International Journal of Psychological and Brain Sciences
2026, Vol. 11, No. 4, pp. 92–99
https://doi.org/10.11648/j.ijpbs.20261104.12

Research Article

Adaptive Support Has Limits: AI-Supported Learning Assistance and the Risk of Never-Skilling

Roland Boettcher[1, 2, *]

[1]Department of Business and Economics, Bochum University of Applied Sciences, Bochum, Germany
[2]DigiTeach Institute for Digital Transformation, Bochum University of Applied Sciences, Bochum, Germany

Abstract

Artificial intelligence (AI)-enabled adaptivity promises to make academic learning support scalable across levels of achievement. This conceptual paper argues that educational value depends on what assistance leaves learners to do, rather than on its quantity alone. It develops a theory-informed synthesis of the assistance dilemma, cognitive load theory, self-regulated learning, and constructive alignment; it does not report a systematic review or an original empirical study. Adaptivity oriented towards challenge preserves a defined competence threshold while varying instructional scaffolding. Relief becomes problematic when it substitutes for the cognitive and metacognitive operations through which that competence is acquired. Randomised studies show that unrestricted and pedagogically constrained AI assistance can produce different learning outcomes, but these findings do not establish a long-term developmental deficit. Building on the term Never-Skilling introduced in medical education, the paper proposes a mechanism in which repeated substitution reduces opportunities for practice and may alter learners' willingness to invest effort. Never-Skilling is defined as persistently lower acquisition of a specified competence than under comparable instruction that preserves learner processing, assessed after AI support is withdrawn. It differs from deskilling, which concerns loss of established competence, and from cognitive offloading, a strategy that can support or impede learning. Assessment arrangements may conceal an acquisition deficit when AI-generated products count as evidence of independent competence. A further possibility, that assessment standards themselves adjust under institutional pressure, is presented as a separate, unconfirmed hypothesis. The paper specifies measurable constructs and four testable hypotheses concerning unsupported transfer, learner processing, assessment validity, and standards drift. Its design recommendation is to adapt routes to a clearly defined standard while checking retention and transfer without substitutive support. These propositions require longitudinal and context-sensitive validation; they do not imply that AI assistance necessarily prevents learning or that an initial deficit is irreversible.

Keywords

Self-Regulated Learning, Cognitive Load, Adaptivity, Generative AI, Higher Education Pedagogy, Constructive Alignment

*Correspondence: Roland Boettcher (roland.boettcher@hs-bochum.de)

Received: 27 August 2026; **Accepted:** 9 November 2026; **Published:** 24 November 2026

Copyright: © The Author(s), 2026. Published by Science Publishing Group. This is an **Open Access** article, distributed under the terms of the Creative Commons Attribution 4.0 License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.

1. The Promise of Adaptivity

The promise of artificial intelligence (AI)-enabled adaptivity is, first and foremost, a promise of equity. Whereas an instructor addressing two hundred students can necessarily accommodate only an average pace and level of challenge, individually scalable AI promises to meet learners where they are. It provides explanations on demand, tailored examples, immediate help with structuring, and assistance without the social inhibition of appearing ignorant before others. Lower-achieving students in particular are expected to benefit because they gain access to the individual support that anonymous mass education often denies them. Adaptivity thus appears not as a technical add-on but as a potential realisation of a pedagogical ideal that conventional university teaching has so far achieved only to a limited extent.

Early studies of generative AI show that assisted task performance and learning need not improve together. In a field experiment with nearly a thousand high-school mathematics students, unrestricted GPT-4 assistance improved practice performance, but performance after withdrawal was roughly 17 per cent below that of a control group; a safeguarded tutor avoided this disadvantage [1]. This is evidence about acquisition within a specific instructional setting, not a demonstration of persistent Never-Skilling in higher education. In a randomised writing study with 117 university students, Fan et al. [2] found greater essay-score improvement with ChatGPT but no significant between-group differences in knowledge gain, transfer, or post-task intrinsic motivation. These findings warrant separating product quality from learning; a null difference in knowledge gain does not by itself establish cognitive suppression. Survey evidence links greater AI use with cognitive offloading and lower critical-thinking scores, but its correlational, self-reported elements cannot establish causation or a developmental trajectory [3].

