PROCEEDINGS
Kalijaga International Conference on Education 2026
Faculty of Tarbiyah and Teacher Training, UIN Sunan Kalijaga
Innovation Toward Education: Shaping the Future of Learning
15 September 2026 | Convention Hall, UIN Sunan Kalijaga, Yogyakarta

CONFERENCE PAPER • KICE 2026 • YOGYAKARTA, INDONESIA

AI, Gamification, and Interactive Learning: Designing PdPC for a New Generation of Learners

M. Saidul Muzakki[1]

[1] Bachelor of Education for Islamic Elementary School Teachers Program UIN Sunan Kalijaga Yogyakarta, Indonesia

[1]Email: m.muzakki@uin-suka.ac.id

Proceedings of the Kalijaga International Conference on Education 2026

Abstract. Generative artificial intelligence (AI) moved from novelty to classroom infrastructure in fewer than three years, while learners arrive with habits shaped by fast, visual, mobile, algorithmically personalized environments. Teachers and lecturers are therefore asked to redesign pengajaran dan pemudahcaraan (PdPC) for a new generation of learners without surrendering critical thinking, creativity, ethical judgment, and independent learning. This paper is a conceptual and integrative review whose objective is to establish what the current evidence supports about AI, gamification, and interactive quizzing in education, and to translate that evidence into usable design guidance for teacher education. Forty-five sources were purposively selected from peer-reviewed meta-analyses, systematic reviews, and official policy frameworks, screened for relevance to instructional design in higher and teacher education, and verified against Crossref and OpenAlex records for correct metadata and absence of retraction notices. Three findings emerge. First, average effects for generative AI, gamification, and game-based student response systems are positive but moderate and highly heterogeneous, and they depend on instructional design rather than on the tool. Second, the same technologies that raise engagement can displace the cognitive effort that produces learning, so design must deliberately protect productive struggle. Third, as execution becomes inexpensive, the teacher's pedagogical judgment becomes more consequential rather than less so. The paper contributes three artifacts: a Human-Centered AI-Interactive Learning Cycle organized into four movements and eight phases; a six-gate decision framework for instructional technology selection; and a selection rubric whose decisive criterion is a subtraction test. The central conclusion is that intelligent technology does not automatically produce intelligent learning.

Keywords: *AI literacy; formative assessment; gamification; generative artificial intelligence; instructional design; teacher education*

Abbreviations: AI, artificial intelligence; GenAI, generative artificial intelligence; HAILC, Human-Centered AI-Interactive Learning Cycle; ICAP, Interactive-Constructive-Active-Passive framework; PdPC, pengajaran dan pemudahcaraan (teaching and facilitation); SDT, self-determination theory; SRL, self-regulated learning; TPACK, technological pedagogical content knowledge; UDL, Universal Design for Learning.

INTRODUCTION

The ecology of teaching has moved through four overlapping shifts. It began teacher-centered, with the lecturer as the sole source of content. It became learner-centered, with the aim shifting from covering material to developing understanding. It became technology-enhanced, where digital media widened access to information and to one another. It is now becoming AI-supported, with a system beside the learner that can generate explanations, examples, feedback, images, quizzes, and entire lesson plans in seconds.

Each shift changed what was scarce. When information was scarce, the teacher who held it was indispensable. When information became abundant through the internet, what became scarce was attention and the ability to judge quality. Now that plausible-sounding answers have become abundant, what has become scarce is something else again: the capacity to formulate a good question, to evaluate an answer critically, to make a defensible decision, and to know when one has actually understood something. The Organization for Economic Co-operation and Development has argued that education systems must move beyond incremental adjustments and reconsider which competencies remain essential, as machines can perform much of what curricula currently reward [1].

This raises an uncomfortable question that every honest educator has now asked privately: if AI can explain, write, illustrate, build presentations, generate assessment items, analyze data, and even help design instruction, what remains distinctly human in education? Two answers are unhelpful. The first is refusal, which mostly redistributes advantage to students who conceal their use well. The second is capitulation, which assumes that more technology equals better learning, an assumption the evidence does not support.

A gap between research and practice sharpens the problem. A systematic review of AI applications in higher education found the field dominated by computer science perspectives, with weak links to pedagogical theory, and asked, pointedly, where the educators were [2]. A subsequent meta-systematic review renewed the call for greater ethical consideration, cross-disciplinary collaboration, and methodological rigor [3], and a further review of opportunities and challenges likewise called for research more closely tied to classroom practice [4]. In Indonesia, adoption studies report that institutional readiness, digital leadership, and perceived usefulness, rather than tool availability, govern whether teachers and universities actually integrate AI [5], [6]. The gap is therefore not one of access but of design capability.

This paper addresses that gap. Its purpose is threefold: to establish what current meta-analytic and policy evidence actually supports about generative AI, gamification, and interactive quizzing in education; to identify the conditions under which these technologies help or harm learning; and to translate those conditions into design artifacts that teachers and teacher educators can apply directly. Two commitments frame the argument. The first is evidential: claims are stated according to the strength of the research supporting them, and contested claims are labeled as such. The second is pedagogical: the unit of analysis is never the tool but the learning experience, defined as what the student is asked to think about, with what support, producing what evidence, judged against what standard.

METHODS

Design of the review

This study is a conceptual and integrative review rather than an empirical investigation. Its purpose is synthesis and theory-building: to assemble the strongest available quantitative evidence about three families of educational technology, to identify the boundary conditions that explain heterogeneous results, and to derive design principles from those conditions. No human participants were involved, and no primary data were collected.

Source selection

Sources were selected purposively rather than exhaustively, following the logic of a critical review in which coverage is judged by the strength of evidence rather than by completeness of retrieval. Searches were conducted in Crossref, OpenAlex, and general academic search engines using combinations of the terms generative artificial intelligence, ChatGPT, gamification, game-based student response system, Kahoot, formative assessment, retrieval practice, self-regulated learning, student agency, and AI literacy, restricted where possible to publications from 2021 onwards.

