The Cognitive Cost of Convenience: AI Tools, Learning, and the Unexamined Tradeoffs in Higher Education

The most consequential finding in emerging AI education research is not whether these tools help students complete tasks—they clearly do—but whether task completion translates into learning. Growing evidence suggests a fundamental tension: AI assistance can improve immediate performance while simultaneously undermining the cognitive processes that build long-term competence. This "productivity-learning paradox" represents perhaps the most significant challenge facing educators as they navigate AI integration, demanding careful attention to when efficiency serves education and when it subverts it.

The research landscape reveals a troubling pattern. Students using GPT-4 for mathematics practice answered **48% more problems correctly** during assisted sessions, (Cuspbeb) yet scored **17% worse** on subsequent unassisted tests compared to students who never had AI access. (PubMed Central) (SSRN) Neuroimaging studies show that LLM-assisted essay writers exhibit the **weakest brain connectivity patterns** compared to those using search engines or working unaided—and these diminished patterns persist even after AI access is removed. (MIT Media Lab) (BRAINI) Meta-analyses confirm medium-to-large positive effects on immediate performance (g = 0.867), but long-term retention and transfer effects remain understudied, and the few longitudinal findings available raise serious concerns about skill atrophy. (nature)

This report synthesizes cognitive psychology foundations, empirical research from 2020-2025, and theoretical frameworks to examine how AI tools affect learning in higher education. The analysis maintains a moderately critical stance: acknowledging genuine benefits while rigorously examining risks that techno-optimist narratives often obscure. The central question is not whether AI can help students perform—it demonstrably can—but whether performance improvements reflect actual learning or merely externally-scaffolded task completion that may leave students less capable when working independently.

How AI reshapes the cognitive architecture of learning

Cognitive Load Theory (CLT), developed by John Sweller and colleagues, provides the most rigorous framework for understanding AI's effects on learning. (MDPI) The theory recognizes that working memory has severe limitations—holding only **2-4 chunks of information** for concurrent processing— (ScienceDirect)and distinguishes three types of cognitive load with fundamentally different implications for learning.

Intrinsic cognitive load represents the inherent complexity of learning material, determined by "element interactivity"—how many elements must be simultaneously processed to understand a concept. This load cannot be reduced through instructional design without simplifying the content itself. **Extraneous cognitive load** results from suboptimal instruction—poor formatting, split attention, redundant information—and contributes nothing to learning. This load should always be minimized. **Germane cognitive load** represents working memory resources devoted to schema construction and automation—the effortful processing essential for building transferable knowledge structures. This is the "productive struggle" that distinguishes genuine learning from mere task completion.

AI tools present a theoretical double-edged sword. When designed well, they can reduce extraneous load—eliminating irrelevant cognitive burden so learners can focus on what matters. However, the same tools can inadvertently reduce germane load—eliminating the cognitive effort required for schema construction and long-term retention. The critical question becomes: **which type of load is being reduced?**

Research published in *Frontiers in Psychology* articulates this "cognitive paradox" directly: "According to Cognitive Load Theory, AI has the potential to maximize the extraneous load through the elimination of redundant work, allowing the student to focus on more important cognitive operations. However, over-reliance on AI has the potential to reduce the germane load, which is needed for deep learning and the development of higher-order thinking skills." (nih)

The **expertise reversal effect** adds crucial nuance. A 2025 meta-analysis by Tetzlaff and colleagues found that low prior knowledge learners benefit from high-assistance instruction (d = **0.505**), while high prior knowledge learners benefit from low-assistance instruction (d = **−0.428**). (ScienceDirect) This asymmetry has profound implications: AI scaffolding appropriate for novices may actively harm learning for advancing students. Static AI assistance that provides the same level of help regardless of learner progress risks impeding precisely those students who have begun developing competence.

The practical implication is that AI tools must be adaptive, fading support as expertise develops. Yet most current implementations provide consistent assistance regardless of learner state, potentially creating a ceiling effect where students never transition to independent capability. As Kalyuga, Rikers, and Paas note: "When instructional guidance is provided to learners who already have sufficient knowledge... an unnecessary extraneous cognitive load could be imposed." (ResearchGate)

When productivity and learning diverge

The most compelling evidence for the productivity-learning tension comes from the University of Pennsylvania study by Bastani and colleagues, which examined approximately 1,000 Turkish high school students in a randomized controlled trial. Students with access to GPT-4 during mathematics practice sessions showed dramatically improved performance—**48% more problems solved correctly** with basic GPT access, **127% improvement** with a safeguarded GPT Tutor version. These are substantial productivity gains by any measure.

