Trends and Reviews of Technological Engineering in Education
Vol. 2, No. 1 (2026) | ISSN 3071-2211 | Open Access

REVIEW ARTICLE

How AI Stimulates Education and the Relevant Creation

Jie Gao[1, *], Feng Fan[2]
[1] Faculty of Business, Information & Human Sciences (FBIHS), Kuala Lumpur University of Science & Technology (KLUST), Block A (FBIHS), Kajang, 43000, Malaysia
[2] College of Journalism and Communication, Hebei Institute of Communications, Shijiazhuang, 051430, China
* Correspondence: 253926198@s.klust.edu.my

Received 24 September 2025 · Accepted 12 February 2026 · Published 30 March 2026

Abstract

Artificial intelligence has moved from a peripheral instructional aid to an infrastructural condition of contemporary education, and its most consequential effect may not be automation but stimulation: the capacity to provoke, extend, and redirect what learners attempt, produce, and understand. This article synthesizes the scholarship on how AI stimulates educational processes and how it stimulates creation, understood as the generation of artefacts, ideas, and interpretations that are novel and valuable within a learner's own developmental frame. Drawing on 136 sources across artificial intelligence in education, learning analytics, intelligent tutoring, computational creativity, human-computer interaction, creativity psychology, computing education, and educational governance, the review is organized around four analytical dimensions: the mechanism of stimulation, the educational locus at which it operates, the form of creation it supports, and the conditions under which it becomes pedagogically productive. Three themes emerge from this narrative synthesis. First, outcomes may depend on task framing, feedback timing, and learners' evaluation of generated output; this review does not establish a universal causal mediation pathway. Second, selected writing and ideation experiments report individual output gains alongside reduced diversity across outputs, but their educational generality remains uncertain. Third, pedagogical and institutional conditions interact with technical capabilities and constraints. The article proposes a conceptual four-layer framework linking capability exposure, pedagogical mediation, productive struggle, and creative consolidation. Its design principles are propositions for educators, institutions, and researchers to examine; the framework has not been validated as a causal model.

Keywords: Artificial Intelligence in Education; Generative AI; Creativity; Co-Creation; Human-AI Interaction; Learning Design; Educational Governance

How to cite: Jie Gao & Feng Fan (2026). How AI Stimulates Education and the Relevant Creation. TRTEE, 2(1), 1–33. https://doi.org/10.63646/LHRG2669 1

1. Introduction

1.1 From Automation to Stimulation

The dominant framing of artificial intelligence in education has for much of its history been substitutive. Early expert systems, and later intelligent tutoring systems, were evaluated against a benchmark of human instruction, and the guiding question was how closely a machine could approximate an attentive tutor. That framing produced a substantial body of effect estimates but narrowed the imagination of the field, since systems were judged by what they could do instead of a teacher rather than by what they made newly possible. The reframing proposed here is that the educationally significant property of contemporary AI is stimulative rather than substitutive: these systems function as provocations, generating material that learners must evaluate, surfacing possibilities they had not considered, lowering the cost of a first attempt, and making iteration cheap enough that revision becomes normal rather than exceptional.

This reframing is supported by the trajectory of the field. Reviews of the research landscape have repeatedly noted that the most cited work concentrated on technical performance while giving little attention to pedagogy, teacher roles, and institutional context [23, 132]. Subsequent syntheses describe a broadening of scope toward the interaction between algorithmic capability and human judgement [31, 54, 88, 133], and longer-run accounts trace the same movement from the evaluation of discrete tools toward the redesign of practice [35, 78, 94]. The meta-level reviews make the pattern visible: ample evidence that AI systems can perform educationally relevant tasks, far less about the arrangements that determine whether such performance translates into learning [16].

The distinction matters because the two framings generate different research questions. A substitutive framing asks whether an AI tutor matches a human tutor and treats learner effort as a cost to be minimized; a stimulative framing asks whether the system changes what learners attempt, how many alternatives they consider, and how rigorously they evaluate their own work. Under the second, a mediocre draft may be more valuable than an excellent one, because it invites critique where the excellent one invites acceptance. That follows from findings that effortful retrieval, generation, and evaluation are most strongly associated with durable learning [12, 24].

1.2 Why Creation Belongs in the Conversation

The second commitment of this review is that creation must be treated as a first-class educational outcome rather than an enrichment activity appended to core instruction. The first ground is conceptual. Creativity research has dismantled the equation of creativity with eminent achievement, replacing it with graded models in which novelty and value are assessed relative to the creator's own frame of reference [9, 61, 102]. Under such models a student's reorganization of a familiar argument is a genuine creative act, and the classroom becomes a legitimate site of creativity research rather than a diluted version of the studio or laboratory.

The second is technological. Generative systems have made plausible text, images, code, and audio inexpensive and immediate, which changes the economics of a first draft. Activities previously gated by production cost, such as composing alternative openings, sketching several visual treatments of a concept, or prototyping a working program before understanding all of its parts, are now open to learners who lack the technical

© 2026 TRTEE. Open Access under CC BY 4.0

fluency they once required [45, 62, 72].

The third is empirical and cautionary. Selected short-story and ideation experiments report improvements in individual outputs alongside lower diversity between outputs [3, 36]. The short-story study found larger gains among participants with lower baseline creativity [36]; these results do not establish gains in lasting creative ability or classroom learning. On some divergent-thinking tasks, model outputs compare favourably with human averages, while the strongest human responses remain ahead [51, 67]. The implications are therefore task- and measure-dependent.

1.3 Research Questions

The review is organized around four questions. The first asks through what mechanisms AI systems stimulate learning processes, and what the evidence supports about the strength and boundary conditions of each. The second asks how AI systems participate in creation across domains, and whether that participation is best modelled as tool use, partnership, or something requiring new conceptual vocabulary. The third asks what conditions determine whether stimulation becomes educationally productive, encompassing task design, assessment, governance, and professional capability. The fourth asks what an integrative account implies for the design of learning environments and for the research agenda of the field.

1.4 Contributions and Structure

The article makes three contributions. It places the learning literature and the creativity literature in a single analytical frame, which the existing reviews largely keep separate. It gives explicit weight to counter-evidence, treating homogenization and metacognitive offloading as findings to be integrated rather than caveats to be acknowledged. And it proposes a four-layer framework specifying where in an instructional sequence AI stimulation is likely to help and where it is likely to displace the effort that produces learning.

2. Conceptual Foundations

2.1 Three Generations of Educational AI

It is analytically useful to distinguish three generations of AI in education, not because the boundaries are sharp but because each embeds a different theory of how machines contribute to learning; Table 1 summarizes the comparison. The first is model-driven and diagnostic, its canonical achievement being knowledge tracing, in which latent mastery of component skills is estimated from a sequence of observed responses [28]. Meta-analyses of the tutoring systems built on it report moderate positive effects, varying by system, domain, and comparison condition [69, 79]; the theoretically important finding is VanLehn's, that what matters is the granularity of interaction rather than the humanity of the tutor [118].

The second generation is data-driven and organizational. Learning analytics reframed the object of inquiry from the learner's knowledge state to the institution's capacity to act on evidence [7, 40, 108, 109], producing dashboards, early-warning systems, and a durable critique of its own practice: few dashboards rested on a defensible model of self-regulated learning [81], landscape reviews found limited evidence of improved outcomes at scale [55, 121], and intervention studies were often too weak to support

© 2026 TRTEE. Open Access under CC BY 4.0

causal claims [70]. The field identified the core problem as a disconnection between measurement and pedagogy [34, 44, 58, 75].

The third generation is generative and open-ended. Large language models do not require a domain model, a task specification, or a curriculum encoding in order to produce educationally relevant output. This is simultaneously their promise and their problem: the system can operate across subjects without engineering effort, and it has no representation of what the learner should be struggling with. The early educational literature converged quickly on a shared inventory of opportunities and risks [4, 60, 74, 116], structured appraisals mapped strengths and threats onto instructional practice [39], and subsequent scoping work systematized the risk side in detail [131].

Table 1. Three generations of educational AI compared.

Generation	Core mechanism	Evidence status	Principal limitation
First: model-driven and diagnostic	Skill mastery inferred from response sequences; adaptive item selection	Strong; multiple meta-analyses	Requires a decomposable domain; scarce where creation is at stake
Second: data-driven and organizational	Institutional prediction, early warning, and dashboard feedback	Mixed; limited outcome evidence	Measurement frequently detached from pedagogical action
Third: generative and open-ended	Open-domain production of text, code, image, and audio on demand	Emerging; short-horizon and contested	No representation of what the learner should be struggling with

2.2 What Generative Models Changed

The technical shift underlying the third generation is worth stating precisely, since imprecision here produces confused pedagogical claims. The transformer architecture enabled training at scales where general linguistic and procedural competence emerges from next-token prediction rather than task-specific supervision [119], and scaling revealed that such models could perform unseen tasks from instructions and a few examples [18]. Parallel developments in image synthesis produced high-fidelity visual output from textual description [47, 101]. A single interface now spans text, code, image, and audio production, which is why the impact on education has been broader and faster than that of any previous educational technology.

Three properties of these systems have direct pedagogical implications. The first is fluency without grounding: output is syntactically and rhetorically well-formed regardless of factual accuracy, and the surveyed literature establishes that this is structural rather than a defect to be engineered away in the near term [56]. The consequence for education is specific, since the surface features learners use to judge credibility are precisely those generative systems produce most reliably. The second is conversational interactivity, which lowers the threshold for iteration but invites a mode of engagement in which the learner supervises rather than produces. The third is stylistic convergence: because models are optimized toward high-probability

© 2026 TRTEE. Open Access under CC BY 4.0

continuations, unprompted output clusters around modal forms, which is the mechanism most plausibly responsible for the observed compression of collective creative diversity [36].

2.3 Conceptualizing Creation

This review uses creation to denote the production of an artefact, argument, or interpretation that is novel and appropriate relative to the creator's frame of reference. The definition follows the standard two-criterion account in creativity research, in which originality alone is insufficient and effectiveness or appropriateness is required [102]. The graded models developed by Beghetto and Kaufman are essential for educational application, since they legitimate the small-scale, personally novel insight as a genuine creative event and thereby make creativity an ordinary rather than exceptional classroom occurrence [9, 61].

Two further theoretical commitments shape the analysis. The first is that creativity is systemic rather than purely individual. Componential and investment theories locate creative outcomes in the interaction of domain skills, task motivation, and context rather than in a trait possessed by individuals [1, 112], and sociocultural extensions make the surrounding actors, audiences, and affordances constitutive of the creative act rather than incidental to it [46]. This matters for AI because a systemic account has a natural place for a non-human participant, whereas a trait account does not.

