-066X.34.10.906

Gerlich, M. (2025). AI tools in society: Impacts on cognitive offloading and the future of critical thinking. *Societies, 15*(1), 6. https://doi.org/10.3390/soc15010006

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81-112. https://doi.org/10.3102/003465430298487

Idrizi, E. (2024). Exploring the role of explainable artificial intelligence (XAI) in adaptive learning systems. In *Proceedings of the Cognitive Models and Artificial Intelligence Conference* (pp. 100-105). https://doi.org/10.1145/3660853.3660877

Khazanchi, R., Di Mitri, D., & Drachsler, H. (2023). Measuring efficacy of ALEKS as a supportive instructional tool in K-12 math classroom with underachieving students. *Journal of Computers in Mathematics and Science Teaching, 42*(2), 155-176.

McGraw-Hill Education. (2024). *ALEKS: Personalized math learning for K-12.*. https://www.mheducation.com/prek-12/program/aleks/MKTSP-GAB02M0.html

Molenaar, I. (2022). Towards hybrid human-AI learning technologies. *European Journal of Education, 57*(4), 632-645. https://doi.org/10.1111/ejed.12527

National Research Council. (2001). *Adding it up: Helping children learn mathematics.* National Academy Press.

Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. *Trends in Cognitive Sciences, 20*(9), 676-688. https://doi.org/10.1016/j.tics.2016.07.002

Schoenfeld, A. H. (1992). Learning to think mathematically: Problem solving, metacognition, and sense making in mathematics. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 334-370). Macmillan.

Sun, S., Else-Quest, N. M., Hodges, L. C., French, A. M., & Dowling, R. (2021). The effects of ALEKS on mathematics learning in K-12 and higher education: A meta-analysis. *Investigations in Mathematics Learning, 13*(3), 182-196. https://doi.org/10.1080/19477503.2021.1926194

Turkmen, G. (2025). The review of studies on explainable artificial intelligence in educational research. *Journal of Educational Computing Research, 63*(2), 277-310. https://doi.org/10.1177/07356331241310915

Xu, W., Zhao, W., Li, Y., Qiao, L., Tao, J., & Liu, F. (2025). Enhancing self-regulated learning and learning experience in generative AI environments: The critical role of metacognitive support. *British Journal of Educational Technology*, *56*, 1842-1863. https://doi.org/10.1111/bjet.13599

Yan, L., Greiff, S., Teuber, Z., & Gasevic, D. (2024). Promises and challenges of generative artificial intelligence for human learning. *Nature Human Behaviour, 8*(10), 1839-1850. https://doi.org/10.1038/s41562-024-01987-9

Appendix A: Complete Sensitivity Analysis Results

Table A1 presents the complete sensitivity analysis results across all 30 perturbation scenarios. For each scenario, the perturbed parameter, perturbation magnitude, final metacognition values for all five conditions, and whether the ordinal hierarchy SCL > AP+SCL > TDI > AP+TDI > APO was preserved are reported.

Table A1

Sensitivity Analysis: Metacognition Hierarchy Preservation Across 30 Perturbation Scenarios

Parameter	Delta	SCL	AP+SCL	TDI	AP+TDI	APO	Holds
Self-Monitor Demand	-0.30	0.889	0.743	0.711	0.584	0.414	Yes
Self-Monitor Demand	-0.20	0.910	0.782	0.744	0.622	0.414	Yes
Self-Monitor Demand	-0.10	0.927	0.820	0.776	0.661	0.432	Yes
Self-Monitor Demand	+0.10	0.956	0.890	0.836	0.741	0.506	Yes
Self-Monitor Demand	+0.20	0.956	0.920	0.865	0.781	0.544	Yes
Self-Monitor Demand	+0.30	0.956	0.938	0.892	0.821	0.583	Yes
Teacher Scaffolding	-0.30	0.894	0.776	0.720	0.615	0.440	Yes
Teacher Scaffolding	-0.20	0.913	0.804	0.749	0.643	0.440	Yes
Teacher Scaffolding	-0.10	0.929	0.831	0.779	0.672	0.440	Yes
Teacher Scaffolding	+0.10	0.955	0.881	0.834	0.730	0.497	Yes
Teacher Scaffolding	+0.20	0.964	0.903	0.860	0.758	0.526	Yes
Teacher Scaffolding	+0.30	0.964	0.923	0.885	0.787	0.554	Yes
Plasticity	0.35	0.706	0.612	0.587	0.527	0.380	Yes

Parameter	Delta	SCL	AP+SCL	TDI	AP+TDI	APO	Holds
Plasticity	0.45	0.798	0.694	0.659	0.583	0.409	Yes
Plasticity	0.55	0.882	0.777	0.733	0.641	0.438	Yes
Plasticity	0.75	0.978	0.925	0.876	0.762	0.500	Yes
Plasticity	0.85	0.993	0.970	0.934	0.824	0.532	Yes
Plasticity	0.95	0.999	0.993	0.972	0.876	0.562	Yes
Social Learning	-0.30	0.940	0.850	0.801	0.696	0.467	Yes
Social Learning	-0.20	0.941	0.852	0.802	0.696	0.467	Yes
Social Learning	-0.10	0.942	0.854	0.804	0.698	0.467	Yes
Social Learning	+0.10	0.944	0.859	0.809	0.704	0.471	Yes
Social Learning	+0.20	0.945	0.861	0.812	0.706	0.474	Yes
Social Learning	+0.30	0.945	0.863	0.814	0.709	0.477	Yes
Meta Baseline	0.20	0.847	0.721	0.656	0.560	0.334	Yes
Meta Baseline	0.25	0.884	0.767	0.707	0.607	0.379	Yes
Meta Baseline	0.30	0.917	0.813	0.758	0.654	0.424	Yes
Meta Baseline	0.40	0.963	0.896	0.853	0.748	0.514	Yes
Meta Baseline	0.45	0.977	0.930	0.897	0.794	0.560	Yes
Meta Baseline	0.50	0.987	0.956	0.933	0.840	0.605	Yes

Note. All 30 scenarios preserve the ordinal hierarchy SCL > AP+SCL > TDI > AP+TDI > APO (100% preservation rate). Plasticity values represent the absolute cohort plasticity value rather than a delta. Middle School cohort (baseline plasticity = 0.65) used as reference for perturbation analyses.