This tension is the paper's point of departure. At the level of assessed performance, the promise of adaptivity seems to deliver precisely what it threatens to undermine at the level of learning. One interpretation treats this merely as a calibration problem: too much help is being provided in the wrong place, and finer dosing would remedy the problem. The following analysis argues that this interpretation is insufficient. It assumes that support can in principle be increased without limit, provided that it is tailored precisely enough to individual learners. That assumption is questionable.

The central thesis is that additional assistance can become counterproductive when it replaces learning-relevant processing. The location of this limit depends on prior knowledge, task demands, timing, and instructional design. Unrestricted access may favour substitution, but its prevalence and long-term consequences must be established empirically. For learners who have not yet acquired a target competence, the relevant risk is reduced acquisition rather than loss of an established ability.

The argument proceeds in four steps. Section 2 shows that the issue is not adaptivity as such but the dimension to which adaptation is directed. Sections 3 and 4 develop the central mechanism: a possible shift towards substitution and the hypothesised coupling of processing and motivation through effort. Sections 5 and 6 examine how assessment can reveal or conceal acquisition gaps when assisted products are treated as evidence of independent competence. Section 7 situates the paper's argument within the emerging category of Never-Skilling before Section 8 formulates an instructional design principle: support should be scaled while the target level remains a constant minimum threshold.

This paper offers a theory-informed conceptual synthesis rather than a systematic literature review or an empirical evaluation. The literature is selected to connect instructional mechanisms, emerging experimental findings, and assessment conditions; no exhaustive search or pooled effect estimate is claimed. The analysis distinguishes three levels: findings established in the cited settings, a proposed mechanism connecting those findings, and hypotheses requiring new tests. Its unit of analysis is the acquisition of a specified competence under a specified arrangement of AI assistance. A study of short-term performance is treated as evidence relevant to this mechanism, not as proof of persistent non-acquisition.

2. The Dimension of Adaptation: Relief Versus Challenge

Adaptivity is not a unitary construct. Advocates and critics of AI-supported learning assistance often talk past one another because they have different forms of adaptation in mind. The decisive question is not merely *whether* teaching is adapted to individual learners but *what* the adaptation targets. Students unquestionably differ in prior knowledge, working speed, language proficiency, and metacognitive confidence, and undifferentiated teaching disadvantages those who need additional structure. This insight legitimises adaptivity, but it does not determine the direction in which adaptive systems should operate.

Aptitude-Treatment Interaction (ATI) [4] treats adaptivity as a match between learner characteristics and instruction. It supports attention to prior knowledge and task demands, rather than a universal quantity of help. The present argument adds a criterion for judging that match: whether learners retain responsibility for the operations needed to acquire the target competence. Supporting learners towards a common threshold and offering enrichment beyond it are compatible with this criterion.

A practical risk arises when adaptation is judged primarily by immediate completion or user satisfaction. In that case, a system may remove demands that are necessary for acquisition. This is a possible design bias rather than an established

property of all adaptive systems. Both additional guidance for novices and enrichment for advanced learners can be productive when they preserve appropriate learner processing.

A second distinction, central to this paper, cuts across conventional approaches to differentiation. These usually ask *by what* instruction is differentiated—content, pace, method, social arrangement, or medium. The relevant boundary here lies on another axis: does adaptation alter only the route to the competence threshold that must be warranted, or does it alter the threshold itself? Support can be adapted while the target level is maintained; alternatively, the task may be simplified until it no longer demonstrates the same competence. This axis distinguishes productive from harmful adaptivity more precisely than the mere question of whether differentiation occurred.