The second commitment is that creativity in education is subject to structural constraints that technology does not remove. Craft's analysis of the limits to creativity in schooling identified assessment regimes, curricular coverage pressure, and institutional risk aversion as the binding constraints [30], and the persistent gap between the rhetorical prominence of creativity and its marginal position in educational psychology has been documented for two decades [93]. A technology that makes creative production cheaper does not by itself make creative work institutionally viable.

2.4 Theoretical Lenses for Stimulation

Four lenses organize the analysis. First, scaffolding supports performance beyond current independent capability; support should adapt as learners gain control [129]. Second, cognitive load theory distinguishes processing demands relevant to learning from avoidable demands [114, 115]. AI assistance may reduce unnecessary work, but can also perform reasoning that learners need to practise. Third, ICAP distinguishes passive, active, constructive, and interactive engagement and predicts different learning outcomes across these modes [24]. This provides a lens for examining what learners do with generated material. Applying ICAP to AI-supported tasks is a theoretical interpretation here, rather than evidence that interaction with a model necessarily produces constructive or interactive engagement.

Fourth, self-regulated learning supplies a vocabulary for planning, monitoring, and evaluating learning activity [90, 135]. Research on cognitive offloading relates the use of external resources to task demands and metacognitive evaluations [99]. It does not establish that all learners delegate whenever assistance is available, or that the consequences are uniformly harmful. Applied to generative AI, this lens motivates examining when assistance supports regulation and when it replaces opportunities to practise, with effects on independent learning requiring direct assessment.

© 2026 TRTEE. Open Access under CC BY 4.0

3. Review Design and Synthesis Strategy

3.1 Review Logic

This article uses a structured narrative review supplemented by integrative synthesis. It brings together learning studies, human-computer interaction research, and policy analyses whose questions, methods, and outcomes differ. An integrative approach can compare these perspectives without pooling incompatible effects [111, 117]. Systematic-review reporting guidance provides a useful transparency reference [89], but the present report is not a reproducible systematic review. Search dates, full database-specific queries, and record-level screening decisions are not supplied. The flow diagram summarizes the reported selection process; it does not independently verify the completeness of retrieval or the application of inclusion criteria.

3.2 Corpus Construction

The corpus was assembled through iterative search of the major indexed databases in education, computing, and psychology, supplemented by backward and forward citation chaining from a seed set of highly cited reviews. Search terms combined technology descriptors with educational and creative descriptors. Priority was given to peer-reviewed journal articles and archival conference proceedings from venues with established review processes, since substantial parts of the generative AI literature circulate as preprints of uneven quality. Figure 1 records the identification, screening, and inclusion sequence.

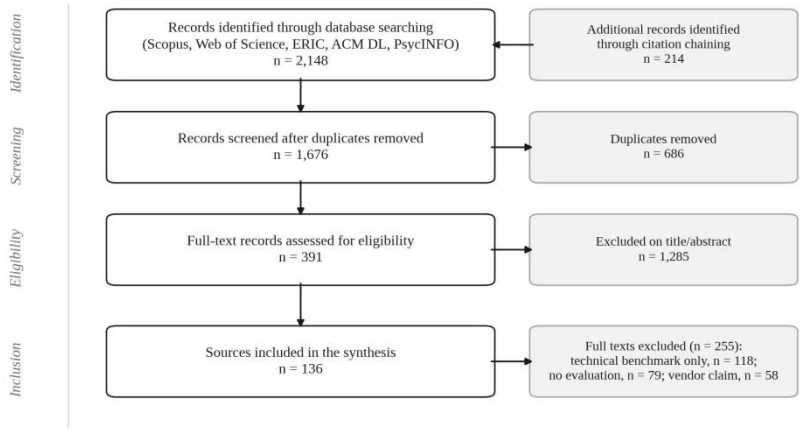


Figure 1.Literature Search and Selection Process

The bibliography contains 136 sources. Figure 2 groups them under seven analytical headings and displays publication years. The cluster bars sum to 136, but the report does not supply a source-level allocation table or explain how overlapping subject coverage was counted. A source may inform more than one discussion theme; this does not establish that the plotted counts use multiple assignment. The figure should therefore be read as a reported descriptive summary, with the cluster coding and counting rule awaiting verification against the underlying records.

© 2026 TRTEE. Open Access under CC BY 4.0

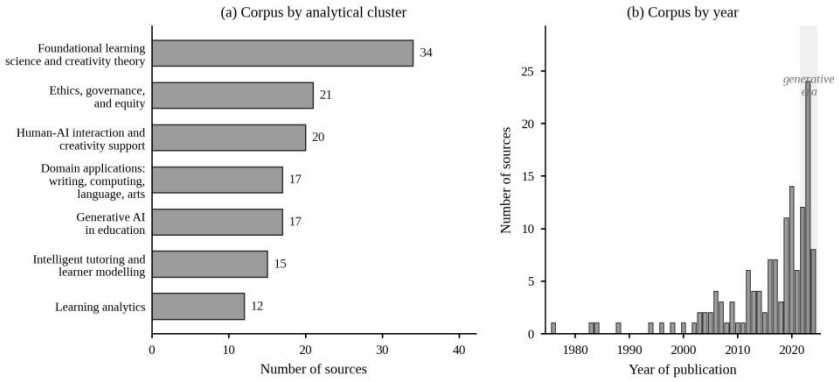


Figure 2. Corpus Composition by Analytical Cluster and Publication Year

3.3 Inclusion and Exclusion Criteria

Sources were included when they met at least one of four criteria: they reported empirical evidence about the effect of an AI system on a learning or creative process; they provided a theoretical account with direct implications for such effects; they synthesized prior work; or they established a governance, ethical, or equity condition bearing on deployment. Sources were excluded when they reported purely technical benchmarks without an educational outcome, described architectures without evaluation, or rested claims of educational impact on vendor documentation.

3.4 Analytical Matrix

Sources were coded along four dimensions. The first is the mechanism of stimulation, comprising adaptation, feedback, dialogue, generation, and orchestration. The second is the educational locus, distinguishing effects on the individual learner, the instructional task, the teacher's practice, and the institution. The third is the form of creation supported, spanning textual, visual, procedural, and conceptual production. The fourth is the condition set governing effectiveness, comprising task design, assessment alignment, professional capability, and governance clarity. Table 2 presents the matrix with representative descriptors.

Table 2. Analytical dimensions used to code the reviewed literature.

Dimension	Categories	Representative descriptor
Mechanism of stimulation	Adaptation; feedback; dialogue; generation; orchestration	What the system does that alters the learner's activity
Educational locus	Learner; task; teacher; institution	Where the effect is realized and who must act for it to occur
Form of creation	Textual; visual; procedural; conceptual	The type of artefact or interpretation the learner produces
Governing condition	Task design; assessment alignment; professional capability; governance clarity	What must hold for the stimulation to become productive

© 2026 TRTEE. Open Access under CC BY 4.0

3.5 Interpretation Criteria and Limitations

The synthesis considers study design, independent replication, and conflicting findings when interpreting evidence. Controlled comparisons can support stronger causal inferences than single-condition deployment reports, subject to their design limitations. These are qualitative interpretive criteria: no study-level quality ratings or reproducible weighting scheme are supplied. Findings about assisted output, independent learning, and diversity between outputs are distinguished because they address different outcomes. Conclusions remain bounded by the tasks and populations studied.

Several limitations constrain the synthesis. The English-language restriction limits its coverage and the reach of equity claims. Much of the generative AI evidence concerns short tasks and context-specific samples, which cannot establish durable learning or programme-level effects. Interpretation and coding involve judgement, and the source-level coding matrix and screening records are not supplied. The cluster allocation rule in Figure 2 also remains unspecified. These reporting gaps limit reproducibility and assessment of selection bias. Finally, results tied to particular model versions may not persist as systems change. The proposed framework should guide further testing rather than be treated as a validated explanation of educational outcomes.

4. How AI Stimulates Learning Processes

4.1 Adaptation and Learner Modelling

Adaptation is the oldest and best-evidenced mechanism by which AI stimulates learning: instruction responsive to an estimate of what a learner currently knows can allocate practice more efficiently than instruction that is not. Knowledge tracing formalized that estimate as a probabilistic model of skill mastery updated from observed performance [28], and the Knowledge-Learning-Instruction framework supplied a principled account of which instructional moves should follow from which inferred knowledge states [65].

The empirical record is favourable but bounded. Meta-analyses report moderate positive effects for intelligent tutoring systems across a range of domains and populations [69, 79]. The magnitude is educationally meaningful without approaching the two-standard-deviation advantage that Bloom attributed to individual human tutoring, and the persistence of that gap has structured the field's ambitions for four decades [14]. VanLehn's analysis is the more useful guide, since it locates the effective ingredient in interaction granularity rather than in adaptation as such [118].

The stimulative reading of this evidence differs from the conventional efficiency reading. Efficiency treats adaptation as valuable because it reduces time spent on already-mastered material; stimulation treats it as valuable because it holds learners at the boundary of their competence, where effortful processing occurs. The two diverge in design implications, since efficiency favours removing difficulty while stimulation favours calibrating it. Research on desirable difficulties indicates that conditions slowing acquisition frequently improve retention and transfer, which means a system optimizing for immediate performance may be optimizing against durable learning [12].

A limitation of this literature is that it presupposes a domain structured enough to support skill decomposition. Mathematics, physics, and introductory programming satisfy that condition; interpretive writing, design, and historical argumentation largely do not. Reviews of the field note the resulting concentration of evidence in well-

© 2026 TRTEE. Open Access under CC BY 4.0

structured domains and the corresponding scarcity of evidence where creation is most at stake [5, 100].

4.2 Feedback and Formative Assessment

Feedback has the strongest theoretical grounding in the learning sciences and the most conditional empirical record. The foundational syntheses establish that it is among the most powerful influences on achievement and simultaneously among the most variable, with effects ranging from strongly positive to negative depending on level, timing, and the learner's disposition toward it [52, 107]. The classic account of formative assessment locates the mechanism in closing a gap between current and desired performance, which requires that the learner both understand the gap and possess a means of closing it [13].

