Journal of Intelligent Decision Making and Information Science

www.jidmis.org

eISSN: 3079-0875



Generativa Between Cognitive Offloading and Thinking: A Theoretical Framework of Metacognitive Tension in the Use of Generative AI

[1]Reinaldo Antonio Guerrero-Chirinos, [2]Miguel Ángel Medina-Romero, [3]Jean Pierre Reyes-Carrión, [4]Fernando Montemayor-Ibarra, [5]Yamira Iraisa Herrera-García, [6]José Darwin Farje-Escobedo

[1]Universidad Técnica Particular de Loja, UTPL (Ecuador) raguerrero12@utpl.edu.ec https://orcid.org/0000-0003-0499-7453

[2]Universidad Michoacana de San Nicolás de Hidalgo, UMSNH (México) miguel.medina.romero@umich.mx https://orcid.org/0000-0003-4067-2816

[3]Universidad Nacional de Educación (UNAE) Ecuador jean.reyes@unae.edu.ec https://orcid.org/0000-0003-4648-4744

[4]Universidad Autónoma de Nuevo León, AUNL (México) fernando.montemayorib@uanl.edu.mx https://orcid.org/0000-0003-3503-0149

[5]Universidad Nacional Toribio Rodríguez de Mendoza de Amazonas, UNTRM (Perú) Yamira.herrera@untrm.edu.pe https://orcid.org/0000-0003-2881-5992

[6]Universidad Nacional Toribio Rodríguez de Mendoza de Amazonas, UNTRM (Perú) jose.farje@untrm.edu.pe https://orcid.org/0000-0002-2164-6682

ARTICLE INFO

Article history:
Received 02 April 2026
Received in revised form 30 April 2026
Accepted 20 May 2026

Keywords: generative artificial intelligence; cognitive discharge; metacognition; desirable difficulties; self-regulation of learning; cognitive load; calibration.

ABSTRACT

Resumen

The mass adoption of generative artificial intelligence in learning revives a classic question: when does the support of an external tool empower thinking and when does it replace it? This conceptual article, aimed at the construction of theory through theoretical synthesis, integrates three bodies of literature, cognitive download, desirable difficulties and metacognition, to theorize the tension between delegating cognition to AI and developing self-regulation. Its distinctive contribution in the face of close frameworks is to specify boundary conditions. The threshold model of metacognitive tension is proposed, which defines a mediating mechanism, the preservation against the suppression of the diagnostic cues of monitoring and the occupation of the metalevel, three moderators, including the learner's expertise, and a threshold conceived as a transition band in the plane of the locus of discharge and metacognitive posture. It derives a typology of four uses and seven propositions that can be contrasted with their evidential status, and identifies implications for pedagogical, tool, and research design.

Abstract

The widespread adoption of generative artificial intelligence in learning revives a classic question: when does support from an external tool enhance thinking, and when does it replace it? This conceptual article, oriented toward theory building through theoretical synthesis, integrates three consolidated bodies of literature, cognitive offloading, desirable difficulties, and metacognition, to theorize the tension between delegating cognition to AI and developing self-regulation. Its distinctive contribution relative to neighboring frameworks is to specify boundary conditions. It proposes the Metacognitive Tension Threshold Model, which defines a mediating mechanism, the preservation versus suppression of the diagnostic cues of monitoring and the occupation of the metalevel, three moderators, including learner expertise, and a threshold located as a transition band within a plane of two determinants, the locus of offloading and metacognitive posture. The model derives a typology of four uses and seven testable propositions with their evidential status, and identifies implications for pedagogical design, tool design, and research.

1. Introduction

All cognitive technology reorganizes the distribution of mental work between the person and his environment. Writing externalizes memory, the calculator externalizes arithmetic, and the search engine outsources access to information. Generative artificial intelligence intensifies this trend because it no longer outsources storage or computation, but the production of text, code, and arguments themselves. This novelty forces us to rethink what is left on the learner's side when the tool can execute the entire task. Research on AI in higher education has grown rapidly in recent years, making it urgent to theoretically order its effects (Crompton & Burke, 2023).

Cognitive psychology describes this externalization as cognitive discharge, that is, the use of physical actions or external resources to reduce the demand for internal processing (Risko & Gilbert, 2016). Discharge is not inherently harmful: the extended mind works by coupling internal and external resources into a single cognitive system (Clark & Chalmers, 1998). The educational problem arises when what is unloaded is precisely what the learner had to learn to do for himself.

Recent evidence suggests that the risk is real. In an experimental study with students who wrote and reviewed essays, those with a generative model produced better texts, but showed no subsequent learning improvements and exhibited what the authors call metacognitive laziness, that is, fewer self-correcting behaviors and less time spent reflecting (Fan et al., 2025). At the same time, other work shows that AI can stimulate thinking when the learner interrogates their outputs instead of simply accepting them.

There is, therefore, an unresolved tension. The same tool can free up resources for higher-order thinking or empty the task of its formative value. The literature documents both poles, but lacks a framework that specifies the conditions that tilt the outcome toward one or the other. This theoretical gap is the one addressed in this work.

