

Review

Retrieval Interruption Framework: AI-Assisted Cognition and Retrieval-Dependent Learning in Higher Education

Christopher Morales 

Pure and Applied Science, Florida SouthWestern State College, Fort Myers, FL 33919, USA;
christopher.morales@fsw.edu

Abstract

Generative artificial intelligence is increasingly used in higher education to provide explanations, solutions, feedback, and organizational support. Although current debates emphasize academic integrity, productivity, and instructional innovation, less attention has been given to how the timing and form of AI assistance may affect learning after that assistance is removed. This conceptual paper develops the Retrieval Interruption Framework (RIF) through a targeted narrative synthesis of research on retrieval practice, productive failure, scaffolding, cognitive load, cognitive offloading, metacognitive monitoring, the expertise reversal effect, the assistance dilemma, and AI-assisted learning. RIF is proposed as an integrative, AI-specific framework rather than a distinct theory of cognition. It distinguishes retrieval-preserving assistance from retrieval-displacing assistance. Retrieval-preserving assistance supports learners after they have attempted task-relevant recall, self-explanation, problem representation, or solution generation. Retrieval-displacing assistance supplies the targeted explanation, solution, or reasoning structure before an initial learner response. The framework predicts that retrieval-displacing assistance may improve immediate performance while weakening delayed unsupported recall, explanation quality, transfer, or metacognitive calibration, particularly in conceptually demanding tasks and among learners with limited prior knowledge. However, early AI guidance may remain productive when it reduces extraneous cognitive load, promotes active processing, fades over time, and is followed by independent performance. RIF reframes AI integration as a sequencing and instructional-design problem. Future research should compare specific forms of AI-first and learner-first assistance using immediate performance measures and delayed unsupported learning outcomes.

Keywords: AI-assisted learning; cognitive offloading; durable learning; generative AI; instructional sequencing; metacognition; retrieval practice; scaffolding



Academic Editors: Eva Ordóñez-Olmedo, Pilar Gómez-Rey, Isotta MacFadden and Lourdes Belén Espejo Villar

Received: 22 May 2026

Revised: 16 July 2026

Accepted: 20 July 2026

Published: 23 July 2026

Copyright: © 2026 by the author.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Generative artificial intelligence (AI) has rapidly entered higher education, reshaping how students access information, construct written responses, receive feedback, and engage with academic tasks. Large language models can generate explanations, summarize readings, organize arguments, solve problems, produce study materials, and simulate tutoring interactions with unprecedented speed (Kasneci et al., 2023; Mollick & Mollick, 2024). Current scholarship has emphasized both the opportunities and risks associated with generative AI, including personalization, accessibility, academic integrity, assessment design, bias, transparency, and responsible implementation (Chan & Hu, 2023; Dwivedi et al., 2023). These concerns are important, but they do not fully address a more specific learning-science

question: how might the timing and form of externally generated explanations, solutions, and reasoning structures influence learning after AI assistance is removed?

This question matters because AI use is not educationally uniform. The same tool may support learning when it provides feedback, clarification, comparison, or alternative explanations after students have attempted a task. It may function differently when it supplies the initial explanation, solution, or reasoning structure before learners have generated their own response. Recent empirical work suggests that the consequences of generative AI depend on how students use it, whether its use is scaffolded, and whether AI-supported performance translates into later unsupported understanding (Bastani et al., 2025; Fan et al., 2025; Xu et al., 2025). The central issue is therefore not whether AI helps or harms learning in general, but when it supports learner-generated cognitive activity and when it substitutes for the activity instruction is intended to develop.

Retrieval is one important component of this problem. In this paper, retrieval refers specifically to the recall or reconstruction of previously learned knowledge from memory. It is conceptually distinct from self-explanation, problem representation, conceptual reconstruction, and metacognitive monitoring, although these processes may interact during complex learning tasks. Research on retrieval practice has shown that actively recalling information can strengthen long-term retention more effectively than passive review alone (Roediger & Karpicke, 2006; Karpicke & Blunt, 2011). Effortful retrieval can also support later accessibility and transfer because learners must reconstruct knowledge rather than merely recognize it. Related work on desirable difficulties and productive struggle suggests that some forms of effort, delay, and initial difficulty can improve long-term learning even when they reduce immediate fluency or performance (Bjork & Bjork, 2011).

Metacognitive monitoring represents a related but separate process. Learners must judge what they know, identify uncertainty, evaluate the adequacy of their reasoning, and decide when additional support is needed. Students may misjudge the depth of their understanding when information feels fluent, familiar, or easy to process, mistaking recognition or polished performance for durable understanding (Bjork et al., 2013; Brown et al., 2014). Generative AI may intensify this problem when it provides coherent and authoritative-sounding explanations before learners have evaluated their own knowledge. Under these conditions, students may move from uncertainty to apparent clarity without identifying what they could recall, explain, or solve independently.

The distinction between immediate performance and later unsupported learning is especially important in higher education. Students are expected to move beyond information access toward explanation, synthesis, transfer, and disciplinary reasoning. Tasks such as explaining a biological mechanism, interpreting evidence, analyzing data, justifying a procedure, or solving an unfamiliar problem may require several distinguishable processes: retrieving relevant knowledge, representing the problem, generating an explanation or solution, monitoring uncertainty, and revising understanding after feedback. AI assistance may support these processes when it provides appropriately timed prompts, modeling, feedback, or comparison. However, when AI supplies the completed cognitive product targeted by instruction before learners have attempted that product themselves, it may improve immediate task completion while reducing opportunities to practice the underlying process.

The present paper develops the Retrieval Interruption Framework (RIF) as an integrative, AI-specific extension of established research on retrieval practice, productive failure, scaffolding, cognitive load, cognitive offloading, metacognitive monitoring, the expertise reversal effect, and the assistance dilemma. RIF does not propose a new cognitive architecture or general learning mechanism. Instead, it organizes established principles around a problem made especially consequential by generative AI: AI systems can supply completed explanations, solutions, and reasoning structures before learners have attempted

the task-relevant cognitive activity that instruction is intended to develop. The framework distinguishes retrieval-preserving assistance, which supports learner activity without replacing the targeted cognitive process, from retrieval-displacing assistance, which performs that process before the learner has generated an initial response.

The aim of this paper is to clarify the conceptual foundations, boundary conditions, and testable predictions of RIF in higher education. Its principal claim is that the educational consequences of AI depend not only on the quality of AI-generated output, but also on the type of assistance provided, its position within the learning sequence, the cognitive target of the task, the learner's prior knowledge, and whether learning is measured during AI-supported performance or through delayed unsupported outcomes. The next section positions RIF in relation to adjacent frameworks in educational psychology and learning science.

2. Conceptual Review Approach

This conceptual paper was developed through a targeted narrative synthesis of literature on retrieval practice, desirable difficulties, productive struggle, scaffolding, cognitive load, cognitive offloading, metacognitive monitoring, expertise reversal, distributed cognition, the assistance dilemma, and AI-assisted learning. Searches were conducted in ERIC, PsycINFO, Scopus, Web of Science, and Google Scholar using combinations of terms including "retrieval practice," "testing effect," "desirable difficulties," "productive struggle," "cognitive offloading," "metacognitive monitoring," "generative AI," "AI-assisted learning," "ChatGPT education," "AI tutoring," and "higher education." Sources were selected when they directly informed the paper's central question: how the timing and form of external cognitive assistance may influence retrieval, learner-generated cognitive processing, feedback, metacognitive calibration, delayed unsupported learning, and transfer. Priority was given to peer-reviewed empirical studies, theoretical syntheses, meta-analyses, and recent research on generative AI in educational settings.

