IIAR JOURNALS

Innovating Knowledge. Advancing Research.

IIAR Journal of Education, Learning, and Human Development

Vol 1, No 1, 2026

Generative AI and the Purpose of Schooling: A Theoretical Framework of Learning, Offloading and the Competencies That Remain Human

OGAH Sunday Ogbo[1], NWOGBOJI Emmanuel Usulor[2], IGWE Emmanuel Sunday[3], BAMIGBOLA Paul Idowu[4], ATOLAGBE Umar Kawu[5]

[1,3,4,5] Veristzon Global Research Lab; [2] International Institute of AI and Research

ABSTRACT

When a system can draft the essay, solve the problem set and explain the concept, the question for education is no longer how to stop students using it but what schooling is for. Debate has been dominated by academic integrity and short-term performance gains, and it lacks a theory explaining why the same tool can raise assisted performance while weakening unaided learning. The objective of this conceptual paper is to develop such a theory and derive testable propositions for curriculum, assessment and classroom practice. The population is the peer-reviewed literature on generative AI in education, the science of learning and cognitive offloading, from which we drew a sample of 33 journal articles: 31 published 2023–2026 and two foundational works cited for the theories they originated. Sources were selected by criterion-based (purposive) sampling, and every reference was verified against Crossref and screened for retractions. The analysis was a conceptual and thematic synthesis: studies were coded for what learner and system did and when learning was measured, the codes were read against learning mechanisms, and propositions were derived where theory and evidence converged. We distinguish three relations a learner can have with a generative system: learning with it, learning from it and offloading to it. Offloading, the default under general-purpose designs, bypasses the processes that build the knowledge and judgement on which the other two relations depend. Four competencies therefore become more rather than less important: evaluative judgement, domain knowledge sufficient to detect error, sustained unaided reasoning and the disposition to verify. Ten falsifiable propositions follow, among them unaided-first sequencing, validity-centred assessment and tool configurations that withhold answers. For African basic education systems, where generative AI is reaching learners faster than teacher preparation, we recommend protecting foundational learning, building teachers' evaluative capacity rather than tool fluency, treating supervised unaided assessment as an asset, and reducing the workload and time pressure that may push pupils towards offloading. The purpose of schooling under generative AI, we conclude, is to produce people who can tell when the machine is wrong.

Keywords: generative AI; curriculum; cognitive offloading; learning theory; competencies

Generative AI and the Purpose of Schooling: A Theoretical Framework of Learning, Offloading and the Competencies That Remain Human

1. INTRODUCTION

Earlier educational technologies left the student something to do: a calculator still required the problem to be set up, and a search engine left reading, comparison and synthesis to the learner. Generative AI differs in kind, producing the finished artefact schools have long used as evidence of learning: the essay, the worked solution, the explanation in the student's own words. Lodge et al. (2023) accordingly argue that the calculator is the wrong analogy. Once the product can be generated without the process, the link between what a student hands in and what the student has learned breaks.

Two responses have dominated: one treats the problem as integrity and asks how to detect misuse; the other sees opportunity in studies where AI-supported students outperform others. Meta-analyses do report gains in performance, motivation and higher-order thinking tendencies (Deng et al., 2025; Wu & Yu, 2024). Neither asks what the gains consist of. In the field experiment that has most shaped the debate, high-school mathematics students given unrestricted GPT-4 access during practice scored far higher while they had it, then lower than students never given access once it was withdrawn (Bastani et al., 2025). A performance gain, then, can also be a learning loss.

The difficulty is partly theoretical. Research on generative AI in education has grown fast, much of it short, small and university-based, and largely apart from two older bodies of work that bear on it: the science of learning, which explains why effortful processing builds durable knowledge (Bjork & Bjork, 2020; Sweller, 2023), and research on cognitive offloading, which explains when and why people hand mental work to external tools (Risko & Gilbert, 2016). Without a framework connecting them, studies reporting benefit and harm sit side by side unexplained.

The stakes are highest where learners can reach generative tools on phones before their teachers are prepared to guide that use. Ghanaian basic school teachers show only moderate AI literacy (Arkorful et al., 2025), and Namibian teachers doubt that adequate professional learning on AI is available (Jatileni et al., 2024). The relation pupils develop by default may, we infer, be the one they keep.

Our objective is a theoretical framework explaining how generative AI shapes learning and what schooling should therefore prioritise, giving curriculum and assessment policy a principled basis for interpreting conflicting evidence. We address four research questions. First, what distinct relations can a learner have with a generative system, and how do they bear on learning? Second, which human competencies does generative AI make more rather than less important, and how are they formed? Third, what testable propositions about curriculum, assessment and classroom practice follow? Fourth, what do these propositions imply for African basic education systems? Being conceptual, the paper advances testable propositions in place of tested hypotheses, each paired with a disconfirming result. Our answer to the question in the title is narrow and demanding: under generative AI, schooling exists to produce people who can tell when the machine is wrong.

2. LITERATURE REVIEW

2.1 CONCEPTUAL REVIEW

Generative AI here means large language models and related systems that produce text, code or explanations from prompts. Two properties matter for learning: outputs are fluent whether or not they are correct, and the process producing them cannot be inspected. Bearman and Ajjawi (2023) accordingly treat AI as a black box whose judgements cannot be traced, and argue that education should prepare learners to work with that opacity rather than dispel it.