The goal is to theorize the tension between cognitive download and metacognitive development, and to specify the conditions under which generative AI strengthens or erodes metacognition. The contribution is twofold: to articulate a mediating mechanism with its moderators and, in the face of close frameworks that describe the phenomenon without limiting it, to offer boundary conditions and verifiable propositions. No new empirical data are presented, but an integrative synthesis intended to guide design and research.

2. Methodological approach: theoretical synthesis

This article adopts the genre of theory construction through theoretical synthesis. Unlike systematic review, whose purpose is to aggregate empirical evidence under exhaustiveness protocols, theoretical synthesis seeks to integrate constructs from different traditions into a coherent and generative conceptual framework. The product is not an inventory of studies, but a model with mechanisms and propositions. The integration of cognitive load theory with self-regulation illustrates this way of proceeding (de Bruin et al., 2020).

The procedure followed three steps. First, three theoretical bodies were identified whose intersection defines the problem: cognitive download, desirable difficulties, and metacognition with self-regulation. Second, seminal sources and high-impact reviews were selected that offered well-defined constructs and explicit mechanisms. Third, the points of friction and complementarity between these traditions were sought in order to derive a model that reconciles them.

The inclusion criteria were the theoretical centrality of the construct and the solidity of the source, with a preference for first-quartile publications of recent years, complemented by seminal works from the field. As it is a narrative and integrative synthesis, and not a systematic review, the selection

of sources entails a component of judgment that is explicitly recognized as a limit. To mitigate confirmation bias, heterogeneous findings and evidence that nuances the model's theses, such as the absence of a single effect of AI on self-regulation, were deliberately incorporated (Lan & Zhou, 2025). The synthesis is deliberately parsimonious and leaves open its empirical extension.

3. Conceptual pillars

3.1 Cognitive download as a metacognitive behavior

Cognitive discharge designates the use of external actions to alter the processing demands of a task and reduce its mental demand (Risko & Gilbert, 2016). The decisive thing about his formulation is that the decision to discharge is governed by metacognitive evaluations: we discharge when we judge that internal demand exceeds our capacities. Recent experimental research confirms that this propensity responds both to the cost of effort of external action and to judgments of trust, which can be erroneous (Chiu & Gilbert, 2024). That decision has measurable consequences: the download improves immediate performance, but it can deteriorate the subsequent memory of delegated information (Grinschgl et al., 2021).

From the perspective of the extended mind, external resources can be integrated into the cognitive system without impairing thinking (Clark & Chalmers, 1998). However, generative AI poses a borderline case, as it can assume not only storage or computation, but also deliberation. Some authors describe this possibility, still speculatively, as a 0 System, an external cognitive layer that precedes human thought and that, if poorly managed, can displace one's own judgment (Chiriatti et al., 2024).

It should be noted that the opposition suggested in the title, between download and thought, is heuristic. Discharge can be part of thinking, as the extended mind thesis maintains, so that the problem is not discharge, but what is discharged and with what posture. This distinction is what the model tries to formalize.

3.2 Metacognition: Construct Hierarchy, Monitoring, and Calibration

Metacognition comprises knowledge about one's own cognition and its regulation (Flavell, 1979). Its classical architecture distinguishes an object level, where the cognition of the task occurs, and a metalevel that monitors and controls that lower level through a feedback loop (Nelson & Narens, 1990). The monitoring informs the control, and the control adjusts the post-processing; The quality of learning depends on the fidelity of that loop. The metacognitive and affective model of self-regulation expands on this architecture by integrating metacognitive experiences, motivation, and affect (Efklides, 2011). A broad meta-analysis of the field, moreover, organizes sixteen self-regulatory constructs and confirms that they are interrelated and that they predict learning across contexts (Sitzmann & Ely, 2011). It is worth specifying the hierarchy of constructs that the model uses, as they are often confused. Self-regulation of learning encompasses metacognition (Zimmerman, 2008), which includes monitoring and control; Calibration is a specific property of monitoring accuracy. The main dependent variables of this model are the accuracy of monitoring, i.e., calibration, and deferred and transferable learning, and subsidiarily the self-regulatory capacity sustained over time. A critical point is that monitoring does not directly access knowledge, but is inferred from cues that the learner generates when processing the task, so that its accuracy depends on the diagnosis of those cues (Koriat, 1997). When these cues are not diagnostic, monitoring becomes inaccurate and regulation is diverted (Baars et al., 2020). The fluency of a response can generate illusions of competence that inflate confidence (Bjork et al., 2013), and the resulting miscalibration impairs learning and retention (Dunlosky & Rawson, 2012). The integration of

cognitive load with self-regulation underscores that the effort invested is both a source of metacognitive information and an object of regulation (de Bruin et al., 2020).

3.3 Desirable difficulties and their relationship with cognitive load

The notion of desirable difficulties holds that certain practice conditions that slow down immediate performance produce more lasting and transferable learning (Bjork & Bjork, 2020). Retrieval, spacing, interleaving, and generation are paradigmatic examples. Recent meta-analytic reviews strongly support this principle: retrieval practice consistently benefits learning in real classrooms (Adesope et al., 2017; Agarwal et al., 2021), and the effect of testing elevates performance across hundreds of studies (Yang et al., 2021). The synthesis of study techniques places the practice of retrieval and the distribution of the study among the most effective (Dunlosky et al., 2013).