Automated writing evaluation is the most mature application of AI to feedback and offers the clearest evidence about boundary conditions. Reviews of efficacy report reliable gains in surface accuracy, including grammar, mechanics, and lexical variety, and weaker evidence for improvement in organization and argumentation [127]. Controlled comparisons find that automated feedback prompts more revision while the revisions concentrate at the surface [113], and meta-analytic evidence on formative assessment in writing indicates the effective ingredient is the learner's engagement in evaluation rather than the delivery of evaluative information [48].

This points to the central design problem for AI feedback. Systems deliver evaluative information at low cost and high volume, which addresses a real constraint in large classes, but the gain depends on what happens after delivery. Feedback literacy is itself a developed competence rather than a precondition [19], and the principles of formative assessment that support self-regulation emphasize learner agency in the feedback loop rather than informational completeness [86]. A system generating comprehensive comments may therefore reduce evaluative work at exactly the point where that work is the learning.

A further consideration concerns explanation. Where automated judgements affect learners materially, the basis of the judgement must be inspectable. The literature on explainable AI in education argues that explanation requirements in this setting are pedagogical as well as ethical, since an explanation that a learner can act on is a form of instruction [64].

4.3 Dialogue and Conversational Agents

Conversational interaction has become the dominant modality of educational AI, and the evidence base has matured rapidly. Systematic reviews of educational chatbots prior to the generative era found positive effects on engagement and motivation with more equivocal effects on measured learning, and identified a persistent problem of narrow dialogue capability that constrained the depth of interaction [68]. Meta-analytic evidence covering more recent systems reports positive overall effects on learning outcomes with substantial heterogeneity attributable to duration, discipline, and the specificity of the pedagogical role assigned to the agent [130].

The mechanism most plausibly responsible is not information delivery but the licensing of low-stakes questioning. Experimental comparison of chatbot and human partners in language learning found participants engaged more readily with the machine partner in

© 2026 TRTEE. Open Access under CC BY 4.0

early interactions, a pattern attributed to reduced social risk [43]. Many learners suppress questions in classrooms for reasons of face rather than comprehension, and an interlocutor with no social standing removes that constraint.

The corresponding risk is that fluent dialogue produces an impression of understanding that outruns the underlying model. Because hallucinated content is generated in the same register as accurate content, learners lack the surface cues ordinarily used to calibrate trust [56], and the educational literature has consistently treated this as a first-order concern [60, 131]. The response supported by the evidence is to make verification an explicit and assessed component of tasks involving generative systems, which converts the failure mode into an occasion for developing evaluative capability.

4.4 Motivation, Engagement, and Self-Regulation

Claims about AI and motivation are frequently made and rarely grounded in motivational theory. Self-determination theory specifies autonomy, competence, and relatedness as the conditions supporting intrinsic motivation [103]. Read through this frame, generative AI has a mixed profile: it plausibly supports competence by making early success attainable, its effect on autonomy is ambiguous because a system that suggests directions may expand or anchor the perceived option space, and its effect on relatedness is likely negative where machine interaction displaces peer collaboration.

Empirical work with students confirms an ambivalent picture. Large-scale surveys of student perceptions report enthusiasm about accessibility, immediacy, and support for language production alongside concern about accuracy, fairness, and the erosion of the skills students are enrolled to develop [22]. Studies of educator perspectives find a parallel ambivalence, with recognition of workload relief coexisting with concern about assessment validity and the loss of diagnostic information that student work previously provided [71].

The self-regulation literature specifies what would need to be true for AI support to be developmentally productive. Models of self-regulated learning describe a cyclical process of forethought, monitoring, and reflection [90, 135]. AI can support each phase, and it can equally short-circuit the cycle by supplying products without requiring goal-setting or monitoring. Work on cognitive offloading indicates the short-circuit is the default rather than the exception, since delegation tracks immediate cost rather than developmental benefit [99]. Self-regulatory demand must therefore be built into task structure, because it will not arise from the availability of a capable assistant.

4.5 Cognitive Load, Scaffolding, and Withdrawal

Cognitive load theory offers the most precise available account of when AI assistance helps and when it harms. The theory distinguishes processing devoted to the intrinsic complexity of the material, processing imposed by poor instructional design, and processing that constitutes schema construction [114, 115]. Assistance that removes the second category is unambiguously beneficial. Assistance that removes the third is harmful in proportion to its effectiveness, since the removed processing is the learning.

The difficulty is that the two are indistinguishable at the interface. When a generative system supplies working syntax to a novice programmer, whether it has removed extraneous load or germane load depends entirely on whether the instructional objective concerns syntax or algorithmic reasoning. This is a design question that cannot be

© 2026 TRTEE. Open Access under CC BY 4.0

answered by the system and must be answered by the task.

The scaffolding literature supplies the corresponding structural requirement. Scaffolding in its original formulation includes progressive withdrawal as a constitutive feature rather than an optional extension [129], and support that persists indefinitely is substitution rather than scaffolding. Most AI assistance deployed in education has no withdrawal mechanism and no assessment of whether the learner has internalized the supported capability, which makes fading among the most tractable and most neglected design opportunities identified here.

4.6 Language Learning and Multilingual Participation

Language learning is the domain where the case for AI stimulation is strongest, for reasons intrinsic to the subject matter. Acquisition requires large volumes of comprehensible input and meaningful output practice, and the binding constraint in most settings is the availability of an interlocutor. Syntheses of dialogue-based computer-assisted language learning document a long tradition addressing exactly that constraint, with consistent findings that conversational practice systems increase production volume [11], while broader reviews report that effectiveness depends more on task design and integration than on the sophistication of the technology [134].

Generative systems extend these affordances substantially, providing patient practice, register adaptation, and immediate corrective feedback across a wide range of languages [66]. The equity implication cuts in two directions. Learners writing in an additional language gain expressive resources previously available only through expensive tutoring, a genuine widening of participation. Yet detection systems used to police AI use exhibit systematic bias against non-native English writing, classifying it as machine-generated at substantially elevated rates [73]. The population that stands to gain most is the population most likely to be wrongly accused.

5. How AI Stimulates Creation

5.1 From Computational Creativity to Co-Creativity

Research on machine creativity long predates generative models and supplies conceptual resources the current educational discussion often lacks. Boden's taxonomy distinguishes combinational, exploratory, and transformational creativity, and remains diagnostically useful, since contemporary systems are strong at combination and exploration within a learned distribution and weak at transformation of the space itself [15]. The evaluation literature established early that assessing creative systems requires attention to process and framing rather than output inspection alone [26, 59], and adversarial approaches to art generation explicitly engineered deviation from learned norms, acknowledging that unmodified generative objectives produce convergence rather than novelty [38].

The more useful frame for education is co-creativity, in which creative outcomes emerge from interaction rather than residing in either party. The question of how computers might participate as partners rather than instruments was posed clearly two decades ago [77]. Empirical study of co-creative drawing agents demonstrated that participants engaged in genuine sense-making with a machine collaborator, adjusting their intentions in response to its contributions [33]. Interaction frameworks developed for co-creative systems now provide vocabulary for the design decisions involved,

© 2026 TRTEE. Open Access under CC BY 4.0

including initiative, turn-taking, and communication channel [98].

This connects to a longer tradition of creativity support tools in human-computer interaction, whose design agenda emphasized exploration, alternative generation, and reflection rather than automated production [105]. Subsequent mapping of the field documented a persistent tension between tools that expand the option space and tools that converge on a solution [42], and human-centred design principles for AI systems apply directly, particularly the requirement that the human retain control over consequential decisions [2, 106].

5.2 Creative Writing and Composition

Writing is the domain where human-AI creation has been studied most systematically. The CoAuthor investigation, which captured detailed interaction traces of writers working with language model suggestions, found that suggestions functioned less as text to be accepted than as prompts that redirected the writer's own thinking, with acceptance frequently followed by substantial revision [72]. This finding is central to the argument of this review, since it identifies the mechanism of stimulation as provocation rather than production.

Work on science writing support reached a compatible conclusion through a different route. Systems designed to generate inspirational sparks rather than finished prose were found useful precisely because their output was incomplete, which required writers to supply the connective reasoning [45]. Narrative support tools that expose high-level control over story arc rather than sentence-level generation similarly preserve authorial intention while reducing the cost of exploring alternatives [25].

The pedagogical inference is specific. Generative assistance appears most productive for creation when it operates at a level of abstraction different from the one at which the learner is working. A system that generates alternatives at the level of structure while the learner works at the level of sentences creates a productive division of labour; a system that generates at the learner's own level substitutes for the learner rather than stimulating them.

5.3 Visual Art, Design, and Music

In visual and musical domains the technical barrier to production has historically been the primary gatekeeper of creative participation. A learner with strong visual ideas and limited rendering skill could not previously realize those ideas, and the years required to acquire technique filtered participation long before conceptual capability was tested. Diffusion-based image synthesis removes this barrier substantially [101], and comparable developments in music generation have altered the relationship between compositional intention and technical execution [17].

The educational consequence is ambivalent and should not be resolved prematurely in either direction. Removing the technical gate widens participation, and the equity argument for doing so is strong, since technique acquisition has always been correlated with access to instruction and instruments. It also removes the extended engagement with material constraint through which aesthetic judgement has traditionally been developed. The generative interface is optimized for immediate output, and the discipline of revision under constraint is exactly what it removes.

The available theoretical resources suggest a resolution rather than a choice. If creativity

© 2026 TRTEE. Open Access under CC BY 4.0

is systemic rather than individual, and if the surrounding structure of audience, critique, and iteration is constitutive rather than incidental [46], then the pedagogical task is to reconstruct the constraint structure that the technology has removed. Constraint can be imposed by the task rather than by the medium, through requirements for justification, comparative critique, and articulation of the relationship between intention and result.

5.4 Computing, Making, and Computational Thinking

Computing education has produced the most rigorous evidence on AI and creation, partly because outcomes are measurable and the community responded quickly. Early evaluation established that code generation models solved a substantial proportion of introductory assessments, performing above the median student on standard problem sets [41]. The disciplinary response recognized the challenge as fundamental rather than incremental [8], and the subsequent working group report gives the most thorough available treatment of how a discipline can restructure its pedagogy in response [95].

A controlled study of 69 novices aged 10–17 found that Codex access improved code-authoring performance [62]. Manual code-modification, unaided post-test, and one-week retention scores did not differ significantly between groups overall. The environment paired authoring with modification, but did not experimentally isolate comprehension-focused framing as the cause of these outcomes. These short-course results do not establish equivalent learning between groups or the effects of prolonged reliance on code generation.