However, when AI access was removed for subsequent testing, students who had used GPT-4 performed **17% worse** than control students who never had AI access. The researchers concluded bluntly: "Students attempt to use GPT-4 as a 'crutch' during practice sessions, and when successful, perform worse on their own." This finding precisely illustrates how performance during assisted practice can diverge dramatically from learning as measured by independent capability.

Critically, the study found that safeguards matter. The "GPT Tutor" condition—which provided hints and Socratic prompts rather than direct answers—largely mitigated these negative learning effects while still improving practice performance. This suggests the problem is not AI assistance per se, but AI assistance that eliminates the cognitive effort required for schema construction.

Neurophysiological evidence deepens these concerns. A 2025 MIT Media Lab study measured brain connectivity via EEG while participants wrote essays under three conditions: LLM assistance (ChatGPT), search engine access only, or brain-only (no external tools). Brain connectivity **systematically scaled down with external support**—the brain-only group exhibited the strongest, widest-ranging neural networks, the search engine group showed intermediate engagement, and LLM assistance elicited the **weakest overall coupling**.

More troubling was what happened when LLM users subsequently wrote without AI assistance: "They did not reflect the same high brain connectivity patterns as the original brain-only group." (Brainmindsociety) The researchers term this accumulated effect "cognitive debt"—a neurological pattern suggesting that habitual AI use may reduce independent thinking capacity even after the tool is removed. Self-reported ownership of essays was lowest in the LLM group, and LLM users struggled to accurately quote their own AI-assisted work.

Meta-analytic evidence provides important context. Wang and Fan's 2025 synthesis of 51 studies (N > 4,000) found substantial positive effects on learning performance (g = **0.867**, a large effect) and moderate effects on higher-order thinking (g = **0.457**). (nature) However, these meta-analyses primarily capture immediate post-test performance rather than delayed retention or transfer—the metrics most relevant to genuine learning. The few studies measuring performance after AI removal show the opposite pattern from immediate performance studies.

The research on student agency by Darvishi and colleagues, examining 1,625 undergraduates across 10 courses, found that "students tend to rely on AI rather than learn from it." When AI assistance was removed, feedback quality declined significantly. Combining AI with self-monitoring checklists—an attempt to preserve metacognitive engagement—did not improve outcomes, suggesting potential cognitive overload when multiple support systems interact. The integration of AI in learning environments, they conclude, "could impact students' agency to take control of their own learning."

The empirical landscape: what controlled studies reveal

Research on AI-assisted learning has expanded rapidly since ChatGPT's 2022 release, though the literature exhibits significant methodological limitations that constrain interpretation. The clearest findings emerge from controlled experiments examining specific learning outcomes.

In mathematics education, the Bastani study remains the most methodologically rigorous, with its randomized design, pre-registration, and large sample. The 17% performance deficit when AI is removed represents a meaningful effect size with practical significance—students who practiced with AI were measurably less competent when working independently.

In programming education, results are more mixed. Studies by Jošt and colleagues found no significant performance differences between ChatGPT users and non-users on exams ($p = 0.856$, $n = 182$). ChatGPT appears to enhance procedural skills and efficiency while showing limited benefits for conceptual mastery. High-performing students benefit disproportionately from AI assistance, raising equity concerns about differential skill development. A quasi-experimental study by Yang and colleagues with high school programming students found that the ChatGPT group showed **lower flow experience, self-efficacy, and learning performance** compared to controls—a rare finding of negative effects on immediate measures. (nature)

In writing and language learning, evidence is similarly complex. AI feedback improves revision frequency and grammatical/structural quality in the short term. However, single instances of GenAI feedback have limited long-term impact on writing proficiency, and students sometimes find AI feedback "too indirect or difficult to understand." Human tutors remain superior for guidance on essay structure and logical flow. One study found that ChatGPT use **significantly reduced students' creative writing abilities**, suggesting domain-specific concerns about AI's effects on generative cognitive processes. (nature)

Effect sizes for intelligent tutoring systems provide useful context. Well-designed ITS produce effects of approximately 0.76 standard deviations compared to no tutoring, with only 0.21 SD difference from human tutors. However, these systems are specifically designed with learning science principles—adaptive difficulty, scaffolded hints, retrieval practice—that differ markedly from how students typically use general-purpose LLMs. The Harvard randomized controlled trial finding that an AI tutor produced median learning gains "more than double" classroom active learning specifically involved a tutor designed to guide rather than answer.