Four inclusion criteria were applied. First, priority was given to meta-analyses and systematic reviews over single studies, because the design question concerns average effects and their moderators. Second, official frameworks and guidance from international organizations were included, which define competencies or set policy expectations. Third, foundational theoretical works were retained where they provide the mechanism explaining a contemporary finding; the template's recency guideline explicitly permits this exception for theoretical work. Fourth, sources had to be verifiable.

Verification procedure

Every reference was checked programmatically against Crossref and OpenAlex records. Titles, author lists, journal, volume, issue, page or article numbers, and publication years were confirmed against the registered metadata; all digital object identifiers were confirmed to resolve, and each record was screened for retraction notices, expressions of concern, and errata. No cited work carries a retraction notice. Two sources are flagged in the text as provisional rather than settled: one cross-sectional study that cannot establish causation, and one working paper that has not completed peer review.

Composition of the evidence base

The final corpus contains 45 sources. Twenty-four were published between 2021 and 2026 and carry the empirical weight of the argument; the remaining 21 are foundational works on motivation, cognition, assessment, and technology integration, retained because they supply explanatory mechanisms that recent meta-analyses assume rather than restate. Of the recent sources, 11 are meta-analyses or systematic reviews, and four are official frameworks or policy analyses.

Analytical approach

Evidence was analyzed in three passes. The first extracted reported effect sizes with confidence intervals where available, together with the moderators the original authors identified. The second examined convergence and divergence across syntheses in the same domain, treating disagreement between meta-analyses as informative rather than as noise. The third mapped the resulting boundary conditions onto

established instructional-design constructs, from which the three design artifacts presented in this paper were derived. Because the review is purposive rather than exhaustive, its conclusions are offered as design guidance rather than as pooled estimates, and the limitations of this approach are stated before the conclusions.

RESULTS AND DISCUSSION

The new generation of learners, without the stereotype

The phrase 'Generation Z' is useful because the learning environment of a 19-year-old in 2026 genuinely differs from that of a 19-year-old in 2006. However, it is dangerous when it slides into the claim that a cohort shares a fixed cognitive profile. That slide has a name in the literature: the *digital native myth*. A review of the evidence found no support for two popular beliefs: that young people who grew up with technology form a homogeneous generation of skilled digital learners, and that they can multitask effectively while learning [7]. Both beliefs are widespread among educators and damage instruction: if teachers assume students already know how to learn with technology, they stop teaching them how to do so.

What is defensible is a description of the environment rather than of the person. That environment is digital and mobile, fast and on-demand, visual and multimedia, social and networked, personalized by algorithms, and increasingly AI-assisted. Two distinctions follow. Being a digital native is not the same as being digitally literate: operating an interface is acquired by exposure, whereas evaluating a source, recognizing a manipulated image, or detecting a fabricated citation must be taught. Being able to use technology is not the same as being able to learn with it, because learning with a tool requires goal-setting, monitoring, self-testing, and revision, which are the components of self-regulated learning [8].

The implication is that this generation needs more explicit instruction, not less. AI literacy in particular is not a technical subject: it has been defined as the set of competencies that enable people to critically evaluate, communicate, and collaborate with AI systems [9], and most conceptualizations combine knowing what AI is, using it, evaluating it, and reasoning about its ethics [10]. The UNESCO competency frameworks make the same point structurally. The student framework organizes competence around a human-centered mindset, the ethics of AI, AI techniques and applications, and AI system design, progressing from understanding to applying to creating [11]; the teacher framework adds AI pedagogy and professional learning [12]. A final caution is distributional: access to devices, connectivity, and paid tools is uneven within a single classroom, so a design that assumes universal access converts a pedagogical choice into an equity problem.

Personal intelligence and artificial intelligence

Personal intelligence is the capacity to understand oneself as a learner: to recognize one's goals, strengths, gaps, motivations, and characteristic learning patterns. In instructional terms, it overlaps closely with metacognition and self-regulated learning [8]. Artificial intelligence, in the sense relevant to classrooms, is the capacity of computational systems to perform tasks that normally require human intelligence. The relationship between them is complementary: personal intelligence helps a person understand themselves, while artificial intelligence helps a person extend what they can do.

This is not a rhetorical flourish; it has a mechanism. A generative model produces fluent output regardless of whether the user understands the topic. A learner with well-developed personal intelligence treats that output as material to evaluate, asking whether it is correct, what it assumes, and how it differs from their own account. A learner without it treats the same output as a substitute for thinking. The same tool, therefore, produces opposite educational effects depending on the learner's metacognitive capacity, which is precisely the variable that teachers can influence. The central proposition of this paper is that more powerful AI does not automatically produce more capable learners. That capability grows only when instructional design deliberately couples human judgment with machine capability.

Generative AI in learning: uses, evidence, and risks

Used deliberately, generative AI supports a wide range of instructional work: brainstorming and framing; explaining concepts at multiple levels; producing worked examples and deliberate non-examples; drafting assessment items and rubrics for teacher review; differentiating tasks, which is where AI intersects most usefully with Universal Design for Learning by making multiple means of representation, engagement, and expression practical rather than aspirational [13]; drafting feedback; generating media, simulations, and role-play partners; and developing teaching materials. For teacher educators, one use deserves emphasis: AI makes it inexpensive to produce imperfect artifacts for students to evaluate. A

flawed lesson plan, a biased assessment item, or a poorly differentiated worksheet is an excellent teaching material, and it used to be costly to create.