Adaptation can thus be distinguished by its functional direction. *Relief* reduces immediate demands; this is beneficial when it removes irrelevant complexity, language barriers, or inaccessible presentation. It becomes substitutive when it removes operations that constitute the learning objective. *Challenge* combines attainable demands with sufficient guidance to sustain learning-relevant processing. The distinction concerns the relation between support and target competence, not a general preference for more difficulty. A reduction in effort or task complexity is not, by itself, evidence of harmful adaptation.

The finding that quantitatively maximising support does not promote learning linearly is well established. The expertise-reversal effect shows that identical supports which benefit novices can impair advanced learners: what gives novices necessary structure forces experts into redundant reconciliation with established schemas and consumes limited working-memory capacity [5]. The effective quantity of help is therefore a dynamic function of expertise. This does not justify unlimited assistance for novices. Empirically supported methods such as worked examples explicitly require novices to follow the operational steps cognitively; they structure processing rather than replace it and thus remain on the side of challenge. If an AI system supplies the final product and the learner does not reconstruct or explain the operations, an opportunity for the processing that scaffolding is meant to support may be lost. The expertise-reversal effect supports structured guidance for novices, not their cognitive substitution.

Research on *desirable difficulties* distinguishes conditions that improve later retention and transfer from those that merely improve performance during practice [6]. Retrieval, spacing, and discrimination can be productive even when practice feels less fluent. Difficulty is desirable only when learners can engage with it and it serves acquisition. Removing unnecessary difficulty is therefore compatible with challenge-oriented adaptivity; removing the opportunity to practise a target operation may not be.

This is not a matter of pedagogical taste. Support preserves the requirement while changing the conditions under which learners address it; substitution replaces the work with a product created without the corresponding processing. In traditional settings, social, temporal, and institutional frictions often helped secure this boundary: learners had to make an attempt, formulate a question, attend office hours, or await feedback. Generative AI changes precisely these conditions.

The proposal complements rather than replaces existing frameworks. ATI addresses which instruction fits which learner [4]; the assistance dilemma addresses when information should be provided or withheld [7]; and cognitive load theory addresses the demands imposed on limited processing capacity [8]. Holstein et al.'s human-AI hybrid adaptivity framework explicitly includes goals, perception, action, decision policies, and timing, and distinguishes immediate performance augmentation from co-learning [9]. The present criterion asks whether a particular allocation of activity preserves acquisition of the specified competence. Constructive alignment then connects that criterion to what assessment actually warrants [10]. The proposed contribution is this connection between activity allocation, acquisition, and institutional visibility, not the discovery that assistance can be excessive.

3. The Assistance Dilemma Under Generative AI

Koedinger and Aleven's assistance dilemma describes the need to balance two instructional risks: too little guidance can leave learners at an unproductive impasse, while too much can anticipate processing they need to perform themselves. The appropriate balance varies with the learner, task, and timing [7]. Generative AI changes the accessibility and scope of help, but it does not remove this dilemma.

An unrestricted large language model (LLM) interface can supply a completed solution before a learner has attempted the task or diagnosed an impasse. When immediate completion is the interaction's objective, pedagogically useful withholding of information may be bypassed. However, this is an affordance of a particular configuration and use, not an invariant objective of every model or application. Teacher-designed tutors can delay answers, request explanations, and provide graduated hints.

Three affordances may increase the opportunity for substitution: low interaction costs, the ability to request assistance ranging from a hint to a complete product, and availability at the moment difficulty is encountered. Their combination can make immediate outsourcing attractive. Whether this leads to over-support depends on the interaction design and learners' practices; low-cost access can also make timely, well-calibrated guidance more widely available.

What is lost is not effort as an end in itself but pedagogically functional friction. It compelled three things:

1) an *initial attempt*, which alone reveals where understanding holds and where it ends;

2) the *formulation of a question*, which first makes the impasse diagnosable;
3) a *needs assessment*: pausing to ask whether help is necessary or another independent step remains possible.

