

Beyond Cognitive Offloading: A Process–Agency Framework for Distinguishing AI Scaffolding from Cognitive Outsourcing in Higher Education

Jiaxuan An

ABSTRACT

Generative artificial intelligence (GenAI) can improve task performance while changing the cognitive activity through which learning occurs. Yet prevailing contrasts between beneficial assistance and harmful dependence often conflate what is delegated, when support enters the task, how the learner responds, and who controls substantive epistemic decisions. This conceptual article develops a process–agency framework for analysing discrete GenAI-use episodes in higher education. Rather than treating cognitive offloading, scaffolding, and outsourcing as equivalent taxonomic categories, it locates them within a hierarchical decision structure. **Cognitive offloading** is the broad externalisation of cognitive work; educationally efficient offloading occurs when peripheral or mastered operations are externalised while focal work remains with the learner. When AI participates in a learning-critical operation, a scaffold-oriented design becomes **enacted AI-supported scaffolding** only if the learner generates, evaluates, revises, and progressively reclaims responsibility. **Cognitive outsourcing** occurs when AI retains primary generative control over the learning-critical operation before meaningful learner engagement. Productive struggle is treated as a quality of enacted learning activity, not a moral requirement or separate classification criterion. The framework separates ex ante conditions, enacted functional profile, in-episode epistemic control, and post-support outcomes. A set of falsifiable propositions specifies how learning-objective criticality, timing, generative control, prior knowledge, regulatory capacity, AI literacy, and assessment design may shape assisted performance and independent learning.

Keywords: generative artificial intelligence; cognitive offloading; cognitive outsourcing; AI scaffolding; learner agency; higher education

1. Introduction

Generative artificial intelligence gives students immediate access to sophisticated linguistic, analytical, and representational assistance. Within a single interface, a learner can request anything from a targeted hint to a complete essay. The same technology may reduce avoidable burden and expose disciplinary practices, or bypass the very operations an activity is intended to develop.

Evidence increasingly suggests that these effects are conditional rather than uniformly beneficial or harmful. ChatGPT-supported instruction improves learning on average, but effects vary by design, duration, outcome, and context (Deng et al., 2025). In a writing experiment, GenAI improved aspects of task performance but changed self-regulated learning sequences without improving knowledge gain or transfer (Fan et al., 2025). In scientific inquiry, large-language-model access reduced mental effort while weakening depth of inquiry (Stadler et al., 2024). In secondary mathematics, an unrestricted GPT tutor improved supported practice but was followed by poorer unaided performance; safeguards that made assistance more learning-oriented largely mitigated this decline (Bastani et al., 2025). Performance *with* AI, therefore, is not evidence of learning *from* AI.

The conceptual vocabulary has not kept pace with this empirical complexity. **Cognitive offloading** denotes the use of external action or representation to reduce internal demand (Risko & Gilbert, 2016), encompassing adaptive practices such as note-taking, diagrams, reminders, and routine calculation. Applied indiscriminately to GenAI, however, it obscures the difference between externalising a peripheral operation and transferring control over a learning-critical one. At the other extreme, dependence, overreliance, and outsourcing are often inferred from use frequency, perceived ease, or weak subsequent performance. Such inferences can pathologise legitimate support, confuse configuration with consequence, and neglect instructional design.

Students recognise GenAI's potential for personalised support while expressing concern about accuracy, ethics, and capability development (Chan & Hu, 2023). Reviews similarly identify opacity, unreliable output, and overreliance, but tend to organise evidence by application or risk rather than by the distribution of cognitive control within an episode (Yan et al., 2024; Zhai et al., 2024). Suazo Galdames et al. (2026) characterise GenAI as both a possible cognitive scaffold and a source of epistemic vulnerability, contingent on pedagogy, AI literacy, and assessment. This distinction is useful but does not yet show how to classify a particular episode. Likewise, the AI Assessment Scale specifies permitted levels of involvement (Furze et al., 2024) but is designed primarily for assessment policy, not process explanation.

This paper asks:

How can conceptually different forms of GenAI assistance be distinguished without classifying them solely by tool, frequency, or outcome?

Through which process and boundary conditions might these forms of use relate differently to assisted performance and independent learning?

The paper makes four contributions. It establishes the **GenAI-use episode**—a learner using a particular system for a particular operation at a particular point in a task—as the unit of analysis. It then places offloading, scaffolding, and outsourcing in a hierarchical decision structure based

on learning-objective criticality, temporal position, generative control, learner engagement, and epistemic control. Separating ex ante conditions, enacted profile, and later outcome prevents outsourcing from being defined by the effect it is meant to explain. Finally, the framework yields falsifiable propositions, measurement strategies, and principles for assessment design.

The argument is conditional. It does not assume that effort is intrinsically educational, that unaided cognition is always preferable, or that all delegation to AI is harmful. It asks whether the distribution of work between learner and system remains aligned with the capability that higher education intends the learner to develop and demonstrate.

2. Conceptual Development Approach

This article uses **theory synthesis** (Jaakkola, 2020) because the focal problem is distributed across constructs usually studied in isolation. It integrates four literatures: (a) cognitive offloading and distributed cognition; (b) scaffolding, cognitive load, and productive struggle; (c) generative learning, self-regulated learning, and epistemic cognition; and (d) empirical and policy research on GenAI in education. It does not claim the coverage of a systematic review or the effect estimation of a meta-analysis.

Sources were purposively sampled and updated through 7 September 2026 through iterative searches combining *generative AI* or *large language models* with *cognitive offloading*, *outsourcing*, *overreliance*, *scaffolding*, *productive struggle*, *self-regulated learning*, *epistemic agency*, *trust*, *assessment*, and *transfer*. Backward and forward citation chaining connected recent GenAI studies to foundational work on offloading (Risko & Gilbert, 2016), scaffolding (Wood et al., 1976), cognitive load (Sweller, 1988), self-regulation (Panadero, 2017; Zimmerman, 2000), and epistemic cognition (Barzilai & Chinn, 2018). Recent peer-reviewed studies were prioritised when they compared AI conditions, examined process, distinguished assisted from independent outcomes, or addressed assessment validity. Evidence from adjacent settings was used only to illuminate relevant mechanisms; for example, research with knowledge workers informs confidence and verification in human-AI interaction (Lee et al., 2025) but is not treated as direct evidence of student learning.

The analysis proceeded in three stages. **Concept differentiation** compared the attributes, antecedents, and consequences of offloading, scaffolding, overreliance, automation bias, and outsourcing. **Mechanism mapping** located these constructs in a temporal sequence comprising use conditions, learner processes, and post-support outcomes. **Proposition development** converted the resulting distinctions into comparative claims specifying predictors, mechanisms or moderators, outcomes, and counterfactual conditions. This sequence keeps predicted outcomes outside the classification itself.