This connects to the computational thinking tradition, which held that the educational value of computing lies in transferable problem formulation and abstraction rather than in syntactic fluency [49, 128]. On that argument, code generation removes precisely the layer that computational thinking advocates regarded as incidental. The constructionist tradition, which held that learning is most durable when learners build shareable artefacts they care about, points the same way [97]: a learner who can produce a working program in an afternoon may reach system-level design questions previously unreachable within a course.

5.5 Inquiry and Academic Knowledge Production

AI participation in research-like activity raises questions about the formation of scholarly capability. Generative systems support literature synthesis, hypothesis generation, analytic code, and drafting, and each is an activity through which research judgement has traditionally been developed. The concern is not that assistance produces poorer immediate output but that it removes the developmental experience of struggling with an unfamiliar literature or an uncooperative dataset.

Experimental evidence from adjacent professional settings establishes the shape of the effect. Controlled study of writing tasks found substantial reductions in time and improvements in rated quality, with the largest gains concentrated among initially lower-performing participants [87]. This compression pattern recurs across the literature and is educationally double-edged: compression of the performance distribution is desirable if it reflects genuine capability gain and undesirable if it reflects convergence on machine-mediated output that no participant fully understands.

The multidisciplinary commentary on generative AI in knowledge work canvasses these tensions across sectors and converges on the position that the technology reorganizes

© 2026 TRTEE. Open Access under CC BY 4.0

rather than eliminates expert work, shifting the locus of expertise toward problem formulation and evaluation [37]. For research education this suggests that supervision should shift emphasis from the production of text toward the formulation of questions and the critique of evidence, which are the capabilities that remain scarce.

5.6 Counter-Evidence: Homogenization, Offloading, and Ownership

Selected experiments distinguish individual from collective outcomes. In an online short-story study, access to generative AI ideas improved assessed story quality and creativity, particularly for less creative writers, while AI-assisted stories became more similar to one another [36]. A separate study of 36 participants found less semantically distinct ideas with ChatGPT than with an alternative creativity-support tool [3]. These findings concern specific tasks and comparators; they do not establish a universal effect on learners or long-term creative development.

Anchoring on model suggestions is one possible explanation for convergence, but a training objective alone does not establish the mechanism in a particular learning context. Assessment focused only on individual product quality may overlook similarities across submissions. Evaluating both individual quality and between-output diversity could detect such a trade-off. Whether repeated assistance narrows students' independent thinking over a programme remains an open longitudinal question, rather than an outcome established by these short-task experiments.

Comparisons on divergent-thinking tasks provide a further boundary on interpretation. Some model outputs match or exceed human average scores, while the best human responses can outperform the models [51, 67]. These are comparisons of generated responses, not experiments showing that assistance reduces the creativity of high-performing learners. The benefit or cost of assistance must be tested directly, with attention to baseline ability, task design, and independent performance. Figure 3 is a conceptual schematic, not a fitted distribution or an empirical effect estimate.

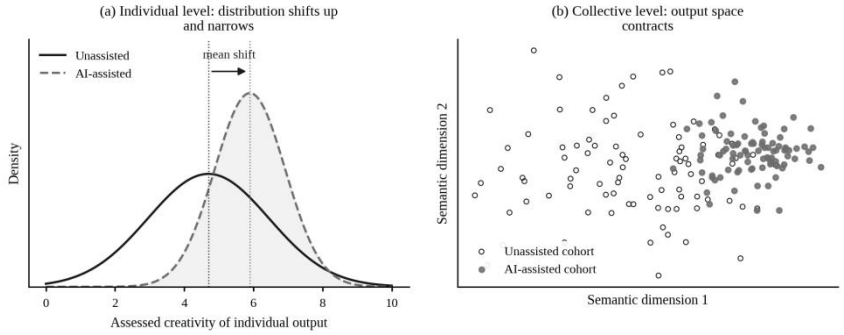


Figure 3. Individual and Collective Creativity Under Generative Assistance

A third strand concerns ownership and authorship. Sustained scholarly engagement with these questions has produced a collaborative rather than substitutive framing, in which the analytically important question is how creative agency is distributed across the interaction rather than whether the machine is creative [123]. For education the practical form of this question is whether learners can articulate what in a submitted work is theirs, which is both an integrity requirement and a metacognitive exercise with independent educational value. Table 3 summarizes the mechanisms reviewed in

© 2026 TRTEE. Open Access under CC BY 4.0

Sections 4 and 5 against their evidence base and boundary conditions.

Table 3. Summary of stimulation mechanisms, evidence base, and boundary conditions.

Mechanism	Primary locus	Evidence base	Principal boundary condition
Adaptation and learner modelling	Learner knowledge state	Strong; meta-analytic	Requires a decomposable domain; concentrated in well-structured subjects
Automated feedback	Task and revision cycle	Moderate; mixed by level	Gains concentrate at surface features; depends on learner feedback literacy
Conversational dialogue	Learner engagement	Moderate; heterogeneous	Fluency decouples from accuracy; requires assessed verification
Generative production	Creative artefact	Emerging; contested	Individual gain coexists with collective convergence
Orchestration and analytics	Teacher and institution	Weak for outcomes	Dashboards frequently lack a pedagogical model

6. Conditions, Constraints, and Consequences

6.1 Academic Integrity and Assessment Redesign

The integrity discussion has consumed a disproportionate share of institutional attention, much of it directed at a solution the evidence does not support. Systematic testing of detection tools reports accuracy insufficient for consequential decisions, degrading sharply under paraphrase and light editing [125], and the bias findings compound the problem because false positives concentrate among writers for whom English is an additional language [73]. An enforcement regime built on unreliable detection with socially patterned errors is not a defensible institutional position.

The constructive literature converges on assessment redesign. Early analyses of the challenge to conventional assessment identified the vulnerability as residing in task design rather than in student conduct, since assessments that ask for reproducible outputs are vulnerable to any reproduction technology [29, 91]. The authentic assessment tradition supplies the alternative, specifying tasks that require engagement with context, process, and defensible judgement rather than the production of a decontextualized artefact [50, 122].

Four redesign strategies recur across the reviewed sources, summarized in Table 4 with the cost each imposes and the vulnerability each leaves open. Their common logic is to shift the object of assessment from an artefact that can be reproduced to activity that is difficult to delegate. Each imposes real costs in staff time, and institutional policy discussions frequently understate these costs when advocating redesign as a solution.

Table 4. Assessment redesign strategies, costs, and residual vulnerabilities.

Redesign	What becomes	Principal cost	Residual vulnerability

© 2026 TRTEE. Open Access under CC BY 4.0

strategy	assessable		
Process-visible assessment	Drafting, revision decisions, and the reasoning behind them	Staff time for staged review	Process artefacts can themselves be generated
Contextual anchoring	Engagement with local, course-specific, or field-collected material	Task authoring and refresh each cycle	Weakens once local material circulates online
Oral defence and live justification	Whether the learner can explain and extend the submitted work	Scheduling and staffing at scale	Disadvantages learners with communication anxiety
Comparative critique of generated output	Evaluative judgement applied to machine-produced material	Requires staff fluency with the systems	Critique itself may be delegated to a second system

Institutional policy frameworks developed specifically for this situation emphasize differentiation by task type rather than blanket permission or prohibition, and they emphasize the need for policy to address staff use as well as student use [21]. The available evidence indicates that policies which prohibit without specifying permitted alternatives produce non-disclosure rather than non-use.

6.2 Bias, Fairness, and Unequal Participation

Algorithmic bias in education has a distinctive structure. Educational models are typically trained on historical records of participation and achievement, which encode the effects of prior inequity, so a model that predicts accurately with respect to that history will reproduce it, and its accuracy will be cited as evidence of its validity. The comprehensive review of algorithmic bias in education documents this pattern across prediction, classification, and recommendation applications, and notes that the demographic attributes required to audit for bias are frequently unavailable in educational datasets [6].

Generative systems introduce a second bias pathway operating through representation rather than prediction. Because training corpora over-represent some varieties of language, some cultural references, and some argumentative conventions, model output normalizes those varieties. In an educational setting this normalization is prescriptive, since students receive it as a model of correct or effective expression. The critical literature on large language models identified this concern early and connected it to the scale and opacity of training data [10].

Access inequity is the third pathway and the most immediately consequential. Digital divide research established long ago that the significant divide is not binary access but differentiated quality of access and use, encompassing device capability, connectivity, and the presence of adults who can support productive use [124, 136]. Generative AI reproduces that structure in sharpened form, since capability differences between free and paid systems are substantial and productive prompting is unevenly distributed and largely untaught.

© 2026 TRTEE. Open Access under CC BY 4.0

6.3 Privacy, Governance, and Institutional Policy

The governance literature in learning analytics anticipated most of the questions now facing institutions deploying generative systems. Early analysis identified the tension between an institution's capacity to act on data and its obligation to respect learner autonomy [110], and subsequent work articulated the obligation to act, arguing that an institution which identifies risk and does not respond has not achieved neutrality [96]. Contemporary treatments frame fairness, transparency, and accountability as design commitments rather than compliance requirements satisfied afterwards [63].

Generative systems add a specific governance problem. Where earlier analytics operated on institutionally held data within institutional infrastructure, generative systems typically transmit content to external providers under terms the institution has not negotiated, which makes the question immediate and concrete as soon as students and staff submit drafts, assessment material, or discussion of individual learners.

The wider AI ethics literature provides principles but limited operational guidance. Comparative analysis of ethics guidelines found convergence on high-level principles alongside divergent interpretation and near-total absence of enforcement [57]. Education-specific frameworks have tried to close this gap by grounding principles in pedagogical relationships rather than generic ethical categories [53, 85], while critical scholarship supplies the corrective that framing adoption as inevitable forecloses questions about whose interests deployment serves [92, 104, 126].

6.4 Teacher Agency, Trust, and AI Literacy

The teacher is the mediating variable through which most of the effects discussed here are realized, and the teacher is under-represented in the research. The observation that educators were largely absent from the AI in education literature was made in the field's most cited systematic review and remains substantially accurate [132]. Systematic review from the teacher's perspective identifies capability, workload, and role clarity as the recurring themes, and finds teachers frequently positioned as implementers of systems they had no part in designing [20].

Trust is the operative construct, and it is neither fixed nor a simple function of accuracy. Intervention research shows that teacher trust in AI-based educational technology is malleable and responds to professional development addressing how systems work and where they fail, rather than to demonstrations of performance [83]. Related work shows that framing evidence as AI-derived affects its perceived credibility independently of its content [32].