Longitudinal evidence remains scarce. Most studies measure immediate post-test performance within 1-8 weeks. The few studies examining effects after AI removal consistently show performance decrements. An operations research course study using AI detection found median 100% AI reliance in final exams, with students unable to fathom learning without AI access—control students showed superior knowledge retention. The critical question of long-term skill development over semesters and years remains essentially unstudied.

Methodological limitations pervade the literature:

- **Short intervention durations** preclude assessment of cumulative effects

- **Immediate outcome measures** conflate task completion with durable learning

- **Selection effects** make observational studies difficult to interpret

- **Conflation of task completion with learning** masks the productivity-learning distinction

- **Limited generalizability** from single-course, single-institution studies

- **Publication bias** toward novel positive findings in an emerging field

---

Historical analogues illuminate the pattern

The current moment is not without precedent. Previous cognitive technologies—calculators, GPS navigation, spell-checkers, and internet search—provide historical evidence for how tool use affects skill development. This research, spanning decades, reveals consistent patterns directly relevant to AI.

Calculator effects on mathematical skills show context-dependent findings. A Vanderbilt study demonstrated that for students with existing multiplication skills, calculator use had no impact. For students without foundational skills, calculator use had **negative impact** on performance. (Vanderbilt University) This expertise-dependent effect parallels CLT's expertise reversal: tools that support experts may harm novices still building foundational competence. Recent findings suggest negative associations between habitual calculator use and mathematical skill development, particularly for students without prior mastery. (PubMed Central)

GPS navigation effects provide the strongest evidence for technology-induced skill atrophy. Dahmani and Bohbot's 2020 longitudinal study found that people with greater lifetime GPS experience had worse spatial memory during self-guided navigation. (nature) (Nature) Critically, **3-year longitudinal follow-up** showed that greater GPS use since initial testing was associated with steeper decline in hippocampal-dependent spatial memory—a dose-dependent relationship. The researchers ruled out reverse causation: "Those who used GPS more did not do so because they felt they had a poor sense of direction, suggesting that extensive GPS use led to a decline in spatial memory rather than the other way around."

This finding has neurological implications. London taxi drivers, who must learn thousands of routes for licensing ("The Knowledge"), show significantly larger posterior hippocampi than non-taxi drivers—spatial cognition exercises the hippocampus. GPS removes requirements to pay attention to surroundings, update internal position, and build cognitive maps. (PubMed Central) The parallels to AI removing requirements for generation, reasoning, and knowledge construction are direct.

Spell-checker effects are less studied but suggestive. Research by Rimbar found that while spell-checkers helped eliminate surface errors, "learners still made the same errors in their spelling after the use of spell-checkers"—the tool corrected output without changing underlying competence. (ResearchGate) Lunsford and Lunsford found that many spelling errors in undergraduate papers resulted from incorrect spell-checker suggestions, demonstrating how tool dependence can create new error types. (UBC Blogs)

The "Google Effect" on memory, documented in Sparrow, Liu, and Wegner's landmark 2011 *Science* paper, demonstrated that people who believed information would be available later had lower recall rates of the information itself. Participants remembered WHERE information was saved better than WHAT the information was—a form of transactive memory where the internet serves as an external storage system. (nature) (MDPI) Critically, explicit instruction to remember material did not significantly improve recall if people believed it was available later.

This finding has direct implications for AI. If students believe AI can provide information or complete tasks on demand, they may invest less in encoding that information or developing those skills. Storm and Stone confirmed that cognitive offloading through internet use is self-reinforcing—each use increases subsequent tendency to offload. (nature) (ScienceDaily) The same pattern likely applies to AI assistance.

Desirable difficulties: why struggle often serves learning

Robert Bjork's research on "desirable difficulties" provides the theoretical foundation for understanding why reducing cognitive effort can impair learning. (Psychology Today) (Ucla) The core insight is paradoxical: **conditions that slow the rate of apparent learning often optimize long-term retention and transfer**, while conditions that make performance improve rapidly often fail to support durable learning. (Learningpit) (ucla)

Bjork distinguishes **storage strength** (how entrenched a memory is) from **retrieval strength** (current accessibility). Current performance reflects only retrieval strength, but storage strength determines long-term retention and relearning. (ucla) If learners interpret current performance as evidence of durable learning, they systematically prefer conditions that feel effective but produce weaker outcomes.