The review is integrative and purposive rather than exhaustive. It establishes conceptual coverage and theoretical contrast, not prevalence or average effects. The resulting framework remains to be validated across disciplines, learner populations, languages, assessment forms, and AI systems.

3. Existing Approaches and the Unresolved Classification Problem

3.1 Cognitive Offloading: Externalisation Without Presumed Educational Value

Cognitive offloading occurs when people use physical action or external resources to reduce internal processing demands (Risko & Gilbert, 2016). Notes externalise prospective memory, diagrams externalise spatial relations, and calculators externalise arithmetic. Offloading is therefore neither inherently beneficial nor harmful; its educational value depends on whether the externalised operation is peripheral or central to the learning objective.

If the objective is disciplinary argumentation, AI-assisted formatting may preserve attention for reasoning. If the objective is statistical interpretation, asking AI to supply that interpretation may remove the target activity even when the prose is accurate. Classification thus depends not on visible AI use but on *the relationship between the delegated operation and the intended capability*.

Task criticality is also developmentally dynamic: an operation central for a novice may be routine for an expert. Cognitive-load theory likewise suggests that support is useful when it reduces demands extraneous to learning but counterproductive when it removes processing needed for schema construction (Sweller, 1988). Whether an operation is appropriately offloaded therefore depends on prior knowledge, course objectives, and timing.

3.2 AI-Supported Scaffolding: Support That Preserves Learner Responsibility

Scaffolding is more specific than assistance. In its foundational formulation, support is responsive to a learner's performance and enables participation in activity the learner could not yet complete alone (Wood et al., 1976). Later scholarship identifies contingency, fading, and transfer of responsibility as defining characteristics (van de Pol et al., 2010). A completed answer that happens to be helpful is not necessarily a scaffold.

GenAI can approximate scaffolding by diagnosing an attempted solution, offering a hint, posing a question, modelling one step, generating a counterexample, or providing feedback for the learner to interpret. Yet the same interface can supply an immediate, polished solution with high apparent authority. Interactivity alone is insufficient; the learner must retain responsibility for generating, monitoring, evaluating, and revising the substantive solution.

The framework distinguishes **scaffolding affordance** from **enacted AI-supported scaffolding**. A system or task provides scaffolding affordance when support (a) responds to an identified or inferred need, (b) remains bounded rather than immediately completing the target operation, and (c) can be reduced, withdrawn, or followed by an independent opportunity. These are design properties, not evidence that scaffolding has occurred. Scaffolding is enacted only when the learner cognitively acts on the support, continues to generate and evaluate substantive content, and progressively reclaims responsibility. A scaffold-oriented system may therefore be enacted as a substitute, whereas a general-purpose system may be used in a scaffolded manner.

Assistance design is therefore a substantive variable. Comparisons between unrestricted and safeguarded GPT tutors show that interface constraints and response design can shift AI from answer engine toward tutor (Bastani et al., 2025). Inquiry-oriented systems similarly organise

interaction around questions and evidence rather than one-shot completion (Li et al., 2025). Although neither design guarantees learning, the form and sequence of assistance matter.

3.3 Cognitive Outsourcing: Transfer of Generative Control Over Learning-Critical Work

The term **cognitive outsourcing** should identify an enacted functional profile, not a learner trait, moral failure, or retrospective outcome:

Cognitive outsourcing is an enacted GenAI-use profile in which the system assumes primary generative control over learning-critical operations before the learner has meaningfully generated, evaluated, or revised a solution, and the learner does not substantively reclaim that control.

Three elements are necessary: the delegated operation is central to the learning objective; the AI contribution is generatively substantive rather than clerical or representational; and the learner has little opportunity to construct or test an initial representation before receiving the AI solution. Poor retention or transfer is not part of the definition but a possible consequence to be tested.

Outsourcing is episode-specific, not a stable learner identity. Within one assignment, the same learner may offload citation formatting, receive scaffolded feedback on an argument, and outsource literature synthesis. Person-level frequency cannot capture these differences.

Overreliance and automation bias are related but distinct. Overreliance concerns excessive acceptance of system recommendations; automation bias is systematic deference to automated advice, often under misplaced trust (Lee & See, 2004; Parasuraman & Riley, 1997). Educational reviews associate overreliance with reduced critical engagement, although the evidence remains heterogeneous and often correlational (Zhai et al., 2024). Cognitive outsourcing instead concerns the location and sequence of generative control relative to a learning objective. A learner may critically inspect an outsourced solution, or overtrust a minor suggestion without outsourcing the target operation.

3.4 A Hierarchical Account of Three Functional Profiles

The three terms do not occupy identical ontological levels. Cognitive offloading is the broad phenomenon of externalising cognitive work; both scaffolded assistance and outsourcing may involve offloading in this general sense. The framework therefore uses **educationally efficient offloading** as the narrower functional profile for episodes in which AI performs peripheral or already-mastered operations while the learner retains focal intellectual work. If the delegated operation is learning-critical, the analysis proceeds to design affordance and learner enactment: bounded, contingent support becomes enacted scaffolding only when the learner uses it to generate, evaluate, and revise, whereas sustained AI control before meaningful learner engagement indicates outsourcing. Table 1 summarises these ideal-typical profiles. Conflicting evidence warrants a mixed or uncertain classification.

Table 1

Ideal-typical distinctions among GenAI-use functional profiles

Feature	Educationally efficient offloading	Enacted AI-supported scaffolding	Cognitive outsourcing
Position in the hierarchy	Peripheral/mastered branch within broader cognitive offloading	Learning-critical branch in which scaffold affordance is enacted	Learning-critical branch in which AI retains primary generative control
Relation to learning objective	Primarily peripheral or already mastered	Learning-critical activity is supported	Learning-critical activity is substantially performed by AI
AI contribution	Storage, retrieval, formatting, routine transformation, bounded assistance	Contingent hint, question, partial modelling, feedback, counterexample	Answer, argument, interpretation, design, or solution with high completeness
Temporal position	Often before or after focal reasoning	Typically after diagnosis or an initial attempt; may recur iteratively	Commonly before meaningful generation or evaluation
Learner's role	Retains focal intellectual operations	Generates, compares, evaluates, revises, progressively assumes responsibility	Primarily selects, accepts, lightly edits, or transmits output
Design and enactment evidence	Focal operation remains learner-controlled	Bounded/contingent affordance plus learner uptake and return of responsibility	High-control output precedes meaningful engagement and control is not reclaimed
Expected effect	Conditional efficiency benefit; no outcome guaranteed	Greater probability of learning when calibrated	Possible assisted-performance gain with risk to independent learning
Ambiguous evidence	Mixed/uncertain	Mixed/uncertain	Mixed/uncertain—not inferred from polish or frequency alone

These are functional profiles of episodes, not categories of tools or prompt types. A request for a summary can be efficient offloading during preliminary source triage, enacted scaffolding when the learner compares it with an independently generated summary, or outsourcing when source comprehension is the assessed operation and the learner never engages the source.