The review was interpretive and theory-building rather than systematic. Its purpose was to integrate established learning-science principles and identify an AI-specific sequencing problem rather than to provide exhaustive coverage of the literature or estimate an overall intervention effect. The synthesis therefore focused on points of convergence, qualification, and tension among retrieval practice, scaffolding, cognitive load, productive failure, metacognition, cognitive offloading, and AI-assisted learning. These bodies of research were used to define the constructs, boundary conditions, and testable propositions of RIF.

3. Relationship to Adjacent Learning Frameworks

RIF established principles from educational psychology and learning science rather than proposing a new cognitive architecture or general theory of learning. Retrieval practice, desirable difficulties, productive failure, scaffolding, cognitive offloading, metacognitive monitoring, Cognitive Load Theory, the expertise reversal effect, and the assistance dilemma already explain important aspects of memory, effort, guidance, external tool use, and learner support. RIF applies these principles to an AI-specific sequencing problem: generative AI can provide a completed explanation, solution, or reasoning structure before learners have attempted the cognitive activity targeted by instruction.

RIF therefore distinguishes among several related but conceptually separate processes. Retrieval refers to accessing previously learned knowledge from memory. Self-explanation refers to generating an account of how or why a concept, mechanism, or solution works. Problem representation involves identifying the structure, conditions, and relevant features of a task. Metacognitive monitoring refers to judging one's knowledge, uncertainty, confidence, or errors. Conceptual reconstruction occurs when learners revise or reorganize understanding after feedback, contradiction, or further evidence. Generative AI

may support or displace one or more of these processes, but RIF does not treat them as interchangeable forms of retrieval.

The contribution of RIF lies in explaining how the timing and form of generative AI assistance may change the sequence among these established processes. The framework distinguishes retrieval-preserving assistance, which supports learners without replacing the task's targeted cognitive operation, from retrieval-displacing assistance, which supplies that operation before learners have generated an initial response. Because generative AI can produce fluent explanations, organized arguments, solutions, and revisions, it may improve immediate task performance without establishing that learners can later recall, explain, transfer, or evaluate the underlying knowledge independently.

Table 1 positions RIF in relation to adjacent frameworks and identifies its narrower contribution: specifying how AI timing, assistance type, task demands, and delayed unsupported outcomes can be examined within an integrated learning-science framework.

Table 1. Positioning Retrieval Interruption Framework Within Adjacent Frameworks.

Adjacent Framework	What the Framework Explains	Limitation in the Context of Generative AI	RIF's Added Explanatory Value
Retrieval practice/ testing effect	Active recall strengthens retention, transfer, and later access to knowledge (Roediger & Karpicke, 2006; Karpicke, 2012).	Explains why recall matters, but not how AI-generated explanations may prevent or reduce the initial retrieval attempt.	Explains how AI may replace the learner's opportunity to attempt recall, self-explanation, or reconstruction before receiving support.
Desirable difficulties	Effort, delay, and challenge can strengthen long-term learning even when they reduce immediate fluency (Bjork & Bjork, 2011).	Explains why some difficulty is productive, but not when AI removes learning-critical effort.	Identifies an AI-mediated difficulty-removal mechanism: fluent assistance may arrive before retrieval effort occurs.
Productive struggle/ productive failure	Initial difficulty, incomplete attempts, or failed problem solving can prepare learners for later instruction (Kapur, 2008).	Explains why initial generation matters, but not AI-generated explanations that eliminate the need for learners to generate an initial response.	Explains when AI short-circuits the learner's first attempt to predict, explain, solve, or represent a problem.
Scaffolding	Support helps learners' complete tasks they could not yet perform independently and gradually transfers responsibility to the learner (Wood et al., 1976).	Can blur the distinction between support that develops independence and support that substitutes for learner cognition.	Distinguishes retrieval-preserving scaffolding from retrieval-displacing substitution. Hints, prompts, and worked examples become interruptive only when they bypass the learner's initial cognitive response.
Cognitive offloading	External tools reduce internal cognitive demands and can free resources for higher-order reasoning (Risko & Gilbert, 2016).	Does not always distinguish offloading extraneous demands from offloading learning-critical processes.	Focuses on the offloading of retrieval, explanation, problem representation, and conceptual reconstruction when those processes are instructional targets.
Metacognitive monitoring	Learners must judge what they know, regulate strategies, and calibrate confidence during learning (Bjork et al., 2013; Metcalfe & Finn, 2008).	Does not fully address how fluent AI output may create apparent understanding before learners evaluate their own knowledge.	Introduces metacognitive bypass, in which AI-generated clarity may reduce opportunities to diagnose uncertainty, recognize gaps, or calibrate confidence.

Table 1. Cont.

Adjacent Framework	What the Framework Explains	Limitation in the Context of Generative AI	RIF's Added Explanatory Value
Cognitive Load Theory/worked-example effect	Novices may benefit from modeling, worked examples, and structured guidance because these supports reduce extraneous load and support schema acquisition (Atkinson et al., 2000; Sweller et al., 2011).	Explains why guidance can help novices, but not when generative AI shifts from useful modeling to substitution for active processing.	Qualifies the retrieval-first claim. AI support is less interruptive when it prompts self-explanation, fades over time, and leads to later independent retrieval.
Expertise reversal effect	Instructional support affects novices and experts differently; supports useful for novices may become redundant as expertise develops (Kalyuga et al., 2003).	Explains why learner knowledge matters, but not how AI-generated explanations may affect learners differently depending on timing and task demands.	Predicts that novices may be more vulnerable to AI-generated fluency, whereas advanced learners may use AI more effectively for critique, comparison, and extension.
Assistance dilemma	Instructional research examines when help should be provided or withheld to optimize learning (Koedinger & Alevan, 2007).	Addresses help timing broadly but predates widely available generative AI systems that instantly generate explanations, solutions, and reasoning structures.	Extends the assistance dilemma to generative AI by theorizing effects on retrieval, error exposure, metacognitive calibration, and delayed unsupported learning.

Note. RIF is an integrative, timing-sensitive framework that applies established learning principles to generative AI. It distinguishes retrieval-preserving assistance from retrieval-displacing assistance and does not assume that scaffolding, worked examples, cognitive offloading, or other forms of instructional support are inherently detrimental.

Table 1 clarifies how RIF extends rather than replaces adjacent learning frameworks. Its central question is whether AI assistance preserves learners' engagement in the cognitive process targeted by a task—such as retrieval, self-explanation, problem representation, or metacognitive monitoring—or performs that process before learners generate an initial response. This distinction guides the following sections, which examine the cognitive foundations, boundary conditions, and empirical implications of the framework.