Four interventions consistently qualify as desirable difficulties:

Varying conditions of learning improves retention over constant conditions. (ResearchGate) Students who studied material in two different rooms outperformed those studying twice in the same room. (ucla) Children who practiced throwing to varying distances outperformed those practicing at a fixed distance—even when tested at the practiced distance. (ucla) Variation creates richer encoding that supports flexible retrieval.

Interleaving instruction (mixing problem types) outperforms blocking (practicing one type repeatedly). (Learningpit) Shea and Morgan found that blocked practice appeared superior during training, but interleaved practice produced far better retention 10 days later. (ucla) Rohrer and Taylor found interleaved math formula learning produced 63% versus 20% correct on delayed tests. (ucla) Crucially, most participants believed blocking was more effective despite worse outcomes.

Spacing study sessions over time outperforms massed practice (cramming). This spacing effect is one of the most robust findings in learning research—cramming supports short-term performance while spacing supports long-term retention. (ucla)

Retrieval practice (testing) enhances learning more than additional study. Roediger and Karpicke found that at 5-minute delay, repeated study appeared superior, but at 2-day and 1-week delays, prior testing produced substantially greater retention. Students in the study condition were more confident despite worse long-term performance.

The **generation effect** is particularly relevant to AI concerns. Information is better remembered when actively generated rather than passively received, with meta-analytic effect sizes of $d = \mathbf{0.40}$ across 86 studies. Neural imaging shows that generation activates broad networks including prefrontal cortex and hippocampal regions. When AI generates content that students would otherwise produce themselves, the generation effect is eliminated.

The **retrieval effort hypothesis** suggests that "difficult but successful retrievals are better for memory than easier successful retrievals"— (Wikipedia) connecting testing effects to the broader desirable difficulties framework. (World Bank Blogs) Effort during retrieval strengthens memory traces.

AI assistance systematically undermines these effects:

- Immediate AI answers eliminate spacing that would occur with delayed solutions

- AI-generated content replaces student generation

- AI provides consistent, optimized responses instead of varied conditions

- AI removes retrieval effort by providing information on demand

- AI creates fluent performance that feels like learning but may not be

Critically, difficulties are only "desirable" when learners can successfully overcome them. If AI is removed from students who have never developed foundational competence without it, the resulting difficulties may be undesirable—overwhelming rather than productive. This creates a dependency trap where early AI use prevents developing the capacity for later independent work.

Theoretical frameworks: extension, offloading, or replacement?

The Extended Mind Thesis, articulated by Andy Clark and David Chalmers in their 1998 *Analysis* paper, proposes that cognitive processes can extend beyond the brain into external objects. (nature) (Wikipedia) If an external resource plays the same functional role as an internal cognitive process, it should count as part of the cognitive system. The classic illustration contrasts Otto, who has Alzheimer's and relies on a notebook for memory, with Inga, who uses biological memory—Clark and Chalmers argue both notebooks and neurons serve equivalent cognitive functions. (ModelThinkers) (Wikipedia)

Clark's 2025 update, published in *Nature Communications*, directly addresses AI. He argues that humans are "natural-born cyborgs" who have always built hybrid thinking systems incorporating non-biological resources. (nature) Fears that AI will make us "dumber" echo historical concerns about writing (Plato's *Phaedrus*), calculators, and internet search. (nature) Results showing reduced unaided abilities after using GPS or search engines, Clark suggests, need not represent "shrinkage and loss" but rather "careful husbanding of our own on-board cognitive capital." (nature)

However, Clark acknowledges a crucial distinction: AI tools can be **"mind-extending" or "mind-replacing"**—the critical question is which. Genuine cognitive extension requires:

- **Extended cognitive hygiene**: Skills to know when to trust and question AI

- **Metacognitive awareness**: Estimating reliability based on subject matter and prompting

- **Synergistic collaboration**: AI altering the creative process rather than replacing it

- **Human remains in the loop**: Active engagement rather than passive acceptance