3.5 Distinctive Contribution of the Process–Agency Framework

The framework does not claim that the broad tension between augmentation and substitution is new. Its intended contribution is to make that tension classifiable and testable at the level where learning activity occurs. Table 2 positions the framework relative to adjacent approaches.

Table 2

Relationship between adjacent approaches and the present framework

Approach	Primary analytical focus	Limitation for analysing a GenAI-use episode	Extension in the present framework
Cognitive offloading	Externalisation of internal cognitive demand	Does not determine whether the delegated operation is educationally peripheral or central	Treats offloading as the superordinate phenomenon and identifies educationally efficient offloading as one functional branch
Overreliance and automation bias	Excessive acceptance of or deference to automated advice	Often operationalised through trust, frequency, or error acceptance rather than learning-objective alignment	Separates reliance from the location and timing of generative control
AI-supported scaffolding	Adaptive support for performance and learning	“Scaffolding” can become a generic positive label for assistance	Separates scaffold-oriented design affordance from scaffolding demonstrated in learner enactment
AI-permission and assessment frameworks	Permitted level of AI involvement in an assessment	Policy categories do not reveal how permitted assistance is enacted	Adds episode-level process evidence and independent demonstrations
Reviews of GenAI’s cognitive effects	Average benefits, risks, and moderators	Aggregate effects obscure heterogeneous task operations and interaction sequences	Separates configuration, mechanism, assisted performance, and post-support learning

The framework’s distinctive move lies in **unit, hierarchy, decomposition, and temporal ordering**. It analyses episodes rather than learners or tools; locates the three profiles within a nested decision structure; separates what AI affords and generates from what the learner does and judges; and distinguishes use conditions, enactment, and outcomes. Mixed or uncertain classifications are retained rather than forced into an evaluative binary.

4. Productive Struggle as the Explanatory Mechanism

4.1 From Effort to Learning-Relevant Engagement

The framework uses **productive struggle** to characterise the quality of enacted learning activity through which different use conditions may produce different outcomes. It is not maximal difficulty but sustained, supportable engagement with ideas, errors, representations, and strategies relevant to the learning objective (Warshauer, 2015). It is manifested in learner generation, monitoring, verification, strategy change, and revision rather than positioned as a temporally prior latent event. Too little engagement restricts knowledge construction; excessive demand can overwhelm cognitive resources or prompt disengagement. Productive struggle occupies the context-dependent region between premature substitution and unproductive overload.

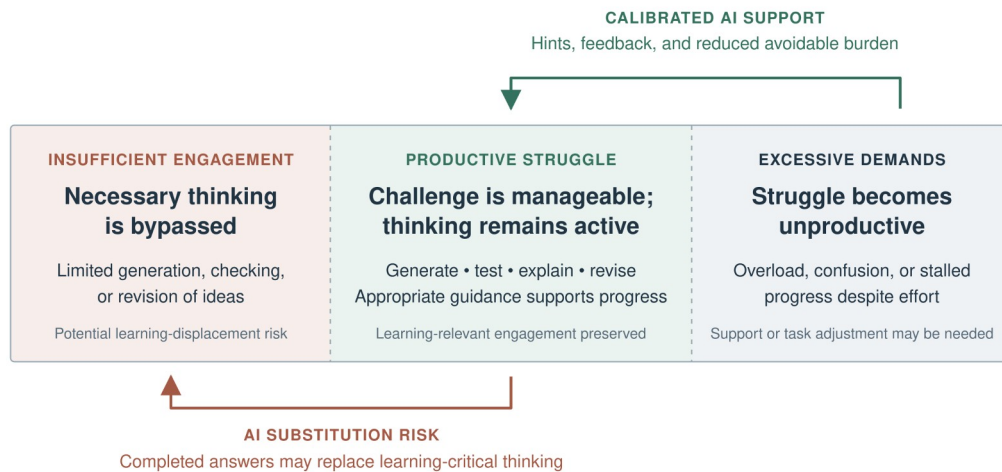
Several theories converge on this mechanism. Generative learning requires learners to select, organise, and integrate information rather than merely receive it (Fiorella & Mayer, 2016; Wittrock, 1974). The ICAP framework predicts deeper learning as engagement moves from passive reception toward constructive and interactive activity (Chi & Wylie, 2014). Desirable-difficulties research distinguishes immediate performance from durable learning (Bjork & Bjork, 1992; Soderstrom & Bjork, 2015). Productive-failure research shows that initially unsuccessful problem solving can prepare learners to understand later instruction when tasks and consolidation are appropriately designed (Kapur, 2008, 2014). Error-based learning highlights the value of detecting and resolving discrepancies (Metcalf, 2017).

These accounts do not imply that learners should struggle unaided. Minimal guidance can impose excessive demands, particularly on novices (Kirschner et al., 2006). The narrower claim is that learning requires learner-owned cognitive activity and that effective support regulates rather than removes it.

4.2 The Dual Role of GenAI

GenAI can move an episode toward productive struggle by reducing avoidable burden, decomposing a task, providing a strategic hint, or offering feedback at an impasse. It can move an episode away from productive struggle by supplying a complete response before the learner has generated or evaluated an idea. These are directional possibilities, not measured thresholds.

Figure 1
Productive struggle and the dual role of generative AI.



Context matters: learning objective, prior knowledge, task demands, and instructional support.

Note. Regions and arrows represent theoretical possibilities, not fixed thresholds or established causal effects.

Low effort on a mastered or simple task does not, by itself, indicate cognitive outsourcing.

Note. The regions are conceptual rather than quantitative. Calibrated support may move excessive demand toward productive engagement; premature substitution may move productive engagement toward insufficient learning-relevant activity. Low effort on an already-mastered or peripheral operation is not evidence of outsourcing.

The explanatory sequence can be stated as:

Ex ante use conditions and scaffolding affordance → enacted learning activity (generation, monitoring, verification, and revision), characterised as productive struggle when appropriately calibrated → enacted epistemic control → independent epistemic performance, retention, and transfer.

Productive struggle is neither a separate mediator preceding these observable processes nor part of the definition of outsourcing. It is a theoretically meaningful quality of the learner's engagement with the target operation. Scaffold-oriented design does not guarantee such engagement: poorly calibrated prompts, inaccurate feedback, or excessive dialogue can still produce confusion or superficial compliance.

5. The Process–Agency Framework

5.1 Unit of Analysis: The GenAI-Use Episode

A **GenAI-use episode** is defined by a learner, task, target operation, AI interaction, and position in the learning sequence. It begins when the learner delegates or seeks support for an operation and ends when the learner accepts, rejects, revises, or acts on the AI contribution. A single assignment may contain differently classified episodes.