Recent syntheses converge on a moderate positive average effect on achievement. A meta-analysis of 37 studies reported an overall effect size g = 0.577 (95% confidence interval: [0.395, 0.759]) [14]. A synthesis of 17 studies with 1,735 students found a medium effect on overall student engagement, with the largest effect on cognitive engagement [15]. A systematic review and meta-analysis of 69 experimental studies found that ChatGPT interventions improved academic performance, affective-motivational states, and higher-order thinking propensities, reduced mental effort, and had no significant effect on self-efficacy [16]. On the self-regulation side, a qualitative systematic review found that AI applications can support the forethought, performance, and reflection phases of self-regulated learning, but outcomes vary depending on whether agency remains human-centered or shifts to the system [17]. A three-level meta-analysis of AI feedback reported positive effects on self-regulated learning in higher education while identifying substantial moderator variation [18].

Four cautions belong beside those numbers. Heterogeneity is large, so that AI works is not a finding, whereas AI worked under these conditions is. Novelty effects are unresolved, since most studies are short and the original authors explicitly caution that positive affective-motivational results may reflect novelty rather than durable change [16]. Post-test design can inflate results: when the outcome is a well-defined task the model can help with, the study may be measuring output quality rather than learning, which is why a shift toward complex, project-based, or proctored assessment has been recommended [16]. Finally, the reduction in mental effort is the finding educators should sit with the longest. Reduced effort is desirable when the effort was wasted, such as formatting a table, and dangerous when the effort was the learning, such as constructing an explanation or retrieving from memory.

The risks are equally concrete. Hallucination, including invented references and identifiers, remains a structural property of generative models rather than a defect that has been repaired. Bias in training data surfaces in examples, images, and evaluative language. Cognitive offloading is a live concern: a cross-sectional study reported a negative association between frequent AI tool use and critical-thinking scores, with cognitive offloading as the mediating variable [19]. The design is correlational and cannot establish causation, but the mechanism it describes warrants attention in the design rather than dismissal. Displaced learning has been observed directly: in a large field experiment with secondary students, access to an unguarded assistant improved performance during practice but reduced performance on subsequent unaided examinations, while a version with pedagogical safeguards mitigated the harm [20]; that study is a working paper and should be treated as provisional, but its shape matters, because help during practice can be help that prevents learning. Academic integrity, undisclosed use, copyright, data privacy for minors, and unequal access complete the list. The operative principle is that AI should augment learning rather than replace it, which means the teacher decides, for each task, which cognitive work belongs to the student and is therefore protected from automation.

From answer machine to thinking partner

Consider two students working on the same assignment. The first writes: AI, write my assignment on media ethics. The second writes: give me three competing positions on whether schools should ban smartphones, with the strongest argument and the weakest point for each; I will evaluate them, choose one, and tell you where I disagree with you. The tool is identical; the cognitive work is not. In the first case, the student contributes a prompt and a copy-paste. In the second, the student sets the frame, judges quality, takes a position, and defends it, with the model supplying raw material for that judgment.

The ICAP framework explains why this distinction predicts learning: outcomes improve as students move from passive reception to active manipulation, to constructive generation of something beyond what was given, to interactive dialogue in which partners build on each other's contributions [21]. Prompt-copy-submit is passive learning in new clothes; structured argumentation with an AI partner followed by peer discussion can be genuinely constructive and interactive. Three moves make this operational. First, require the trace rather than only the product: the prompt sequence, the output, the student's critique of it, and the revised version, where the critique is the assessed learning. Second, assign the model a role that requires resistance, such as a skeptical reviewer, the strongest objection, or a generator of three explanations, one of which contains an error. Third, make permitted AI use explicit and graduated; an AI assessment scale that specifies, for each task, the permitted level of involvement from none, through assisted planning or editing, to full collaboration with disclosure, provides that clarity [22]. Students most often misuse AI when the rules are ambiguous, and the stakes are high.

Gamification: evidence and boundary conditions

Gamification is the use of game design elements in non-game contexts [23]. It differs from game-based learning, in which the game itself is the learning environment. The distinction matters in practice: gamification is inexpensive to add and easy to do badly, whereas game-based learning is costly to build and harder to align. Most teachers will gamify; few will build games.

Meta-analyses agree that, on average, gamification helps, but disagree about by how much. Across 30 interventions, an effect size of g = 0.504 (95% confidence interval: [0.284, 0.723]) was reported for academic performance [24]. Across 30 independent studies with 3,083 participants, an effect size of g = 0.464 (95% confidence interval: [0.244, 0.684]) was reported [25]. A further meta-analysis found significant but small-to-moderate effects on cognitive, motivational, and behavioral outcomes [26]. Examining the motivational claim directly through a self-determination lens produced the most instructive pattern: a small effect on intrinsic motivation, g = 0.257, a moderate effect on perceived autonomy, g = 0.638, a large effect on relatedness, g = 1.776, and minimal impact on perceived competence, g = 0.277 [27].

Read together, these results tell a more interesting story than gamification works. Gamification reliably improves how connected and autonomous students feel and modestly improves performance, but it does not reliably improve perceived competence, and its effect on intrinsic motivation is small. In terms of self-determination, competence is the need most directly tied to durable motivation [28], and it is the need that points and badges address least well. A cautionary case reinforces the point: a semester-long study found that a gamified course produced lower intrinsic motivation, satisfaction, and empowerment over time, and lower final examination performance, than its non-gamified counterpart [29]. The most plausible explanation was established in a meta-analysis decades earlier, namely that tangible, expected, performance-contingent rewards can undermine intrinsic motivation for an activity a person already finds interesting [30].

Three practical corollaries follow. Points should be tied to the intended behavior for repetition rather than to speed alone, because a leaderboard that rewards fast guessing will produce fast guessing. Competition is not universally motivating, and for students with low perceived competence, a public ranking can reinforce a negative self-image; therefore, team-based scoring, personal-best tracking, and anonymous participation are alternatives rather than compromises. Short bursts outperform permanent regimes because effects attenuate as novelty fades.