AI literacy provides the organizing construct for what teachers and learners need. Competency frameworks specify recognition of AI, understanding of its mechanisms, evaluation of its outputs, and ethical reasoning about its use [76, 84]. Applied to teachers, this maps onto technological pedagogical content knowledge, in which effective use requires the integration of technological, pedagogical, and subject knowledge rather than their addition [82]. The data literacy literature shows why the integration is not optional, having documented that teachers presented with data they cannot relate to their own instructional decisions will not use it [80].

Acceptance research supplies a final consideration. Established models identify performance expectancy, effort expectancy, social influence, and facilitating conditions as the determinants of technology use [120]. Facilitating conditions are the component

© 2026 TRTEE. Open Access under CC BY 4.0

most often neglected in educational AI deployment, and they encompass the time, infrastructure, and institutional permission required to change practice.

7. An Integrative Framework: Stimulation, Mediation, and Creation

7.1 Framework Overview

The review proposes a four-layer conceptual framework for examining how AI capability may contribute to educational value. The layers organize design questions, rather than establish a tested causal sequence or necessary conditions for benefit. Figure 4 summarizes these proposed relationships, which require empirical validation.

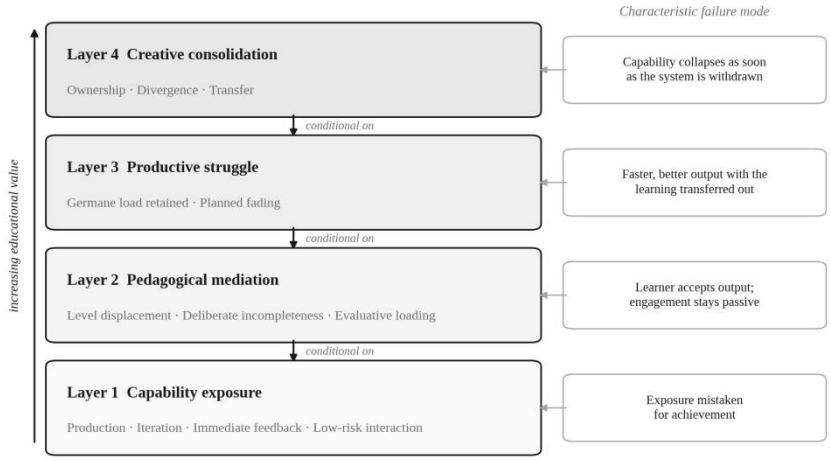


Figure 4. Four-Layer Framework for AI-Supported Creative Learning

7.2 Layer One: Capability Exposure

The first layer concerns what the system makes possible that was not possible before. Capability exposure is necessary but weakly informative, since it establishes only that a barrier has been removed, and the characteristic failure here is to treat exposure as achievement. Four exposures recur: production without technical prerequisite, iteration without proportional cost, feedback without waiting, and interaction without social risk, documented respectively in creative domains where technique gated participation [17, 101], in writing [72], in large classes [127], and in language learning [43].

7.3 Layer Two: Pedagogical Mediation

The second layer concerns how exposed capability is embedded in instructional design, and this is where the reviewed evidence locates most of the variance in outcomes. The same generative system produces different results depending on whether learners are asked to accept, critique, compare, or extend its output, and the ICAP ordering predicts the direction of the difference [24]. The programming evidence provides the cleanest demonstration, since assistance did not degrade subsequent independent performance when the task required comprehension of the generated code [62].

Three mediation moves recur across successful cases in the corpus. Level displacement assigns the system and the learner to different levels of abstraction, as in narrative tools that generate structure while the learner controls execution [25]. Deliberate incompleteness supplies partial material that requires the learner to furnish connective

© 2026 TRTEE. Open Access under CC BY 4.0

reasoning [45]. Evaluative loading makes assessment of generated output the assessed activity, which converts the reliability problem into pedagogical content [56, 64].

7.4 Layer Three: Productive Struggle

The third layer concerns whether the mediated activity preserves the effort that constitutes learning. This is where well-intentioned design most frequently fails, because the failure mode resembles success: learners complete more work, complete it faster, and produce better artefacts, while the processing that would have built durable capability has been transferred to the system.

Cognitive load theory specifies the distinction between load that should be removed and load that constitutes learning [115], and the desirable difficulties literature establishes that conditions slowing acquisition frequently improve retention and transfer [12]. Research on cognitive offloading establishes that learners will delegate whenever delegation is available and cheap, independently of developmental cost [99]. The design consequence is that productive struggle must be structurally required rather than merely encouraged. Scaffolding theory supplies the mechanism in the form of planned withdrawal, which is present in the theory and largely absent from deployed systems [129].

7.5 Layer Four: Creative Consolidation

The fourth layer concerns whether stimulated activity produces creative capability that persists beyond the supported context, and the evidence supports three requirements. Ownership requires that learners identify and justify what in a work is theirs, which the co-creativity literature treats as a question about the distribution of creative agency [98, 123]. Divergence requires deliberate counteraction of convergence pressure, given the documented compression of collective diversity [3, 36]. Transfer requires that capability be demonstrable without the supporting system.

The systemic view of creativity supports a specific design response to the divergence problem. If creative outcomes emerge from the interaction of person, process, product, place, and audience [46], diversity can be engineered at the level of the surrounding structure even when the generative system exerts convergent pressure. Assigning differentiated constraints, audiences, or evaluative criteria across a cohort reintroduces variance that the technology removes.

7.6 Design Principles

Eight principles follow from the framework and the reviewed evidence. Table 5 summarizes them with their evidential basis and characteristic failure mode.

Table 5. Design principles for AI-supported stimulation of learning and creation.

Design principle	Evidential basis	Characteristic failure mode
Specify the target of struggle before introducing assistance	Cognitive load theory; distinction between extraneous and germane processing	Assistance removes the processing that constitutes the learning objective
Displace the level at which the system	Human-AI writing and narrative studies; division of	System and learner compete at the same level; learner

© 2026 TRTEE. Open Access under CC BY 4.0

operates	labour by abstraction	becomes an editor
Build withdrawal into the design	Scaffolding theory; fading as a definitional requirement	Permanent support presented as scaffolding; capability never transfers
Assess process alongside product	Authentic and sustainable assessment; process-visible task design	Product-only assessment is vulnerable to any reproduction technology
Require verification as assessed work	Hallucination literature; explainability in education	Accuracy warnings issued as advice rather than as graded requirement
Engineer divergence at cohort level	Homogenization findings; systemic models of creativity	Individual scores rise while cohort variance silently collapses
Invest in teacher capability before deployment	Trust intervention research; TPACK; data literacy	Systems deployed to staff positioned as implementers, not designers
Establish data governance prospectively	Learning analytics ethics; AI governance frameworks	External transmission of learner content under unnegotiated terms

8. Discussion

8.1 Theoretical Implications

Three theoretical implications follow from the synthesis. The first concerns the unit of analysis for creativity research in education. The homogenization findings show that individual and collective creative outcomes can move in opposite directions under the same intervention, so a design measuring only individual outcomes is structurally incapable of detecting an important class of effects. Creativity research in education has been overwhelmingly individual in its measurement tradition, and the systemic frameworks that would support collective measurement have not been matched by corresponding instrumentation [1, 46].

The second concerns the theoretical status of scaffolding. The reviewed evidence indicates that the withdrawal condition, definitional in the original formulation [129], has been quietly dropped in most applications to AI-supported learning. Restoring it is not terminological, since a support structure without withdrawal has a different predicted effect on capability development, and the field would benefit from explicit theoretical work on fading in systems whose capability is general rather than domain-specific.

The third concerns the relationship between fluency and understanding. Generative systems decouple the production of well-formed output from the understanding such output has historically indexed, and educational assessment has relied on that coupling for as long as written assessment has existed. The theoretical work required is a specification of what written work now evidences, and the reviewed literature has posed this question more clearly than it has answered it [27, 131].

8.2 Implications for Practice and Policy

© 2026 TRTEE. Open Access under CC BY 4.0

For teachers, the synthesis supports a shift in the locus of instructional design effort from content delivery toward specifying what learners must do with material they did not produce. That is a substantial change in professional practice and requires time and support rather than exhortation; the evidence on teacher trust indicates such support works best when it addresses the mechanisms and failure modes of systems rather than their features [20, 83].

For institutions, three priorities emerge. Assessment redesign should be resourced as a curriculum project rather than delegated to individual staff, since the redesign strategies supported by evidence are labour-intensive and require coordination. Detection-based enforcement should be de-emphasized in favour of design-based approaches, given the documented reliability and equity problems [73, 125]. Data governance should be established prospectively, and policies should differentiate by task type and address staff use alongside student use [21].

For policy, the equity implications require attention they are not currently receiving. The access structure documented in digital divide research reproduces itself in the generative AI context with sharper gradients, since capability differences between free and paid systems are substantial and prompting skill is unevenly distributed and largely untaught [124, 136]. Policy attention has concentrated on integrity and safety while access has received comparatively little, and the distributional consequences are predictable.

8.3 A Research Agenda

Six priorities follow from the gaps identified in the corpus, set out in Table 6 with the open question and a suitable design for each. They share a common shape: the effects that matter most for the argument advanced here are cumulative, collective, or contextual, and none of the three is well served by the short-horizon, individual-outcome, single-site study that dominates the current evidence base.

Table 6. Research agenda: priorities, open questions, and suitable designs.

Research priority	Open question	Suitable design
Longitudinal effects	Does sustained assistance produce capability atrophy across a programme of study?	Multi-year cohort tracking with periodic withdrawal probes
Collective-level measurement	How can the diversity of cohort output be instrumented and monitored?	Corpus-level semantic dispersion analysis across cohorts
Fading in general-capability systems	When and how should support be withdrawn when the system is not domain-specific?	Design-based research with staged support conditions
Teacher mediation	Which instructional moves convert generated output into constructive engagement?	Classroom experiments coding interaction against ICAP
Cross-context replication	Do the findings hold outside well-resourced English-medium institutions?	Coordinated multi-site replication with shared protocols
Assessment validity	What does written and produced work now evidence	Validity studies comparing assessed and independently

© 2026 TRTEE. Open Access under CC BY 4.0

	about underlying capability?	demonstrated capability

8.4 Limitations

The limitations stated in Section 3.5 constrain the conclusions in specific ways. The English-language restriction weakens the equity analysis most directly, since the populations whose access is most constrained are least represented in the corpus. The recency of the generative AI literature means the framework's third and fourth layers rest on shorter-term evidence than its first and second. And the pace of capability change means that any claim tied to a specific system generation should be read as provisional.