This unit avoids inferring cognition from AI presence, treating a polished product as evidence of understanding or misconduct, and assuming that equal use frequencies represent equivalent activity. Frequency becomes interpretable only after the underlying episodes have been classified.

5.2 Layer 1: Ex Ante Conditions and Scaffolding Affordance

Three variables specify the ex ante conditions without using learner-process or later-outcome evidence:

Learning-objective criticality: the extent to which the AI-performed operation is the capability the activity intends to develop or assess.

Temporal position: whether AI is consulted before initial generation, after an attempt, at impasse, during revision, or after completion.

AI generative control: the extent to which the system determines substantive components of the solution, including content, reasoning path, criteria, or final form.

For learning-critical operations, a fourth ex ante feature is **scaffolding affordance**: whether assistance is contingent, bounded, and designed to permit withdrawal or an independent attempt. Affordance indicates what the design makes possible, not what the learner actually does. Prompt length cannot substitute for any of these variables: a long prompt may request a complete answer, whereas a short request for a counterexample may trigger extensive evaluation. Coding should draw on task specifications, system behaviour, and interaction sequence rather than lexical proxies alone.

5.3 Layer 2: Learner Enactment and Functional Profile

Layer 2 records how the opportunity is enacted. **Learner regulatory engagement** comprises goal setting, self-generation, monitoring, verification, error detection, comparison, revision, and strategic help-seeking. When these activities constitute sustained, supportable work on the target operation, productive struggle is preserved or calibrated; when AI pre-empts them, it is displaced. **Enacted epistemic control** is the learner's observable exercise of substantive judgment—for example, setting criteria, checking a claim, rejecting a suggestion, choosing among alternatives, or revising reasoning.

The functional profile is assigned only after Layer 1 and Layer 2 evidence are considered, but before later outcomes are known. Educationally efficient offloading applies when the delegated operation is peripheral or mastered and focal control remains with the learner. Enacted AI-supported scaffolding applies when a scaffold-oriented affordance is taken up through learner generation, evaluation, revision, and return of responsibility. Cognitive outsourcing applies when AI assumes primary control over a learning-critical operation before meaningful learner engagement and the learner does not substantively reclaim that control. Evidence that does not support one ideal type is coded as mixed or uncertain.

5.4 Layer 3: Assisted and Post-Support Outcomes

Layer 3 separates performance achieved with AI from capability demonstrated after support. The term **independent epistemic performance** is used for a single post-support demonstration. The stronger inference of **demonstrated epistemic agency** requires consistent evidence across tasks or occasions that a learner can initiate inquiry, explain and defend judgments, adapt knowledge claims, and assume responsibility for them (Barzilai & Chinn, 2018; Greene et al., 2016).

Outcomes must be separated by support condition and time:

- **AI-assisted performance:** quality or speed while AI is available;
- **immediate unaided performance:** capability shortly after support is removed;
- **delayed retention:** capability after a meaningful delay;
- **near and far transfer:** application to structurally similar and dissimilar tasks;
- **calibration:** correspondence between confidence and actual performance; and
- **independent epistemic performance:** explanation, defence, adaptation, and accountable judgment on a specific post-support task; and
- **demonstrated epistemic agency:** consistent independent epistemic performance across tasks, occasions, or novel conditions.

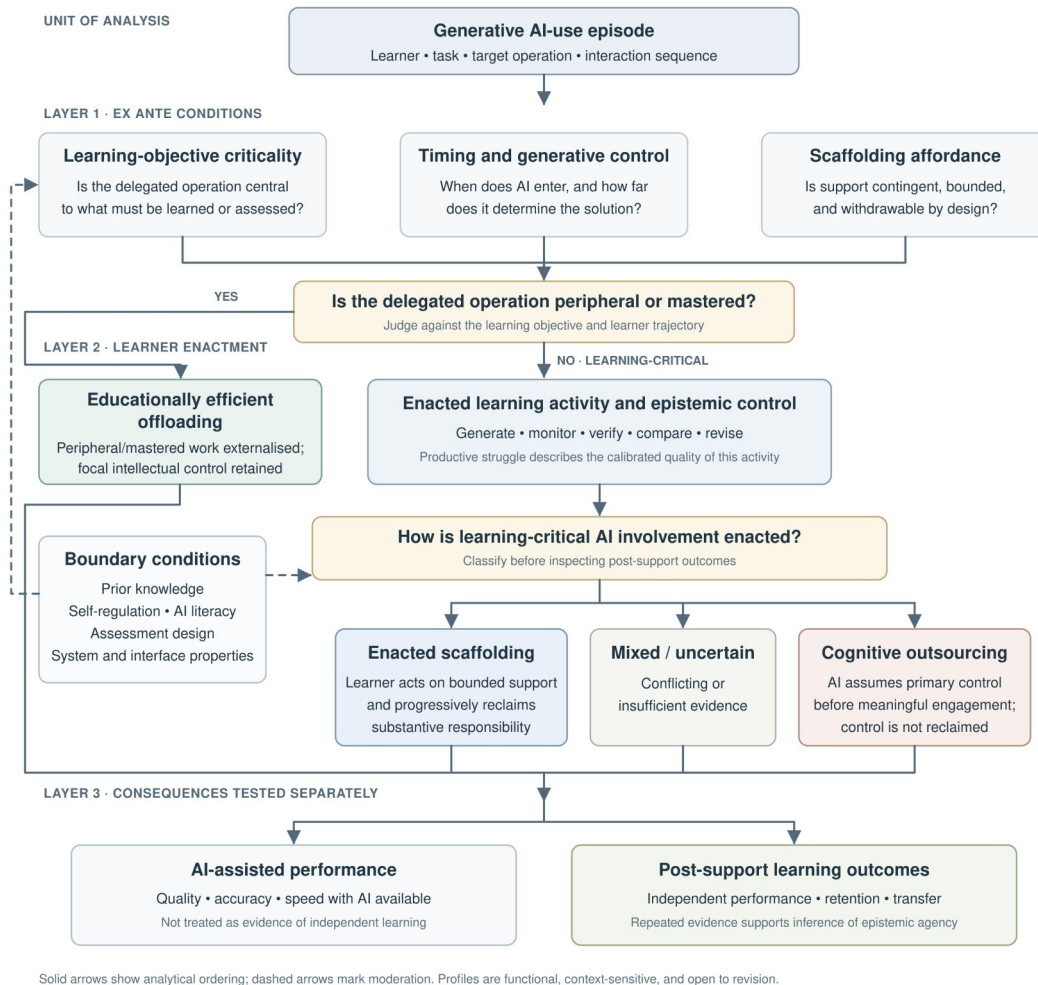
An excellent AI-assisted product may coexist with weak independent understanding, whereas an imperfect product produced through active generation and revision may support learning. Independent probes are therefore a method of validating learning claims, not a normative demand that every professional capability must ultimately be performed without AI. Performance and learning are related but not interchangeable (Soderstrom & Bjork, 2015).