Cognitive offloading is the use of physical action or external resources to reduce the cognitive demands of a task (Risko & Gilbert, 2016). It is often adaptive, but its educational significance depends on what is

© IIARJournal. All rights reserved.

offloaded and by whom. Children as young as four offload more when tasks are harder, but also offload too little or too much, with limited selectivity, and seldom on their own initiative (Armitage & Gilbert, 2025). Knowing when to offload is itself learned, and young learners judge it poorly.

Evaluative judgement is the capability to judge the quality of one's own and others' work. Bearman et al. (2024) argue that generative AI makes it urgent, because learners must now judge both AI outputs and the processes that produced them, and that assessment should help them calibrate their judgements of quality so that humans remain its arbiters.

Over-reliance, or AI dependency, is use in which the learner accepts outputs without the scrutiny the task requires. Position papers anticipated it: Kasneci et al. (2023) argued that learners would need new competencies to evaluate model outputs critically, and Yan et al. (2024) warned that generative AI could become a crutch that undermines intellectual abilities. Zhai et al. (2024), reviewing 14 studies, link it to weakened decision-making, critical thinking and analytical reasoning, and explain it as a preference for fast, convenient answers.

Put together, over-reliance is offloading the learner cannot justify, and evaluative judgement is what would allow the learner to justify or decline it. The framework turns on that relationship.

2.2 THEORETICAL REVIEW

Three bodies of theory explain why the learner's relation with a generative system matters more than the fact of use.

Cognitive load theory starts from the asymmetry between a severely limited working memory and an effectively unlimited long-term memory organised into schemas (Sweller, 2023). Learning, on this account, is the construction and automation of schemas. For generative AI, it follows that removing extraneous load helps, as when AI clarifies a confusing explanation, but removing the processing that constitutes schema construction does not, even though it feels like help, because nothing is then stored. Assistance therefore has different effects at different levels of expertise: what an expert can delegate, a novice may never have acquired.

The desirable-difficulties tradition sharpens the point. Conditions that slow apparent progress, such as retrieval, spacing, interleaving and generating one's own answers, tend to improve long-term retention and transfer, while conditions that make practice feel fluent often do not (Bjork & Bjork, 2020). The complication is metacognitive: learners read effort as a sign of poor learning and fluency as a sign of good learning, and so avoid the conditions that would serve them. De Bruin et al. (2023) show how perceived effort and perceived learning govern whether students start and persist with effortful strategies. Wang et al. (2023) found that teaching students about retrieval practice increased their knowledge of it and intention to use it, but that its perceived cost in time and effort weighed against actual use. Generative AI lowers perceived effort precisely when effort does the work of learning.

The third body of theory concerns the human–AI relation itself. Lodge et al. (2023) propose a typology on two dimensions, individual versus collaborative and offloading versus extending, yielding four relations: cognitive offloading, extended mind, co-regulation and hybrid learning. It rightly refuses a single metaphor but does not say which relation serves which learning goal or what each requires of the learner beforehand. Macnamara et al. (2024) supply part of the missing account, arguing that AI assistance may accelerate skill decay in experts and hinder skill acquisition in learners, and that in both cases the person may not notice. If the costs of assistance are invisible, self-regulation cannot be relied on to limit them.

Together these theories predict that the effect of generative AI depends on whether it removes extraneous load or constitutive processing, on prior knowledge, and on whether the learner perceives the difference. None, however, specifies the competencies schools should now prioritise, which is the gap this framework addresses.

2.3 EMPIRICAL REVIEW

The empirical literature is young and points two ways. On the positive side, a meta-analysis of 69 experimental studies associated ChatGPT use with better academic performance, greater motivation, stronger higher-order thinking tendencies and lower mental effort, with no change in self-efficacy (Deng

© IIARJournal. All rights reserved.

et al., 2025). An earlier meta-analysis of 24 randomised studies of AI chatbots found a large overall effect on learning outcomes (Wu & Yu, 2024). In a randomised trial in undergraduate physics, students learned more in less time from a custom AI tutor built on research-based pedagogical principles than from in-class active learning (Kestin et al., 2025). ChatGPT-generated hints for mathematics problems produced learning gains comparable to human tutors' hints (Pardos & Bhandari, 2024).

The same studies carry warnings that their headlines obscure. Wu and Yu (2024) found effects larger in higher education than in primary and secondary education, and larger for short interventions than long ones, a pattern they attribute to novelty. Deng et al. (2025) themselves criticise the evidence's quality, and reduced mental effort is a benefit only if the effort removed was extraneous. The Kestin et al. (2025) tutor was built to walk students through each problem step by step, and its authors credit that design for the gains. And about a third of the ChatGPT-generated help in Pardos and Bhandari (2024) failed quality checks before an error-reduction procedure. A learner unable to notice such errors would have learned from them.