Interactive quizzing: the Kahoot! case

Kahoot! is treated here as a case study of a class of tools, game-based student response systems, rather than as a product recommendation; the analysis applies equally to Quizizz, Mentimeter, Wordwall, Blooket, Socrative, or a form with instant feedback. A review of 93 studies concluded that such systems can positively affect learning performance, classroom dynamics, attitudes of students and teachers, and student anxiety, while also noting studies in which the effect was small or absent [31]. A meta-analysis of 43 experimental studies, comprising 1,706 experimental and 1,647 control participants, reported a significant positive effect on academic achievement, d = 0.772, 95% confidence interval [0.540, 1.004], a large effect on retention, 1.492, a large effect on motivation, 0.960, a medium effect on attitude, 0.678, and a small negative effect on anxiety, meaning a reduction, of -0.338 [32]. A further review concluded that the platform improves learning outcomes and enhances interaction when used appropriately [33]. Two words carry the weight in that conclusion: appropriately used.

Placement decides purpose. Before learning, a short quiz surfaces prior knowledge and misconceptions; this is the most underused and arguably most valuable application, because it changes what the teacher does next and requires items whose incorrect options each correspond to a specific misconception, so that the result becomes diagnostic information rather than a score. During learning, a single well-designed item answers whether an explanation landed, and a deliberately contentious item that splits the class becomes a discussion trigger; the productive sequence is pose, vote, discuss in pairs, re-vote, explain. After learning, the strongest evidence applies, and it has little to do with games: retrieval practice, the act of pulling information out of memory rather than reading it again, produces durable learning [34], and a systematic meta-analytic review confirmed that classroom quizzing boosts learning [35]. Immediate feedback is the second mechanism, with meta-analytic evidence indicating a medium-sized effect, and information-rich feedback substantially outperforming mere marking or praise [36]. A platform that displays only red or green delivers a fraction of the available benefit; the teacher's brief explanation after each item is where the pedagogy occurs.

The same literature documents the weaknesses [31]. Competition creates pressure for some students, and public ranking can demotivate those who most need encouragement. Speed is not depth, since scoring

by response time rewards both recall and quick guessing, and penalizes the reflective student. Leaderboards do not suit every context, particularly small classes, sensitive topics, or cohorts with wide ability gaps. Connectivity and device access are real constraints, and a lesson that collapses without stable internet is fragile. Multiple-choice items have a ceiling and cannot assess argumentation, design, performance, collaboration, or professional judgment. Teachers also report difficulty accurately gauging item difficulty, which determines whether the activity diagnoses or merely entertains. The design response is not to abandon the tool but to bind it, combining interactive quizzing with discussion, argumentation, projects, reflection, demonstration, performance assessment, and authentic assessment.

A conceptual model: the Human-Centered AI-Interactive Learning Cycle

Combining generative AI, gamified retrieval, and interactive assessment without a design logic produces a busy classroom rather than a learning one. The model proposed here organizes these components into a cycle that maps onto the phases of self-regulated learning [8] and keeps human judgment at the points where it matters. The Human-Centered AI-Interactive Learning Cycle has eight phases grouped into four movements: Know, Think, Make, and Grow.

In the Know movement, Diagnose uses a short interactive quiz to surface prior knowledge and misconceptions, with AI generating distractors targeted at known misconceptions; the human decision is what counts as a misconception worth addressing. Personalize routes for students to different entry points based on that diagnosis, with AI making differentiated materials feasible at scale [13]; the human decision is which differences matter and which are noise.

In the Think movement, Engage introduces gamified elements that raise participation and lower the social cost of being wrong; this is where gamification earns its place, through relatedness and autonomy rather than deep intrinsic motivation [27]. Explore has students investigate a genuine problem, using AI as a source of perspectives, counter-arguments, and worked examples rather than answers. Exploration here is guided because minimally guided instruction underperforms for novices [37], so the phase requires a driving question, explicit criteria, a time box, and available scaffolds.

In the Make movement, Create has students produce something that did not previously exist, with AI's contribution explicitly bounded by a declared level of permitted use [22]. Evaluating has students and teachers judge the product against public criteria; evaluating an AI-generated artifact against the same rubric is among the most efficient ways to teach criteria.

In the Grow movement, Reflect has students articulate what they did, what they found difficult, where they used AI, what they would change, and what they now understand; this phase is the one most often skipped and the one that converts activity into learning. Improve feeds by incorporating teacher, peer, and AI feedback into a revision, since feedback is most effective when it is informative and acted upon [36].

The eight phases map onto the three-phase model of self-regulation, with forethought in Diagnose and Personalize, performance in Engage, Explore, and Create, and self-reflection in Evaluate, Reflect, and Improve [8]. They also climb the revised taxonomy of educational objectives from remembering and understanding, through applying and analyzing, to evaluating and creating [38], and they satisfy the ICAP progression, because the later phases are constructive and interactive by construction [21]. Two rules keep the cycle honest. First, at least one phase in every implementation must be completed without AI assistance, usually Explore or Evaluate; this is the phase that produces the desirable difficulty on which learning depends. Second, the cycle must never stop before Reflect: a cycle truncated after Create produces artifacts rather than learners, so when time is short, the phase to cut is Engage, not Reflect.

Three design scenarios

The cycle is illustrated below in a teacher-education course on instructional media development, with the protected phase marked. Two further scenarios follow in condensed form.

Table 1. The Human-Centered AI-Interactive Learning Cycle applied to a teacher-education course on instructional media development.