This tradition distinguishes between performance and learning. Performance is what is observed during practice; learning is the relatively permanent change in ability, which only manifests itself in a delayed manner and in new contexts (Soderstrom & Bjork, 2015). The consequence is counterintuitive: momentary fluency can be a bad sign of real learning, and inflate the learner's confidence to the point of deteriorating their calibration.

Desirable difficulties and cognitive load theory maintain a recognized friction: the former recommends introducing difficulty, while the latter prescribes reducing the non-task load (Sweller et al., 2019; Sweller, 2023). The model reconciles them by distinguishing the type of load affected. Introducing desirable difficulty operates on the schema-building processing, while reducing the burden is benign only when it falls on the extrinsic. This distinction will be central to the mechanism of the model.

3.4 Generative AI, metacognitive laziness, and overdependence

Research on generative AI in learning has identified two recurring findings, although with results that are still heterogeneous. The first is that generative assistance improves the immediate output, but may not translate into learning, and is associated with a reduction in self-regulatory behaviors that the authors call metacognitive laziness (Fan et al., 2025). The second is that confidence in the tool shapes its use independently of its accuracy. Along these lines, a study on AI assistance found that students tended to depend on help rather than learn from it, and that self-regulation scaffolds replaced it only partially (Darvishi et al., 2024). From a hybrid intelligence perspective, it has been argued that AI can externalize, modify, or extend human cognition, so that the design problem is not the tool in the abstract, but which of these modes predominates in each use (Cukurova, 2025). A complementary line conceptualizes the relationship as a hybrid regulation between human and AI, in which the tool can assume part of the planning, monitoring, and reflection that previously fell to the learner (Molenaar, 2022). The study of how this regulation is distributed, supported by multimodal data, shows that AI can both sustain and displace self-regulation according to the degree of control that the learner retains (Molenaar et al., 2023). Field reviews confirm that AI can encourage or hinder self-regulation, without a single effect (Lan and Zhou, 2025).

Experimental evidence on AI systems shows that the mere knowledge that a piece of advice comes from an AI induces people to follow it even when it contradicts one's own judgment, resulting in overdependence and measurable costs (Klingbeil et al., 2024). It is advisable to read this result with caution, as it comes from a limited number of studies, but its mechanism is clear: if judgment is delegated, the opportunity to regulate learning is also delegated. Recent meta-analyses are heterogeneous: generative AI can improve performance and higher-order thinking while reducing the mental effort invested (Deng et al., 2025; Wang & Fan, 2025) and, in inquiry tasks, engage the depth of reasoning (Stadler et al., 2024). These findings share an underlying mechanism. When the

tool absorbs generation and judgment, the learner stops producing the diagnostic cues that sustain monitoring. Metacognitive laziness and overdependence are not separate phenomena, but manifestations of the same process of suppressing clues. The model it follows formalizes that intuition.

4. The threshold model of metacognitive tension

4.1 Premises

The model is based on five premises derived from the previous pillars. First, cognitive discharge is a behavior governed by metacognitive judgment and, therefore, regulable (Risko & Gilbert, 2016). Second, lasting learning depends on preserving some productive effort, i.e., desirable difficulties (Bjork & Bjork, 2020; Soderstrom & Bjork, 2015). Third, monitoring is nourished by cues that the learner generates when processing the task (Baars et al., 2020). Fourth, poorly calibrated confidence increases discharge regardless of accuracy (Klingbeil et al., 2024). Fifth, the threshold is not universal, but shifts with the skill of the learner, so that the same use can be augmentative for an expert and substitute for a novice (Kalyuga et al., 2003; Sweller, 2023).

4.2 Mediating mechanism and moderators

The model carefully distinguishes a mediating mechanism from a set of moderators. The mediating mechanism, that through which the discharge affects metacognition, is the preservation against suppression of the diagnostic cues that the learner generates when processing the task. When the trainee delegates production, he or she stops experiencing the effort and errors that served as a signal to judge his or her understanding, and monitoring loses its raw material, which favors overconfidence based on the fluidity of the output (Baars et al., 2020; Nelson & Narens, 1990).

The occupation of the metalevel requires this mechanism. In healthy use, the download frees the learner from the object level to invest their effort in evaluating, integrating, and deciding on the output of the AI. In erosive use, the tool also occupies the metalevel: the learner not only stops producing, but also stops judging, approaching the total delegation that some describe as System 0 (Chiriatti et al., 2024). Three moderators operate on this mechanism that displace the threshold, different from the two dimensions that constitute its plane, the locus of discharge and the metacognitive posture (Figure 1). The first moderator is task design and preservation of desirable difficulties: cognitive load theory provides the criterion that offloading extrinsic load is usually benign, while offloading intrinsic processing, i.e., resources aimed at constructing schemas, erodes learning (Sweller, 2023). The second is the calibration of trust: overconfidence pushes towards unjustified delegation and verification contains it, although in a different way perceived incompetence can induce compensatory delegation, so that it is uncalibrated confidence, in either direction, that shifts use towards the substitute zone (Klingbeil et al., 2024); useful confidence is the one that is calibrated to the actual reliability of the tool (Lee & See, 2004). The third moderator is the learner's expertise. Due to the effect of reversal by expertise, the depth of discharge that an expert tolerates without prejudice is erosive for a novice, whose learning still depends on the processing that the AI would absorb (Kalyuga et al., 2003); consequently, explicit guidance helps the novice, but loses value as prior knowledge grows (Kirschner et al., 2006). The locus of effort and the metacognitive posture are not moderators here, but the two dimensions whose combination defines the position of each use with respect to the threshold. The three moderators do not operate in isolation: their joint configuration shifts the band and repositions each episode of use.