The evidence of harm is more direct about mechanism. Nearly a thousand Turkish high-school students practised mathematics with a standard GPT-4 interface or a version prompted to safeguard learning (Bastani et al., 2025); both raised practice performance, by 48% and 127% respectively. When access was withdrawn, students who had used the standard interface scored 17% lower than controls who never had it, while the safeguarded version largely avoided the loss. Darvishi et al. (2024), in a randomised experiment with 1,625 students across ten courses, found that students tended to rely on AI assistance in peer review rather than learn from it, and that when it was removed, self-regulation checklists only partly compensated. Stadler et al. (2024) assigned 91 university students to research a socio-scientific question with ChatGPT or a conventional search engine; the ChatGPT group reported lower cognitive load and produced weaker reasoning and argumentation. In a randomised study of 117 students, Fan et al. (2025) found that ChatGPT support improved essays more than human-expert or analytics support without improving knowledge gain or transfer, and that the groups' self-regulated learning processes differed.

Metacognitive evidence adds a layer. Across two studies of adults answering logical-reasoning items, Fernandes et al. (2026) found that ChatGPT raised scores but participants substantially overestimated their performance. Their confidence barely distinguished right answers from wrong, and higher AI literacy went with more confidence but less accurate self-assessment. In a Nigerian sample of 529 pre-service teachers, Ayanwale et al. (2024) found no support for the assumption that active AI use improves the ability to detect AI. Familiarity is not judgement, and the same holds for teachers: Celik (2023) found that technological knowledge alone does not yield sound pedagogical integration of AI.

Correlational studies agree but deserve less weight. Gerlich (2025) reported, among 666 participants, that frequent AI use was associated with weaker critical thinking, with cognitive offloading as a mediator and younger users the most reliant. Abbas et al. (2024), in two surveys of university students in Pakistan (one of them three-wave), found that workload and time pressure predicted ChatGPT use, and that use predicted procrastination, self-reported memory loss and lower grades. Zhang et al. (2024) linked dependency to academic stress and performance expectations. These designs cannot establish direction but are consistent with pressure increasing use.

Most studies are short, involve university students in wealthier countries, and measure performance during assistance, not learning afterwards. Very few follow learners after withdrawal, the only design that separates learning from assisted performance, and those that do (Bastani et al., 2025; Darvishi et al., 2024) find costs. That gap supports the argument of Dawson et al. (2024) that validity, not cheating, is the central problem of assessment.

3. MATERIALS AND METHODS

3.1 RESEARCH DESIGN

This theory-building conceptual paper reports no new data. It brings theories from the science of learning to bear on findings about generative AI,

© IIARJournal. All rights reserved.

states the resulting account as a framework, and expresses its commitments as testable propositions.

3.2 POPULATION AND SAMPLE

The population is the peer-reviewed literature on generative AI in education, the science of learning and cognitive offloading, from which sources were selected purposively against the criteria in Section 3.3. The sample comprises 33 peer-reviewed journal articles: 31 published 2023–2026 and two foundational works (Risko & Gilbert, 2016; Bjork & Bjork, 2020). Six come from African settings and two from other middle-income settings (Pakistan and Türkiye).

3.3 INCLUSION AND EXCLUSION CRITERIA

Inclusion required a peer-reviewed journal article from 2023–2026 with a working DOI and a matching Crossref journal-article record (title, authors, journal, volume, issue, pagination); only the foundational works fall outside the date window. Excluded were preprints, conference papers, book chapters and reports, among them a preprint on an AI mathematics tutor in Ghana; a 2025 ChatGPT meta-analysis retracted in 2026; a source lacking a retrievable Crossref record; and sources whose abstracts did not confirm the claim they would have supported.

3.4 DATA COLLECTION PROCEDURE AND TOOLS

Searches ran in October 2026 through OpenAlex and general web search, combining terms for generative AI, ChatGPT and chatbots with learning outcomes, over-reliance, metacognition, cognitive load, critical thinking, assessment, teachers, AI literacy and African country names. Cognitive offloading, desirable difficulties and retrieval practice were searched without reference to AI. Metadata, corrections and retraction notices came from the Crossref API, retraction flags and abstracts from OpenAlex, and missing abstracts from publisher pages.

3.5 DATA ANALYSIS

First, each empirical study was coded for what the learner did, what the system did, whether learning was measured during or after assistance, and who the learners were. A second pass read the codes against the mechanisms of cognitive load theory, desirable difficulties and offloading research; the with–from–offloading distinction emerged as the smallest set of categories separating studies reporting durable gains from those reporting hidden costs. A third wrote propositions wherever theory and at least one empirical finding converged.

3.6 BASES OF JUDGEMENT

Judgement rests on theoretical coherence and empirical plausibility, not statistical aggregation. Randomised and withdrawal designs carry more weight than correlational and self-report findings; where evidence conflicts, a proposition states the conditions expected to produce each outcome.

3.7 RELIABILITY AND VALIDITY

Every reference was verified against Crossref twice, during drafting and again independently, and screened for retractions, corrections and expressions of concern; two carry published corrections and none is retracted. Each citation was independently re-checked against its source abstract and overstated claims narrowed; an independent review checked consistency across sections; and every proposition carries a disconfirming result (Table 3). Limits remain: the sources are mostly short, concentrated in higher education and high-income settings, and partly self-reported, and the African implications are inferences from surveys, a scoping review and general mechanisms, since no classroom trial in African schools met the criteria.