Phase	Activity	Permitted AI use
Diagnose	Six-item quiz on media selection principles; distractors built from documented misconceptions	Teacher-side only
Personalise	Three entry tasks assigned based on the diagnosis result	Yes, for materials
Engage	Team challenge ranking five sample artifacts against criteria; the fastest correct justification scores, not the fastest click.	No
Explore	Analyze a real Year-4 science topic and identify the specific learning difficulty the media must address	No AI: protected phase
Create	Build the media artifact in a presentation or design application	Yes, for assets and copy-editing, not for the pedagogical rationale
Evaluate	Peer review against a rubric; each team also critiques one AI-generated artifact using the same rubric	Yes, as an object of critique

Phase	Activity	Permitted AI use
Reflect	Written reflection on what AI made easier, what it made worse, and what the student decided	No
Improve	Revision informed by peer feedback; second submission	Yes

The protected phase is deliberately the diagnosis of the learning difficulty because that is the professional judgment that defines a teacher; AI accelerates production, not analysis. In a second scenario, on digital literacy and media ethics in Islamic education, an opening poll asks which of four posts students would share, with the distribution rather than a correct answer as the data; groups then receive different case files, learn verification techniques, and verify their case without AI in the protected phase, before producing a community-facing verification guide, cross-reviewing it against criteria of accuracy, fairness, clarity, and adab in disagreement, and reflecting on their responsibility for what they transmit. In a third scenario, in Arabic language learning, retrieval practice through interactive quizzing consolidates weekly vocabulary and forms; AI then generates a dialogue containing deliberate errors that students identify, correct, and explain, inverting the usual dependency because error identification requires the rule to be known; students subsequently compose and perform their own dialogue, with AI use limited to vocabulary checking rather than composition.

The changing role of the teacher

The teacher was once primarily a source of information. Where information is abundant, and generation is free, that role has little scarcity value. What replaces it is not a lesser role but a more demanding one: the teacher as learning designer, and with it the functions of curator, facilitator, mentor, evaluator, ethical guide, and AI-literacy facilitator.

The paradox is worth stating plainly: the more capable AI becomes, the more a teacher's pedagogical knowledge matters. AI can generate content, but it cannot decide why a particular student should learn a particular thing at a particular time; what is worth learning given a curriculum, a community, and a set of values; how an experience should be sequenced so that difficulty is productive; how well is well enough, and against whose standard; or how learning connects to being a decent human being in a specific society. Those decisions constitute teaching, and each becomes more consequential as execution becomes inexpensive. This is also where the Indonesian adoption evidence converges with the international literature: intention to use AI in schools is driven by digital leadership, digital readiness, and perceived usefulness [6], and institutional adoption in public universities depends on organizational conditions rather than on the availability of tools [5]. Building design capability, not procuring technology, is therefore the higher-leverage intervention.

Anchoring practice in pedagogical models

Naming theories is easy; the point is the decision each one makes. TPACK insists that technological, pedagogical, and content knowledge interact rather than add up [39], and the decision it changes is which tool to use for this content: the content-specific difficulty is stated first, and the tool follows from it. The substitution-augmentation-modification-redefinition model is memorable and widely used, but a critical review found it under-theorized, insensitive to context, and misleading in implying that redefinition is always superior [40]; it is therefore best used as a conversation starter about whether a technology substitutes or transforms, not as an evaluative rubric, since substitution is often the correct answer.

Universal Design for Learning changes how many routes into a task there are, and generative AI makes its multiple means of representation, engagement, and expression genuinely achievable at the the classroom scale [13]. Outcome-based planning changes the direction of design, requiring that assessment and activity be derived from the intended outcome rather than the reverse; this discipline prevents the use of a platform from being reported as a learning outcome. The revised taxonomy changes the cognitive level of questions. If AI can complete every task at the remember-and-understand level, such tasks are no longer sufficient as assessment, although they remain useful as practice [38]. Constructivist and social-constructivist accounts change who does the talking and provide the theoretical basis for pair discussion between the first and second votes on a quiz item.

Student agency changes how much choice a task contains. A systematic review of higher education research found that agency is linked to individual learning outcomes, while noting that evidence on broader societal outcomes remains thin [41]. Agency is not the same as freedom: it is choice within structure, including choice of topic, format, sequence, and permitted level of AI involvement. Formative assessment changes what is done with results: assessment is formative only if it alters the next instructional move [42], so a quiz whose results are recorded but not acted upon is summative assessment with a friendlier interface.

Ethics as a design constraint

Ethics is not a closing section but a constraint operating from the first decision. On academic integrity, the problem is not that students use AI but that assessment designs built for an AI-free world no longer discriminate between learning and output; the durable responses are structural, including assessing process as well as product, requiring oral defense for high-stakes judgments, using portfolios and drafts, and stating permitted AI use per task rather than per course [22]. On transparency, disclosure should be normal, specific, and low-stakes, because a disclosure culture that punishes honesty produces dishonesty.

On privacy, student work, images, voices, and identifying details should not be entered into consumer services without consent and a lawful basis, with additional caution where learners are minors; international guidance calls for data-protection safeguards, age-appropriate validation before deployment, and institutional policy rather than ad hoc teacher decisions [43]. On bias and equity, bias appears in examples, names, images, and evaluative language. It is not neutral across the languages and cultures of Southeast Asian classrooms. In contrast, equity of access is the more immediate issue, since any design that assumes paid tools or reliable home connectivity disadvantages the students who already have the least. Regarding copyright, training-data provenance is contested, and generated material may reproduce protected works, so institutional guidance should be explicit about permitted uses in published output.

On human oversight, a community-wide ethical framework for AI in education has been argued for because the field has borrowed its ethics from data science without addressing distinctly educational questions about autonomy, purpose, and who decides [44]. A useful corrective to techno-optimism adds that the debate should include the possibility that some applications ought not to be built at all [45]. Neither position argues for rejection; both argue for educators exercising judgment rather than receiving tools. The practical principle is that AI should be used to strengthen the learning process, not to conceal its absence.