4.3 The threshold and conditions for strengthening

The threshold is defined by the observable configuration of the use, not by its outcome, which avoids circularity. A use is above the threshold when two observable features concur: the discharge reaches the nuclear generative processes and the learner adopts a passive posture, with little verification or editing of the outputs. Below the threshold, the discharge is augmentative and frees up resources that the learner reinvests in higher-order judgment. The effect on metacognition, miscalibration or its improvement is measured independently, which makes the relationship between configuration and result falsifiable. The threshold is not a single value on a single variable, but a transition band in the plane defined by its two main determinants, the locus of discharge and the metacognitive posture (Figure 1). The underlying continuum is the degree of metacognitive substitution, which grows as the discharge occupies nuclear processes and the learner becomes accepting. The band is not fixed: the moderators, especially the expertise, move it for each apprentice and task. Generative AI strengthens metacognition when four conditions are met together. First, the learner retains the monitoring and control loop. Second, the effort is relocated to higher-order processes instead of being eliminated. Third, trust is calibrated and is accompanied by verification behaviors. Fourth, the task preserves at least one desirable difficulty, adjusted to the skill of the apprentice. Erosion occurs when the discharge crosses the threshold and the monitoring keys are no longer generated by the learner.

4.4 Types of uses

The transition band organizes a typology of four uses (Figure 1). The horizontal axis represents the locus of discharge, from extrinsic charge to nuclear generative processes. The vertical axis represents the metacognitive posture, from an active and verifying attitude to a passive and accepting one. It should be noted that both determinants are conceptually separable, but they tend to correlate in practice, so that the cells are not equiprobable.

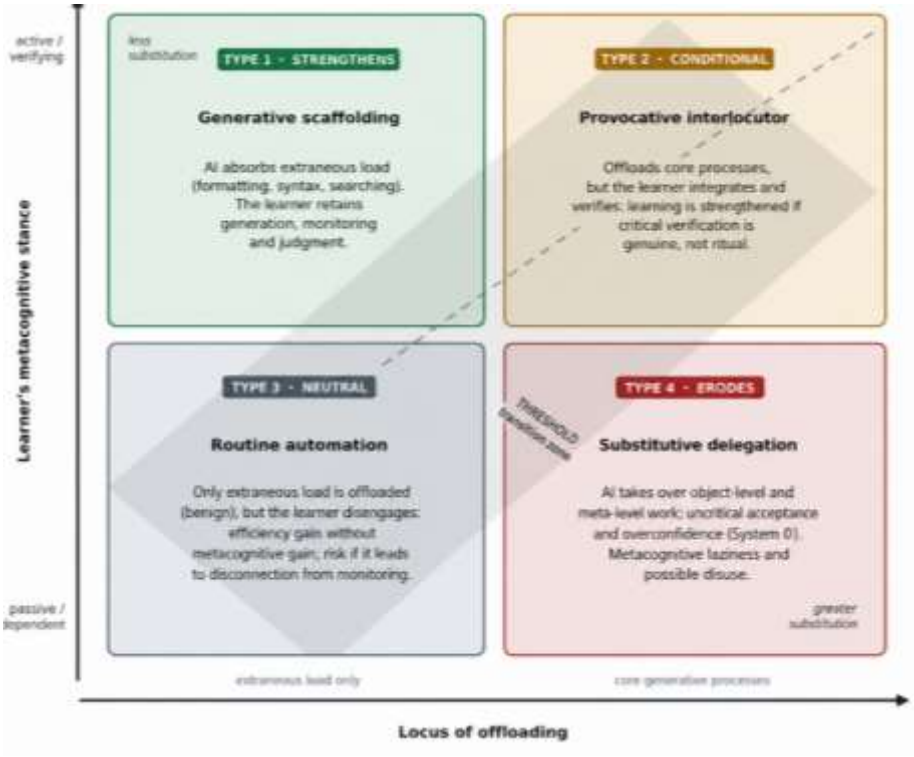


Figure 1. Typology of generative AI uses according to the threshold model. The diagonal band represents the threshold as a transition zone on the metacognitive substitution continuum.

Type 1, generative scaffolding, falls below the threshold: AI takes over formatting, syntax, or search, and the learner retains generation, monitoring, and judgment, strengthening metacognition. Type 4, substitute delegation, comes first: AI occupies the object level and the metalevel, with uncritical acceptance and inflated confidence, which produces metacognitive laziness and possible disuse. They are the two poles of the substitution continuum.