4. Retrieval, Feedback, and Durable Learning

Retrieval refers specifically to the recall or reconstruction of previously learned knowledge from memory. It is not merely an assessment event that reveals what has already been learned; it can also modify the accessibility and future usability of knowledge. Karpicke (2012) identified retrieval as a central mechanism of meaningful learning because each retrieval attempt can strengthen later access to information. Roediger et al. (2011) similarly described testing as a learning event that can improve retention, knowledge organization, and transfer when used instructionally rather than only evaluatively. Within the RIF, retrieval is therefore treated as a distinct memory process that may contribute to, but should not be conflated with, self-explanation, problem representation, conceptual revision, or metacognitive monitoring.

Research on effective learning strategies supports the importance of providing opportunities for independent recall. Dunlosky et al. (2013) identified practice testing and distributed practice as high-utility learning techniques, whereas rereading and highlighting were judged less effective under many conditions. Students nevertheless often rely on strategies that feel fluent without producing durable recall. Generative AI may amplify this problem by making organized and apparently complete explanations immediately available, thereby reducing the perceived need to reconstruct knowledge independently.

This distinction is connected to the difference between immediate performance and durable learning. Soderstrom and Bjork (2015) argued that observable performance during instruction can be an unreliable indicator of long-term learning. Conditions that improve immediate accuracy or fluency may not produce stronger retention or transfer, whereas conditions that initially slow performance may support later learning. RIF therefore treats immediate AI-supported performance and delayed unsupported learning as separate outcomes rather than assuming that improvement in one demonstrates improvement in the other.

Retrieval attempts may also be diagnostically valuable because they expose uncertainty, incomplete knowledge, and misconceptions. Butler (2010) found that repeated testing can support transfer more effectively than repeated studying, while Kornell et al. (2009) showed that unsuccessful retrieval attempts can enhance subsequent learning under some conditions. Pretesting research similarly indicates that an unsuccessful attempt before instruction can prepare learners to benefit from later information (Richland et al., 2009). These findings do not suggest that difficulty is always productive or that assistance should always be delayed. They indicate that an initial attempt may create a specific knowledge state against which later explanation, correction, or feedback can be evaluated.

Feedback provided after an initial attempt may therefore function differently from information supplied before learners have generated a response. Post-attempt feedback can address a specific error, refine a partially developed explanation, or correct an inaccurate reconstruction. Information provided before an attempt may still support comprehension or schema formation, but it does not reveal what the learner could retrieve or generate independently. Hattie and Timperley (2007) argued that effective feedback helps learners understand their current performance, desired goals, and next steps. Shute (2008) similarly emphasized that formative feedback should be specific and connected to learner performance. Within RIF, the distinction is not that feedback before retrieval is inherently ineffective, but that feedback can operate more diagnostically when it responds to an identifiable learner attempt.

Metacognitive monitoring is related to retrieval but remains conceptually distinct from it. Retrieval concerns access to knowledge, whereas metacognitive monitoring concerns learners' judgments about what they know, how confident they should be, and whether additional support is needed. Metcalfe and Finn (2008) showed that judgments of learning can be shaped by retrieval fluency, while Bjork et al. (2013) demonstrated that learners' beliefs about effective learning strategies may conflict with evidence from learning science. Fluent AI-generated output may therefore influence confidence even when independent retrieval remains weak. This possibility should be examined through measures that compare confidence judgments with delayed unsupported recall, explanation, or transfer rather than by treating metacognitive monitoring as another form of retrieval.

RIF builds on this evidence by asking whether AI assistance preserves or replaces the learner's opportunity to perform the cognitive process targeted by instruction. In a retrieval task, the relevant process is recall from memory. In an explanation task, it is the generation of a causal or conceptual account. In a problem-solving task, it may involve representing the problem and producing a solution strategy. AI assistance is retrieval-preserving when it supports these processes without supplying the completed target response before the learner has attempted it. It becomes retrieval-displacing when it performs the target process and responsibility is not subsequently returned to the learner through faded guidance, independent practice, or delayed unsupported assessment.

RIF therefore does not claim that all learning depends on unsupported recall. Its narrower claim is that the educational consequences of AI assistance should be evaluated according to the process targeted by the task, the timing and completeness of the assistance, and the learner's later ability to perform without that assistance.

5. AI-Assisted Cognition and Cognitive Offloading

Artificial intelligence belongs to a broader history of cognitive tool use in education. Learners have long relied on books, notes, diagrams, calculators, search engines, worked examples, and peer discussion to extend cognition beyond unaided memory. These tools can reduce unnecessary cognitive demands, support organization, and allow learners to devote attention to more complex reasoning. From this perspective, external assistance is not inherently incompatible with learning. Cognitive activity frequently develops through interactions among learners, tools, representations, and environments rather than within the individual mind alone (Clark & Chalmers, 1998; Kirsh, 1995). The central question for the RIF is therefore not whether learners use external tools, but what cognitive activity those tools perform, when assistance is introduced, and whether responsibility is ultimately returned to the learner.

Cognitive offloading refers to using physical or digital tools to reduce internal cognitive demands (Risko & Gilbert, 2016). Offloading can be adaptive when it helps learners manage complexity, store information externally, or direct limited cognitive resources toward higher-order reasoning. Writing down a formula, sketching a diagram, using a graphing tool, or consulting notes may reduce unnecessary memory demands while preserving the reasoning required by a task. The instructional consequences become different, however, when a tool performs the cognitive process that learners are expected to develop. RIF therefore distinguishes between offloading auxiliary demands and offloading a task's central cognitive target. The latter might include recalling relevant knowledge, generating an explanation, representing a problem, constructing an argument, or revising a conceptual model.

Generative AI extends earlier forms of cognitive support because it can produce fluent explanations, organize arguments, generate examples, propose solutions, compare ideas, and simulate dialogue in real time. Earlier tools often provided information, storage, representation, or computational support. Generative AI can participate more directly in producing the cognitive product itself. When students ask AI to summarize a reading, explain a mechanism, draft an argument, or solve a problem, the system may perform part of the generative work that the task was designed to elicit from the learner. Whether this is instructionally beneficial depends on the task's learning objective. AI-generated organization may reduce extraneous demands in one context while replacing the targeted reasoning process in another.

Research on artificial intelligence in education has emphasized the potential of intelligent systems to support tutoring, adaptive instruction, feedback, and personalized learning when they are designed and implemented carefully (Holmes et al., 2019; Luckin et al., 2016; Zawacki-Richter et al., 2019). Generative AI extends these possibilities by making conversational explanation and natural-language assistance widely available. In higher education, AI may help students clarify terminology, compare interpretations, generate practice questions, access alternative explanations, and receive formative feedback outside formal class time.

This distinction matters because AI-supported performance may not correspond directly to independent understanding. A learner may produce a coherent explanation, revise a sophisticated response, or participate in a discussion using AI-supported language without being able to recall, explain, or apply the underlying concepts independently. Bastani et al. (2025), for example, found that generative AI could improve supported performance during practice while producing weaker later performance when students worked without AI, unless the system incorporated instructional guardrails. This finding does not establish that AI assistance generally harms learning. It demonstrates why immediate AI-supported performance and delayed unsupported learning should be measured separately.

