Jurnal Pustaka Indonesia (JPI)
2026, Vol. 6, No. 1, 728–744
https://doi.org/10.62159/jpi.v6i1.3320

JPI JURNAL PUSTAKA INDONESIA

OPEN ACCESS Research Article

# Designing Critical Human–AI Learning in Higher Education: The SC-AIM–FAVIC Framework for Analytical and Logical Reasoning

Ahmad Syaikhu*
Universitas Terbuka, Indonesia

**Abstract**
Generative artificial intelligence (AI) is widely used in university study, yet unstructured use may substitute for the reasoning that courses aim to develop. This article reports a conceptual model-development study that produced the Symbiotic Critical Human–AI Learning Model (SC-AIM) and its five-stage FAVIC framework (Formulation, Augmentation, Verification, Integration, Creation). Of the five ADDIE phases, Analysis and Design were completed through a targeted literature-based needs analysis, and Development was completed only to the level of draft artefacts: a stage specification, an activity protocol, a process-sensitive rubric, and an instrument blueprint. Expert review, prototype testing, implementation, and evaluation are proposed and have not been conducted. The design outputs are an evidence-to-design matrix that separates empirical findings from theoretical propositions and author's interpretations, a comparison with existing frameworks, a FAVIC specification with return rules and scaffold-fading criteria, and a prospective multi-class evaluation protocol. Symbiosis is defined as role complementarity in which AI contributes breadth while the human retains problem framing, verification, and judgement. No expert-validation, pilot, or learner-outcome data are reported, so the effectiveness, feasibility, and generalizability of SC-AIM remain untested. The model is intended for pilot use in Indonesian higher education after expert judgement and prototype testing.

**ARTICLE HISTORY**
Received : 10 January 2026
Revised : 29 March 2026
Accepted : 21 April 2026

**KEYWORDS**
Generative AI; critical thinking; cognitive offloading; instructional design; higher education; FAVIC

**PUBLISHER'S NOTE**
This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY 4.0) license.

**CORRESPONDING AUTHOR**
***Ahmad Syaikhu**, Universitas Terbuka, Indonesia. Email: ahmadsyaikhu@gmail.com

## Introduction

Universities are expected to graduate people who can analyse complex problems, evaluate arguments, and reason logically. Generative AI has entered this mission abruptly. Large language models are used for searching, drafting, summarising, and problem solving, and scholarly work describes both promise for personalised support and concerns about academic integrity and students' own learning (Chan, 2023; Cotton et al., 2024). An early systematic review of 146 articles on AI in higher education, covering 2007–2018, reported little critical reflection on risks and a weak connection to pedagogical theory (Zawacki-Richter et al., 2019). That review predates generative AI, so its characterisation cannot be carried over unchanged. More recent work does ground generative-AI use in pedagogical or cognitive theory, for example through metacognition (Tankelevitch et al., 2024), hybrid human–AI learning (Molenaar, 2022), and institutional policy (Chan, 2023; UNESCO, 2023). The gap addressed here is narrower: a classroom-level, assessable sequence for reasoning with generative AI, compared explicitly with these frameworks in the Result section.

The intended population is undergraduate students in Indonesian universities, including open and distance settings such as Universitas Terbuka, where independent, asynchronous study is common. Evidence specific to this population is thin in the sources reviewed. One qualitative study of seven master's students at two Indonesian universities found that students valued AI for academic

research and theory scrutiny but worried about echo chambers and limited nuance (Darwin et al., 2024). It is perceptual evidence from a small, self-reported sample. No controlled study was found of generative AI and reasoning in an Indonesian open and distance setting, so the design rests on international evidence and on assumptions about this context that remain to be examined.