Types 2 and 3 are located on the transition band and their effect is conditional. Type 2, provocative interlocutor, unloads core processes, but retains an active stance: asking AI to question a thesis or generate counterexamples relocates the desirable difficulty in critical verification and strengthens metacognition, as long as that verification is genuine and not a ritual procedure. Type 3, automation of routines, discharges only extrinsic cargo, which in itself is benign (Sweller, 2023); Its risk does not come from the depth of the discharge, but from the passive posture missing the monitoring opportunity, with a net neutral effect unless it leads to sustained disconnection.

4.5 Integrative synthesis

Figure 2 integrates the above elements. The use of generative AI acts as an antecedent of cognitive download, which affects metacognition through the mediating mechanism of preservation versus key suppression and metalevel occupation. The three moderators position each use with respect to the threshold, from which two trajectories fork. The augmentative trajectory leads to strengthened metacognition and feeds a virtuous loop, in which the best calibration makes the subsequent discharge more judicious. The substitutionary trajectory leads to eroded metacognition and feeds an erosive loop: the fluidity of the outputs inflates confidence and impairs calibration, which increases the propensity to delegate to subsequent episodes.

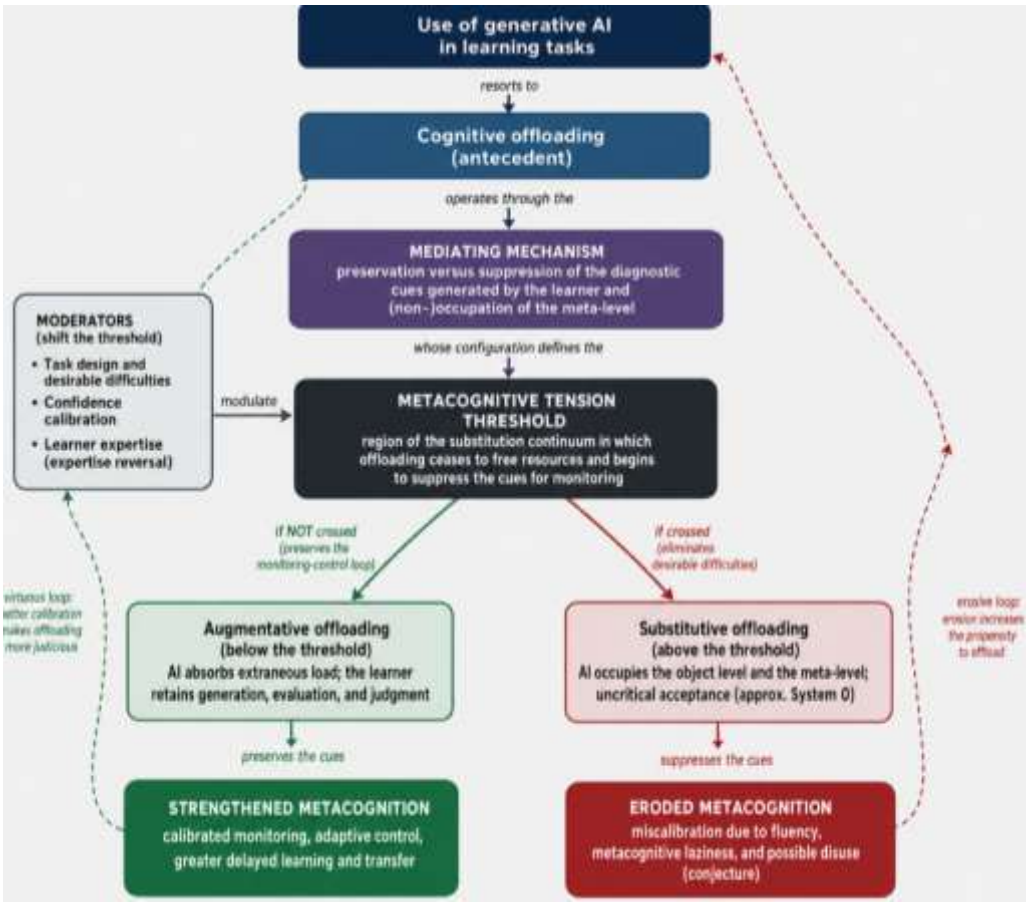


Figure 2. Conceptual map of the threshold model of metacognitive tension, with the mediating mechanism, moderators, and virtuous and erosive loops.

The map makes it visible that the outcome does not depend on the tool itself, but on the joint configuration of the mechanism and moderators. The same AI can be placed on either of the two trajectories depending on how the activity is designed, the confidence is calibrated and the demand is adjusted to the expertise.

4.6 Relationship with upcoming frameworks

The model dialogues with three recent frameworks and is distinguished from them by specifying boundary conditions. The notion of System 0 describes an external cognitive layer that precedes thinking, but does not specify when it is beneficial or harmful (Chiriatti et al., 2024). Human-AI hybrid regulation characterizes the distribution of regulatory functions, although it leaves the conditions for this distribution partially open (Molenaar, 2022; Molenaar et al., 2023). Metacognitive laziness empirically documents the reduction of self-regulatory behaviors, but as a finding, not as a framework of conditions (Fan et al., 2025). The present model provides the mechanism, moderators and threshold that articulate these observations in conditional predictions (Table 1).