5.5 Interpreting the Framework

Figure 2 summarises the hierarchical logic. Layer 1 identifies learning-objective criticality, timing, substantive solution control, and scaffolding affordance. Layer 2 records learner enactment and assigns a provisional functional profile. Layer 3 measures assisted performance separately from post-support learning. Outcomes test the consequences of a profile but do not determine it retrospectively.

Figure 2

Hierarchical process–agency framework for generative AI use.



Note. Solid arrows show analytical ordering; dashed arrows mark moderation by learner, task, assessment, and system conditions. The profiles are functional and context-sensitive, are assigned before post-support outcomes are inspected, and may remain mixed when evidence conflicts.

5.6 Boundary Conditions

Prior knowledge. Novices may need more guidance but are less able to detect fluent errors. Experts can often offload routine operations without losing conceptual control. The same AI action can therefore have different implications across knowledge levels.

Self-regulated learning. Learners differ in planning, monitoring, strategy adjustment, and reflection (Panadero, 2017; Zimmerman, 2000). Strong regulation may transform a high-control

response into an object of critique; weak regulation may turn feedback-oriented support into passive acceptance. Recent mapping work shows rapid growth but continuing fragmentation in research at the intersection of AI and self-regulated learning (Banihashem et al., 2025).

AI literacy and confidence. AI literacy includes understanding limitations, evaluating outputs, and using AI ethically. Excessive trust may suppress verification, whereas task-specific self-confidence may support evaluation. People may also display algorithm appreciation under some judgment conditions, which cautions against assuming either uniform distrust or uniform deference (Logg et al., 2019). Among knowledge workers, higher confidence in GenAI was associated with less reported critical-thinking effort, while greater self-confidence was associated with more (Lee et al., 2025). Student research likewise shows that trust shapes reliance on and resistance to GenAI, while AI literacy and self-regulated learning are associated with writing performance in AI-supported settings (Shi et al., 2025; Wang et al., 2025). These findings do not establish causality but support treating trust, literacy, and regulatory capacity as distinct boundary conditions.

Instructional and assessment design. Learning objectives determine which operations are critical. Assessment conditions determine whether assisted products or independent capabilities are rewarded. Policies stating only whether AI is allowed cannot resolve this relationship. Task-level guidance, process evidence, and independent demonstrations align more closely with learning (Furze et al., 2024; Luo, 2024).

AI system properties. Response completeness, error rates, citation behaviour, interface defaults, adaptivity, and whether a system asks questions or delivers answers influence control. Studies should report model, version, interface, assistance design, and system prompt where feasible (Kasneci et al., 2023).

6. Propositions

The propositions are comparative, level-specific, and open to disconfirmation. Task difficulty should be measured separately from learning-objective criticality and included as a control or experimentally crossed factor where relevant.

P1: Learning-objective criticality × generative control. The negative association between AI generative control and learning-relevant engagement will become stronger as the delegated operation becomes more central to the learning objective, controlling for task difficulty.

P2a: Attempt-first affordance. On learning-critical tasks, attempt-first, hint-oriented support will elicit more learner generation, monitoring, verification, and revision than answer-first, complete-output support, controlling for prior knowledge and time on task.

P2b: Affordance-enactment distinction. Scaffold-oriented affordance will predict independent learning only to the extent that learners enact it through generation, evaluation, revision, and progressive reclamation of responsibility.

P3a: Assisted-performance advantage. Relative to no AI support, answer-first, complete-output support will improve immediate performance while AI remains available.

P3b: Post-support disadvantage. Relative to attempt-first, hint-oriented support, answer-first, complete-output support will produce weaker unaided retention and transfer, even when it produces equal or better assisted performance.

P4a: Configuration-process relationship. For learning-critical operations, high AI generative control introduced before an initial attempt will be associated with displaced productive struggle, whereas contingent support introduced after an attempt or at impasse will be associated with preserved or calibrated productive struggle.

P4b: Process-control relationship. Productive struggle, manifested in learner generation, monitoring, verification, strategy change, and revision, will be positively associated with enacted epistemic control.

P4c: Indirect learning relationship. Enacted epistemic control will mediate the relationship between the enacted functional profile and independent epistemic performance, retention, and transfer.

P5a: Prior-knowledge moderation. The negative association between AI generative control and enacted epistemic control will be stronger at lower levels of domain-specific prior knowledge.

P5b: Self-regulation moderation. Self-regulated learning will weaken the negative association between AI generative control and enacted epistemic control by increasing strategic monitoring, help-seeking, and revision.

P5c: Critical AI-literacy moderation. Performance-based critical AI literacy will weaken the negative association between AI generative control and enacted epistemic control by increasing the detection, verification, and correction of flawed output.

P6: Assessment design. Compared with product-only assessment, designs requiring traceable reasoning, justified AI-related decisions, and subsequent unaided performance will reduce the proportion of outsourcing episodes and strengthen the validity of inferences about learner capability.

P1–P4c specify the explanatory sequence and relevant counterfactuals. P5a–P5c distinguish domain knowledge, self-regulation, and critical AI literacy as separate moderators. P6 concerns a course- or programme-level intervention and therefore requires multilevel testing when the outcome is classified at episode level.

7. Research Agenda and Operationalisation

7.1 Measurement and Study Design

Generic AI-use frequency is an inadequate primary predictor because equal frequencies may represent fundamentally different cognitive activity. Measurement should combine task analysis, interaction traces, and independent outcomes.