Cognitive offloading research by Risko and Gilbert examines when delegating cognitive tasks to external tools helps versus harms. Grinschgl, Papenmeier, and Meyerhoff's experimental work demonstrated a fundamental tradeoff: cognitive offloading **boosts immediate performance but diminishes memory formation**. (PubMed Central) (Sage Journals) "What users learn in this context appears to be how to effectively utilize the offloading device rather than solving the problem at hand with one's own cognitive abilities." This effect persists for long-term memory, though informing participants of a later memory test can partially mitigate negative effects. (PubMed Central)

Distributed cognition theory, developed by Edwin Hutchins through studies of naval navigation and airline cockpits, expands the unit of cognitive analysis to entire sociotechnical systems. (Wikipedia) Cognitive work becomes "sedimented" in tools and artifacts—"the cartographer has done much of the reasoning for the navigator who uses a map." Artifacts don't merely amplify cognition; they fundamentally reorganize cognitive processes. (ScienceDirect)

Applied to human-AI teams, this suggests AI agents should be incorporated as knowledge sources in a team's **Transactive Memory System**—the group-level system for knowing "who knows what." (Frontiers) Research on ICU physicians working with AI found that in higher-performing teams, accessing AI information was positively linked to generating novel hypotheses. (PubMed) However, this framework assumes expertise that allows strategic tool use—novices may lack the metacognitive capacity to treat AI as a genuine team member rather than an oracle.

The theoretical literature converges on distinguishing beneficial extension from harmful replacement:

Beneficial Extension	Harmful Replacement
User maintains metacognitive control	User passively accepts outputs
Tool used strategically based on task demands	Tool used habitually regardless of context
Cognitive effort redirected to higher-order thinking	Cognitive effort eliminated entirely
Skills transfer to unaided contexts	Dependency prevents independent performance
"Productive struggle" preserved where valuable	Struggle eliminated, learning diminished
User can function without tool	User cannot perform basic tasks independently

Messeri and Crockett's 2024 *Nature* paper warns that AI can create "cognitive monocultures"—cementing certain approaches while impeding alternatives. Like agricultural monocultures, this improves efficiency but increases vulnerability. (nature) Applied to education, if all students converge on AI-mediated approaches, diversity of thinking—and resilience when AI is unavailable—may be compromised.

When AI might genuinely support learning

The evidence does not support blanket condemnation of AI in education. Under specific conditions, AI assistance can enhance rather than undermine learning. Identifying these conditions is essential for evidence-based integration.

Socratic prompting preserves productive struggle. The Bastani study found that "GPT Tutor"—designed to provide hints and guiding questions rather than direct answers—largely mitigated the learning decrements observed with standard GPT-4 access while still improving practice performance. This aligns with cognitive load theory: Socratic scaffolding reduces extraneous load (confusion about how to approach problems) while preserving germane load (the cognitive work of generating solutions).

Retrieval practice enhancement shows promise. Research by An and colleagues found that LLM-generated multiple-choice questions for retrieval practice improved test accuracy by **16 percentage points** (89% vs. 73%). Using AI to create practice materials that then require student retrieval inverts the typical dynamic—AI supports the preparation of learning conditions rather than replacing learning itself.

Intelligent tutoring systems with principled design demonstrate consistent benefits. Well-designed ITS produce effect sizes around 0.76 standard deviations through adaptive difficulty, scaffolded hints, spaced repetition, and retrieval practice. The key distinction is that these systems are engineered for learning rather than task completion—they deliberately introduce productive difficulties rather than eliminating them. Systems like ASSISTments and Carnegie Learning's MATHia show replicable benefits with high evidence ratings.

Expert users with foundational competence may safely offload. The expertise reversal effect and calculator research converge: tools that harm novices may be neutral or beneficial for experts. An experienced writer using AI to accelerate drafting differs fundamentally from a student who has never learned to structure arguments independently. The critical variable is whether foundational skills exist prior to tool adoption.