These activities can contribute to learning when they are matched to the learner and task. Kapur's productive-failure study shows benefits of an instructional sequence that combines initial problem solving with subsequent consolidation [11]. It does not establish that unguided failure is always necessary or superior. The relevant implication is to preserve opportunities for generative processing where they are productive, while providing sufficient explanation and feedback.

Experimental findings support treating assistance design as consequential. Bastani et al. [1] compared unrestricted and safeguarded GPT-4-based assistance within one high-school setting; only unrestricted assistance produced a subsequent performance deficit. In a separate university-physics study, Kestin et al. [12] found greater immediate learning gains with a custom AI tutor than with an active-learning classroom. The latter comparison involved a different setting and instructional package and did not isolate the initial attempt as the sole causal feature. Together, the studies show that AI support need not have a uniform effect; they do not identify one universally sufficient safeguard or establish long-term acquisition trajectories.

4. The Dual Role of Effort

The proposed mechanism links the cognitive and motivational roles of effort. Learning requires appropriate processing, while willingness to undertake that processing depends partly on its perceived costs and returns. If a learner repeatedly obtains an assessed product without performing the target operations, opportunities for acquisition may decrease. Effort here means engagement in learning-relevant operations, not time spent, subjective strain, or difficulty for its own sake.

Cognitive load theory provides an important qualification [8, 13, 14]. Reducing extraneous demands can free capacity for learning; even a completed worked example can support acquisition when learners explain and reconstruct it. Harm is therefore hypothesised where assistance replaces the processing needed for the specified objective, not wherever it reduces load or supplies content. Product quality alone cannot show whether such processing occurred.

Two motivational pathways are plausible but require direct testing. First, when an assessed outcome can be obtained with little independent processing, learners may judge that processing to have low instrumental value. Second, an assisted product may alter perceived competence without a corresponding change in independent mastery. Self-determination theory highlights competence and autonomy as relevant motivational needs [15], but does not establish that AI-generated success necessarily satisfies them by proxy. Fan et al.'s absence of a post-task intrinsic-motivation difference [2] cautions against treating motivational decline as an established result.

The hypothesised feedback loop is that reduced practice limits acquisition, making later unsupported tasks harder; this difficulty may then increase reliance on substitution. Goal setting, monitoring, strategy selection, and self-evaluation are potential sites of this process because they are central to self-regulated learning (SRL) [16]. The reverse trajectory is also possible: appropriately scaffolded success may increase willingness and capacity to work independently. Longitudinal evidence is needed to discriminate these pathways.

The proposed coupling is therefore a candidate mechanism, not a demonstrated causal cycle. Cognitive processing, perceived effort, motivation, and independent competence must be measured separately. Reduced effort accompanied by equal or better delayed transfer would indicate greater efficiency rather than support the proposed acquisition deficit.

5. When the Proxy Displaces the Purpose

Whether conserving effort is rational depends entirely on the objective. If the priority is to *pass*, conserving effort is rational. If the objective is competence, whose development consists precisely in that effort, conservation defeats the purpose. The rationality of offloading therefore rests on a displacement of the objective embedded in the assessment logic students follow.

A grade is intended to indicate attainment of defined learning outcomes. When production of the assessable artefact becomes the operative goal, it can diverge from the competence it is meant to indicate. This risk is related to Goodhart's analysis of targeted measures [17] and Strathern's application to educational audit [18]. Generative AI can widen the gap when a credible product is accepted without sufficient evidence of the learner's own capabilities. Neither product quality nor AI use alone establishes the presence or absence of competence.

The visibility problem concerns assessment validity. Formative feedback can support planning, explanation, and revision, but a formative label does not guarantee active processing: a learner may also delegate practice tasks completely. Conversely, summative assessment can legitimately include AI use if the stated outcome concerns competent work with AI and the assessment verifies that outcome. For claims about independent competence, additional evidence of independent reasoning is needed. The distinction is thus between what the activity elicits and what the assessment claims to warrant, rather than a categorical division between formative and summative formats.

The proposed displacement arises when product completion becomes the operative objective without corresponding acquisition. The next section considers when assessment can expose this discrepancy and when institutional responses might conceal it.