Table 3
Integrated operationalisation and identification matrix

Construct	Type and empirical level	Recommended operationalisation	Required empirical distinction	Most relevant propositions/designs
Learning-objective criticality	Task/operation-level contextual property	Pre-register the target operation; obtain independent disciplinary ratings of alignment with the stated learning outcome	Distinguish from task difficulty, workload, and learner-perceived importance	P1; cross criticality with generative-control conditions and include task difficulty separately
Temporal position	Episode-level categorical attribute	Code AI entry before attempt, after attempt, at impasse, during revision, or after completion from time-stamped traces	Distinguish timing from duration and number of prompts	P2a, P4a; manipulate attempt-first versus answer-first support
AI generative control	Episode-level formative configuration	Code how far AI determines substantive claims, criteria, reasoning path, and final form	Distinguish from output length, AI-authored word proportion, and use frequency	P1, P3a–b, P4a, P5a–c; manipulate response completeness and decision authority
Scaffolding affordance	System/task-level design property	Code contingency, boundedness, diagnostic responsiveness, and opportunity for withdrawal or independent attempt	Distinguish what support permits from how the learner enacts it	P2a–b; compare scaffold-oriented and complete-output designs with equivalent content coverage
Enacted learning activity / productive struggle	Episode-level process quality	Code generation, monitoring, verification, strategic errors, strategy change, and revision on the target operation	Do not treat time, effort, frustration, or generic activity as productive struggle	P1, P2a–b, P4a–b; trace analysis, think-aloud, and stimulated recall
Enacted epistemic control	Episode-level behavioural construct	Code criteria setting, source evaluation, rejection/adaptation of output, and accountable substantive choice	Distinguish substantive judgment from visible editing or prompt sophistication	P4b–c, P5a–c; interaction traces and decision logs
AI-assisted performance	Episode/task-level outcome with AI available	Blind scoring of quality and accuracy; record speed and completeness	Keep separate from unaided capability and learning	P3a–b; supported-task performance
Independent epistemic performance	Post-support task-level outcome	Use an unaided explanation, defence, adaptation, or justification task	A single probe supports a task-specific performance inference, not stable agency	P2b, P4c; immediate or delayed unaided task
Retention and transfer	Post-support learning outcomes	Delayed unaided test and structurally near/far parallel tasks	Avoid tests that merely reproduce the supported task	P3b, P4c; delayed and transfer designs
Demonstrated epistemic agency	Cross-task or longitudinal capability	Require repeated independent initiation, explanation, defence, adaptation, and accountable judgment	Distinguish from one oral response, verbal fluency, and language confidence	P4c; repeated tasks, occasions, or novel conditions
Domain-specific prior knowledge	Pre-episode learner-level moderator	Use an unaided pretest matched to the target operation	Distinguish from general attainment, AI experience, confidence, and task difficulty	P5a; adequately powered learner × episode interaction
Self-regulated learning	Learner- and state-level moderator	Combine a validated measure with trace-based planning, monitoring, help-seeking, and revision	Distinguish capacity from the same in-episode behaviours used as outcomes	P5b; multilevel interaction and within-learner sensitivity analyses
Critical AI literacy	Learner-level performance-based moderator	Assess detection of flawed output, source verification, limitation recognition, and correction	Distinguish from trust, confidence, declarative familiarity, and use frequency	P5c; seeded-error and verification tasks

Factorial experiments can manipulate timing (attempt-first vs. answer-first), response form (hint vs. complete output), and learning-objective criticality while measuring task difficulty separately. Delayed post-tests can distinguish transient performance gains from durable learning. Process mining can compare sequences such as prompt–accept–submit and attempt–prompt–compare–revise–retry, while think-aloud and stimulated-recall methods can explain acceptance or

rejection. Multilevel designs are required when course-level assessment conditions are used to explain episode-level behaviour.

Studies should include at least one unaided outcome and report both benefits and costs. Analysing time saved and product quality alongside retention, transfer, and confidence calibration prevents either dismissing efficiency or mislabelling it as learning.

Longitudinal studies should examine whether learners acquire independent capability, remain reliant on persistent assistance, or develop hybrid expertise. The framework does not require every capability to be performed without AI; it requires explicit decisions about which capabilities are non-delegable, which may be distributed, and how those decisions change over time.

The profiles should initially function as coding categories, not as a validated diagnostic scale. Protocols should define the target operation before inspecting traces, use at least two trained raters, report agreement, and retain an uncertain/mixed category. Classification is strongest when task specifications, draft histories, interactions, learner explanations, and unaided follow-up performance converge; no single prompt, fluent product, or hesitant oral response is sufficient.

7.2 Empirical Identifiability and Falsification

The framework will be useful only if its constructs can be distinguished empirically. This requires more than administering self-report scales and demonstrating conventional discriminant validity. As Table 3 shows, learning-objective criticality is a task-level property; temporal position is a categorical event attribute; AI generative control is a formative episode configuration; scaffolding affordance is a design property; productive struggle characterises an enacted process; enacted epistemic control is an in-episode behavioural construct; and learning and agency are post-support inferences. Treating these variables as interchangeable reflective traits would erase the distinctions the framework is designed to make and create avoidable construct-validity problems (Flake & Fried, 2020).

Construct definition should precede statistical modelling. Researchers should specify the target operation and causal estimand before viewing learner–AI traces, ideally using a directed acyclic graph to identify hypothesised confounders, mediators, and moderators (Rohrer, 2018). Measurement must then match empirical level and temporal location. P1 requires learning-objective criticality to be rated independently of difficulty and crossed with generative control. P4a–c require ordered evidence: ex ante conditions, enacted activity, epistemic control, and post-support outcomes. Productive struggle should be operationalised through the quality and calibration of generation, monitoring, verification, strategy change, and revision, not entered as a separate mediator measured with the same behaviours. P5a–c require domain knowledge, self-regulation, and critical AI literacy to be modelled as distinct moderators rather than collapsed into learner readiness.

Three tests are especially important. First, **measurement separation** should be evaluated within, not across, construct classes. Reflective multi-item measures may be examined with factor models; contextual ratings and formative episode indices instead require content validity, rater reliability, and sensitivity analysis. Second, **temporal separation** is indispensable:

functional profiles must be assigned before post-support outcomes are inspected, and one independent probe must not be overinterpreted as stable epistemic agency. Third, **incremental and rival-model tests** should establish whether the hierarchy predicts enacted process and unaided outcomes beyond AI-use frequency, task difficulty, assisted product quality, prior attainment, and confidence. Relevant alternatives include a non-hierarchical three-category model, models that collapse learning activity and epistemic control, and models in which assisted performance rather than enacted profile explains later outcomes. Indirect effects should be estimated with resampling procedures and interpreted against these rival structures (MacKinnon et al., 2004).

These requirements clarify what would count against the framework. Persistent inability to distinguish criticality from difficulty, generative control from assistance quantity, scaffolding affordance from learner enactment, or in-episode control from independent performance would indicate construct redundancy. The hierarchy would also require revision if a simpler non-hierarchical model performs equally well, if the predicted interactions in P1 and P5a–c fail to replicate, or if the profiles add no explanatory value beyond use frequency and task difficulty. Falsification therefore applies both to individual propositions and to the framework’s conceptual decomposition.

7.3 Boundary and Counterexample Cases

The framework is most informative when an apparently simple behaviour changes classification once the learning objective, learner state, or interaction sequence is specified. Table 4 provides analytic examples, not validated scoring rules.

Table 4
Boundary cases for episode-level classification

GenAI-use episode	Provisional profile	Evidence that could change the classification
AI corrects grammar after a student independently develops an essay argument	Educationally efficient offloading	Becomes mixed or outsourcing if AI rewrites the claims, evidence relations, or interpretation being assessed
A student attempts a mathematics problem and requests one strategic hint	Enacted AI-supported scaffolding	Becomes outsourcing if the “hint” reveals the complete solution path and the student merely transcribes it
AI generates possible research questions that a student compares with self-generated questions	Enacted scaffolding or mixed	Becomes outsourcing if question formulation is the focal objective and AI options determine the final choice without independent criteria
An expert asks AI to generate routine code that the expert tests and integrates	Educationally efficient offloading	The same episode may be outsourcing for a novice whose objective is to learn the underlying code structure
A student critiques a complete AI-generated essay rather than composing one	Enacted scaffolding or mixed	Classification depends on whether critical evaluation or original composition is the stated learning objective

GenAI-use episode	Provisional profile	Evidence that could change the classification
A student uses AI-generated source summaries before deciding what to read	Efficient offloading or mixed	Becomes outsourcing when accurate source comprehension and synthesis are required but original sources are not consulted

No prompt form has a fixed educational meaning. Classification requires an explicit learning objective and evidence of who generated and evaluated substantive content; testing consequences additionally requires an independent post-support measure.