AI-assisted cognition also raises questions about reliance and judgment. Research on automation has shown that users may become overly reliant on automated systems,

particularly when outputs appear authoritative or reduce the effort required for task completion (Parasuraman & Riley, 1997). Research on automation bias similarly suggests that individuals may accept automated recommendations even when independent judgment is warranted (Skitka et al., 1999). Educational AI differs from automated decision systems in aviation, medicine, and other professional settings, but the underlying issue remains relevant. Learners may accept a fluent AI-generated explanation without determining whether it is accurate, complete, applicable to the task, or consistent with their own knowledge. The concern is therefore not only whether learners use AI, but whether they retain sufficient disciplinary knowledge and evaluative responsibility to question its output.

AI use also interacts with self-regulated learning. Effective learners plan, monitor, evaluate, and adjust their strategies over time (Azevedo & Hadwin, 2005; Winne & Hadwin, 1998). AI can support these processes by asking learners to articulate an approach, justify reasoning, evaluate confidence, identify uncertainty, or revise an initial response. It may weaken self-regulatory engagement when it supplies direction before learners have clarified the task, generated an approach, or evaluated their own understanding. Fan et al. (2025) describe this concern as “metacognitive laziness,” whereas Xu et al. (2025) show that explicit metacognitive support in generative AI environments can strengthen self-regulated learning. Together, these findings indicate that the educational consequences of AI depend partly on whether its design prompts learner monitoring or bypasses it.

RIF does not imply that all assistance must follow unsupported retrieval. Learners with limited prior knowledge, language barriers, incomplete schemas, or an unclear understanding of the task may require early modeling or scaffolding before meaningful independent performance is possible. Scaffolding can help learners engage in tasks they could not yet complete independently and should gradually transfer responsibility back to the learner (Wood et al., 1976).

6. Retrieval Interruption Framework (RIF)

The framework’s central distinction is between retrieval-preserving assistance and retrieval-displacing assistance. Retrieval-preserving assistance occurs when AI provides prompts, cues, feedback, comparison, correction, or elaboration without replacing the learner’s opportunity to generate an initial task-relevant response. Retrieval-displacing assistance occurs when AI supplies the completed explanation, solution, reasoning structure, or organizational product before the learner has attempted the process that instruction is intended to develop. The distinction is functional rather than purely chronological. Assistance provided early in learning is not necessarily displacing, and assistance provided later is not necessarily productive. Its classification depends on what cognitive work the AI performs and whether responsibility is subsequently transferred back to the learner.

Within RIF, retrieval interruption is defined narrowly as the reduction or bypassing of a learner’s attempt to access previously learned knowledge from memory before receiving external assistance. Retrieval is therefore distinct from self-explanation, problem representation, conceptual revision, and metacognitive monitoring. AI may also displace these processes, but such displacement should not automatically be described as retrieval interruption. For example, an AI-generated explanation may replace self-explanation even when no prior knowledge was available to retrieve, whereas an AI-generated answer may interrupt retrieval when the learner possessed relevant knowledge but did not attempt to access it. This distinction improves the framework’s conceptual precision and allows each process to be measured independently.

Not all guidance constitutes retrieval-displacing assistance. Learners with limited prior knowledge may require modeling, worked examples, cues, or structured scaffolding before meaningful independent performance is possible. Early assistance is more likely

to remain productive when it reduces extraneous cognitive load, highlights relevant task features, prompts active processing or self-explanation, fades as competence develops, and is followed by an independent attempt. It becomes more likely to displace learning when it supplies the completed target response, removes the need for learner generation, and does not subsequently return cognitive responsibility to the learner.

RIF also distinguishes immediate AI-supported performance from delayed unsupported learning. This difference should be treated as a predicted performance–learning dissociation, not as a separate theoretical construct. AI may improve the accuracy, fluency, completeness, or speed of an immediate response while leaving uncertain whether learners can later recall, explain, solve, or transfer the underlying knowledge without assistance. Evidence consistent with RIF would therefore involve stronger immediate performance in an AI-first condition accompanied by weaker delayed unsupported performance than in a learner-first condition. Conversely, equivalent or stronger delayed unsupported learning following AI-first assistance would challenge the claim that the assistance was retrieval-displacing under those conditions.

Reduced metacognitive monitoring is one possible mechanism through which retrieval-displacing assistance may affect learning, but it should not be presented as a separate construct called metacognitive bypass. When AI supplies a fluent response before learners evaluate their own understanding, they may have fewer opportunities to identify uncertainty, detect errors, or calibrate confidence. These outcomes can be tested through confidence judgments, confidence accuracy, error detection, strategy selection, and reflections completed before and after AI use. Treating metacognitive monitoring as a measurable mechanism rather than a separate label makes the framework more parsimonious. Figure 1 illustrates the proposed retrieval-preserving and retrieval-displacing pathways.

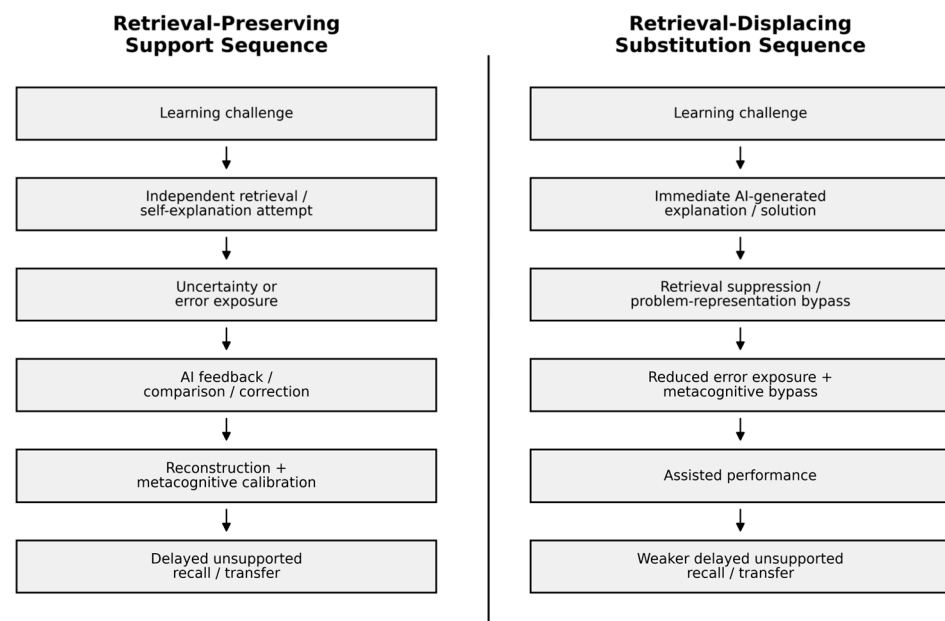


Figure 1. Retrieval-preserving support and retrieval-displacing substitution in AI-assisted learning. Note. The model contrasts retrieval-preserving support with retrieval-displacing substitution. AI preserves learning when it follows an initial learner attempt but may displace retrieval when it supplies an explanation or solution before that attempt.