3.8 VARIABLE MEASUREMENT

No variables are measured. The framework's dimension, the location of the target process, is defined by who performs the cognitive work a task is designed to exercise (Table 1); the four competencies are defined in Table 2. The propositions specify measurable indicators (Table 3): unaided performance after withdrawal, delayed retention and transfer, detection of seeded errors, confidence–accuracy gaps, correlations between unsupervised products and supervised scores, and offloading under varying workload.

3.9 AVAILABILITY OF MATERIALS

The reference list with DOIs is in Section 7; the verification log and coding notes are available from the corresponding author on request, without restriction.

© IIARJournal. All rights reserved.

3.10 AI USAGE DISCLOSURE

Generative AI tools assisted literature retrieval, drafting and formatting under the authors' direction; the authors verified all sources against publisher and Crossref records and take full responsibility for the content.

3.11 OTHER DISCLOSURES

With no human participants, ethical approval was not required. There was no external funding and there are no conflicts of interest.

4. RESULTS AND DISCUSSION

4.1 RESULTS

The with–from–offloading framework. The synthesis yields three relations (Table 1), distinguished not by tool or prompt but by where the cognitive work the task was designed to exercise takes place. We call that work the *target process*: retrieving a fact, constructing an argument, setting up an equation, judging which source to trust.

In *learning with* AI, the learner keeps the target process and the system acts on it from outside, questioning, critiquing, offering counter-arguments or checking what the learner has produced. This corresponds to the co-regulation and hybrid relations of Lodge et al. (2023). In *learning from* AI, the system supplies explanations, worked examples or feedback that the learner must process, so the target process is shifted rather than surrendered: the learner studies rather than produces. Cognitive load theory predicts that this helps novices when the material is accurate and well sequenced (Sweller, 2023), and evidence on AI hints agrees (Pardos & Bhandari, 2024; for a tutor combining both relations, Kestin et al., 2025). In *offloading to* AI, the system performs the target process and the learner receives the product. Offloading is efficient and frequently rational (Risko & Gilbert, 2016), and becomes an educational problem only when the offloaded process is the one the learner was meant to acquire.

The central claim of the framework is asymmetric. Learning with and learning from both depend on capacities the learner brings: domain knowledge to judge an explanation, evaluative judgement to recognise a weak argument, persistence to keep thinking when the answer is one prompt away. They are built by doing the target process unaided, which is what offloading skips, so offloading practised as a default erodes the preconditions of the other two relations. General-purpose interfaces make offloading the path of least resistance, which learners take because it feels like learning (Bjork & Bjork, 2020; de Bruin et al., 2023).

The boundaries blur in three ways, and the blurring is part of the theory. One interaction can drift between relations, as when a request for an explanation becomes a request for the answer. The same request can fall into different relations, because what counts as the target process depends on the learning goal: asking for a grammatical correction is offloading in a grammar lesson and harmless in a history essay. And the relation depends on expertise: an expert who offloads a routine step risks gradual, unnoticed decay of a skill already held, whereas a novice who offloads the same step may never acquire it (Macnamara et al., 2024). The framework is therefore a tool for analysing tasks, not for classifying behaviours.

Table. 1. Three relations between learner and generative AI, defined by the location of the target process

Relation	Who performs the target process	Typical form	Learning mechanism engaged or bypassed	Illustrative evidence
Learning with	Learner; AI acts on the learner's work	AI questions, critiques or checks a learner-produced answer	Engages generation, retrieval and self-explanation; adds external regulation	Safeguarded tutor limited post-withdrawal loss (Bastani et al., 2025); step-by-step tutor spanning learning with and from outperformed active learning (Kestin et al., 2025)

© *IIARJournal. All rights reserved.*

Relation	Who performs the target process	Typical form	Learning mechanism engaged or bypassed	Illustrative evidence
Learning from	Shifted: AI explains, learner studies and processes	Explanations, worked examples, hints, feedback	Engages schema construction if content is accurate and processed; risks fluency illusion	AI-generated hints matched tutor-authored hints, but a third failed quality checks before correction (Pardos & Bhandari, 2024)
Offloading to	AI; learner receives the product	Complete answers, essays, solutions	Bypasses retrieval, generation and effortful processing; weakens calibration	Gains reversed after withdrawal (Bastani et al., 2025); reliance rather than learning (Darvishi et al., 2024); better essays without better transfer (Fan et al., 2025)

Competencies that generative AI makes more important. When competent-looking work is available on demand, the scarce human contribution is judging whether it is right. Four competencies follow, best understood as a dependent structure (Table 2).

Evaluative judgement, judging quality against standards, is the competency the others serve. It requires an explicit sense of what good work looks like and how AI processes go wrong (Bearman & Ajjawi, 2023; Bearman et al., 2024), but it draws on *domain knowledge sufficient to detect error*, which is the foundation of the structure. Errors cannot be recognised without a schema against which they register (Sweller, 2023), and fluent outputs are wrong often enough for this to matter (Pardos & Bhandari, 2024). *Sustained unaided reasoning*, holding a line of thought through difficulty without external completion, is both the means by which knowledge is built and the faculty offloading leaves unpractised (Bastani et al., 2025; Stadler et al., 2024). The *disposition to verify*, a habitual reluctance to accept fluent output unchecked, is motivational rather than cognitive. It activates the other three; without it, knowledge and judgement sit unused, as the uniform overconfidence reported by Fernandes et al. (2026) suggests.