8. Implications for Teaching, Assessment, and Policy

8.1 Teaching and Learning Design

Five principles follow:

Name the focal operation. Students need to know whether a course is developing idea generation, evidence evaluation, modelling, disciplinary interpretation, writing, coding, or another capability.

Match assistance to learner state. Offer hints, examples, decomposition, or feedback when they reduce excessive demands while preserving target activity.

Create an attempt point. A tentative thesis, prediction, solution path, source judgment, or design sketch before high-control assistance gives learners a representation to compare and revise.

Build verification into the task. Ask learners to identify a rejected AI claim, document a source check, compare alternatives, or explain why a revision improved the work.

Fade support and test transfer. A parallel unaided task can show whether responsibility has moved back to the learner.

These principles do not require ritualised struggle. A diagnostic attempt may be brief, and support should arrive before frustration becomes unproductive. Human–AI coordination may itself be a legitimate objective, in which case assessment should target coordination, verification, and accountable judgment.

8.2 Assessment Design

Assessment validity depends on the intended inference. Extensive AI generation threatens a claim that an essay demonstrates independent argumentation, but prohibiting AI may reduce authenticity when the target capability is directing and evaluating AI-supported analysis. Authenticity alone is insufficient: GenAI can produce work that experienced assessors judge to satisfy ostensibly authentic tasks (Kofinas et al., 2025). Valid inference instead requires alignment among learning outcomes, permitted support, process evidence, and strategically sampled independent performance.

A defensible brief specifies which operations students must perform themselves, where AI support may be bounded or extensive, what process evidence must be retained, how independent performance will be sampled, and what disclosure is required. The AI Assessment Scale communicates degrees of permitted use (Furze et al., 2024); the process–agency

framework asks how that permission was enacted. Permission level and cognitive function are not equivalent.

8.3 Institutional Policy

Policies should move beyond universal lists of allowed and prohibited tools: capabilities change rapidly, and the same tool can perform peripheral or learning-critical operations. Policy should require local specification of learning objectives, proportionate disclosure, and appropriate evidence of capability. University policies vary substantially, revealing the limits of reactive, integrity-centred responses (Luo, 2024). Learning-centred policy must address assessment validity, equity, privacy, disciplinary norms, and staff capability alongside misconduct.

Institutions should not use this framework as an automated detection instrument. Neither prompt logs nor linguistic features independently establish outsourcing. The framework is for educational interpretation and research, not inferring intent or imposing sanctions without due process.

9. Discussion

GenAI use is better understood as a distribution of cognitive and epistemic work than as a binary state of tool presence. The framework no longer treats offloading, scaffolding, and outsourcing as ontologically equivalent categories. Cognitive offloading is the superordinate externalisation phenomenon; educationally efficient offloading, enacted scaffolding, and cognitive outsourcing are context-sensitive functional profiles produced by the relationship among learning-objective criticality, design affordance, and learner enactment.

Emerging syntheses already identify the tension between cognitive scaffolding and epistemic vulnerability (Suazo Galdames et al., 2026). The present framework makes that tension analytically tractable by locating it within an episode, arranging its functional profiles hierarchically, decomposing use by criticality, timing, control, affordance, and learner enactment, and separating assisted from unaided outcomes. Retaining mixed cases further avoids a moralised binary.

Scaffolding must remain distinguishable from generic assistance. Contingency, boundedness, and withdrawability establish a scaffolding affordance, but learner generation, evaluation, revision, and reclamation of responsibility establish its enactment. A scaffold-oriented system may fail through misdiagnosis, excessive completeness, or passive uptake, while a general-purpose system may be enacted strategically. Design labels alone cannot determine the functional profile.

Nor does the framework assume that cognition should remain entirely internal. Cognitive work has long been socially and technologically distributed, and external resources may form part of effective thinking (Clark & Chalmers, 1998). The relevant question is whether a given arrangement enables learners to acquire, exercise, and demonstrate the judgment for which they remain accountable.

The framework is neither an academic-integrity detector nor a stable learner typology or universal acceptability scale. Its classifications are provisional, require convergent evidence, and remain separate from institutional judgments about authorship or misconduct.

10. Limitations

The framework remains unvalidated. Distinguishing peripheral from critical operations depends on explicit learning objectives, and cross-disciplinary disagreement may be substantive rather than measurement error. Epistemic agency is also difficult to observe: verbal fluency may mask weak understanding, whereas language anxiety or disability may suppress valid explanation. Multiple forms of evidence are therefore necessary.

The focus on individual learning underrepresents collaborative and tool-mediated academic practice. Future work should extend episode-level analysis to teams and human–AI assemblages without dissolving accountability. Normative misuse is another risk: productive struggle must not justify inaccessible design, unnecessary workload, or withdrawal of accommodations. Appropriate offloading and persistent assistive support may be essential to equitable participation.

Finally, the evidence base is temporally fragile because models, interfaces, and learner practices change faster than research cycles. Findings tied to one system version should not be generalised without reporting system properties and instructional conditions.

11. Conclusion

The educational significance of GenAI cannot be inferred from its presence, frequency of use, or the polish of an assisted product. It depends on how cognitive and epistemic work is distributed within a learning episode.

Cognitive offloading is the broad externalisation of cognitive work. Within this broader class, educationally efficient offloading delegates peripheral or mastered operations while preserving focal control. For learning-critical operations, scaffold-oriented affordance becomes enacted scaffolding only when learners generate, evaluate, revise, and reclaim responsibility; cognitive outsourcing occurs when AI retains primary generative control before meaningful learner engagement. Defining these profiles before later outcomes are known makes their relationships with retention, transfer, and epistemic agency empirically testable.

Productive struggle characterises the calibrated quality of learner generation, monitoring, verification, strategy change, and revision; it is not a redundant mediator placed before those behaviours. Prior knowledge, self-regulation, AI literacy, assessment design, and system properties define boundary conditions. The resulting agenda shifts attention from generic adoption to learning-objective alignment, sequence, substantive solution control, enactment, and independent capability—allowing higher education to use GenAI’s efficiency without mistaking assisted production for learning.

Declarations

Funding. No funding was received for the preparation or publication of this article.

Conflict of Interest. The authors declare that they have no financial or non-financial conflicts of interest related to this article.