Framework	Unit of analysis	Central mechanism	Border conditions?
System 0 (Chiriatti et al., 2024)	Human cognitive system and AI	External cognitive layer that precedes thought	It does not specify them
Hybrid regulation (Molenaar, 2022; Molenaar et al., 2023)	Sharing of self-regulation	Transferring regulatory functions to AI	Partially
Metacognitive laziness (Fan et al., 2025)	Intra-task learning process	Reduction of self-regulatory behaviors	No (empirical finding)
Threshold Model (this job)	Download Episode	Preservation vs. Key Deletion and Metalevel Occupation	Yes: threshold, three moderators and typology

Table 1. Comparison of the threshold model with nearby frames.

5. Theoretical propositions

The model is specified in seven testable propositions. Next to each one, its main dependent variable and its current evidential status are indicated, distinguishing the partially supported ones from the conjectures that require new evidence. It should be noted that this partial support comes largely from the literature on cognitive download, metamemory and calibration outside of AI; direct evidence on the accuracy of monitoring in generative AI contexts is still scarce, so calibration-focused propositions should be read as testable extrapolations and not as established findings in this domain.

P1 *(calibration; partial backing).* With the same task, the download that eliminates the generation of diagnostic keys reduces the accuracy of the trainee's monitoring, even if it improves the quality of the immediate product.

P2 *(deferred learning; partial support).* The discharge that relocates the effort to the metalevel, through evaluation, integration or contrast of the outputs of the AI, produces greater delayed learning than the discharge of equal intensity that does not relocate that effort.

P3 *(discharge intensity; partial backup).* Poorly calibrated confidence in AI increases the intensity of the discharge regardless of the accuracy of the tool, shifting the use to the replacement area.

P4 *(learning; conceptual support).* The preservation of at least one desirable difficulty in the task moderates the relationship between download and learning, attenuating metacognitive erosion.

P5 *(threshold boundary; conjecture derived from cognitive load).* The threshold shifts with expertise: the greater the prior knowledge, the greater the depth of discharge tolerable without erosion, due to the effect of reversal by expertise.

P6 *(self-regulatory capacity; conjecture).* Sustained substitute use is associated with a possible impairment due to disuse of the self-regulatory capacity, observable when the tool is removed; it requires longitudinal evidence.

P7 *(system dynamics; conjecture).* There is an erosive loop and, symmetrically, a virtuous one: the calibration resulting from one episode modulates the intensity and character of the discharge in the following episodes.

5.1 Indicators for Testing

The proposals admit an initial operationalization. The crossing of the threshold can be estimated from the proportion of the response generated by the AI to the learner and the presence of verification behaviors, such as substantive edits, contrast queries, or justifications. The diagnosis of the keys is approximated by the divergence between learning judgments and deferred performance, that is, by calibration. The locus of effort is inferred from the distribution of time and activity between production and evaluation, accessible through multimodal process data (Molenaar et al., 2023). Delayed learning is measured with unassisted testing after a delay (Soderstrom & Bjork, 2015). To separate this mechanism from a rival explanation for mere reduction of practice, it is convenient to equalize the amount of recovery and vary only the generation of diagnostic cues: the model predicts that calibration deteriorates even at the same practice.

6. Implications

6.1 For pedagogical design

The model suggests that the task, and not the prohibition of the tool, is the main lever. A derivative and non-obvious recommendation is to require learner generation or prediction before exposing the AI output, so that diagnostic cues that would otherwise be suppressed by the tool are preserved. Another is to grade friction according to expertise: removing scaffolding for the expert and preserving difficulty for the novice, consistent with reversion by expertise (Kalyuga et al., 2003).

A third implication is to decouple the assessment of assisted performance from actual learning, through deferred testing without AI that reveals the distance between the sense of mastery and effective learning (Soderstrom & Bjork, 2015), with feedback that corrects monitoring (Baars et al., 2020; Hattie & Timperley, 2007). At this level, cultivating literacy in artificial intelligence, understood as the ability to understand and critically evaluate these tools, is a condition for the learner to remain within the monitoring and control loop (Ng et al., 2021). Explicitly teaching self-regulation strategies, and not just implicitly activating them, improves learning and prepares the learner for a more judicious use of AI (Dignath & Veenman, 2021; Donker et al., 2014).

6.2 For Tool Design

The model favors tools that provoke thought rather than replace it. Interfaces that ask questions, expose their uncertainty, offer competing alternatives, or introduce deliberate friction before delivering an answer push usage into the augmentative zone and contain experimentally documented overdependence (Klingbeil et al., 2024). These friction strategies are related to the idea of cognitive forcing functions and the principle of appropriate trust, according to which the design should adjust dependence to the actual reliability of the system rather than maximize acceptance (Lee & See, 2004). Designing the tool to ask the learner for prior judgment or explicit verification moves the use of Type 4 to Type 2.

6.3 For future research

The seven propositions define an agenda. Designs are required that separate the effect on immediate performance from the effect on delayed learning, that measure calibration in addition to performance, that manipulate the locus of discharge and the preservation of desirable difficulties, and that incorporate expertise as a moderator. Longitudinal studies are especially necessary to test the conjectures of disuse and feedback loops, which today have no direct evidence.