Meta-analytic evidence suggests structured integration works. Wang and Fan's meta-analysis found the largest effects when ChatGPT was used as an "intelligent tutor" (g = 0.945 for higher-order thinking) rather than as a simple answer provider. Problem-based learning contexts showed particularly strong effects (g = 1.113). Optimal intervention duration was 4-8 weeks—long enough for meaningful engagement but potentially before dependency patterns solidify. (nature)

Frameworks for productive AI integration have emerged from learning science:

Cornell's three-tier framework offers practical guidance:

- **Prohibit** where AI interferes with developing foundational understanding

- **Allow with attribution** where AI serves as a resource requiring student accountability

- **Encourage** where AI enables higher-level objectives and creative exploration (Cornell Teaching)

The NSTA framework positions AI as a "discussion partner" rather than answer provider—students should evaluate, question, and interpret AI information rather than accept it uncritically. (NSTA)

Evidence-based principles for AI tutoring systems include:

1. Adaptive teaching based on demonstrated understanding

2. Active recall practice for long-term retention

3. Spaced repetition with personalized schedules

4. Preventing cognitive overload through strategic presentation

5. Encouraging source verification and critical evaluation

6. Maintaining human expertise as partner, not replacement

The common thread is that beneficial AI use **preserves or enhances cognitive engagement** rather than eliminating it. AI becomes harmful when it removes the struggle that builds competence; it becomes beneficial when it creates conditions for productive struggle or handles genuinely extraneous tasks.

What we don't yet know

The research literature on AI in education exhibits significant gaps that constrain evidence-based policy. Acknowledging uncertainty is essential for intellectual honesty.

Longitudinal studies measuring skill development over years remain essentially absent. Most research tracks outcomes over weeks; the few studies measuring performance months later show concerning patterns, but multi-year effects on cognitive development are unknown. Students who complete degrees with extensive AI assistance may differ systematically from previous cohorts—we cannot yet measure this.

Transfer assessments are rare. Whether AI-assisted learning transfers to novel contexts or real-world applications is almost entirely unstudied. Students may perform well on tasks similar to AI-assisted practice while failing to apply knowledge in new situations. The fundamental purpose of education—preparing students for contexts they haven't yet encountered—requires transfer, but we lack data on AI's effects on it.

Differential engagement modes are underexplored. Research rarely compares generation (asking AI to create content), editing (using AI to improve student work), critique (using AI to evaluate student output), and explanation (asking AI to clarify concepts). These may have very different cognitive implications. A student who writes a draft and uses AI for feedback differs from one who asks AI to generate the draft.

Expertise-level interactions require more study. How do novices, intermediates, and experts differentially benefit or suffer from AI assistance? At what point does scaffolding become counterproductive? Can we identify threshold competencies that predict when AI assistance helps versus harms?

Discipline-specific effects are incompletely mapped. Mathematics, programming, writing, and clinical reasoning may respond differently to AI assistance. Some domains may involve skills more vulnerable to atrophy; others may benefit more from AI scaffolding. Comparative studies across disciplines are needed.

Skill decomposition is underdeveloped. Which specific sub-skills—procedural versus conceptual, surface versus deep, generative versus evaluative—are most affected by AI assistance? Understanding this would enable targeted integration that protects vulnerable skills while leveraging AI for appropriate tasks.

Optimal scaffolding conditions are unknown. Under what precise conditions can AI assistance be designed to promote rather than undermine learning? How should scaffolding fade as expertise develops? What interaction patterns maximize learning while preserving efficiency benefits?

Neurological effects require replication. The MIT EEG findings are provocative but preliminary—not yet peer-reviewed, with modest sample sizes. The suggestion of persistent "cognitive debt" demands replication with larger samples and longer follow-up.

Publication bias likely affects the literature. Novel positive findings attract attention in an emerging field; null results and negative findings may be underreported. The true effect distribution may be less favorable than published meta-analyses suggest.

Equity implications of cognitive offloading

The differential effects of AI on cognitive development raise significant equity concerns that extend beyond traditional access questions.

The third digital divide, as articulated by Michael Trucano at Brookings, represents a new equity dimension. The first divide concerned access to devices; the second concerned skills to use technology beneficially; the third concerns access to AI and—critically—**skilled support in using AI for learning**. Trucano poses the uncomfortable question: "Is it possible to imagine a future where the rich have access to technology, increasingly powered by AI, and to teachers to help them use this technology as part of their learning, while poor kids just have access to the technology?"

If AI differentially affects skill development based on how it is used, students in well-resourced environments with sophisticated pedagogical integration may develop stronger independent capabilities than students in under-resourced environments where AI serves as an unsupervised shortcut. This could create a **two-tiered education system** where learning outcomes vary not by AI access but by quality of AI integration.

Algorithmic bias intersects with cognitive effects. AI systems trained on data not reflecting cultural and socioeconomic diversity may provide differentially useful assistance. Students whose backgrounds are underrepresented in training data may receive less effective scaffolding, compounding existing inequities.