The experimental evidence is instructive because performance with AI and learning without AI can diverge. In a field experiment with nearly a thousand high school mathematics students, access to an unrestricted GPT-4 interface improved practice performance but reduced later exam performance once access was removed, whereas a tutor version with learning safeguards largely mitigated this harm (Bastani et al., 2025). In a randomised laboratory study with 117 university students, ChatGPT support improved essay scores without significantly greater knowledge gain or transfer and was associated with reduced metacognitive activity, termed metacognitive laziness (Fan et al., 2025). Students researching a socio-scientific issue with ChatGPT reported lower cognitive load than those using a search engine but produced lower-quality reasoning (Stadler et al., 2024). A systematic review of 14 articles concluded that over-reliance on AI dialogue systems can affect decision-making, critical thinking, and analytical reasoning (Zhai et al., 2024). These studies differ in age, task, and context. School mathematics tasks are largely procedural and teacher-directed, whereas university analytical tasks are open-ended and more autonomous, so the findings are treated as indirect support for mechanisms, not as direct evidence for Indonesian university students. Because gains measured while AI is available do not show unaided reasoning, the model targets, and its evaluation protocol prioritises, unaided performance.

Survey-based studies point the same way but carry weaker warrant. A mixed-methods study of 666 participants found that more frequent AI-tool use was negatively associated with critical-thinking scores, mediated by cognitive offloading (Gerlich, 2025), and a survey of 319 knowledge workers found that greater confidence in generative AI was associated with less self-reported critical thinking (Lee et al., 2025). These are associations from self-report or cross-sectional data, not causal learning effects, and neither sampled students in a controlled instructional setting. They are used only to frame risks and to generate hypotheses, not to derive requirements on their own. Cognitive offloading, the use of action to reduce the mental demands of a task, is not harmful in itself (Risko & Gilbert, 2016). The design question is which operations may be delegated and which must remain with the learner.

A second concern is reliability. Language models can generate fluent text that is unfaithful to sources or factually wrong, widely documented as hallucination (Ji et al., 2023). It is theoretically plausible that finished answers invite a surface approach to learning, in the sense of Marton and Säljö (1976), but this is a hypothesised mechanism and not direct evidence that generative AI induces such an approach.

This article presents the Symbiotic Critical Human–AI Learning Model (SC-AIM) and its operational framework, FAVIC. The term symbiotic is used in a restricted, design sense drawn from Licklider's (1960) account of cooperative human–computer interaction: the human and the AI take interdependent, complementary roles, and each party's contribution is conditioned on the other's. Three conditions make the use of the term operational: the human commits to a position before consulting AI, the AI supplies what the human is unlikely to produce alone (breadth, alternative framings, counter-arguments), and the human re-enters to verify and integrate. This differs from a conventional human-in-the-loop workflow, in which a person supervises or corrects a machine

process (Mosqueira-Rey et al., 2023), because here the object of development is the human's reasoning and the AI's output is the object of evaluation. The term is not offered as a new theory of cognition.

The article separates development questions, answered here, from effectiveness questions, deferred. The development questions are: (1) which design requirements follow from the reviewed evidence, and at what level of warrant (unit of analysis: the reviewed sources); (2) how can the requirements be specified as a stage-by-stage model with a draft assessment framework (unit of analysis: the model's stages and artefacts); and (3) how would the model be evaluated (unit of analysis: the elements of a prospective protocol). None requires expert, teacher, or student evaluation data. The deferred questions are whether SC-AIM improves unaided analytical and logical reasoning, whether it is feasible and acceptable to lecturers and students, and whether it generalises across courses and institutions. These need data that this article does not report.

## Method

The study is a conceptual model-development study that follows the ADDIE structure of analysis, design, development, implementation, and evaluation (Branch, 2009; Molenda, 2003). Table 1 states which phases were completed and which remain proposed. The Development phase was completed only to the level of draft artefacts. Iterative review cycles, decision gates, and prototype testing have not occurred, and no claim of tested development is made.

**Table 1.** Status of ADDIE phases in this article

| Phase | Status | Deliverables reported here | Pending |
|---|---|---|---|
| Analysis | Completed (literature-based) | Evidence-to-design matrix (Table 4) | Needs survey of students and lecturers |
| Design | Completed | FAVIC specification, principles, Figure 1, comparison with frameworks | Expert critique |
| Development | Draft artefacts only | Stage activity protocol, verification protocol, draft rubric, instrument blueprint | Expert review, revision, prototype test |
| Implementation | Proposed | Prospective protocol (Table 3) | All activities |
| Evaluation | Proposed | Prospective protocol (Table 3) | All activities |