Testable Propositions

Whether AI functions as retrieval-preserving or retrieval-displacing assistance is expected to depend on the type and completeness of the assistance, its position within the learning sequence, the cognitive target of the task, learner prior knowledge, the degree to

which guidance is faded, and the outcome used to measure learning. The following propositions specify conditions under which the framework could be supported or challenged:

Proposition 1. *AI timing and delayed unsupported learning. In tasks requiring learners to explain, solve, or represent previously taught content, a completed AI-generated response provided before an initial learner attempt will produce equal or stronger immediate task performance but weaker delayed unsupported recall, explanation quality, or transfer than AI feedback provided after an initial learner-generated response. This proposition would be challenged if learners in the AI-first condition demonstrate equivalent or stronger delayed unsupported performance after prior knowledge, instructional time, and task exposure are controlled.*

Proposition 2. *AI timing and metacognitive calibration. In tasks for which learners can generate an initial response, AI feedback provided after that response will produce more accurate metacognitive calibration than a completed AI-generated response provided before learner effort. Calibration may be measured through the correspondence between confidence and accuracy, identification of errors or knowledge gaps, and selection of appropriate study strategies. This proposition would be challenged if AI-first assistance produces equivalent or stronger calibration on delayed unsupported measures.*

Proposition 3. *Novice learners and productive guidance. Among learners with limited prior knowledge, completed AI-generated explanations that remove the need for active processing will produce weaker delayed independent performance than AI-supported worked examples or scaffolds that direct attention, prompt self-explanation, fade over time, and culminate in an unsupported attempt. This proposition does not predict that all early assistance will be harmful. It would be challenged if completed AI-generated responses produce equivalent or stronger delayed unsupported learning than actively processed and faded guidance.*

Proposition 4. *Task-specific effects. The difference between AI-first and learner-first sequencing will be greater in tasks whose instructional targets include recall, explanation, problem representation, argument generation, conceptual integration, or transfer than in tasks involving editing, formatting, translation, accessibility, or other auxiliary support. This proposition would be challenged if sequencing produces comparable effects across task types or if stronger negative effects consistently occur in auxiliary tasks rather than in tasks targeting learner-generated cognition.*

These propositions distinguish immediate AI-supported performance from delayed unsupported learning and identify patterns that could either support or challenge RIF. Conversely, evidence that AI-first assistance consistently produces equal or stronger delayed learning would require revision of the framework's proposed sequencing effect.

7. Boundary Conditions of the Retrieval Interruption Framework

RIF does not propose that AI-supported assistance weakens learning across all contexts. Its predictions depend on boundary conditions that determine whether AI preserves or displaces the cognitive process targeted by instruction. Because retrieval refers specifically to accessing previously learned knowledge from memory, the term retrieval interruption should be reserved for situations in which AI reduces or bypasses an available retrieval attempt. When AI replaces self-explanation, problem representation, or argument generation, it should be described more broadly as retrieval-displacing assistance rather than assumed to involve retrieval directly.

Learner prior knowledge and assistance design are central boundary conditions. Novice learners often possess incomplete schemas and less developed standards for evaluating whether an explanation or solution is accurate, complete, and applicable. Research on the expertise reversal effect indicates that instructional supports may function differently

as prior knowledge develops (Kalyuga et al., 2003). More advanced learners may use AI to compare, critique, verify, or extend existing understanding, whereas novices may accept fluent AI-generated responses without possessing the conceptual knowledge needed to evaluate them. However, RIF does not predict that novices should begin every task through unsupported retrieval.

Cognitive Load Theory and worked-example research qualify any general retrieval-first principle. Novices may benefit from modeling, worked examples, and structured guidance that reduce extraneous cognitive load and support schema acquisition (Atkinson et al., 2000; Sweller et al., 2011). Early AI assistance is therefore more likely to remain productive when it highlights relevant features, models a process, prompts active processing or self-explanation, fades as competence develops, and culminates in an independent attempt. It becomes more likely to displace learning when it supplies the completed target response, permits passive adoption, and does not subsequently return responsibility to the learner. The relevant distinction is therefore not between assistance and no assistance, but between guidance that develops independent performance and assistance that replaces it.

Task type and cognitive target represent a second set of boundary conditions. AI assistance is more likely to affect learning when it performs the same cognitive operation that instruction is intended to develop. Tasks involving recall, explanation, problem representation, conceptual integration, argument generation, disciplinary reasoning, or transfer require learners to generate and reorganize knowledge. When AI supplies the initial explanation, argument, or solution structure, learners may lose an opportunity to practice the targeted process. By contrast, AI may be less disruptive when it supports auxiliary aspects of a task without replacing its central learning objective. Editing grammar, improving formatting, translating instructions, organizing notes, supporting accessibility, or reducing unnecessary workload may preserve the reasoning that learners are expected to perform. These uses should not automatically be categorized as retrieval-preserving, however. Their classification depends on the instructional objective. For example, AI-supported editing may be auxiliary in a science explanation task but may replace the targeted process in a course designed to develop sentence-level writing or revision skills.

Instructional sequencing and feedback design also shape the effects of AI assistance. Learners may first generate an explanation, prediction, solution, diagram, or justification and then use AI to identify omissions, compare alternatives, generate counterexamples, or revise their response. In this sequence, AI functions as feedback acting on an identifiable learner attempt. Effective feedback helps learners understand their current performance, desired goals, and next steps for improvement (Hattie & Timperley, 2007). A completed AI-generated response provided before learner effort may still support comprehension or schema formation, but it should not be treated as feedback in the same sense because it does not respond to evidence of the learner's current thinking.

Self-regulation, motivation, and tolerance for uncertainty influence how learners seek and use AI assistance. Students with stronger self-regulatory skills may use AI to challenge their reasoning, request feedback on an attempted explanation, or generate practice questions. Other learners may use AI primarily to accelerate completion, avoid uncertainty, or reduce effort. Confusion and cognitive disequilibrium can sometimes support conceptual change and deeper processing when they occur under appropriate instructional conditions (D'Mello & Graesser, 2012; Lodge et al., 2018). RIF does not assume that uncertainty is inherently beneficial or that learners should remain confused without support. The relevant question is whether learners receive enough time and structure to identify a knowledge gap, generate an initial response when appropriate, and use subsequent assistance to revise their understanding.

AI reliability and learner evaluation capacity constitute another boundary condition. Generative AI can produce responses that are plausible but inaccurate or incomplete (Ji et al., 2023). Learners with limited prior knowledge may be especially likely to accept such responses because fluency can obscure errors or unsupported claims. This creates two potentially interacting risks: AI may displace learner-generated cognition while also introducing information that the learner cannot adequately evaluate. RIF therefore intersects with AI literacy as a cognitive as well as ethical concern. Learners need sufficient disciplinary knowledge, source-evaluation skills, and criteria for determining whether AI-generated content is accurate, relevant, and adequately supported.

Disciplinary context determines which forms of cognition learners are expected to internalize. RIF is therefore not limited to science, technology, engineering, and mathematics. Its predictions apply wherever AI performs the same cognitive work that an instructional task is intended to develop in the learner.

Finally, RIF should be interpreted developmentally. The same learner may require different forms of AI assistance at different stages of knowledge development. Early instruction may require modeling and structured guidance to support initial schema formation. As learners acquire relevant knowledge, independent generation and retrieval may become increasingly important for exposing gaps and strengthening performance. At more advanced stages, AI may function as a tool for efficiency, critique, elaboration, or extension without displacing foundational learning. Productive AI use therefore depends on the learner's current knowledge state, the cognitive objective of the task, the form of assistance provided, and whether responsibility is progressively transferred back to the learner.