Digital literacy disparities affect metacognitive engagement. Students with stronger AI literacy may use tools strategically, preserving learning benefits while gaining efficiency. Students without such literacy may use AI in ways that undermine skill development. If schools differ in their capacity to teach critical AI engagement, these differences may amplify over time.

The irony of "democratizing" education through AI becomes apparent: if AI assistance impairs skill development in some usage patterns, widespread adoption without sophisticated pedagogical support may disadvantage precisely the students it aims to help. Well-resourced students may receive guidance preventing harmful dependency; under-resourced students may not.

Practical implications for educators and institutions

The evidence supports neither wholesale adoption nor blanket prohibition of AI in education. Instead, it demands careful, context-specific integration guided by learning science principles.

For course design:

- Distinguish assignments where AI would interfere with learning objectives from those where AI appropriately enhances capability

- Use Cornell's prohibit/allow/encourage framework to make explicit decisions for each assignment

- Design assessments that require demonstration of understanding, not just output production

- Include AI-free practice opportunities for schema consolidation

- Consider whether foundational skills are at risk of atrophy with current AI policies

For AI tool selection:

- Prefer tools designed to scaffold rather than replace thinking

- Socratic tutoring approaches show better learning outcomes than answer-providing systems

- Adaptive difficulty and fading scaffolds align with expertise reversal research

- Tools requiring student generation before AI assistance preserve generation effects

For teaching practice:

- Explicitly teach metacognitive skills for evaluating when AI is appropriate

- Help students distinguish productive struggle from unproductive frustration

- Model critical engagement with AI outputs rather than passive acceptance

- Create opportunities to practice skills without AI assistance

- Monitor for signs of dependency (inability to perform basic tasks independently)

For institutional policy:

- Develop explicit, comprehensive AI policies addressing governance, pedagogy, and operations

- Invest in faculty development for evidence-based AI integration

- Address equity through both access and pedagogical support

- Establish monitoring mechanisms for long-term effects

- Avoid detection-focused approaches in favor of learning-focused design

For assessment:

- Design tasks that reveal understanding rather than output quality

- Include components requiring independent demonstration of competence

- Consider process-focused assessment alongside product assessment

- Use oral examinations or in-class demonstrations where appropriate

- Recognize that AI-assisted work may not predict independent capability

---

Conclusions: navigating the productivity-learning tension

The evidence reviewed here does not support simple conclusions. AI tools demonstrably improve task performance, and meta-analyses confirm substantial positive effects on immediate learning metrics. Yet the theoretical frameworks from cognitive psychology—cognitive load theory, desirable difficulties, the generation and testing effects—predict that eliminating cognitive effort will impair long-term learning, and emerging empirical evidence supports these predictions.

The central insight is that learning and task completion are not synonymous. Students who complete assignments faster with AI assistance are not necessarily learning more—and may be learning less. Performance during assisted practice diverges from capability during independent performance. (World Bank Blogs) (ResearchGate) Neural engagement decreases with AI assistance and may not recover when assistance is removed. Historical analogues from GPS, calculators, and internet search show consistent patterns of tool dependency and skill atrophy.

The type of AI assistance matters critically. Socratic scaffolding that preserves productive struggle shows different effects than answer-providing assistance that eliminates cognitive effort. Intelligent tutoring systems designed with learning science principles differ fundamentally from general-purpose LLMs used without pedagogical structure. The question is not whether to use AI but how to use it in ways that support rather than undermine learning.

Expertise level moderates effects. Assistance appropriate for novices may harm advancing learners (expertise reversal effect), and assistance beneficial for experts may harm novices still building foundational competence. Static AI assistance that provides the same help regardless of learner state cannot optimize for both populations simultaneously.

Research gaps constrain confident conclusions. Longitudinal studies over years, transfer assessments, and skill decomposition analyses are largely absent. The true long-term effects of AI-assisted education on cognitive development remain uncertain. Epistemic humility is warranted.

For educators, the practical implication is vigilance. The efficiency benefits of AI are real and should not be dismissed—but neither should they be conflated with learning benefits. Every AI integration decision should ask: Does this preserve the cognitive engagement necessary for schema construction? Does this eliminate struggle that would build competence? Can students demonstrate capability without AI assistance?