### *Analysis procedure*

The needs analysis was a targeted narrative review, not a systematic review. Eligible sources were peer-reviewed articles or archival conference papers on generative AI or large language models in relation to learning, reasoning, cognitive effort, metacognition, or output accuracy, published from 2022 onward, together with foundational theory of any date on offloading, scaffolding, metacognition, and cognitive load, and intergovernmental guidance. Opinion pieces, preprints, and tool-acceptance surveys were excluded. Sources were located through publisher and indexing platforms, using combinations of the terms generative AI, ChatGPT, large language model, critical thinking, reasoning, metacognition, cognitive offloading, hallucination, and higher education, and through citation chaining [authors to insert search dates and the platforms actually used]. No protocol was registered, and the numbers of records screened were not logged, so the evidence base is the set of studies in Table 4. It is not exhaustive. Each source was checked against the publisher

record for retraction or correction. One widely cited meta-analysis of ChatGPT and learning was excluded for this reason because it was retracted in April 2026 (Wang & Fan, 2026).

Each source was assigned a warrant level: direct experimental evidence, systematic synthesis, associational (survey or self-report) evidence, perceptual (qualitative) evidence, or theoretical proposition. Experimental studies that measured unaided outcomes were given most weight, and conflicting findings were reconciled by asking which outcome was measured, with or without AI, and in which population. Methodological quality was judged informally from design, sample, and outcome measure. Publication bias was not assessed formally and probably favours positive or striking findings. For each source, the mechanism was recorded it implies, whether that mechanism was tested or inferred, the resulting design requirement, and the FAVIC stage it informed. Table 4 is the resulting decision matrix and audit trail. [Author to confirm how coding and design decisions were made, including any independent review.]

### *Design foundations and constructs*

The Design phase used scaffolding, in which support is withdrawn as competence grows (Wood et al., 1976); the expertise reversal effect, in which supports that help novices can hinder more proficient learners (Kalyuga et al., 2003); metacognitive monitoring and control (Flavell, 1979; Tankelevitch et al., 2024); cognitive load (Sweller, 1988); and productive struggle before support (Kapur, 2016). Critical thinking follows the expert consensus of interpretation, analysis, evaluation, inference, explanation, and self-regulation (Facione, 1990) and reasonable reflective thinking about what to believe or do (Ennis, 1993). Table 2 distinguishes the constructs and links each to observable indicators, tasks, and measures.

**Table 2.** Operationalisation of constructs

| Construct | Working definition | Observable indicators | Learning task | Outcome measure |
|---|---|---|---|---|
| Analytical reasoning | Decomposing a problem and judging evidence and assumptions | Identifies assumptions; separates claim from evidence; rates source quality | Formulation memo; claim-check table | Unaided test subscale; rubric F and V |
| Logical reasoning | Drawing valid or strong inferences | Labels steps as deductive or inductive; detects fallacies | Argument map; defence | Unaided test subscale; rubric I and C |
| Critical thinking (umbrella) | Reflective judgement on what to believe or do | Self-correction after failed verification | Reflection on what changed and why | Composite of the above |
| AI literacy and responsible use | Knowing what AI can and cannot do and using it accountably | Bounded prompts; honest disclosure | Prompt log; disclosure statement | Rubric A and C; disclosure audit |

### *Prospective evaluation protocol*

The protocol below is prospective. None of it has been conducted, and ethics approval has not yet been obtained; it will be sought from the university research ethics committee before any data collection. Table 3 summarises the protocol, and the text gives the decisions that the table cannot.

**Table 3.** Prospective evaluation protocol (not yet conducted)

| Phase | Status | Design and instruments | Analysis and decision criteria |
|---|---|---|---|
| Expert review | Planned | 5–7 independent experts rate model and instruments on a 4-point scale | Aiken's V ≥ .80 retained; lower items revised and re-rated once, then dropped |
| Pilot of instruments | Planned | Small trial outside the main sample; two parallel test forms; rubric | Difficulty .30–.70; discrimination ≥ .30; reliability ≥ .70; ICC ≥ .75 |
| Main pilot study | Planned | Four intact classes, eight weeks, four FAVIC cycles; pre-test, unaided post-test, delayed test, far-transfer task | Multilevel model; Hedges' g with CI; exploratory interpretation |
| Qualitative strand | Planned | Interviews and focus groups with purposive sample | Thematic analysis with negative-case analysis |
| Integration | Planned | Joint display of strands | Convergence, divergence, expansion |

Setting and population. The pilot will use a foundational general-education course taught online through the university's learning management system, with asynchronous activities and scheduled tutorials. Students will be recruited from all sections enrolled in the term and will be invited through the course site. A baseline questionnaire will record prior AI experience, access to devices, and connectivity.