6. The Assessment Regime as a Condition of the Effective Limit

Assessment arrangements can conceal or expose an acquisition deficit; they are not a necessary cause of every learning loss. A regime that accepts an AI-generated product as sufficient evidence of independent competence may fail to detect substitution. Supervised performance, probing dialogue, or independently completed transfer tasks can provide additional evidence, although none is automatically valid or reliable. Constructive alignment requires matching that evidence to the intended outcome [10]. The claim here concerns institutional visibility and incentives, while the acquisition mechanism can also operate outside graded work.

One possible safeguard is an AI-restricted assessment of independent performance. Such an assessment can reveal a discrepancy between assisted output and acquired competence if its demands and marking standard remain stable. The separate question is whether that reference point is maintained under institutional pressure.

The proposed endogenous-anchor mechanism is explicitly conditional: if growing failure rates lead institutions to reduce task demands or alter marking thresholds, an apparent recovery in pass rates may conceal a persistent acquisition gap. Completion incentives, resource constraints, and concerns about failure rates are possible pathways to investigate, not demonstrated causes in the present analysis. Dohmen and Krempkow [19] provide background on German higher-education financing, not evidence that AI causes standards to fall. Research on grade inflation in the United States and United Kingdom documents changing grade distributions and discusses possible explanations [20, 21]; it does not establish an AI-induced mechanism, a German pattern, or that every grade increase reflects weaker standards.

If the hypothesised adjustment occurs, two kinds of masking could coincide: an assisted grade may overstate individual independent competence, and a changing standard may obscure the acquisition gap at cohort level. Stable pass rates alone cannot distinguish this scenario from effective teaching or improved intake. Independent performance measures with stable scoring criteria are needed to test it.

A further question concerns learners' responses to difficulty. Monitoring errors, attributing them to modifiable strategies, and revising an approach are themselves aspects of SRL [16]. Some learners may respond to an unsupported setback by practising independently; others may increase substitution. Neither response should be assumed from failure rates or age alone. Repeated measures of strategy choice and attribution would be needed to examine this part of the proposed loop.

7. Never-Skilling

Cognitive offloading, deskilling, and Never-Skilling describe different analytical objects. Cognitive offloading is the use of external action or resources to reduce a task's internal processing demands [22]. It can be useful, can occur before expertise is acquired, and does not inherently imply deterioration. Deskilling concerns a decline in an established ability. Never-Skilling concerns reduced acquisition of an ability that was not yet established. Offloading may contribute to either outcome, or to neither; the strategy must therefore be distinguished from its developmental consequences.

For students who have not yet acquired a specified competence, a decline from earlier mastery is an inappropriate diagnostic criterion. Goal setting, metacognitive monitoring, and structuring unfamiliar problems can develop through guided and increasingly independent practice. When AI repeatedly performs those operations instead, their acquisition may be reduced. This is a hypothesis of non-emergence relative to an attainable learning trajectory, not a claim that a learner has no initial ability or cannot later learn.

Ke et al. [23] introduced Never-Skilling in medical education and distinguished it from deskilling and mis-skilling, the acquisition of flawed knowledge through uncritical acceptance of AI errors. Their framework proposes AI-independent baseline competence, critical calibration, and supervised integration. The present paper adopts their terminology rather than claiming to originate it. It extends the question to general higher-education competences and proposes an effort-mediated acquisition mechanism coupled with an assessment-visibility problem. Both extensions remain to be validated outside clinical training.

Operationally, Never-Skilling denotes persistently lower acquisition of a specified competence than under comparable instruction that preserves learner processing, observed across repeated assessments after substitutive support is withdrawn. A baseline measure is still possible and necessary: it estimates initial competence, including partial knowledge. What is counterfactual is the subsequent acquisition trajectory. A single weak post-test, frequent AI use, or absence of prior mastery is insufficient. The term does not imply irreversible failure or a neurological critical period.