Two design instruments for practitioners

The first instrument is a six-gate framework for instructional design decisions. The gates run in order, and a design that fails the Test gate cannot be rescued by a better tool, while a design that fails the Trust gate should not be run at all.

Table 2. The six-gate framework for instructional technology decisions.

Gate	Question	Failure signal
Think	What is the learning outcome, stated as something the student can do?	The outcome names an activity rather than a capability.
Target	Who are these learners in terms of prior knowledge, language, devices, and confidence?	The design assumes a homogeneous class.
Technology	What technology is genuinely required by the outcome and the learners?	The tool was chosen before the outcome.
Test	What evidence of learning will be accepted, and can AI produce it without the student learning?	The evidence is a product that AI can generate.
Trust	Is this ethical, private, fair, accessible, and honestly disclosed?	Only some students can participate fully.
Transform	What competence has actually changed, and how is that known?	Only satisfaction or participation was measured.

The second instrument is a selection rubric applied before a tool is chosen. It asks what the learning outcome is; what learning activity produces it; what these learners need; what will count as evidence; whether the technology adds value that the activity cannot otherwise obtain; what the risks are; and, decisively, whether the activity would still be meaningful if the technology were removed. This last item is a subtraction test: if removing the tool leaves nothing, the tool was the lesson; if it leaves a slower, sound-based activity, the tool is doing legitimate work. Table 3 applies the rubric to common categories.

Table 3. Fit and risk profile of common instructional technologies.

Technology	Best-fit purpose	Weak fit	Principal risk
Generative AI	Perspectives, examples, differentiation, feedback drafts, artifacts for critique	Producing the student's own reasoning or final judgment	Displaced cognition; hallucination; undisclosed use
Quiz platform	Diagnosis, retrieval practice, misconception checks, participation	Assessing argument, design, or professional judgment	Speed rewarded over depth; anxiety from public ranking
Video	Demonstration, procedure, phenomena unavailable live	Content that the class could construct itself	Passive consumption without a task
Learning management system	Sequencing, distribution, submission, and feedback records	Generating engagement by itself	Becoming a file repository
Simulation	Practicing decisions that are risky or costly in reality	Simple factual recall	Fidelity mistaken for validity
Collaborative tools	Co-construction, peer review, shared artifacts	Individual accountability	Free-riding without role structure

Against technology for technology's sake

The most common design error in educational technology is not the misuse of a tool, but a reversed sequence: a tool becomes available, enthusiasm follows, and a justification is constructed afterward. The defensible sequence runs from learning outcome to learner to pedagogy to assessment to technology; the indefensible one begins with the technology and searches for a reason to use it. Three habits preserve the first sequence. Name the difficulty first, meaning the specific thing students find hard. Choose the smallest tool that solves it, since a paper card sort may outperform an application. Measure the outcome rather than the enthusiasm, because satisfaction with an activity is weakly related to how much was learned from it, and the activities that feel most effortful are often those that teach most.

Limitations

Three limitations bound these conclusions. First, the review is purposive rather than exhaustive; it privileges meta-analyses and official frameworks over primary studies, and a systematic search with formal screening and quality appraisal could surface evidence that contradicts parts of the synthesis. Second, the evidence base itself is uneven: most generative-AI studies are short, few report delayed post-tests, and publication bias is a recognized concern in educational technology, so reported average effects should be treated as upper bounds rather than as settled values. Third, the three artifacts proposed here, the cycle, the six-gate framework, and the selection rubric, are theoretically derived and have not yet been tested empirically. Their validation requires design-based research in authentic teacher-education settings, with delayed and transfer measures rather than immediate post-tests, and with explicit comparison between implementations that protect a phase from automation and those that do not. Replication in Indonesian and wider Southeast Asian contexts, where device access, connectivity, and institutional readiness vary substantially, is a necessary next step.

CONCLUSION

Personal intelligence is the human capacity to know oneself as a learner; artificial intelligence is the capacity to extend what a person can do. Placed together without design, the second erodes the first, because the fluent answer arrives before the question has been properly formed. Placed together by design, the second strengthens the first because AI supplies material, alternatives, and speed, while the student supplies the question, the judgment, and the decision.

The evidence reviewed here is therefore best read as conditional rather than triumphant. Gamification yields small-to-moderate gains and can backfire when it substitutes external rewards for intrinsic motivation. Interactive quiz platforms primarily succeed because retrieval practice and immediate, informative feedback are effective. Generative AI produces moderate average gains and can reduce mental effort in ways that help or harm, depending on which effort is removed. In every case, the differentiating variable is instructional design rather than the technology itself.

For teacher education, the practical implication is direct. The generation now entering classrooms does not need teachers who use more technology; it needs teachers who understand human beings, pedagogy, technology, and context well enough to decide what should be automated and what must remain difficult. The three artifacts offered here are intended to make that decision explicit and repeatable: a cycle that protects at least one phase from automation and never stops before reflection; six gates that place ethics inside the design rather than after it; and a subtraction test that asks whether the activity would survive the removal of the tool. Intelligent technology does not automatically produce intelligent learning. Learning becomes intelligent when technology, pedagogy, and human intelligence are deliberately designed to enable students to think, create, collaborate, reflect, and grow.

REFERENCES

[1] OECD, What Should Teachers Teach and Students Learn in a Future of Powerful AI? (OECD Education Spotlights). Paris, France: OECD Publishing, 2025, doi: 10.1787/ca56c7d6-en.

[2] O. Zawacki-Richter, V. I. Marín, M. Bond, and F. Gouverneur, "Systematic review of research on artificial intelligence applications in higher education – where are the educators?" Int. J. Educ. Technol. High. Educ., vol. 16, no. 1, Art. no. 39, 2019, doi: 10.1186/s41239-019-0171-0.