Design, sample, and allocation. A single class per condition would confound treatment with class and tutor, and covariate adjustment cannot remove that confounding. The protocol therefore uses four intact classes, two taught with SC-AIM and two as comparison, with each of two tutors teaching one class in each condition. Classes will be allocated to conditions by lot within tutor. With about 25–30 students per class (N ≈ 100–120) and an assumed intraclass correlation of .05, the design effect is about 2.2, giving an effective sample near 50 and sensitivity only to large effects (about d = 0.8 at 80% power). The study is therefore a pilot, and estimates will be interpreted as exploratory and not as confirmatory causal evidence. Both students and classes will be reported, and with four clusters, cluster-level inference is limited.

Comparison condition and fidelity. The comparison classes follow the same syllabus, tasks, platform, and scheduled time, and may use freely available AI tools, but they receive no FAVIC templates, artefacts, or rubric. Time on task will be logged. Both tutors receive the same briefing on the AI policy, and SC-AIM tutors additionally receive training in the protocol. Classes use separate course sites, and a short questionnaire will check for sharing of materials across classes. Fidelity will be checked through artefact completion (target ≥ 80%), tutor logs, and observation of a sample of tutorial sessions. Intended instruction lasts eight weeks, in four two-week FAVIC cycles.

Outcomes. The primary outcome is a composite unaided reasoning score, with the analytical and logical subscales as secondary outcomes, tested with Holm correction. Retention (a delayed unaided test four weeks after the course) and far transfer (an unfamiliar task in a different domain) are pre-specified secondary outcomes, because they bear on whether reasoning is learned instead of outsourced. Both groups take the same independently supervised assessments, in person at regional study centres where possible or in supervised synchronous online sessions. Parallel forms A and B will be counterbalanced across pre-test and post-test and equated in the pilot, with at least eight

weeks between administrations to limit practice effects. The model does not predict differential gains in analytical and logical reasoning. They are treated as distinct but correlated, and if the pilot shows a single dominant factor, only the composite will be reported.

Analysis. Post-test scores will be analysed with a linear mixed model with students nested in classes, a random intercept for class, condition as a fixed effect, and the pre-test and tutor as covariates, using restricted maximum likelihood with a small-sample correction. Residual normality, linearity, and homogeneity will be inspected, and effect sizes will follow cluster-adjusted procedures for Hedges' g (Hedges, 2007). Baseline imbalance will be reported as standardised differences, with values above 0.25 treated as a concern. Attrition will be reported by condition; missing data will be handled by multiple imputation under a missing-at-random assumption with a complete-case sensitivity analysis. Treating students as independent would be pseudoreplication, which the random effect is meant to avoid (Shadish et al., 2002).

Hypotheses. H1: adjusted unaided composite scores are higher in SC-AIM classes. H2: verification and integration rubric scores rise over the course in SC-AIM classes. To make early and late artefacts comparable, tasks will be rotated across time points by Latin square within class, and artefacts will be anonymised, shuffled, and scored blind to condition and time. H3: experts rate the model and instruments as having acceptable content validity. Content validity is distinct from classroom feasibility. Feasibility will be judged from the pilot against stated criteria: at least 80% artefact completion, a median of no more than four hours of student work per cycle, and no more than 20 minutes of lecturer scoring per student per cycle.

Expert panel and rater reliability. Experts must hold a doctorate or equivalent, have at least five years in instructional design, assessment, or AI in education, have at least one peer-reviewed publication, and be independent of the author. They rate relevance, clarity, completeness, and feasibility, and content validity is summarised with Aiken's V (Aiken, 1985). Two trained raters will score every artefact independently and blind. Agreement will be estimated as ICC(2,1), two-way random effects with absolute agreement, with values of at least .75 acceptable (Koo & Li, 2016). A difference of one level or more on any dimension is resolved by a third rater, and the final score is the mean. Internal consistency will be reported with Cronbach's alpha, interpreted with the cautions in Taber (2018).