The proposed mechanism draws on the view of SRL as a set of practices that can be developed through instruction, feedback, and increasingly independent activity [16]. It predicts risk when learners repeatedly delegate those practices without subsequently reconstructing or exercising them. Observational evidence of delegation is insufficient on its own; it must be connected to lower gains on independently assessed target operations.

Beginning students may have particular support needs because university study increases demands for independent regulation. General evidence on continuing development of executive functions [24] does not establish an age-specific vulnerability to AI-induced Never-Skilling. Age, prior knowledge, educational history, and initial SRL should therefore be examined separately. The trainability of SRL among university students [25] also argues against assuming a fixed developmental window or an acquisition deficit that higher

education cannot remedy.

8. Recommendations for Adaptive Support

This conceptual analysis does not imply that higher education should abandon adaptive technologies. It proposes a functional criterion for system design that requires empirical evaluation. The criterion is whether assistance preserves the processing needed for the target competence while reducing demands that do not serve that objective.

Three design implications follow:

1) First, instructional tasks must remain within the range of *desirable difficulties*: appropriate to the target competence and prior knowledge, with opportunities for explanation, retrieval, and transfer, while sufficiently scaffolded to prevent motivational withdrawal.
2) Second, pedagogically functional friction must be intentionally preserved in interaction design. The initial independent attempt, precise verbal formulation of the problem, and metacognitive assessment of need should remain the learner's responsibility because they are constitutive of learning rather than dysfunctional barriers to it.
3) Third, adaptive support should integrate content guidance with *metacognitive* prompts where these serve the learning objective. AI should help learners plan, monitor, and critically evaluate their own approach instead of algorithmically supplying the final answer whose independent production would have developed those control processes.

The principle requires a clear separation of instructional support (*Scaffolding*) and academic target level (*Target Competence*). Only the instructional path—the amount, mode, and duration of temporary scaffolds—adapts dynamically. The target remains invariant: a minimum threshold lowered for no one. Adaptive support brings weaker learners to that threshold; adaptive *Enrichment* takes stronger learners beyond it. Disadvantaged learners thus receive more support, more time, and finer-grained intermediate steps, but towards the same competence threshold rather than a reduced one.

Support that fades as competence grows can help learners assume responsibility for the target operations. Releasing a final product without opportunities for explanation or reconstruction may instead bypass practice. These are design expectations: the same example, hint, or feedback message can be productive or substitutive depending on how the learner engages with it.

Typical higher-education contexts illustrate the principle (Table 1). The comparison is illustrative, not exhaustive, and does not replace empirical evaluation.

Table 1. *Substitutive versus supportive AI use in typical higher-education contexts (illustrative).*

Learning context	Substitutive AI use	Supportive AI use
Writing	AI produces a finished essay	AI probes the structure of the argument
Problem solving	AI supplies every solution step	AI gives graduated hints after an initial attempt
Exam preparation	AI produces a summary without learner input	AI quizzes, diagnoses gaps, and demands justification
Feedback	AI rewrites the work and the learner accepts it without review	AI identifies an issue; the learner revises and explains the change

The recommendation distinguishes instructional support from the competence being certified. Where independent performance is the outcome, assessment should include evidence obtained without substitutive AI assistance, using stable criteria and appropriate conditions. Where effective work with AI is an explicit outcome, assessment may permit AI while examining verification, judgement, and accountable use. These outcomes can coexist in a programme but should not be conflated. Accessibility measures and appropriate accommodations should preserve access to a valid demonstration of the intended competence; standardisation should concern the inference being made rather than insistence on identical conditions for every learner. The effectiveness and equity of these arrangements require evaluation.

9. Limitations and Research Agenda

This paper contributes propositions rather than new causal evidence. Its selective conceptual synthesis may omit alternative explanations and does not estimate the prevalence or magnitude of Never-Skilling. Short-term experiments, self-reported use, and grade distributions cannot establish a persistent acquisition deficit. Four hypotheses make the proposed mechanism open to disconfirmation.