6.4 Socio-cultural and ethical dimension

The model is both individual and cognitive, which constitutes both its scope and its limit. A natural extension is shared regulation, in which monitoring and control are distributed among peers and tools, and not just within a learner (Molenaar, 2022). Added to this are considerations of equity, since unequal access to these tools can translate metacognitive tension into a learning gap, and academic integrity, given that the boundary between legitimate scaffolding and covert substitution is precisely what the model tries to delimit.

7. Limitations

This work is conceptual and its propositions await empirical contrast. The synthesis is parsimonious and prioritizes theoretical centrality over exhaustive coverage, so it does not exhaust the literature on AI in education. Several empirical anchors rely on a small number of studies, so directional statements should be read as hypotheses and not as conclusions. The threshold is defined as a band and not as a precise value, which requires specific measurement developments. Finally, most of the available evidence comes from writing and memory tasks, so generalization to other domains, such as programming or mathematical solving, remains an open question.

8. Conclusion

The relevant question is not whether generative artificial intelligence helps or harms thinking, but under what conditions it does one thing or another. The threshold model of metacognitive tension proposes an articulated response: the discharge strengthens metacognition while preserving the keys to monitoring and relocating effort to judgment, and erodes it when it crosses the threshold at which the tool also occupies the metalevel, a threshold that is displaced by the learner's expertise. Distinguishing between offloading execution and offloading thought allows for the design of tasks, tools, and research that keep the learner as an agent of his or her own learning.

References

[1] Adesope, O. O., Trevisan, D. A., y Sundararajan, N. (2017). Rethinking the use of tests: A meta-analysis of practice testing. *Review of Educational Research, 87*(3), 659–701. https://doi.org/10.3102/0034654316689306

[2] Agarwal, P. K., Nunes, L. D., y Blunt, J. R. (2021). Retrieval practice consistently benefits student learning: A systematic review of applied research in schools and classrooms. *Educational Psychology Review, 33*(4), 1409–1453. https://doi.org/10.1007/s10648-021-09595-9

[3] Baars, M., Wijnia, L., de Bruin, A., y Paas, F. (2020). The relation between students' effort and monitoring judgments during learning: A meta-analysis. *Educational Psychology Review, 32*(4), 979–1002. https://doi.org/10.1007/s10648-020-09569-3
[4] Bjork, R. A., y Bjork, E. L. (2020). Desirable difficulties in theory and practice. *Journal of Applied Research in Memory and Cognition, 9*(4), 475–479. https://doi.org/10.1016/j.jarmac.2020.09.003
[5] Bjork, R. A., Dunlosky, J., y Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology, 64*, 417–444. https://doi.org/10.1146/annurev-psych-113011-143823
[6] Chiriatti, M., Ganapini, M. B., Panai, E., Ubiali, M., y Riva, G. (2024). The case for human–AI interaction as system 0 thinking. *Nature Human Behaviour, 8*, 1829–1830. https://doi.org/10.1038/s41562-024-01995-5
[7] Chiu, G., y Gilbert, S. J. (2024). Influence of the physical effort of reminder-setting on strategic offloading of delayed intentions. *Quarterly Journal of Experimental Psychology, 77*(6), 1295–1311. https://doi.org/10.1177/17470218231199977
[8] Clark, A., y Chalmers, D. (1998). The extended mind. *Analysis, 58*(1), 7–19. https://doi.org/10.1093/analys/58.1.7
[9] Crompton, H., y Burke, D. (2023). Artificial intelligence in higher education: The state of the field. *International Journal of Educational Technology in Higher Education, 20*, Artículo 22. https://doi.org/10.1186/s41239-023-00392-8
[10] Cukurova, M. (2025). The interplay of learning, analytics and artificial intelligence in education: A vision for hybrid intelligence. *British Journal of Educational Technology, 56*(2), 469–488. https://doi.org/10.1111/bjet.13514
[11] Darvishi, A., Khosravi, H., Sadiq, S., Gašević, D., y Siemens, G. (2024). Impact of AI assistance on student agency. *Computers & Education, 210*, Artículo 104967. https://doi.org/10.1016/j.compedu.2023.104967
[12] de Bruin, A. B. H., Roelle, J., Carpenter, S. K., y Baars, M. (2020). Synthesizing cognitive load and self-regulation theory: A theoretical framework and research agenda. *Educational Psychology Review, 32*(4), 903–915. https://doi.org/10.1007/s10648-020-09576-4
[13] Deng, R., Jiang, M., Yu, X., Lu, Y., y Liu, S. (2025). Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. *Computers & Education, 227*, Artículo 105224. https://doi.org/10.1016/j.compedu.2024.105224
[14] Dignath, C., y Veenman, M. V. J. (2021). The role of direct strategy instruction and indirect activation of self-regulated learning: Evidence from classroom observation studies. *Educational Psychology Review, 33*(2), 489–533. https://doi.org/10.1007/s10648-020-09534-0
[15] Donker, A. S., de Boer, H., Kostons, D., Dignath van Ewijk, C. C., y van der Werf, M. P. C. (2014). Effectiveness of learning strategy instruction on academic performance: A meta-analysis. *Educational Research Review, 11*, 1–26. https://doi.org/10.1016/j.edurev.2013.11.002
[16] Dunlosky, J., y Rawson, K. A. (2012). Overconfidence produces underachievement: Inaccurate self-evaluations undermine students' learning and retention. *Learning and Instruction, 22*(4), 271–280. https://doi.org/10.1016/j.learninstruc.2011.08.003
[17] Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., y Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest, 14*(1), 4–58. https://doi.org/10.1177/1529100612453266
[18] Efklides, A. (2011). Interactions of metacognition with motivation and affect in self-regulated learning: The MASRL model. *Educational Psychologist, 46*(1), 6–25. https://doi.org/10.1080/00461520.2011.538645
[19] Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., Shen, Y., Li, X., y Gašević, D. (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.13544
[20] 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
[21] Grinschgl, S., Papenmeier, F., y Meyerhoff, H. S. (2021). Consequences of cognitive offloading: Boosting performance but diminishing memory. *Quarterly Journal of Experimental Psychology, 74*(9), 1477–1496. https://doi.org/10.1177/17470218211008060
[22] Hattie, J., y Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112. https://doi.org/10.3102/003465430298487
[23] Kalyuga, S., Ayres, P., Chandler, P., y Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist, 38*(1), 23–31. https://doi.org/10.1207/S15326985EP3801_4
[24] Kirschner, P. A., Sweller, J., y Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41*(2), 75–86. https://doi.org/10.1207/s15326985ep4102_1