The goal is not to reject tools that enhance human capability but to distinguish tools that genuinely extend cognition from those that merely replace it. The difference between a scaffold that supports independent development and a crutch that prevents it may determine whether AI in education represents an advancement or a regression—whether the next generation of students is more capable or merely more assisted.

The stakes are significant. If AI assistance during education impairs the development of independent cognitive capability, students may graduate less prepared for contexts where AI is unavailable, unreliable, or inappropriate. The productivity gains during education would come at the cost of capability thereafter. Whether this tradeoff is acceptable depends on assumptions about the future that educators cannot verify.

What can be said with confidence is that the question deserves far more attention than it currently receives. The enthusiasm for AI integration often proceeds without rigorous examination of learning effects. The research reviewed here suggests such enthusiasm is premature. Until we understand better when AI helps and when it harms, the precautionary principle may warrant more conservative integration than current adoption patterns reflect.

Key sources and further reading

Cognitive Load Theory Foundations

- Sweller, J. (2010). Element Interactivity and Intrinsic, Extraneous, and Germane Cognitive Load. *Educational Psychology Review*, 22, 123-138.

- Sweller, J., van Merriënboer, J.J.G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, 10(3), 251-296.

- Tetzlaff, L., Peters, T., & Simonsmeier, B.A. (2025). A cornerstone of adaptivity – A meta-analysis of the expertise reversal effect. *Learning and Instruction*.

Desirable Difficulties and Memory Research

- Bjork, R.A. & Bjork, E.L. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In *Psychology and the Real World* (pp. 56-64). Worth Publishers.

- Bjork, R.A. & Bjork, E.L. (2020). Desirable difficulties in theory and practice. *Journal of Applied Research in Memory and Cognition*, 9(4), 475-479.

- Roediger, H.L. & Karpicke, J.D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, 17, 249-255.

- Bertsch, S. et al. (2007). The generation effect: A meta-analytic review. *Memory & Cognition*, 35, 201-210.

Empirical Studies on AI in Education

- Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, Ö., & Mariman, R. (2024). Generative AI Can Harm Learning. *SSRN Working Paper*, The Wharton School.

- Kosmyna, N. et al. (2025). Your Brain on ChatGPT: Accumulation of Cognitive Debt when Using an AI Assistant for Essay Writing Task. *arXiv preprint*, arXiv:2506.08872.

- Darvishi, A., Khosravi, H., Sadiq, S., Gašević, D., & Siemens, G. (2024). Impact of AI assistance on student agency. *Computers & Education*, 210, 104967.

- Wang, J. & Fan, W. (2025). The effect of ChatGPT on students' learning performance, learning perception, and higher-order thinking: insights from a meta-analysis. *Humanities and Social Sciences Communications*, 12, Article 621.

Historical Analogues

- Sparrow, B., Liu, J., & Wegner, D.M. (2011). Google effects on memory: Cognitive consequences of having information at our fingertips. *Science*, 333, 776-778.

- Dahmani, L. & Bohbot, V.D. (2020). Habitual use of GPS negatively impacts spatial memory during self-guided navigation. *Scientific Reports*, 10, 6310.

Theoretical Frameworks

- Clark, A. & Chalmers, D. (1998). The extended mind. *Analysis*, 58(1), 7-19.

- Clark, A. (2025). Extending Minds with Generative AI. *Nature Communications*, 16, 4627.

- Risko, E.F. & Gilbert, S.J. (2016). Cognitive offloading. *Trends in Cognitive Sciences*, 20(9), 676-688.

- Grinschgl, S., Papenmeier, F., & Meyerhoff, H.S. (2021). Consequences of cognitive offloading: Boosting performance but diminishing memory. *Quarterly Journal of Experimental Psychology*, 74(9), 1477-1496.

- Gerlich, M. (2025). AI Tools in Society: Impacts on Cognitive Offloading and the Future of Critical Thinking. *Societies*, 15(1), 6.

Institutional Frameworks and Policy

- UNESCO (2023). Guidance for generative AI in education and research. Paris: UNESCO.

- EDUCAUSE (2024). 2024 EDUCAUSE Action Plan: AI Policies and Guidelines.

- Cornell Center for Teaching Innovation (2023). CU Committee Report: Generative Artificial Intelligence for Education and Pedagogy.

- Jose, B. et al. (2025). The cognitive paradox of AI in education: between enhancement and erosion. *Frontiers in Psychology*, 16, 1550621.