These boundary conditions clarify the limited and conditional scope of RIF. AI assistance is neither inherently retrieval-preserving nor inherently retrieval-displacing. Its classification depends on what cognitive work the system performs, whether that work is the instructional target, what the learner can already do, how the assistance is processed, and whether the learner later performs independently. Table 2 summarizes the framework's boundary conditions and corresponding research implications.

Table 2. Boundary Conditions and Predicted Effects in Retrieval Interruption Framework.

Boundary Condition	Conditions Making Retrieval-Displacing Substitution More Likely	Conditions Making Retrieval-Preserving Support More Likely	Research Implication
Learner expertise/ prior knowledge	Learners use AI-generated explanations before developing foundational schemas, evaluation standards, or the ability to detect errors.	AI provides modeling, worked examples, or scaffolding that reduces extraneous load, prompts self-explanation, and gradually moves learners toward independent retrieval. Advanced learners use AI to critique, extend, or compare existing understanding.	Test prior knowledge as a moderator of AI-assisted learning outcomes, including whether AI-supported worked examples help novices when followed by faded guidance and delayed retrieval.
Task cognitive target	AI performs the same cognitive operation the task is intended to develop, such as explanation, transfer, problem representation, argument generation, or conceptual reconstruction.	AI supports lower-level or auxiliary processes, such as editing, formatting, translation, accessibility, organization, or reducing extraneous workload, without replacing the target cognitive process.	Compare AI effects across tasks with different cognitive targets and determine whether interruption is stronger when AI performs the task's core learning operation.

Table 2. *Cont.*

Boundary Condition	Conditions Making Retrieval-Displacing Substitution More Likely	Conditions Making Retrieval-Preserving Support More Likely	Research Implication
Instructional sequencing	AI is used before students attempt independent retrieval, self-explanation, prediction, problem representation, or solution generation.	AI is introduced after an initial learner attempt and is used for comparison, feedback, elaboration, or correction.	Compare AI-first and retrieval-first instructional designs using both immediate performance and delayed unsupported learning outcomes.
Feedback design	AI provides completed answers, explanations, or reasoning structures before learners generate an initial response.	AI provides feedback, counterexamples, corrective guidance, or comparison after learner effort.	Examine how feedback timing affects delayed recall, explanation quality, transfer, and metacognitive calibration.
Self-regulation and metacognitive monitoring	Learners use AI to avoid uncertainty, accelerate completion, bypass monitoring, or accept fluent output as evidence of understanding.	Learners use AI to test, revise, justify, and evaluate their own understanding, including through confidence ratings or reflection prompts.	Measure confidence accuracy, strategy use, metacognitive monitoring, and whether AI use produces calibration or miscalibration.
AI reliability and learner evaluation capacity	AI-generated output is plausible but incomplete or inaccurate, and learners lack the knowledge needed to detect errors.	Learners have sufficient AI literacy, disciplinary knowledge, or instructor-provided evaluation criteria to interrogate AI output critically.	Examine whether retrieval interruption is amplified when learners cannot identify AI errors, omissions, or unsupported claims.
Disciplinary context	The discipline requires internalization of foundational concepts, mechanisms, procedures, evidence evaluation, reasoning patterns, or disciplinary voice.	AI supports later-stage synthesis, comparison, critique, communication, or revision after learners have developed initial disciplinary understanding.	Examine whether retrieval interruption varies by discipline, task structure, and the type of cognition students are expected to internalize.
Developmental stage of learning	AI substitutes for early retrieval, self-explanation, or problem representation before knowledge structures are stable.	AI supports elaboration, refinement, transfer, or productivity after foundational understanding has begun to develop.	Study AI effects across introductory, intermediate, and advanced coursework, including transitions from modeling to faded support to independent retrieval.

Note. These are conceptual predictions, not confirmed causal relationships. The boundary conditions identify when AI assistance may preserve retrieval-dependent learning and when it may displace retrieval, self-explanation, problem representation, metacognitive monitoring, or conceptual reconstruction.

These boundary conditions delimit the scope of RIF. The framework does not treat AI assistance as inherently beneficial or disruptive; its effects depend on the learner's prior knowledge, the cognitive target of the task, the form and timing of assistance, and whether responsibility is subsequently returned to the learner. RIF therefore asks when AI functions as retrieval-preserving assistance and when it becomes retrieval-displacing assistance. This conditional distinction provides a basis for empirical studies comparing immediate AI-supported performance with delayed unsupported recall, explanation, transfer, and metacognitive calibration.

8. Implications for AI-Assisted Learning Design and Future Research

RIF has implications for instructional design and empirical research in higher education. Because AI may function as either retrieval-preserving assistance or retrieval-displacing assistance, its use should be examined as a question of task design, assistance type, instructional sequencing, and learner readiness rather than only as an issue of policy, academic integrity, or access. The central design question is whether AI preserves learners' engagement with the cognitive process targeted by instruction while supporting feedback, clarification, comparison, and revision. This question cannot be answered by timing alone. Instructors must also consider what the AI provides, what learners already know, whether active processing is required, and whether students are later expected to perform independently.

One instructional approach is a learner-first AI workflow in which students generate an initial task-relevant response before receiving AI assistance. Depending on the task, this response may involve recalling prior knowledge, explaining a concept, representing a problem, making a prediction, constructing an argument, or proposing a solution. AI can then be used to identify omissions, compare alternatives, generate counterexamples, provide corrective feedback, or support revision. This sequence is consistent with research indicating that feedback can be especially useful when it responds to an identifiable learner performance that can be compared with a goal, standard, or model (Hattie & Timperley, 2007; Shute, 2008). Within RIF, the value of this sequence lies not in delaying assistance for its own sake, but in ensuring that AI feedback acts on learner-generated thinking rather than replacing it.

A learner-first workflow is not appropriate under every condition. Learners who lack sufficient prior knowledge or do not yet understand the structure of a task may first require modeling, worked examples, cues, or guided practice. In these cases, productive AI assistance should reduce extraneous demands, direct attention to relevant features, prompt active processing or self-explanation, and gradually fade as learners become more capable of independent performance. Figure 2 therefore presents learner-first sequencing as one retrieval-preserving design option rather than as a universal requirement for AI-assisted learning.

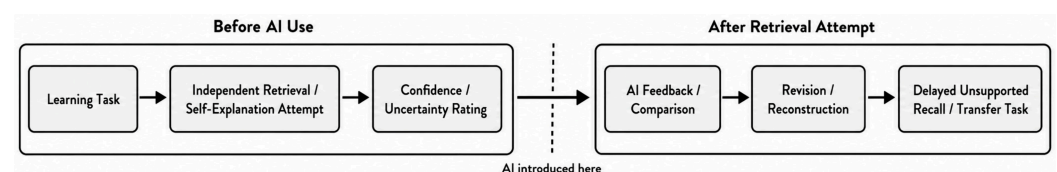


Figure 2. Learner-First AI Workflow for Retrieval-Preserving Assistance. Note. The workflow illustrates one instructional sequence in which learners produce an initial task-relevant response. The dashed vertical line indicates the point at which AI assistance is introduced, after the learner has completed an independent retrieval or self-explanation attempt and rated their confidence.