Qualitative strand and ethics. Interview participants (about 12–16 students and all tutors) will be sampled purposively for maximum variation in AI experience, course performance, and class. Interviews take place at mid-course and after the post-test, and focus groups after the post-test. Contradictory accounts will be retained through negative-case analysis and reported, and the qualitative findings will be set beside the quantitative results in a joint display (Braun & Clarke, 2006; Creswell & Plano Clark, 2018). Participation is voluntary and has no effect on grades. Prompt logs and AI-chat excerpts are submitted by students, may be redacted by them, are used only with consent, are de-identified, stored encrypted, and deleted after five years. No student is required to hold a paid AI account or to give personal data to an AI vendor.

## Result and Discussion

### *Result*

The results are conceptual design products: the evidence-to-design matrix, the SC-AIM model and its comparison with other frameworks, the FAVIC specification, and the draft assessment framework

and instrument blueprint. They are not empirical findings from implementation. The hypotheses, analyses, and thresholds in the Method are prospective and are not results.

### *Evidence-to-design matrix*

Table 4 records how each source was converted into a design interpretation and at what warrant. Only the first three rows are direct experimental evidence on learning with generative AI, and none was conducted with Indonesian university students. Each interpretation in the third column is the author's inference, and no mechanism is treated as established for the target population.

**Table 4.** Evidence-to-design matrix with levels of warrant

| Source and warrant | Reported finding | Design interpretation (author's inference) | Stage |
|---|---|---|---|
| Bastani et al. (2025); field experiment, ≈1,000 school students; direct | Unrestricted AI raised practice scores but lowered later unaided exam scores; safeguards mitigated this | Require own attempt before AI; bound AI behaviour | F, A |
| Fan et al. (2025); randomised lab study, 117 university students; direct | Higher essay scores, no greater knowledge gain or transfer; less metacognitive activity | Prompt and assess monitoring and evaluation; require reflection | V, I |
| Stadler et al. (2024); experiment, 91 students; direct | Lower cognitive load but lower reasoning quality than with search | Keep demanding reasoning with the student before and after AI | F, I |
| Zhai et al. (2024); systematic review of 14 articles; synthesis | Over-reliance linked to weaker decision-making and analytical thinking | Limit substitution; define bounded AI roles | A |
| Gerlich (2025); Lee et al. (2025); survey and self-report; associational | AI use or confidence associated with less critical thinking; effort shifts to verification and integration | Hypothesis only: assess verification, integration, and stewardship explicitly | V, I, C |
| Ji et al. (2023); survey of the hallucination literature; synthesis | Generated text can be unfaithful or false | Check claims against independent sources | V |
| Sweller (1988); Kalyuga et al. (2003); Wood et al. (1976); Kapur (2016); theoretical | Scaffolds aid novices and can hinder proficient learners; struggle before support can be productive | Fade scaffolds by mastery; place Formulation before AI | F, all |
| Flavell (1979); Tankelevitch et al. (2024); theoretical | Monitoring and control are central; generative AI adds metacognitive demands | Embed assessed monitoring prompts | V, I |
| Darwin et al. (2024); qualitative, 7 Indonesian master's students; perceptual | AI helps research and theory scrutiny; concerns about echo chambers | Ask AI for diverse and opposing views; context cue only | A |

### *The SC-AIM model compared with existing frameworks*

SC-AIM rests on five principles. First, human-first cognition: the student frames the problem and states a position before consulting AI. Second, AI as bounded dialogic partner: AI broadens perspectives and tests reasoning and does not perform the operation the task assesses. Third, mandatory verification: no AI claim enters final work unchecked. Fourth, visible process: assessment

values questions, checking, and synthesis as well as the product. Fifth, facilitated autonomy: scaffolds fade by explicit rule. Table 5 compares the model with related frameworks, and the comparison reflects the author's reading of those sources. The precise additions are an authenticable human-first commitment, a minimum verification protocol with return rules, stage-linked artefacts scored against observable anchors, and fading criteria.