9.1. Measurable Constructs and Testable Hypotheses

The primary outcome should be change from baseline on

independently completed tasks aligned with a specified competence, followed by delayed retention and novel transfer assessments. SRL can be assessed through planning quality, confidence-accuracy calibration, error detection, and strategy revision, using both task traces and performance measures. Exposure should distinguish requested final solutions, graduated hints, and reconstructed explanations rather than simply count AI sessions. Learning-relevant processing can be coded from independent attempts, explanations, and revisions; time on task and self-reported effort are complementary indicators, not substitutes for evidence of processing.

H1 - Acquisition. Among learners with comparable initial competence, repeated access to substitutive final-answer assistance will produce smaller gains on delayed unsupported transfer than assistance that preserves target processing. Immediate product quality may nevertheless improve. Equal or better delayed transfer under substitution would challenge this prediction.

H2 - Processing and regulation. Differences in acquisition will be partly mediated by opportunities actually taken for explanation, retrieval, monitoring, and revision. The effect may vary with initial competence and SRL. These mediators should precede outcome measurement; motivation and perceived competence should be assessed repeatedly rather than inferred from product quality. A short-term reduction in strain accompanied by preserved transfer would not count as Never-Skilling.

H3 - Assessment visibility. Under product-only assessment, learners receiving substitutive assistance will show a larger discrepancy between assessed product quality and independent competence than under assessment that additionally samples independent reasoning. Measures should target the same learning outcome and use comparable scoring criteria. The prediction concerns validity of the competence inference, not a universal superiority of any one examination format.

H4 - Endogenous standards. If institutions respond to AI-related acquisition gaps by easing assessment demands, stable or improving pass rates will coexist with weakening performance on independently scored anchor tasks. Stable anchor performance, or stable task demands despite changing pass rates, would count against this explanation. Grade inflation alone is insufficient evidence; changes in intake, teaching, task difficulty, and grading must be examined separately.

9.2. Design and Boundary Conditions

A suitable test would compare unrestricted assistance, graduated assistance, and an appropriate instructional comparison with equivalent curriculum and access to teaching. Baseline, immediate, delayed, and transfer measures would distinguish initial differences, temporary dependence, and persistent differences in acquisition. Random assignment where feasible, or a justified longitudinal comparison with adjustment for initial differences, is preferable to a cross-sectional comparison of frequent and infrequent users. Mediation estimates would still require explicit causal assumptions. Attrition, outside AI access, accessibility needs, and unequal time on task require attention.

The institutional hypothesis needs a separate multi-cohort design with stable anchor tasks, independent scoring, and records of task and marking changes. The clinical framework of Ke et al. [23] and the general competences considered here should be tested in their own domains rather than assumed to share identical acquisition processes. Neither age nor a lack of initial mastery is a sufficient risk marker. SRL remains trainable at university [25]; remedial instruction and later acquisition remain possible. Findings may therefore support, restrict, or reject the proposed mechanism without implying an irreversible deficit.

Abbreviations

AI	Artificial Intelligence
ATI	Aptitude-Treatment Interaction
LLM	Large Language Model
SRL	Self-regulated Learning

Author Contributions

Roland Boettcher: Conceptualization, Methodology, Writing – original draft, Writing – review & editing

Conflicts of Interest

The author declares no conflicts of interest.

References

[1] 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*, 122(26), e2422633122. https://doi.org/10.1073/pnas.2422633122

[2] Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., Shen, Y., Li, X., & Gašević, D. (2025). Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. *British Journal of Educational Technology*, 56(2), 489-530. https://doi.org/10.1111/bjet.13544

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

[4] Cronbach, L. J., & Snow, R. E. (1977). Aptitudes and instructional methods: A handbook for research on interactions. Irvington.