9. Conclusion

This narrative review proposes stimulation as one perspective on educational AI. The reviewed literature suggests that systems can alter what learners attempt and produce, while the educational value of those changes depends on the task, learner, instructional setting, and technical conditions. The synthesis does not establish a universal mediation pathway or rank institutional factors above technical ones. Immediate improvements in assisted performance should be distinguished from durable learning.

Selected short-story and ideation experiments report benefits to individual output alongside reduced diversity between outputs. These results motivate cohort-level measurement; they do not demonstrate that educational programmes produce less differentiated graduates. Establishing the durability and educational relevance of this trade-off requires longitudinal studies with independent creative tasks and explicit comparisons of assistance conditions.

The four-layer framework identifies candidate areas for intervention. Capability exposure depends on functionality, access, infrastructure, and institutional choices. Pedagogical mediation, productive struggle, and creative consolidation raise further design questions. These are conceptual propositions, not a validated sequence in which failure at one layer necessarily prevents benefit at another. Testing requires measures of supported and independent performance.

A practical question is how educational settings can make generated material an occasion for learners to explain, evaluate, revise, and create. Assessment design, teacher capability, governance, and access all deserve attention alongside the capabilities and limitations of the systems used. Improvements in underlying models may change the opportunities and risks, but do not by themselves establish educational value. That value requires evaluation in the intended learning context over appropriate timescales.

© 2026 TRTEE. Open Access under CC BY 4.0

Acknowledgements

The authors thank colleagues who commented on earlier outlines of this review. Any errors of interpretation remain the responsibility of the authors.

Author Contributions

J.G.: Conceptualization, Methodology, Literature Screening, Writing — Original Draft, Project Administration. F.F.: Theoretical Framing, Data Curation, Visualization, Validation, Writing — Review & Editing. Both authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

The authors declare no conflicts of interest.

Generative AI Statement

Generative AI was used solely for language editing and readability improvement; the authors remain fully responsible for the content.

© 2026 TRTEE. Open Access under CC BY 4.0

References

[1] Amabile, T. M. (1983). The social psychology of creativity: A componential conceptualization. Journal of Personality and Social Psychology, 45(2), 357–376. https://doi.org/10.1037/0022-3514.45.2.357

[2] Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P. N., Inkpen, K., Teevan, J., Kikin-Gil, R., & Horvitz, E. (2019). Guidelines for human-AI interaction. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, 1–13. https://doi.org/10.1145/3290605.3300233

[3] Anderson, B. R., Shah, J. H., & Kreminski, M. (2024). Homogenization effects of large language models on human creative ideation. Proceedings of the 16th Conference on Creativity & Cognition, 413–425. https://doi.org/10.1145/3635636.3656204

[4] Baidoo-Anu, D., & Owusu Ansah, L. (2023). Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. Journal of AI, 7(1), 52–62. https://doi.org/10.61969/jai.1337500

[5] Baker, R. S. (2016). Stupid tutoring systems, intelligent humans. International Journal of Artificial Intelligence in Education, 26(2), 600–614. https://doi.org/10.1007/s40593-016-0105-0

[6] Baker, R. S., & Hawn, A. (2022). Algorithmic bias in education. International Journal of Artificial Intelligence in Education, 32(4), 1052–1092. https://doi.org/10.1007/s40593-021-00285-9

[7] Baker, R. S., & Inventado, P. S. (2014). Educational data mining and learning analytics. In J. A. Larusson & B. White (Eds.), Learning analytics: From research to practice (pp. 61–75). Springer. https://doi.org/10.1007/978-1-4614-3305-7_4

[8] Becker, B. A., Denny, P., Finnie-Ansley, J., Luxton-Reilly, A., Prather, J., & Santos, E. A. (2023). Programming is hard — or at least it used to be: Educational opportunities and challenges of AI code generation. Proceedings of the 54th ACM Technical Symposium on Computer Science Education, 500–506. https://doi.org/10.1145/3545945.3569759

[9] Beghetto, R. A., & Kaufman, J. C. (2007). Toward a broader conception of creativity: A case for mini-c creativity. Psychology of Aesthetics, Creativity, and the Arts, 1(2), 73–79. https://doi.org/10.1037/1931-3896.1.2.73

[10] Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 610–623. https://doi.org/10.1145/3442188.3445922

[11] Bibauw, S., François, T., & Desmet, P. (2019). Discussing with a computer to practice a foreign language: Research synthesis and conceptual framework of dialogue-based CALL. Computer Assisted Language Learning, 32(8), 827–877. https://doi.org/10.1080/09588221.2018.1535508

[12] 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

[13] Black, P., & Wiliam, D. (1998). Assessment and classroom learning. Assessment in Education: Principles, Policy & Practice, 5(1), 7–74.

© 2026 TRTEE. Open Access under CC BY 4.0

https://doi.org/10.1080/0969595980050102
[14] Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. Educational Researcher, 13(6), 4–16. https://doi.org/10.3102/0013189X013006004
[15] Boden, M. A. (2009). Computer models of creativity. AI Magazine, 30(3), 23–34. https://doi.org/10.1609/aimag.v30i3.2254
[16] Bond, M., Khosravi, H., De Laat, M., Bergdahl, N., Negrea, V., Oxley, E., Pham, P., Chong, S. W., & Siemens, G. (2024). A meta systematic review of artificial intelligence in higher education: A call for increased ethics, collaboration, and rigour. International Journal of Educational Technology in Higher Education, 21, 4. https://doi.org/10.1186/s41239-023-00436-z
[17] Briot, J.-P., & Pachet, F. (2020). Deep learning for music generation: Challenges and directions. Neural Computing and Applications, 32(4), 981–993. https://doi.org/10.1007/s00521-018-3813-6
[18] Brown, T. B., et al. (2020). Language models are few-shot learners. arXiv. https://doi.org/10.48550/arXiv.2005.14165

[19] Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. Assessment & Evaluation in Higher Education, 43(8), 1315–1325. https://doi.org/10.1080/02602938.2018.1463354
[20] Celik, I., Dindar, M., Muukkonen, H., & Järvelä, S. (2022). The promises and challenges of artificial intelligence for teachers: A systematic review of research. TechTrends, 66(4), 616–630. https://doi.org/10.1007/s11528-022-00715-y
[21] Chan, C. K. Y. (2023). A comprehensive AI policy education framework for university teaching and learning. International Journal of Educational Technology in Higher Education, 20, 38. https://doi.org/10.1186/s41239-023-00408-3
[22] Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. International Journal of Educational Technology in Higher Education, 20, 43. https://doi.org/10.1186/s41239-023-00411-8
[23] Chen, X., Xie, H., Zou, D., & Hwang, G.-J. (2020). Application and theory gaps during the rise of artificial intelligence in education. Computers and Education: Artificial Intelligence, 1, 100002. https://doi.org/10.1016/j.caeai.2020.100002
[24] Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. Educational Psychologist, 49(4), 219–243. https://doi.org/10.1080/00461520.2014.965823
[25] Chung, J. J. Y., Kim, W., Yoo, K. M., Lee, H., Adar, E., & Chang, M. (2022). TaleBrush: Sketching stories with generative pretrained language models. Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, 1–19. https://doi.org/10.1145/3491102.3501819
[26] Colton, S., & Wiggins, G. A. (2012). Computational creativity: The final frontier? Proceedings of the 20th European Conference on Artificial Intelligence, 21–26. https://doi.org/10.3233/978-1-61499-098-7-21
[27] Cope, B., Kalantzis, M., & Searsmith, D. (2021). Artificial intelligence for education: Knowledge and its assessment in AI-enabled learning ecologies. Educational Philosophy and Theory, 53(12), 1229–1245.

© 2026 TRTEE. Open Access under CC BY 4.0

https://doi.org/10.1080/00131857.2020.1728732
[28] Corbett, A. T., & Anderson, J. R. (1994). Knowledge tracing: Modeling the acquisition of procedural knowledge. User Modeling and User-Adapted Interaction, 4(4), 253–278. https://doi.org/10.1007/BF01099821
[29] Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. Innovations in Education and Teaching International, 61(2), 228–239. https://doi.org/10.1080/14703297.2023.2190148
[30] Craft, A. (2003). The limits to creativity in education: Dilemmas for the educator. British Journal of Educational Studies, 51(2), 113–127. https://doi.org/10.1111/1467-8527.t01-1-00229
[31] Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education: The state of the field. International Journal of Educational Technology in Higher Education, 20, 22. https://doi.org/10.1186/s41239-023-00392-8
[32] Cukurova, M., Luckin, R., & Kent, C. (2020). Impact of an artificial intelligence research frame on the perceived credibility of educational research evidence. International Journal of Artificial Intelligence in Education, 30(2), 205–235. https://doi.org/10.1007/s40593-019-00188-w
[33] Davis, N., Hsiao, C.-P., Yashraj Singh, K., Li, L., & Magerko, B. (2016). Empirically studying participatory sense-making in abstract drawing with a co-creative cognitive agent. Proceedings of the 21st International Conference on Intelligent User Interfaces, 196–207. https://doi.org/10.1145/2856767.2856795
[34] Dawson, S., Joksimovic, S., Poquet, O., & Siemens, G. (2019). Increasing the impact of learning analytics. Proceedings of the 9th International Conference on Learning Analytics & Knowledge, 446–455. https://doi.org/10.1145/3303772.3303784
[35] Dillenbourg, P. (2016). The evolution of research on digital education. International Journal of Artificial Intelligence in Education, 26(2), 544–560. https://doi.org/10.1007/s40593-016-0106-z
[36] Doshi, A. R., & Hauser, O. P. (2024). Generative AI enhances individual creativity but reduces the collective diversity of novel content. Science Advances, 10(28), eadn5290. https://doi.org/10.1126/sciadv.adn5290
[37] Dwivedi, Y. K., et al. (2023). Opinion paper: So what if ChatGPT wrote it? Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. International Journal of Information Management, 71, 102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642

[38] Elgammal, A., Liu, B., Elhoseiny, M., & Mazzone, M. (2017). CAN: Creative adversarial networks generating art by learning about styles and deviating from style norms. arXiv. https://doi.org/10.48550/arXiv.1706.07068
[39] Farrokhnia, M., Banihashem, S. K., Noroozi, O., & Wals, A. (2024). A SWOT analysis of ChatGPT: Implications for educational practice and research. Innovations in Education and Teaching International, 61(3), 460–474. https://doi.org/10.1080/14703297.2023.2195846
[40] Ferguson, R. (2012). Learning analytics: Drivers, developments and challenges. International Journal of Technology Enhanced Learning, 4(5–6), 304–317. https://doi.org/10.1504/IJTEL.2012.051816