Data Availability. No new dataset was generated or analysed in this conceptual article. All sources discussed in the article are cited in the reference list.

Ethics Statement. Ethical approval was not applicable because this conceptual article did not involve human participants, personal data, or animal subjects.

Generative AI Disclosure. Generative AI tools were used during manuscript preparation for language editing, structural refinement, and figure-layout assistance. The authors reviewed and verified all arguments, interpretations, citations, and final wording and remain fully responsible for the content of the article.

References

- Banihashem, S. K., Bond, M., Bergdahl, N., Khosravi, H., & Noroozi, O. (2025). A systematic mapping review at the intersection of artificial intelligence and self-regulated learning. *International Journal of Educational Technology in Higher Education*, 22, Article 50.
- Barzilai, S., & Chinn, C. A. (2018). On the goals of epistemic education: Promoting apt epistemic performance. *Journal of the Learning Sciences*, 27(3), 353–389.
- Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, Ö., & Mariman, R. (2025). Generative AI without guardrails can harm learning: Evidence from high school mathematics. *Proceedings of the National Academy of Sciences of the United States of America*, 122(26), Article e2422633122.
- Bjork, R. A., & Bjork, E. L. (1992). A new theory of disuse and an old theory of stimulus fluctuation. In A. F. Healy, S. M. Kosslyn, & R. M. Shiffrin (Eds.), *From learning processes to cognitive processes: Essays in honor of William K. Estes* (Vol. 2, pp. 35–67). Erlbaum.
- Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education*, 20, Article 43.
- Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, 49(4), 219–243.
- Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, 58(1), 7–19.
- Deng, R., Jiang, M., Yu, X., Lu, Y., & Liu, S. (2025). Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. *Computers & Education*, 227, Article 105224.
- Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., Shen, Y., Li, X., & 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.
- Fiorella, L., & Mayer, R. E. (2016). Eight ways to promote generative learning. *Educational Psychology Review*, 28(4), 717–741.
- Flake, J. K., & Fried, E. I. (2020). Measurement schmeasurement: Questionable measurement practices and how to avoid them. *Advances in Methods and Practices in Psychological Science*, 3(4), 456–465.
- Furze, L., Perkins, M., Roe, J., & MacVaugh, J. (2024). The AI assessment scale (AIAS) in action: A pilot implementation of GenAI-supported assessment. *Australasian Journal of Educational Technology*, 40(4), 38–55.
- Greene, J. A., Sandoval, W. A., & Bråten, I. (Eds.). (2016). *Handbook of epistemic cognition*. Routledge.
- Jaakkola, E. (2020). Designing conceptual articles: Four approaches. *AMS Review*, 10(1–2), 18–26.
- Kapur, M. (2008). Productive failure. *Cognition and Instruction*, 26(3), 379–424.
- Kapur, M. (2014). Productive failure in learning math. *Cognitive Science*, 38(5), 1008–1022.
- Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., ... Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, Article 102274.
- Kirschner, P. A., Sweller, J., & 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.
- Kofinas, A. K., Tsay, C.-H., & Pike, D. (2025). The impact of generative AI on academic integrity of authentic assessments within a higher education context. *British Journal of Educational Technology*, 56(6), 2522–2549.
- Lee, H.-P., Sarkar, A., Tankelevitch, L., Drosos, I., Rintel, S., Banks, R., & Wilson, N. (2025). The impact of generative AI on critical thinking: Self-reported reductions in cognitive effort and confidence effects from a survey of knowledge workers. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (Article 1121, pp. 1–22). Association for Computing Machinery.

- Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, 46(1), 50–80.
- Li, P. H., Lee, H. Y., Lin, C. J., Wang, W. S., & Huang, Y. M. (2025). InquiryGPT: Augmenting ChatGPT for enhancing inquiry-based learning in STEM education. *Journal of Educational Computing Research*, 62(8), 2157–2186.
- Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, 151, 90–103.
- Luo, J. (2024). A critical review of GenAI policies in higher education assessment: A call to reconsider the “originality” of students’ work. *Assessment & Evaluation in Higher Education*, 49(5), 651–664.
- MacKinnon, D. P., Lockwood, C. M., & Williams, J. (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research*, 39(1), 99–128.
- Metcalfe, J. (2017). Learning from errors. *Annual Review of Psychology*, 68, 465–489.
- Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for research. *Frontiers in Psychology*, 8, Article 422.
- Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, 39(2), 230–253.
- Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. *Trends in Cognitive Sciences*, 20(9), 676–688.
- Rohrer, J. M. (2018). Thinking clearly about correlations and causation: Graphical causal models for observational data. *Advances in Methods and Practices in Psychological Science*, 1(1), 27–42.
- Shi, J., Liu, W., & Hu, K. (2025). Exploring how AI literacy and self-regulated learning relate to student writing performance and well-being in generative AI-supported higher education. *Behavioral Sciences*, 15(5), Article 705.
- Soderstrom, N. C., & Bjork, R. A. (2015). Learning versus performance: An integrative review. *Perspectives on Psychological Science*, 10(2), 176–199.
- Stadler, M., Bannert, M., & Sailer, M. (2024). Cognitive ease at a cost: LLMs reduce mental effort but compromise depth in student scientific inquiry. *Computers in Human Behavior*, 160, Article 108386.
- Suazo Galdames, I. C., Santiesteban Velázquez, M., Saracosti, M., & Chaple Gil, A. M. (2026). Between cognitive offloading and critical autonomy: A systematic review of the epistemic implications of generative AI in higher education. *Frontiers in Education*, 11, Article 1878665.
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257–285.
- van de Pol, J., Volman, M., & Beishuizen, J. (2010). Scaffolding in teacher–student interaction: A decade of research. *Educational Psychology Review*, 22(3), 271–296.
- Wang, F., Li, N., Cheung, A. C. K., & Wong, G. K. W. (2025). In GenAI we trust: An investigation of university students’ reliance on and resistance to generative AI in language learning. *International Journal of Educational Technology in Higher Education*, 22, Article 59.
- Warshauer, H. K. (2015). Productive struggle in middle school mathematics classrooms. *Journal of Mathematics Teacher Education*, 18(4), 375–400.
- Wittrock, M. C. (1974). Learning as a generative process. *Educational Psychologist*, 11(2), 87–95.
- Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, 17(2), 89–100.
- Yan, L., Sha, L., Zhao, L., Li, Y., Martinez-Maldonado, R., Chen, G., Li, X., Jin, Y., & Gašević, D. (2024). Practical and ethical challenges of large language models in education: A systematic scoping review. *British Journal of Educational Technology*, 55(1), 90–112.
- Zhai, C., Wibowo, S., & Li, L. D. (2024). The effects of over-reliance on AI dialogue systems on students’ cognitive abilities: A systematic review. *Smart Learning Environments*, 11, Article 28.
- Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13–39). Academic Press.