None of these is new. What has changed is their status: once by-products of schoolwork, they must now be its deliberate objects, because the schoolwork can be done without them.

Table. 2. Competencies that generative AI makes more important

Competency	What it is	Why generative AI raises its importance	How it is formed	How default offloading erodes it
Evaluative judgement	Capability to judge the quality of one's own and others' work, including AI output and process	Humans must remain arbiters of quality	Comparing work against standards, feedback, critique of exemplars	Learners accept products they never judged
Domain knowledge sufficient to detect error	Organised schemas in a subject that let errors register as errors; the foundation of the other three	Outputs are fluent whether or not correct	Retrieval, practice and effortful processing over time	Removes the processing through which schemas are built
Sustained unaided reasoning	Holding and developing a line of thought without external completion	The answer is always one prompt away	Productive struggle and desirable difficulties	Effort is avoided at the point where it does its work

© IIARJournal. All rights reserved.

Competency	What it is	Why generative AI raises its importance	How it is formed	How default offloading erodes it
Disposition to verify	Habitual reluctance to accept output without checking	Overconfidence accompanies AI-assisted work	Norms, modelling, routines that make checking normal	Fluency is read as correctness, and checking feels unnecessary

Propositions. Ten propositions follow, each with a disconfirming result (Table 3).

Curriculum. **P1.** Learners who offload the target process to generative AI during practice will perform worse on later unaided assessment of it than learners who practise without AI, even when their assisted performance was higher. Withdrawal designs have produced this pattern (Bastani et al., 2025; Darvishi et al., 2024). **P2.** The benefit of learning with and from generative AI will be moderated by prior domain knowledge: below a threshold, AI use will raise the quality of products without raising learners' ability to detect errors in them. Errors register only against schemas (Sweller, 2023), and fluent errors are common (Pardos & Bhandari, 2024). **P3.** In basic education, curricula that secure foundational literacy, numeracy and core domain knowledge before introducing open-ended generative tools will produce stronger evaluative judgement later than curricula that introduce them early. Children's offloading shows limited selectivity (Armitage & Gilbert, 2025), and chatbot effects appear weaker in primary and secondary education than in higher education (Wu & Yu, 2024).

Assessment. **P4.** Where unsupervised products may be generated with unrestricted AI, their scores will correlate less strongly with supervised unaided measures of the same construct; pairing them with supervised or oral components will restore the relationship. This is the empirical form of the validity argument of Dawson et al. (2024). **P5.** Tasks requiring learners to find and correct seeded errors in AI output will develop evaluative judgement more than tasks requiring only production of work with AI. Judgement develops as learners calibrate their judgements of quality (Bearman et al., 2024). **P6.** Requiring learners to commit to an answer and a confidence rating before consulting AI, and then to compare, will reduce the overconfidence associated with AI-assisted work. The predicted mechanism is restoring a performance signal that assisted fluency suppresses (Fernandes et al., 2026; see also the altered self-regulation reported by Fan et al., 2025).

Classroom practice. **P7.** An unaided attempt before AI access will yield better retention and transfer than AI-first sequencing, at the cost of lower immediate performance and higher perceived effort. Students are predicted to misjudge those costs as signs of poor learning (Bjork & Bjork, 2020; de Bruin et al., 2023). **P8.** Tools configured to withhold final answers and prompt learner generation will produce smaller withdrawal deficits than general-purpose chat interfaces. The safeguarded tutor largely avoided the withdrawal loss (Bastani et al., 2025); answer-withholding is the predicted active ingredient. **P9.** Offloading will increase with workload, time pressure and the stakes attached to products, independently of attitudes towards AI. Pressure predicted use in survey studies (Abbas et al., 2024; Zhang et al., 2024), and the perceived cost of effortful strategies is negatively associated with their use (Wang et al., 2023). **P10.** Where teachers' AI literacy and evaluative capacity are low, pupils' use of generative AI will default to offloading, and professional development that builds teachers' own evaluative judgement of AI output will move classroom use towards learning with and from more than training in tool operation. Teacher AI literacy and confidence in professional learning are limited in the Ghanaian and Namibian surveys (Arkorful et al., 2025; Jatileni et al., 2024), and technological knowledge alone does not produce sound pedagogical integration (Celik, 2023).

Table. 3. Propositions and the evidence that would count against them

No.	Domain	Proposition (short form)	Disconfirming result
P1	Curriculum	Offloading during practice lowers later unaided performance	No post-withdrawal deficit relative to unaided practice

© IIARJournal. All rights reserved.