[3] M. Bond, H. Khosravi, M. De Laat, N. Bergdahl, V. Negrea, E. Oxley, P. Pham, S. W. Chong, and G. Siemens, "A meta systematic review of artificial intelligence in higher education: a call for increased ethics, collaboration, and rigour," Int. J. Educ. Technol. High. Educ., vol. 21, no. 1, Art. no. 4, 2024, doi: 10.1186/s41239-023-00436-z.

[4] T. K. F. Chiu, Q. Xia, X. Zhou, C. S. Chai, and M. Cheng, "Systematic literature review on opportunities, challenges, and future research recommendations of artificial intelligence in education," Comput. Educ. Artif. Intell., vol. 4, Art. no. 100118, 2023, doi: 10.1016/j.caeai.2022.100118.

[5] Helmiatin, A. Hidayat, and M. R. Kahar, "Investigating the adoption of AI in higher education: a study of public universities in Indonesia," Cogent Educ., vol. 11, no. 1, Art. no. 2380175, 2024, doi: 10.1080/2331186X.2024.2380175.

[6] H. G. R. Harsanti, N. Sudibjo, S. Riady, and P. Yu, "Exploring Indonesian teachers' intention to use artificial intelligence in schools: the impact of digital leadership, digital readiness, and perceived usefulness," Cogent Soc. Sci., vol. 11, no. 1, Art. no. 2593596, 2025, doi: 10.1080/23311886.2025.2593596.

[7] P. A. Kirschner and P. De Bruyckere, "The myths of the digital native and the multitasker," Teach. Teach. Educ., vol. 67, pp. 135–142, 2017, doi: 10.1016/j.tate.2017.06.001.

[8] B. J. Zimmerman, "Becoming a self-regulated learner: an overview," Theory Pract., vol. 41, no. 2, pp. 64–70, 2002, doi: 10.1207/s15430421tip4102_2.

[9] D. Long and B. Magerko, "What is AI literacy? Competencies and design considerations," in Proc. 2020 CHI Conf. Human Factors in Computing Systems, Honolulu, HI, USA, 2020, pp. 1–16, doi: 10.1145/3313831.3376727.

[10] D. T. K. Ng, J. K. L. Leung, S. K. W. Chu, and M. S. Qiao, "Conceptualizing AI literacy: an exploratory review," Comput. Educ. Artif. Intell., vol. 2, Art. no. 100041, 2021, doi: 10.1016/j.caeai.2021.100041.

[11] UNESCO, AI Competency Framework for Students. Paris, France: UNESCO, 2024. [Online]. Available: https://unesdoc.unesco.org/ark:/48223/pf0000391105

[12] UNESCO, AI Competency Framework for Teachers. Paris, France: UNESCO, 2024. [Online]. Available: https://unesdoc.unesco.org/ark:/48223/pf0000391104

[13] CAST, Universal Design for Learning Guidelines Version 3.0. Lynnfield, MA, USA: CAST, 2024. [Online]. Available: https://udlguidelines.cast.org

[14] Z. Liu, H. Zuo, and Y. Lu, "The impact of ChatGPT on students' academic achievement: a meta-analysis," J. Comput. Assist. Learn., vol. 41, no. 4, Art. no. e70096, 2025, doi: 10.1111/jcal.70096.

[15] Y. M. E. Heung and T. K. F. Chiu, "How ChatGPT impacts student engagement from a systematic review and meta-analysis study," Comput. Educ. Artif. Intell., vol. 8, Art. no. 100361, 2025, doi: 10.1016/j.caeai.2025.100361.

[16] R. Deng, M. Jiang, X. Yu, Y. Lu, and S. Liu, "Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies," Comput. Educ., vol. 227, Art. no. 105224, 2025, doi: 10.1016/j.compedu.2024.105224.

[17] M. Lan and X. Zhou, "A qualitative systematic review on AI empowered self-regulated learning in higher education," npj Sci. Learn., vol. 10, no. 1, Art. no. 21, 2025, doi: 10.1038/s41539-025-00319-0.

[18] C.-Q. Huang, L.-N. Lu, Q.-H. Huang, Y. Zhang, T. He, Y.-F. Tu, and G.-J. Hwang, "Effects of artificial intelligence feedback on students' self-regulated learning in higher education: a three-level meta-analysis," Educ. Psychol. Rev., vol. 38, no. 1, 2026, doi: 10.1007/s10648-026-10166-z.

[19] M. Gerlich, "AI tools in society: impacts on cognitive offloading and the future of critical thinking," Societies, vol. 15, no. 1, Art. no. 6, 2025, doi: 10.3390/soc15010006.

[20] H. Bastani, O. Bastani, A. Sungu, H. Ge, Ö. Kabakcı, and R. Mariman, "Generative AI can harm learning," SSRN Working Paper 4895486, 2024, doi: 10.2139/ssrn.4895486.

[21] M. T. H. Chi and R. Wylie, "The ICAP framework: linking cognitive engagement to active learning outcomes," Educ. Psychol., vol. 49, no. 4, pp. 219–243, 2014, doi: 10.1080/00461520.2014.965823.

[22] M. Perkins, L. Furze, J. Roe, and J. MacVaugh, "The Artificial Intelligence Assessment Scale (AIAS): a framework for ethical integration of generative AI in educational assessment," J. Univ. Teach. Learn. Pract., vol. 21, no. 6, Art. no. e810, 2024, doi: 10.53761/q3azde36.

[23] S. Deterding, D. Dixon, R. Khaled, and L. Nacke, "From game design elements to gamefulness: defining 'gamification'," in Proc. 15th Int. Academic MindTrek Conf., Tampere, Finland, 2011, pp. 9–15, doi: 10.1145/2181037.2181040.