© 2026 TRTEE. Open Access under CC BY 4.0

[41] Finnie-Ansley, J., Denny, P., Becker, B. A., Luxton-Reilly, A., & Prather, J. (2022). The robots are coming: Exploring the implications of OpenAI Codex on introductory programming. Proceedings of the 24th Australasian Computing Education Conference, 10–19. https://doi.org/10.1145/3511861.3511863
[42] Frich, J., MacDonald Vermeulen, L., Remy, C., Biskjaer, M. M., & Dalsgaard, P. (2019). Mapping the landscape of creativity support tools in HCI. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, 1–18. https://doi.org/10.1145/3290605.3300619
[43] Fryer, L. K., Ainley, M., Thompson, A., Gibson, A., & Sherlock, Z. (2017). Stimulating and sustaining interest in a language course: An experimental comparison of chatbot and human task partners. Computers in Human Behavior, 75, 461–468. https://doi.org/10.1016/j.chb.2017.05.045
[44] Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. TechTrends, 59(1), 64–71. https://doi.org/10.1007/s11528-014-0822-x
[45] Gero, K. I., Liu, V., & Chilton, L. (2022). Sparks: Inspiration for science writing using language models. Proceedings of the 2022 ACM Designing Interactive Systems Conference, 1002–1019. https://doi.org/10.1145/3532106.3533533
[46] Glăveanu, V. P. (2013). Rewriting the language of creativity: The five A's framework. Review of General Psychology, 17(1), 69–81. https://doi.org/10.1037/a0029528
[47] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2020). Generative adversarial networks. Communications of the ACM, 63(11), 139–144. https://doi.org/10.1145/3422622
[48] Graham, S., Hebert, M., & Harris, K. R. (2015). Formative assessment and writing: A meta-analysis. The Elementary School Journal, 115(4), 523–547. https://doi.org/10.1086/681947
[49] Grover, S., & Pea, R. (2013). Computational thinking in K–12: A review of the state of the field. Educational Researcher, 42(1), 38–43. https://doi.org/10.3102/0013189X12463051
[50] Gulikers, J. T. M., Bastiaens, T. J., & Kirschner, P. A. (2004). A five-dimensional framework for authentic assessment. Educational Technology Research and Development, 52(3), 67–86. https://doi.org/10.1007/BF02504676
[51] Haase, J., & Hanel, P. H. P. (2023). Artificial muses: Generative artificial intelligence chatbots have risen to human-level creativity. Journal of Creativity, 33(3), 100066. https://doi.org/10.1016/j.yjoc.2023.100066
[52] Hattie, J., & Timperley, H. (2007). The power of feedback. Review of Educational Research, 77(1), 81–112. https://doi.org/10.3102/003465430298487
[53] Holmes, W., Porayska-Pomsta, K., Holstein, K., Sutherland, E., Baker, T., Shum, S. B., Santos, O. C., Rodrigo, M. T., Cukurova, M., Bittencourt, I. I., & Koedinger, K. R. (2022). Ethics of AI in education: Towards a community-wide framework. International Journal of Artificial Intelligence in Education, 32(3), 504–526. https://doi.org/10.1007/s40593-021-00239-1
[54] Hwang, G.-J., Xie, H., Wah, B. W., & Gašević, D. (2020). Vision, challenges, roles and research issues of artificial intelligence in education. Computers and Education: Artificial Intelligence, 1, 100001. https://doi.org/10.1016/j.caeai.2020.100001
[55] Ifenthaler, D., & Yau, J. Y.-K. (2020). Utilising learning analytics to support study

© 2026 TRTEE. Open Access under CC BY 4.0

success in higher education: A systematic review. Educational Technology Research and Development, 68(4), 1961–1990. https://doi.org/10.1007/s11423-020-09788-z

[56] Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y. J., Madotto, A., & Fung, P. (2023). Survey of hallucination in natural language generation. ACM Computing Surveys, 55(12), 1–38. https://doi.org/10.1145/3571730

[57] Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature Machine Intelligence, 1(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2

[58] Joksimović, S., Kovanović, V., & Dawson, S. (2019). The journey of learning analytics. HERDSA Review of Higher Education, 6, 37–63. https://herdsa.org.au/herdsa-review-higher-education-vol-6/37-63

[59] Jordanous, A. (2012). A standardised procedure for evaluating creative systems: Computational creativity evaluation based on what it is to be creative. Cognitive Computation, 4(3), 246–279. https://doi.org/10.1007/s12559-012-9156-1

[60] 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, 102274. https://doi.org/10.1016/j.lindif.2023.102274

[61] Kaufman, J. C., & Beghetto, R. A. (2009). Beyond big and little: The four C model of creativity. Review of General Psychology, 13(1), 1–12. https://doi.org/10.1037/a0013688

[62] Kazemitabaar, M., Chow, J., Ma, C. K. T., Ericson, B. J., Weintrop, D., & Grossman, T. (2023). Studying the effect of AI code generators on supporting novice learners in introductory programming. Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, 1–23. https://doi.org/10.1145/3544548.3580919

[63] Khalil, M., Prinsloo, P., & Slade, S. (2023). Fairness, trust, transparency, equity, and responsibility in learning analytics. Journal of Learning Analytics, 10(1), 1–7. https://doi.org/10.18608/jla.2023.7983

[64] Khosravi, H., Buckingham Shum, S., Chen, G., Conati, C., Tsai, Y.-S., Kay, J., Knight, S., Martinez-Maldonado, R., Sadiq, S., & Gašević, D. (2022). Explainable artificial intelligence in education. Computers and Education: Artificial Intelligence, 3, 100074. https://doi.org/10.1016/j.caeai.2022.100074

[65] Koedinger, K. R., Corbett, A. T., & Perfetti, C. (2012). The Knowledge-Learning-Instruction framework: Bridging the science-practice chasm to enhance robust student learning. Cognitive Science, 36(5), 757–798. https://doi.org/10.1111/j.1551-6709.2012.01245.x

[66] Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. RELC Journal, 54(2), 537–550. https://doi.org/10.1177/00336882231162868

[67] Koivisto, M., & Grassini, S. (2023). Best humans still outperform artificial intelligence in a creative divergent thinking task. Scientific Reports, 13, 13601. https://doi.org/10.1038/s41598-023-40858-3

[68] Kuhail, M. A., Alturki, N., Alramlawi, S., & Alhejori, K. (2023). Interacting with educational chatbots: A systematic review. Education and Information

© 2026 TRTEE. Open Access under CC BY 4.0

Technologies, 28(1), 973–1018. https://doi.org/10.1007/s10639-022-11177-3
[69] Kulik, J. A., & Fletcher, J. D. (2016). Effectiveness of intelligent tutoring systems: A meta-analytic review. Review of Educational Research, 86(1), 42–78. https://doi.org/10.3102/0034654315581420
[70] Larrabee Sønderlund, A., Hughes, E., & Smith, J. (2019). The efficacy of learning analytics interventions in higher education: A systematic review. British Journal of Educational Technology, 50(5), 2594–2618. https://doi.org/10.1111/bjet.12720
[71] Lee, D., Arnold, M., Srivastava, A., Plastow, K., Strelan, P., Ploeckl, F., Lekkas, D., & Palmer, E. (2024). The impact of generative AI on higher education learning and teaching: A study of educators' perspectives. Computers and Education: Artificial Intelligence, 6, 100221. https://doi.org/10.1016/j.caeai.2024.100221
[72] Lee, M., Liang, P., & Yang, Q. (2022). CoAuthor: Designing a human-AI collaborative writing dataset for exploring language model capabilities. Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, 1–19. https://doi.org/10.1145/3491102.3502030
[73] Liang, W., Yuksekgonul, M., Mao, Y., Wu, E., & Zou, J. (2023). GPT detectors are biased against non-native English writers. Patterns, 4(7), 100779. https://doi.org/10.1016/j.patter.2023.100779
[74] Lo, C. K. (2023). What is the impact of ChatGPT on education? A rapid review of the literature. Education Sciences, 13(4), 410. https://doi.org/10.3390/educsci13040410
[75] Lodge, J. M., & Corrin, L. (2017). What data and analytics can and do say about effective learning. npj Science of Learning, 2, 5. https://doi.org/10.1038/s41539-017-0006-5
[76] Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, 1–16. https://doi.org/10.1145/3313831.3376727
[77] Lubart, T. (2005). How can computers be partners in the creative process: Classification and commentary on the special issue. International Journal of Human-Computer Studies, 63(4–5), 365–369. https://doi.org/10.1016/j.ijhcs.2005.04.002
[78] Luckin, R., & Cukurova, M. (2019). Designing educational technologies in the age of AI: A learning sciences-driven approach. British Journal of Educational Technology, 50(6), 2824–2838. https://doi.org/10.1111/bjet.12861
[79] Ma, W., Adesope, O. O., Nesbit, J. C., & Liu, Q. (2014). Intelligent tutoring systems and learning outcomes: A meta-analysis. Journal of Educational Psychology, 106(4), 901–918. https://doi.org/10.1037/a0037123
[80] Mandinach, E. B., & Gummer, E. S. (2016). What does it mean for teachers to be data literate: Laying out the skills, knowledge, and dispositions. Teaching and Teacher Education, 60, 366–376. https://doi.org/10.1016/j.tate.2016.07.011
[81] Matcha, W., Uzir, N. A., Gašević, D., & Pardo, A. (2020). A systematic review of empirical studies on learning analytics dashboards: A self-regulated learning perspective. IEEE Transactions on Learning Technologies, 13(2), 226–245. https://doi.org/10.1109/TLT.2019.2916802
[82] Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. Teachers College Record, 108(6), 1017–