RIF therefore distinguishes between completed-response generation and learner-process support. Completed-response generation provides an explanation, solution, argument, organizational structure, or written product that learners can adopt or revise. Learner-process support instead prompts students to recall relevant knowledge, clarify the task, generate an initial approach, explain their reasoning, identify uncertainty, evaluate evidence, or revise their own response. AI systems could support these processes by asking learners what they already know, requesting an initial explanation or prediction, eliciting confidence judgments, providing hints rather than completed answers, or asking learners to compare their response with an AI-generated model. These designs align with self-regulated learning models that emphasize planning, monitoring, strategy use, and

reflection (Winne & Hadwin, 1998; Zimmerman, 2002). They also reflect emerging evidence that explicit metacognitive support can shape whether generative AI is used for learning rather than only for task completion (Xu et al., 2025).

The framework also has implications for assessment. As generative AI becomes more available, an unsupervised written product may provide less direct evidence of what a learner can recall, explain, evaluate, or transfer independently. This does not eliminate the educational value of essays, reports, problem sets, or other written assignments. Instead, instructors may need to supplement AI-supported products with evidence of the processes and knowledge underlying them. Oral explanations, annotated drafts, process notes, laboratory debriefs, concept maps, confidence judgments, version histories, reflective comparisons, and delayed unsupported tasks can help distinguish externally supported performance from independent understanding. Assessment should therefore examine not only the quality of the final product, but also the learner's reasoning process and ability to perform the targeted cognitive activity after assistance is removed.

Existing AI-literacy frameworks emphasize responsible use, source evaluation, transparency, and the ability to interact effectively with AI systems (Long & Magerko, 2020; Ng et al., 2021). RIF extends this discussion by proposing a cognitive dimension of AI literacy: understanding how the timing, form, and completeness of AI assistance may affect learner effort, confidence, monitoring, retrieval, and transfer. Students should learn not only to determine whether AI-generated content is accurate, biased, supported, or ethically usable, but also to recognize when AI is supporting their thinking and when it is performing the thinking that a task is intended to develop.

Operationalizing Retrieval Interruption Framework

RIF can be operationalized through four variables: the cognitive process targeted by the task, the form of AI assistance, the point at which assistance enters the learning sequence, and the outcome used to evaluate learning. The cognitive target may involve recall, self-explanation, problem representation, argument generation, conceptual integration, or transfer. AI assistance may take the form of a cue, hint, worked example, partial scaffold, corrective feedback, or completed response. Its timing may occur before an initial learner attempt, after an attempt, or during a scaffolded sequence in which support is gradually faded.

Retrieval interruption should be operationalized narrowly. It can be tested when learners possess relevant prior knowledge, the task requires access to that knowledge from memory, and AI assistance is provided before learners attempt recall. When AI replaces explanation, argument generation, or problem representation rather than recall, the outcome should be described more broadly as displacement of the task's targeted cognitive process. This distinction prevents retrieval from being used as an umbrella term for conceptually separate forms of learner activity.

A direct experimental test of RIF could compare at least two instructional conditions. In an AI-first condition, learners would receive a completed AI-generated explanation, solution, or reasoning structure before generating an initial response. In a learner-first condition, learners would produce an initial explanation, prediction, solution, or problem representation before receiving AI-generated feedback or comparison. Studies involving novices could add a third guided-AI condition in which learners receive a worked example, prompts, or partial scaffolding that requires self-explanation, fades over time, and culminates in independent performance.

Immediate outcomes may include accuracy, response quality, completion time, perceived difficulty, and perceived usefulness. Delayed unsupported outcomes should correspond directly to the task's cognitive target. These may include recall accuracy, explanation

quality without AI, transfer to an unfamiliar problem, independent problem-solving accuracy, argument quality, error detection, and confidence calibration. Because immediate performance may not reliably represent durable learning, studies should prespecify a delayed assessment completed without AI rather than relying only on the quality of AI-supported products (Roediger & Karpicke, 2006; Soderstrom & Bjork, 2015).

Evidence would support RIF's proposed performance–learning dissociation if AI-first learners demonstrated equal or stronger immediate performance but weaker delayed unsupported performance than learner-first learners. Evidence that AI-first learners demonstrated equivalent or stronger delayed recall, explanation, or transfer under otherwise comparable conditions would challenge the proposed sequencing effect. Bastani et al. (2025) provide an emerging model for this distinction by showing that generative AI may improve supported task performance without necessarily improving later performance after assistance is removed.

Learner prior knowledge should be tested as a moderator rather than treated only as a categorical novice–expert distinction. For novices, studies should compare completed AI-generated answers with worked examples, hints, self-explanation prompts, faded scaffolding, and post-attempt feedback. RIF would be supported if actively processed and faded guidance produced stronger delayed unsupported performance than completed-response generation. It would be challenged if completed AI responses consistently produced equivalent or stronger delayed learning.

Metacognitive calibration should be evaluated separately from retrieval or task accuracy. Researchers could collect confidence judgments before AI use, immediately after AI use, and before or after delayed unsupported testing. Calibration can be operationalized as the correspondence between confidence and actual performance, the ability to identify errors or knowledge gaps, and the selection of appropriate study strategies. Increased confidence without corresponding gains in delayed recall, explanation, or transfer would be consistent with a fluency-based illusion of understanding. By contrast, stronger confidence accuracy following appropriately scaffolded AI use would indicate that AI can support metacognitive monitoring rather than displace it (Fan et al., 2025; Xu et al., 2025).

Task type should also be manipulated directly. Researchers could compare learner-first and AI-first conditions across tasks requiring recall, explanation, argument generation, problem representation, or transfer and tasks involving editing, formatting, translation, accessibility, or organization. RIF predicts larger sequencing effects when AI performs the same cognitive process that the task is intended to develop. Comparable effects across task types, or consistently stronger negative effects in auxiliary tasks, would challenge this prediction.

Laboratory experiments should be complemented by classroom-based and design-based research. AI-use logs, process drafts, think-aloud protocols, classroom observations, confidence judgments, student reflections, and delayed assessments can document when learners consult AI, what assistance they request, how they process its output, and whether they can later perform independently. These methods would allow researchers to examine AI assistance as a dynamic instructional process rather than as a binary measure of tool use. These operational criteria make RIF empirically contestable. The framework should not be treated as supported merely because students use AI before completing a task or because immediate performance improves.

9. Conclusions

Generative AI is becoming part of the cognitive infrastructure of higher education. Instructors and researchers must also examine what cognitive work AI performs, when assistance enters the learning sequence, and whether learners can later perform the targeted

process independently. RIF offers an integrative, AI-specific extension of established research on retrieval practice, scaffolding, cognitive offloading, metacognition, Cognitive Load Theory, productive failure, and the assistance dilemma. It does not propose a new cognitive architecture or universal theory of learning. Its narrower contribution is to distinguish retrieval-preserving assistance, which supports learner-generated cognitive activity, from retrieval-displacing assistance, which supplies the task's completed cognitive product before learners have attempted it and does not subsequently return responsibility to them.