No.	Domain	Proposition (short form)	Disconfirming result
P2	Curriculum	Prior knowledge moderates benefit; below threshold, products improve but error detection does not	Low-knowledge AI users detect seeded errors as well as high-knowledge users
P3	Curriculum	Foundations before open tools yields stronger later judgement	Early-introduction cohorts equal or exceed later cohorts on error-detection tasks
P4	Assessment	Unrestricted AI weakens product–supervised score correlation; mixed designs restore it	Correlations unchanged under unrestricted AI
P5	Assessment	Seeded-error critique builds evaluative judgement more than AI-assisted production	No difference in judgement measures between task types
P6	Assessment	Predict-then-compare reduces AI-related overconfidence	Calibration gap unchanged by commitment before consultation
P7	Classroom	Unaided-first sequencing improves retention and transfer	AI-first sequencing equal or better on delayed tests
P8	Classroom	Answer-withholding configurations reduce withdrawal deficits	Deficits equal across configurations
P9	Classroom	Workload, time pressure and stakes increase offloading	Offloading unrelated to pressure once attitudes are controlled
P10	Classroom and system	Low teacher evaluative capacity → pupil offloading; evaluative professional development shifts use towards with and from more than tool training	Pupil offloading unrelated to teacher capacity, or evaluative professional development no more effective than tool-operation training

4.2 DISCUSSION

The synthesis answers the Introduction's four research questions. On the first, a learner can learn with, learn from or offload to a generative system; the relations differ in where the target process takes place, and the first two depend on capacities the third leaves unbuilt. On the second, four competencies gain importance, all formed through effortful unaided work and resting on domain knowledge. The propositions in Table 3 answer the third; the fourth is taken up below. No proposition is accepted or rejected here; they are offered for testing through withdrawal designs with delayed unaided tests (P1, P7, P8), seeded-error and calibration tasks (P2, P5, P6), cohort comparisons (P3), correlations between assessment components (P4), and studies of offloading under varying workload and teacher capacity (P9, P10).

Read against the empirical review, the framework reconciles apparently contradictory findings. Positive effects (Deng et al., 2025; Wu & Yu, 2024) come largely from studies measuring performance while the tool is present or using systems designed to keep the learner generating, negative findings mostly from withdrawal designs and general-purpose tools (Bastani et al., 2025; Darvishi et al., 2024; Fan et al., 2025). Both can be correct: they measure different relations, and sometimes assisted performance is called learning.

As for the theories in Section 2.2, the framework is consistent with cognitive load theory and the desirable-difficulties account, applying their prediction that removing constitutive processing impairs learning to a tool that removes it on request. Where the typology of Lodge et al. (2023) maps relations by dimension, the framework ranks them by dependency. It extends the account of invisible skill decay by Macnamara et al. (2024) into schooling,

© IIARJournal. All rights reserved.

where the learner has no prior skill to decay and the loss is of skills never acquired. And it gives the evaluative-judgement agenda (Bearman et al., 2024) a developmental floor. Judgement of AI output cannot be taught in the abstract to learners lacking the domain knowledge against which it is judged. Here we depart from versions of AI literacy centred on tool operation: the findings that AI literacy went with greater confidence but less accurate self-assessment (Fernandes et al., 2026) and that active use was not found to improve the ability to detect AI (Ayanwale et al., 2024) suggest that familiarity can mask a deficit in judgement.

One objection is that the argument defends drudgery against a technology that could free learners for higher-order work. The objection assumes higher-order work can proceed without the knowledge it operates on, which cognitive load theory denies (Sweller, 2023); the framework opposes not AI use but the default relation in which the learner's thinking is replaced. Another is that well-designed tutors make the problem disappear. They may reduce it (Bastani et al., 2025; Kestin et al., 2025), but learners do not use only such tutors, and general-purpose systems are one tab away.

Implications for African basic education. The fourth question concerns systems where teacher preparation lags the technology; the evidence there comes mostly from surveys of teachers, pre-service teachers and university students. In Ghana, basic school teachers' AI literacy was only moderate in knowledge, application, evaluation and ethics (Arkorful et al., 2025). In Namibia, in-service teachers' intention to teach AI depended on perceived relevance, attitude, social good and confidence rather than on readiness (Jatileni et al., 2024). At a South African university, pre-service science teachers' intentions to use generative AI depended on their AI literacy, perceived usefulness and perceived control (Ramnarain et al., 2025). South Africa's introduction of Coding and Robotics into schools has run ahead of teacher training and infrastructure, and a scoping review flags over-dependence alongside equity among the risks of using generative AI to fill the gap (Tshidi & Dewa, 2024). University students' attitudes towards generative tools differ significantly across Nigeria, Kenya and South Africa (Oyelere & Aruleba, 2025), which cautions against continental generalisation.

Four implications follow. First, foundational literacy, numeracy and core subject knowledge are the precondition of productive AI use, not a competitor for curriculum time (P2, P3); where many pupils have not secured them, early unstructured exposure risks producing fluent products over unformed knowledge. Second, where professional development budgets are thin, teacher preparation should prioritise teachers' own evaluative judgement of AI output in their subjects over operational tool training (P10; Celik, 2023). Third, the supervised, unaided examinations still central in many African systems, often criticised as narrow, gain new value as guarantors of validity (P4). Fourth, large classes, high stakes and time pressure are widely reported features of many African classrooms, although none of the sources reviewed here measures them. Survey evidence from university students elsewhere links pressure with greater use (Abbas et al., 2024), as P9 predicts. Adding AI to overloaded classrooms without changing workload may produce the relation policy least wants.