© 2026 TRTEE. Open Access under CC BY 4.0

1054. https://doi.org/10.1111/j.1467-9620.2006.00684.x
[83] Nazaretsky, T., Ariely, M., Cukurova, M., & Alexandron, G. (2022). Teachers' trust in AI-powered educational technology and a professional development program to improve it. British Journal of Educational Technology, 53(4), 914–931. https://doi.org/10.1111/bjet.13232
[84] Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. Computers and Education: Artificial Intelligence, 2, 100041. https://doi.org/10.1016/j.caeai.2021.100041
[85] Nguyen, A., Ngo, H. N., Hong, Y., Dang, B., & Nguyen, B.-P. T. (2023). Ethical principles for artificial intelligence in education. Education and Information Technologies, 28(4), 4221–4241. https://doi.org/10.1007/s10639-022-11316-w
[86] Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. Studies in Higher Education, 31(2), 199–218. https://doi.org/10.1080/03075070600572090
[87] Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. Science, 381(6654), 187–192. https://doi.org/10.1126/science.adh2586
[88] Ouyang, F., & Jiao, P. (2021). Artificial intelligence in education: The three paradigms. Computers and Education: Artificial Intelligence, 2, 100020. https://doi.org/10.1016/j.caeai.2021.100020
[89] Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ, 372, n71. https://doi.org/10.1136/bmj.n71
[90] Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for research. Frontiers in Psychology, 8, 422. https://doi.org/10.3389/fpsyg.2017.00422
[91] Perkins, M. (2023). Academic integrity considerations of AI large language models in the post-pandemic era: ChatGPT and beyond. Journal of University Teaching and Learning Practice, 20(2), 07. https://doi.org/10.53761/1.20.02.07
[92] Perrotta, C., & Selwyn, N. (2020). Deep learning goes to school: Toward a relational understanding of AI in education. Learning, Media and Technology, 45(3), 251–269. https://doi.org/10.1080/17439884.2020.1686017
[93] Plucker, J. A., Beghetto, R. A., & Dow, G. T. (2004). Why isn't creativity more important to educational psychologists? Potentials, pitfalls, and future directions in creativity research. Educational Psychologist, 39(2), 83–96. https://doi.org/10.1207/s15326985ep3902_1
[94] Popenici, S. A. D., & Kerr, S. (2017). Exploring the impact of artificial intelligence on teaching and learning in higher education. Research and Practice in Technology Enhanced Learning, 12, 22. https://doi.org/10.1186/s41039-017-0062-8
[95] Prather, J., Denny, P., Leinonen, J., Becker, B. A., Albluwi, I., Craig, M., Keuning, H., Kiesler, N., Kohn, T., Luxton-Reilly, A., MacNeil, S., Petersen, A., Pettit, R., Reeves, B. N., & Savelka, J. (2023). The robots are here: Navigating the generative AI revolution in computing education. Proceedings of the 2023

© 2026 TRTEE. Open Access under CC BY 4.0

Working Group Reports on Innovation and Technology in Computer Science Education, 108–159. https://doi.org/10.1145/3623762.3633499
[96] Prinsloo, P., & Slade, S. (2017). An elephant in the learning analytics room: The obligation to act. Proceedings of the Seventh International Learning Analytics & Knowledge Conference, 46–55. https://doi.org/10.1145/3027385.3027406
[97] Resnick, M., Maloney, J., Monroy-Hernández, A., Rusk, N., Eastmond, E., Brennan, K., Millner, A., Rosenbaum, E., Silver, J., Silverman, B., & Kafai, Y. (2009). Scratch: Programming for all. Communications of the ACM, 52(11), 60–67. https://doi.org/10.1145/1592761.1592779
[98] Rezwana, J., & Maher, M. L. (2023). Designing creative AI partners with COFI: A framework for modeling interaction in human-AI co-creative systems. ACM Transactions on Computer-Human Interaction, 30(5), 1–28. https://doi.org/10.1145/3519026
[99] 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
[100] Roll, I., & Wylie, R. (2016). Evolution and revolution in artificial intelligence in education. International Journal of Artificial Intelligence in Education, 26(2), 582–599. https://doi.org/10.1007/s40593-016-0110-3
[101] Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 10684–10695. https://doi.org/10.1109/CVPR52688.2022.01042
[102] Runco, M. A., & Jaeger, G. J. (2012). The standard definition of creativity. Creativity Research Journal, 24(1), 92–96. https://doi.org/10.1080/10400419.2012.650092
[103] Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. American Psychologist, 55(1), 68–78. https://doi.org/10.1037/0003-066X.55.1.68
[104] Selwyn, N. (2022). The future of AI and education: Some cautionary notes. European Journal of Education, 57(4), 620–631. https://doi.org/10.1111/ejed.12532
[105] Shneiderman, B. (2007). Creativity support tools: Accelerating discovery and innovation. Communications of the ACM, 50(12), 20–32. https://doi.org/10.1145/1323688.1323689
[106] Shneiderman, B. (2020). Human-centered artificial intelligence: Reliable, safe & trustworthy. International Journal of Human-Computer Interaction, 36(6), 495–504. https://doi.org/10.1080/10447318.2020.1741118
[107] Shute, V. J. (2008). Focus on formative feedback. Review of Educational Research, 78(1), 153–189. https://doi.org/10.3102/0034654307313795
[108] Siemens, G. (2013). Learning analytics: The emergence of a discipline. American Behavioral Scientist, 57(10), 1380–1400. https://doi.org/10.1177/0002764213498851
[109] Siemens, G., & Baker, R. S. J. d. (2012). Learning analytics and educational data mining: Towards communication and collaboration. Proceedings of the 2nd International Conference on Learning Analytics and Knowledge, 252–254. https://doi.org/10.1145/2330601.2330661
[110] Slade, S., & Prinsloo, P. (2013). Learning analytics: Ethical issues and dilemmas. American Behavioral Scientist, 57(10), 1510–1529.

© 2026 TRTEE. Open Access under CC BY 4.0

https://doi.org/10.1177/0002764213479366
[111] Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. Journal of Business Research, 104, 333–339. https://doi.org/10.1016/j.jbusres.2019.07.039
[112] Sternberg, R. J., & Lubart, T. I. (1996). Investing in creativity. American Psychologist, 51(7), 677–688. https://doi.org/10.1037/0003-066X.51.7.677
[113] Stevenson, M., & Phakiti, A. (2014). The effects of computer-generated feedback on the quality of writing. Assessing Writing, 19, 51–65. https://doi.org/10.1016/j.asw.2013.11.007
[114] Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. Cognitive Science, 12(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4
[115] 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
[116] Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel? ChatGPT as a case study of using chatbots in education. Smart Learning Environments, 10, 15. https://doi.org/10.1186/s40561-023-00237-x
[117] Torraco, R. J. (2005). Writing integrative literature reviews: Guidelines and examples. Human Resource Development Review, 4(3), 356–367. https://doi.org/10.1177/1534484305278283
[118] VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. Educational Psychologist, 46(4), 197–221. https://doi.org/10.1080/00461520.2011.611369
[119] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. arXiv. https://doi.org/10.48550/arXiv.1706.03762
[120] Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. MIS Quarterly, 27(3), 425–478. https://doi.org/10.2307/30036540
[121] Viberg, O., Hatakka, M., Bälter, O., & Mavroudi, A. (2018). The current landscape of learning analytics in higher education. Computers in Human Behavior, 89, 98–110. https://doi.org/10.1016/j.chb.2018.07.027
[122] Villarroel, V., Bloxham, S., Bruna, D., Bruna, C., & Herrera-Seda, C. (2018). Authentic assessment: Creating a blueprint for course design. Assessment & Evaluation in Higher Education, 43(5), 840–854. https://doi.org/10.1080/02602938.2017.1412396
[123] Vinchon, F., Lubart, T., Bartolotta, S., Gironnay, V., Botella, M., Bourgeois-Bougrine, S., Burkhardt, J.-M., Bonnardel, N., Corazza, G. E., Glaveanu, V., Hanchett Hanson, M., Ivcevic, Z., Karwowski, M., Kaufman, J. C., Okada, T., Reiter-Palmon, R., & Gaggioli, A. (2023). Artificial intelligence & creativity: A manifesto for collaboration. The Journal of Creative Behavior, 57(4), 472–484. https://doi.org/10.1002/jocb.597
[124] Warschauer, M., & Matuchniak, T. (2010). New technology and digital worlds: Analyzing evidence of equity in access, use, and outcomes. Review of Research in Education, 34(1), 179–225. https://doi.org/10.3102/0091732X09349791
[125] Weber-Wulff, D., Anohina-Naumeca, A., Bjelobaba, S., Foltýnek, T., Guerrero-

© 2026 TRTEE. Open Access under CC BY 4.0

Dib, J., Popoola, O., Šigut, P., & Waddington, L. (2023). Testing of detection tools for AI-generated text. International Journal for Educational Integrity, 19, 26. https://doi.org/10.1007/s40979-023-00146-z

[126] Williamson, B., & Eynon, R. (2020). Historical threads, missing links, and future directions in AI in education. Learning, Media and Technology, 45(3), 223–235. https://doi.org/10.1080/17439884.2020.1798995

[127] Wilson, J., & Roscoe, R. D. (2020). Automated writing evaluation and feedback: Multiple metrics of efficacy. Journal of Educational Computing Research, 58(1), 87–125. https://doi.org/10.1177/0735633119830764

[128] Wing, J. M. (2006). Computational thinking. Communications of the ACM, 49(3), 33–35. https://doi.org/10.1145/1118178.1118215

[129] Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. Journal of Child Psychology and Psychiatry, 17(2), 89–100. https://doi.org/10.1111/j.1469-7610.1976.tb00381.x

[130] 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

[131] Yan, L., Sha, L., Zhao, L., Li, Y., Martinez-Maldonado, R., Chen, G., Li, X., Jin, Y., & Gašević, D. (2024). Practical and ethical challenges of large language models in education: A systematic scoping review. British Journal of Educational Technology, 55(1), 90–112. https://doi.org/10.1111/bjet.13370

[132] Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education — where are the educators? International Journal of Educational Technology in Higher Education, 16, 39. https://doi.org/10.1186/s41239-019-0171-0

[133] Zhai, X., Chu, X., Chai, C. S., Jong, M. S. Y., Istenic, A., Spector, M., Liu, J.-B., Yuan, J., & Li, Y. (2021). A review of artificial intelligence (AI) in education from 2010 to 2020. Complexity, 2021, 8812542. https://doi.org/10.1155/2021/8812542

[134] Zhang, R., & Zou, D. (2022). Types, purposes, and effectiveness of state-of-the-art technologies for second and foreign language learning. Computer Assisted Language Learning, 35(4), 696–742. https://doi.org/10.1080/09588221.2020.1744666

[135] 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

[136] van Dijk, J. A. G. M. (2006). Digital divide research, achievements and shortcomings. Poetics, 34(4–5), 221–235. https://doi.org/10.1016/j.poetic.2006.05.004

© 2026 TRTEE. Open Access under CC BY 4.0