[24] S. Bai, K. F. Hew, and B. Huang, "Does gamification improve student learning outcome? Evidence from a meta-analysis and synthesis of qualitative data in educational contexts," Educ. Res. Rev., vol. 30, Art. no. 100322, 2020, doi: 10.1016/j.edurev.2020.100322.

[25] R. Huang, A. D. Ritzhaupt, M. Sommer, J. Zhu, A. Stephen, N. Valle, J. Hampton, and J. Li, "The impact of gamification in educational settings on student learning outcomes: a meta-analysis," Educ. Technol. Res. Dev., vol. 68, no. 4, pp. 1875–1901, 2020, doi: 10.1007/s11423-020-09807-z.

[26] M. Sailer and L. Homner, "The gamification of learning: a meta-analysis," Educ. Psychol. Rev., vol. 32, no. 1, pp. 77–112, 2020, doi: 10.1007/s10648-019-09498-w.

[27] L. Li, K. F. Hew, and J. Du, "Gamification enhances student intrinsic motivation, perceptions of autonomy and relatedness, but minimal impact on competency: a meta-analysis and systematic review," Educ. Technol. Res. Dev., vol. 72, no. 2, pp. 765–796, 2024, doi: 10.1007/s11423-023-10337-7.

[28] R. M. Ryan and E. L. Deci, "Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being," Amer. Psychol., vol. 55, no. 1, pp. 68–78, 2000, doi: 10.1037/0003-066X.55.1.68.

[29] M. D. Hanus and J. Fox, "Assessing the effects of gamification in the classroom: a longitudinal study on intrinsic motivation, social comparison, satisfaction, effort, and academic performance," Comput. Educ., vol. 80, pp. 152–161, 2015, doi: 10.1016/j.compedu.2014.08.019.

[30] E. L. Deci, R. Koestner, and R. M. Ryan, "A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation," Psychol. Bull., vol. 125, no. 6, pp. 627–668, 1999, doi: 10.1037/0033-2909.125.6.627.

[31] A. I. Wang and R. Tahir, "The effect of using Kahoot! for learning – a literature review," Comput. Educ., vol. 149, Art. no. 103818, 2020, doi: 10.1016/j.compedu.2020.103818.

[32] O. Özdemir, "Kahoot! game-based digital learning platform: a comprehensive meta-analysis," J. Comput. Assist. Learn., vol. 41, no. 1, Art. no. e13084, 2025, doi: 10.1111/jcal.13084.

[33] Q. Zhang and Z. Yu, "A literature review on the influence of Kahoot! on learning outcomes, interaction, and collaboration," Educ. Inf. Technol., vol. 26, no. 4, pp. 4507–4535, 2021, doi: 10.1007/s10639-021-10459-6.

[34] H. L. Roediger III and J. D. Karpicke, "Test-enhanced learning: taking memory tests improves long-term retention," Psychol. Sci., vol. 17, no. 3, pp. 249–255, 2006, doi: 10.1111/j.1467-9280.2006.01693.x.

[35] C. Yang, L. Luo, M. A. Vadillo, R. Yu, and D. R. Shanks, "Testing (quizzing) boosts classroom learning: a systematic and meta-analytic review," Psychol. Bull., vol. 147, no. 4, pp. 399–435, 2021, doi: 10.1037/bul0000309.

[36] B. Wisniewski, K. Zierer, and J. Hattie, "The power of feedback revisited: a meta-analysis of educational feedback research," Front. Psychol., vol. 10, Art. no. 3087, 2020, doi: 10.3389/fpsyg.2019.03087.

[37] P. A. Kirschner, J. Sweller, and R. E. Clark, "Why minimal guidance during instruction does not work: an analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching," Educ. Psychol., vol. 41, no. 2, pp. 75–86, 2006, doi: 10.1207/s15326985ep4102_1.

[38] L. W. Anderson and D. R. Krathwohl, Eds., A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives. New York, NY, USA: Longman, 2001.

[39] P. Mishra and M. J. Koehler, "Technological pedagogical content knowledge: a framework for teacher knowledge," Teach. Coll. Rec., vol. 108, no. 6, pp. 1017–1054, 2006, doi: 10.1177/016146810610800610.

[40] E. R. Hamilton, J. M. Rosenberg, and M. Akcaoglu, "The substitution augmentation modification redefinition (SAMR) model: a critical review and suggestions for its use," TechTrends, vol. 60, no. 5, pp. 433–441, 2016, doi: 10.1007/s11528-016-0091-y.

[41] M. H. Stenalt and B. Lassesen, "Does student agency benefit student learning? A systematic review of higher education research," Assess. Eval. High. Educ., vol. 47, no. 5, pp. 653–669, 2022, doi: 10.1080/02602938.2021.1967874.

[42] P. Black and D. Wiliam, "Assessment and classroom learning," Assess. Educ. Princ. Policy Pract., vol. 5, no. 1, pp. 7–74, 1998, doi: 10.1080/0969595980050102.

[43] UNESCO, Guidance for Generative AI in Education and Research. Paris, France: UNESCO, 2023. [Online]. Available: https://unesdoc.unesco.org/ark:/48223/pf0000386693

[44] W. Holmes, K. Porayska-Pomsta, K. Holstein, E. Sutherland, T. Baker, S. Buckingham Shum, O. C. Santos, M. T. Rodrigo, M. Cukurova, I. I. Bittencourt, and K. R. Koedinger, "Ethics of AI in education: towards a community-wide framework," Int. J. Artif. Intell. Educ., vol. 32, no. 3, pp. 504–526, 2022, doi: 10.1007/s40593-021-00239-1.

[45] N. Selwyn, "The future of AI and education: some cautionary notes," Eur. J. Educ., vol. 57, no. 4, pp. 620–631, 2022, doi: 10.1111/ejed.12532.