RIF treats retrieval as distinct from self-explanation, problem representation, argument generation, conceptual revision, and metacognitive monitoring. The framework therefore does not assume that all AI assistance should follow unsupported retrieval. Novices may require modeling, worked examples, prompts, or structured guidance when these supports reduce extraneous cognitive load, promote active processing, fade as competence develops, and culminate in independent performance. The framework's central empirical prediction concerns the relationship between immediate AI-supported performance and delayed unsupported learning. Evidence consistent with RIF would occur when completed AI assistance produces equal or stronger immediate performance but weaker later recall, explanation, transfer, problem solving, or metacognitive calibration than learner-first or actively scaffolded assistance. Evidence that AI-first assistance produces equivalent or stronger delayed unsupported learning under comparable conditions would challenge the proposed sequencing effect and require refinement of the framework.

RIF therefore positions AI integration as a conditional instructional-design problem rather than a binary choice between adoption and prohibition. Its value will depend on whether future research can identify reliable interactions among assistance type, instructional timing, task demands, prior knowledge, guidance fading, and delayed independent outcomes. Such evidence can support more precise decisions about when generative AI extends learner cognition and when it performs cognitive work that learners still need to develop.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable. This conceptual review did not involve human participants, human data, or animal subjects.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The author declares no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
RIF	Retrieval Interruption Framework

References

- Atkinson, R. K., Derry, S. J., Renkl, A., & Wortham, D. (2000). Learning from examples: Instructional principles from the worked examples research. *Review of Educational Research*, 70, 181–214. [\[CrossRef\]](#)
- Azevedo, R., & Hadwin, A. F. (2005). Scaffolding self-regulated learning and metacognition: Implications for the design of computer-based scaffolds. *Instructional Science*, 33, 367–379. [\[CrossRef\]](#)

- 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, e2422633122. [\[CrossRef\]](#) [\[PubMed\]](#)
- Bjork, R. A., & Bjork, E. L. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. M. Hough, & J. R. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 56–64). Worth Publishers.
- Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology*, 64, 417–444. [\[CrossRef\]](#) [\[PubMed\]](#)
- Brown, P. C., Roediger, H. L., III, & McDaniel, M. A. (2014). *Make it stick: The science of successful learning*. Harvard University Press.
- Butler, A. C. (2010). Repeated testing produces superior transfer of learning relative to repeated studying. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36, 1118–1133. [\[CrossRef\]](#) [\[PubMed\]](#)
- Chan, C. K., & Hu, W. (2023). Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education*, 20, 43. [\[CrossRef\]](#)
- Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, 58, 7–19. [\[CrossRef\]](#)
- D'Mello, S., & Graesser, A. (2012). Dynamics of affective states during complex learning. *Learning and Instruction*, 22, 145–157. [\[CrossRef\]](#)
- Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & 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, 4–58. [\[CrossRef\]](#) [\[PubMed\]](#)
- Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., . . . Wright, R. (2023). "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, 71, 102642. [\[CrossRef\]](#)
- 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, 489–530. [\[CrossRef\]](#)
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77, 81–112. [\[CrossRef\]](#)
- Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial intelligence in education: Promises and implications for teaching and learning*. Center for Curriculum Redesign.
- Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y. J., Madotto, A., & Fung, P. (2023). Survey of hallucination in natural language generation. *ACM Computing Surveys*, 55, 248. [\[CrossRef\]](#)
- Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist*, 38, 23–31. [\[CrossRef\]](#) [\[PubMed\]](#)
- Kapur, M. (2008). Productive failure. *Cognition and Instruction*, 26, 379–424. [\[CrossRef\]](#)
- Karpicke, J. D. (2012). Retrieval-based learning: Active retrieval promotes meaningful learning. *Current Directions in Psychological Science*, 21, 157–163. [\[CrossRef\]](#)
- Karpicke, J. D., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science*, 331, 772–775. [\[CrossRef\]](#) [\[PubMed\]](#)
- 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, 102274. [\[CrossRef\]](#)
- Kirsh, D. (1995). The intelligent use of space. *Artificial Intelligence*, 73, 31–68. [\[CrossRef\]](#)
- Koedinger, K. R., & Alevan, V. (2007). Exploring the assistance dilemma in experiments with cognitive tutors. *Educational Psychology Review*, 19, 239–264. [\[CrossRef\]](#)
- Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35, 989–998. [\[CrossRef\]](#) [\[PubMed\]](#)
- Lodge, J. M., Kennedy, G., Lockyer, L., Arguel, A., & Pachman, M. (2018). Understanding difficulties and resulting confusion in learning: An integrative review. *Frontiers in Education*, 3, 49. [\[CrossRef\]](#)
- Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI conference on human factors in computing systems* (pp. 1–16). Association for Computing Machinery. [\[CrossRef\]](#)
- Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence unleashed: An argument for AI in education*. Pearson.
- Metcalfe, J., & Finn, B. (2008). Evidence that judgments of learning are causally related to study choice. *Psychonomic Bulletin & Review*, 15, 174–179. [\[CrossRef\]](#) [\[PubMed\]](#)

- Mollick, E. R., & Mollick, L. (2024). *Instructors as innovators: A future-focused approach to new AI learning opportunities, with prompts*. The Wharton School, University of Pennsylvania. [\[CrossRef\]](#)
- Ng, D. T., Leung, J. K., Chu, S. K., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence*, 2, 100041. [\[CrossRef\]](#)
- Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, 39, 230–253. [\[CrossRef\]](#)
- Richland, L. E., Kornell, N., & Kao, L. S. (2009). The pretesting effect: Do unsuccessful retrieval attempts enhance learning? *Journal of Experimental Psychology: Applied*, 15, 243–257. [\[CrossRef\]](#) [\[PubMed\]](#)
- Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. *Trends in Cognitive Sciences*, 20, 676–688. [\[CrossRef\]](#) [\[PubMed\]](#)
- Roediger, H. L., III, & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, 17, 249–255. [\[CrossRef\]](#) [\[PubMed\]](#)
- Roediger, H. L., III, Putnam, A. L., & Smith, M. A. (2011). Ten benefits of testing and their applications to educational practice. In J. P. Mestre, & B. H. Ross (Eds.), *Psychology of learning and motivation* (Volume 55, pp. 1–36). Academic Press. [\[CrossRef\]](#)
- Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78, 153–189. [\[CrossRef\]](#)
- Skitka, L. J., Mosier, K. L., & Burdick, M. (1999). Does automation bias decision-making? *International Journal of Human-Computer Studies*, 51, 991–1006. [\[CrossRef\]](#)
- Soderstrom, N. C., & Bjork, R. A. (2015). Learning versus performance: An integrative review. *Perspectives on Psychological Science*, 10, 176–199. [\[CrossRef\]](#) [\[PubMed\]](#)
- Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. Springer.
- Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 277–304). Lawrence Erlbaum Associates.
- Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, 17, 89–100. [\[CrossRef\]](#) [\[PubMed\]](#)
- Xu, X., Qiao, L., Cheng, N., Liu, H., & Zhao, W. (2025). Enhancing self-regulated learning and learning experience in generative AI environments: The critical role of metacognitive support. *British Journal of Educational Technology*, 56, 1842–1863. [\[CrossRef\]](#)
- Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education: Where are the educators? *International Journal of Educational Technology in Higher Education*, 16, 39. [\[CrossRef\]](#)
- Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory into Practice*, 41, 64–70. [\[CrossRef\]](#) [\[PubMed\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.