[25] Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General, 126*(4), 349–370. https://doi.org/10.1037/0096-3445.126.4.349
[26] Klingbeil, A., Grützner, C., y Schreck, P. (2024). Trust and reliance on AI: An experimental study on the extent and costs of overreliance on AI. *Computers in Human Behavior, 160*, 108352. https://doi.org/10.1016/j.chb.2024.108352
[27] Lan, M., y Zhou, X. (2025). A qualitative systematic review on AI empowered self-regulated learning in higher education. *npj Science of Learning, 10*, Artículo 21. https://doi.org/10.1038/s41539-025-00319-0
[28] Lee, J. D., y See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors, 46*(1), 50–80. https://doi.org/10.1518/hfes.46.1.50_30392
[29] Molenaar, I. (2022). The concept of hybrid human-AI regulation: Exemplifying how to support young learners' self-regulated learning. *Computers and Education: Artificial Intelligence, 3*, 100070. https://doi.org/10.1016/j.caeai.2022.100070
[30] Molenaar, I., de Mooij, S., Azevedo, R., Bannert, M., Järvelä, S., y Gašević, D. (2023). Measuring self-regulated learning and the role of AI: Five years of research using multimodal multichannel data. *Computers in Human Behavior, 139*, 107540. https://doi.org/10.1016/j.chb.2022.107540
[31] Nelson, T. O., y Narens, L. (1990). Metamemory: A theoretical framework and new findings. En G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 26, pp. 125–173). Academic Press.
[32] Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., y Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence, 2*, 100041. https://doi.org/10.1016/j.caeai.2021.100041
[33] Risko, E. F., y Gilbert, S. J. (2016). Cognitive offloading. *Trends in Cognitive Sciences, 20*(9), 676–688. https://doi.org/10.1016/j.tics.2016.07.002
[34] Sitzmann, T., y Ely, K. (2011). A meta-analysis of self-regulated learning in work-related training and educational attainment: What we know and where we need to go. *Psychological Bulletin, 137*(3), 421–442. https://doi.org/10.1037/a0022777
[35] Soderstrom, N. C., y Bjork, R. A. (2015). Learning versus performance: An integrative review. *Perspectives on Psychological Science, 10*(2), 176–199. https://doi.org/10.1177/1745691615569000
[36] Stadler, M., Bannert, M., y Sailer, M. (2024). Cognitive ease at a cost: LLMs reduce mental effort but compromise depth in student scientific inquiry. *Computers in Human Behavior, 160*, Artículo 108386. https://doi.org/10.1016/j.chb.2024.108386
[37] Sweller, J. (2023). The development of cognitive load theory: Replication crises and incorporation of other theories can lead to theory expansion. *Educational Psychology Review, 35*, Artículo 95. https://doi.org/10.1007/s10648-023-09817-2
[38] Sweller, J., van Merriënboer, J. J. G., y Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review, 31*(2), 261–292. https://doi.org/10.1007/s10648-019-09465-5
[39] Wang, J., y Fan, W. (2025). The effect of ChatGPT on students' learning performance, learning perception, and higher-order thinking: Insights from a meta-analysis. *Humanities and Social Sciences Communications, 12*(1), 1–21. https://doi.org/10.1057/s41599-025-04787-y
[40] Yang, C., Luo, L., Vadillo, M. A., Yu, R., y Shanks, D. R. (2021). Testing (quizzing) boosts classroom learning: A systematic and meta-analytic review. *Psychological Bulletin, 147*(4), 399–435. https://doi.org/10.1037/bul0000309
[41] Zimmerman, B. J. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *American Educational Research Journal, 45*(1), 166–183. https://doi.org/10.3102/0002831207312909