Where qualified teachers are scarce, however, accurate explanations and answer-withholding tutors on basic phones could widen access to learning from without the withdrawal deficits of general-purpose tools, if P8 holds. But access does not decide the relation; teachers and the design of tools and tasks do.

Future research should test the propositions in withdrawal designs of realistic duration, with school-age learners and in African classrooms, where evidence is thinnest; observe the location of the target process directly rather than infer it from the tool; and test whether the four competencies form the proposed dependent structure.

5. CONCLUSION AND RECOMMENDATION

The question generative AI puts to schooling is what relation to it schooling will cultivate. Because the productive relations depend on knowledge, judgement and persistence that only unaided work builds, a system that lets offloading become the default undermines its capacity to use the technology well. The competencies that remain human let a person stand outside the machine's output and assess it.

© IIARJournal. All rights reserved.

Curriculum authorities should make the four competencies explicit curriculum aims and sequence open-ended generative tools after foundational learning rather than alongside it in the early grades. *Examination and assessment bodies* should audit assessments for validity under AI rather than detectability, retaining supervised unaided components and redesigning take-home school-based assessment around staged, in-class and oral evidence. *Teacher education institutions and ministries*, particularly in African systems, should direct scarce professional development towards teachers' subject-specific evaluation of AI output, using seeded-error tasks they can reuse with pupils, rather than interface training, and ministries should treat workload and assessment stakes as levers on pupils' AI use (P9). *Teachers* can adopt an attempt-then-consult routine, ask learners to commit to an answer and confidence before consulting AI, and set tasks requiring pupils to find what the machine got wrong. *Developers and procurement agencies* should prefer configurations that withhold final answers and prompt generation, and require evidence of learning after withdrawal, not just engagement or assisted performance.

6. CONTRIBUTION TO EXISTING KNOWLEDGE

We make three contributions. Theoretically, the with–from–offloading framework defines learner–AI relations by the location of the target process, which helps explain why studies of one technology report both gains and losses, and we give a dependent account of the four competencies generative AI makes more important, with domain knowledge as their foundation. Practically, we translate the framework into curriculum, assessment and classroom designs, with attention to African basic education. For research, we set out ten falsifiable propositions and the results that would count against them. The framework rests on a young evidence base, mostly from higher education outside Africa, and its value will be settled by whether the propositions survive testing.

7. REFERENCES

Abbas, M., Jam, F. A., & Khan, T. I. (2024). Is it harmful or helpful? Examining the causes and consequences of generative AI usage among university students. *International Journal of Educational Technology in Higher Education*, *21*(1), Article 10. https://doi.org/10.1186/s41239-024-00444-7

Arkorful, V., Arthur, F., Boateng, E., Ofosu-Koranteng, M., Salifu, I., Attom, E. R., Tetteh, S. A., Quayson, E., Asare-Bediako, S., & Nortey, S. A. (2025). Exploring artificial intelligence literacy among basic school teachers in Ghana. *Discover Education*, *4*(1), Article 250. https://doi.org/10.1007/s44217-025-00630-3

Armitage, K. L., & Gilbert, S. J. (2025). The nature and development of cognitive offloading in children. *Child Development Perspectives*, *19*(2), 108–115. https://doi.org/10.1111/cdep.12532

Ayanwale, M. A., Adelana, O. P., Molefi, R. R., Adeeko, O., & Ishola, A. M. (2024). Examining artificial intelligence literacy among pre-service teachers for future classrooms. *Computers and Education Open*, *6*, Article 100179. https://doi.org/10.1016/j.caeo.2024.100179

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), Article e2422633122. https://doi.org/10.1073/pnas.2422633122

Bearman, M., & Ajjawi, R. (2023). Learning to work with the black box: Pedagogy for a world with artificial intelligence. *British Journal of Educational Technology*, *54*(5), 1160–1173. https://doi.org/10.1111/bjet.13337

Bearman, M., Tai, J., Dawson, P., Boud, D., & Ajjawi, R. (2024). Developing evaluative judgement for a time of generative artificial intelligence. *Assessment & Evaluation in Higher Education*, *49*(6), 893–905. https://doi.org/10.1080/02602938.2024.2335321

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

Celik, I. (2023). Towards Intelligent-TPACK: An empirical study on teachers' professional

© IIARJournal. All rights reserved.

knowledge to ethically integrate artificial intelligence (AI)-based tools into education. *Computers in Human Behavior*, *138*, Article 107468. https://doi.org/10.1016/j.chb.2022.107468

Darvishi, A., Khosravi, H., Sadiq, S., Gašević, D., & Siemens, G. (2024). Impact of AI assistance on student agency. *Computers & Education*, *210*, Article 104967. https://doi.org/10.1016/j.compedu.2023.104967

Dawson, P., Bearman, M., Dollinger, M., & Boud, D. (2024). Validity matters more than cheating. *Assessment & Evaluation in Higher Education*, *49*(7), 1005–1016. https://doi.org/10.1080/02602938.2024.2386662

de Bruin, A. B. H., Biwer, F., Hui, L., Onan, E., David, L., & Wiradhany, W. (2023). Worth the effort: The start and stick to desirable difficulties (S2D2) framework. *Educational Psychology Review*, *35*(2), Article 41. https://doi.org/10.1007/s10648-023-09766-w