[5] Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist*, 38(1), 23-31. https://doi.org/10.1207/S15326985EP3801_4

[6] 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. https://doi.org/10.1016/j.jarmac.2020.09.003

[7] Koedinger, K. R., & Aleven, V. (2007). Exploring the assistance dilemma in experiments with cognitive tutors. *Educational Psychology Review*, 19(3), 239-264. https://doi.org/10.1007/s10648-007-9049-0

[8] Sweller, J., van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review*, 31(2), 261-292. https://doi.org/10.1007/s10648-019-09465-5

[9] Holstein, K., Aleven, V., & Rummel, N. (2020). A conceptual framework for human-AI hybrid adaptivity in education. In I. I. Bittencourt, M. Cukurova, K. Muldner, R. Luckin, & E. Millán (Eds.), *Artificial Intelligence in Education* (Lecture Notes in Computer Science, Vol. 12163, pp. 240-254). Springer. https://doi.org/10.1007/978-3-030-52237-7_20

[10] Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education*, 32(3), 347-364. https://doi.org/10.1007/bf00138871

[11] Kapur, M. (2008). Productive failure. *Cognition and Instruction*, 26(3), 379-424. https://doi.org/10.1080/07370000802212669

[12] Kestin, G., Miller, K., Klales, A., Milbourne, T., & Ponti, G. (2025). AI tutoring outperforms in-class active learning: An RCT introducing a novel research-based design in an authentic educational setting. *Scientific Reports*, 15, 17458. https://doi.org/10.1038/s41598-025-97652-6

[13] Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257-285. https://doi.org/10.1207/s15516709cog1202_4

[14] Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review*, 22(2), 123-138. https://doi.org/10.1007/s10648-010-9128-5

[15] Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, 11(4), 227-268. https://doi.org/10.1207/s15327965pli1104_01

[16] Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory Into Practice*, 41(2), 64-70. https://doi.org/10.1207/s15430421tip4102_2

[17] Goodhart, C. A. E. (1984). Problems of monetary management: The U.K. experience. In *Monetary Theory and Practice* (pp. 91-121). Palgrave Macmillan. (Reprinted from 1975.) https://doi.org/10.1007/978-1-349-17295-5_4

[18] Strathern, M. (1997). "Improving ratings": Audit in the British university system. *European Review*, 5(3), 305-321. https://doi.org/10.1017/S1062798700002660

[19] Dohmen, D., & Krempkow, R. (2014). *Die Entwicklung der Hochschulfinanzierung - von 2000 bis 2025* [The development of higher-education financing from 2000 to 2025]. Konrad-Adenauer-Stiftung. https://www.kas.de/c/document_library/get_file?groupId=252038&uuid=b727520f-98ad-da23-c622-59660eb84fbd

[20] Rojstaczer, S., & Healy, C. (2012). Where A is ordinary: The evolution of American college and university grading, 1940-2009. *Teachers College Record*, 114(7), 1-23. https://doi.org/10.1177/016146811211400707

[21] Bachan, R. (2017). Grade inflation in UK higher education. *Studies in Higher Education*, 42(8), 1580-1600. https://doi.org/10.1080/03075079.2015.1019450

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

[23] Ke, Y., Jin, L., Ong, J. C. L., Thirunavukarasu, A. J., Car, J., Cheung, C. Y., Tham, Y. C., Ting, D. S. W., Ong, M. E. H., Compton, S., Narayan, A., Keane, P. A., Wong, T. Y., Bates, D. W., Tan, P., & Liu, N. (2026). AI-induced never-skilling in medical education. *Nature Medicine*, 32(6), 1997-2006. https://doi.org/10.1038/s41591-026-04438-y

[24] Arain, M., Haque, M., Johal, L., Mathur, P., Nel, W., Rais, A., Sandhu, R., & Sharma, S. (2013). Maturation of the adolescent brain. *Neuropsychiatric Disease and Treatment*, 9, 449-461. https://doi.org/10.2147/NDT.S39776

[25] Theobald, M. (2021). Self-regulated learning training programs enhance university students' academic performance, self-regulated learning strategies, and motivation: A meta-analysis. *Contemporary Educational Psychology*, 66, 101976. https://doi.org/10.1016/j.cedpsych.2021.101976