Deng, R., Jiang, M., Yu, X., Lu, Y., & Liu, S. (2025). Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. *Computers & Education*, *227*, Article 105224. https://doi.org/10.1016/j.compedu.2024.105224

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

Fernandes, D., Villa, S., Nicholls, S., Haavisto, O., Buschek, D., Schmidt, A., Kosch, T., Shen, C., & Welsch, R. (2026). AI makes you smarter but none the wiser: The disconnect between performance and metacognition. *Computers in Human Behavior*, *175*, Article 108779. https://doi.org/10.1016/j.chb.2025.108779

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

Jatileni, C. N., Sanusi, I. T., Olaleye, S. A., Ayanwale, M. A., Agbo, F. J., & Oyelere, P. B. (2024). Artificial intelligence in compulsory level of education: Perspectives from Namibian in-service teachers. *Education and Information Technologies*, *29*(10), 12569–12596. https://doi.org/10.1007/s10639-023-12341-z

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., . . . Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, *103*, Article 102274. https://doi.org/10.1016/j.lindif.2023.102274

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*(1), Article 17458. https://doi.org/10.1038/s41598-025-97652-6

Lodge, J. M., Yang, S., Furze, L., & Dawson, P. (2023). It's not like a calculator, so what is the relationship between learners and generative artificial intelligence? *Learning: Research and Practice*, *9*(2), 117–124. https://doi.org/10.1080/23735082.2023.2261106

Macnamara, B. N., Berber, I., Çavuşoğlu, M. C., Krupinski, E. A., Nallapareddy, N., Nelson, N. E., Smith, P. J., Wilson-Delfosse, A. L., & Ray, S. (2024). Does using artificial intelligence assistance accelerate skill decay and hinder skill development without performers' awareness? *Cognitive Research: Principles and Implications*, *9*(1), Article 46. https://doi.org/10.1186/s41235-024-00572-8

Oyelere, S. S., & Aruleba, K. (2025). A comparative study of student perceptions on generative AI in programming education across Sub-Saharan Africa. *Computers and Education Open*, *8*, Article 100245. https://doi.org/10.1016/j.caeo.2025.100245

© *IIARJournal. All rights reserved.*

Pardos, Z. A., & Bhandari, S. (2024). ChatGPT-generated help produces learning gains equivalent to human tutor-authored help on mathematics skills. *PLOS ONE*, *19*(5), Article e0304013. https://doi.org/10.1371/journal.pone.0304013

Ramnarain, U., Ogegbo, A. A., Penn, M., Ojetunde, S., & Mdlalose, N. (2025). Pre-service science teachers' intention to use generative artificial intelligence in inquiry-based teaching. *Journal of Science Education and Technology*, *34*(6), 1272–1285. https://doi.org/10.1007/s10956-024-10159-z

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

Stadler, M., Bannert, M., & Sailer, M. (2024). Cognitive ease at a cost: LLMs reduce mental effort but compromise depth in student scientific inquiry. *Computers in Human Behavior*, *160*, Article 108386. https://doi.org/10.1016/j.chb.2024.108386

Sweller, J. (2023). The development of cognitive load theory: Replication crises and incorporation of other theories can lead to theory expansion. *Educational Psychology Review*, *35*(4), Article 95. https://doi.org/10.1007/s10648-023-09817-2

Tshidi, M., & Dewa, A. (2024). The promise and peril of Coding & Robotics education in South Africa: A scoping review of teacher preparation and generative artificial intelligence's potential for delivering equity. *Journal of Education*, (96), 141–164. https://doi.org/10.17159/2520-9868/i96a08

Wang, L., Muenks, K., & Yan, V. X. (2023). Interventions to promote retrieval practice: Strategy knowledge predicts intent, but perceived cost predicts usage. *Journal of Educational Psychology*, *115*(8), 1070–1086. https://doi.org/10.1037/edu0000813

Wu, R., & Yu, Z. (2024). Do AI chatbots improve students learning outcomes? Evidence from a meta-analysis. *British Journal of Educational Technology*, *55*(1), 10–33. https://doi.org/10.1111/bjet.13334

Yan, L., Greiff, S., Teuber, Z., & Gašević, D. (2024). Promises and challenges of generative artificial intelligence for human learning. *Nature Human Behaviour*, *8*(10), 1839–1850. https://doi.org/10.1038/s41562-024-02004-5

Zhai, C., Wibowo, S., & Li, L. D. (2024). The effects of over-reliance on AI dialogue systems on students' cognitive abilities: A systematic review. *Smart Learning Environments*, *11*(1), Article 28. https://doi.org/10.1186/s40561-024-00316-7

Zhang, S., Zhao, X., Zhou, T., & Kim, J. H. (2024). Do you have AI dependency? The roles of academic self-efficacy, academic stress, and performance expectations on problematic AI usage behavior. *International Journal of Educational Technology in Higher Education*, *21*(1), Article 34. https://doi.org/10.1186/s41239-024-00467-0

© IIARJournal. All rights reserved.