**Table 5.** Comparison of SC-AIM with related frameworks

| Framework | Primary focus | Not specified at classroom level | SC-AIM addition |
|---|---|---|---|
| AI literacy (Long & Magerko, 2020; Ng et al., 2021) | Competencies for understanding, using, and evaluating AI | Sequence of activities and assessment evidence for reasoning | Stage protocol, artefacts, rubric |
| Self-regulated learning (Zimmerman, 2002) | Forethought, performance, and reflection phases for any task | Limits on AI functions and verification of AI output | Permitted and not permitted AI use per stage |
| Metacognitive demands of generative AI (Tankelevitch et al., 2024) | Metacognition as a lens, mainly for system design | Instruction and grading | Assessed monitoring prompts |
| Hybrid human–AI learning (Molenaar, 2022) | Distribution of control between humans and AI | Student tasks for reasoning with generative AI | Human-first commitment; return rules |
| Human-in-the-loop learning (Mosqueira-Rey et al., 2023) | Human supervision of a machine process | Human cognition as the learning goal | AI output as object of evaluation |

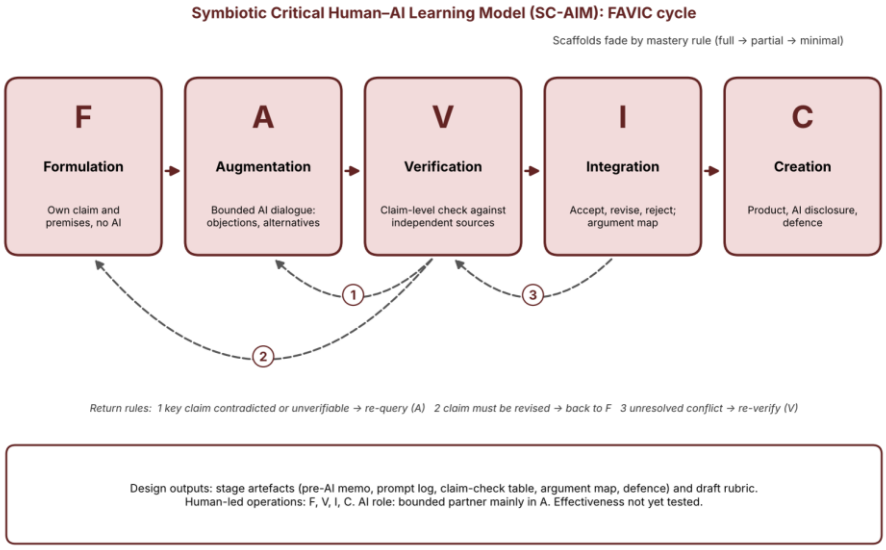



**Figure 1.** The SC-AIM model operationalised through the FAVIC cycle, with return paths when verification or integration fails.

### *FAVIC specification*

FAVIC is an iterative cycle with a fixed forward order and defined return paths, not a purely linear sequence (Figure 1). Table 6 specifies each stage with entry and exit criteria, permitted and not permitted AI use with representative prompts, the evidence artefact, and the return rule. The

Formulation stage is authenticated without surveillance by a short, time-windowed entry in the learning management system with a declaration, low grade weight, and a consistency check against the later defence. Students with connectivity limits may submit a handwritten memo as a photograph or complete it in a tutorial. No camera proctoring is used.

**Table 6.** FAVIC stage specification

| Stage | Student task and exit criterion | AI use: permitted and not permitted | Artefact and return rule |
|---|---|---|---|
| Formulation | Restate the problem, list assumptions, and state a claim with at least two premises; exit at Basic or above | Not permitted: any AI use | Pre-AI memo; incomplete memo is resubmitted before Augmentation opens |
| Augmentation | Use AI to obtain alternative framings, objections, and missing considerations; at least three distinct prompts | Permitted: "What is the strongest objection to my claim and what assumption does it expose?" Not permitted: "Write my analysis of X." | Prompt log; only answer-seeking prompts means repeat with prompt patterns |
| Verification | Split output into discrete claims; check each against at least two independent sources opened by the student; label supported, contradicted, or unverifiable; audit logic and bias | Permitted: AI suggests search terms. Not permitted: AI as the only verifier; AI-supplied citations not opened in the original source | Claim-check table; a failed key claim returns to Augmentation (narrower query) or Formulation (revise claim) |
| Integration | For each premise, accept, revise, or reject with a reason; draw an argument map labelling steps as deductive or inductive | Permitted: AI critique of the map, with the student deciding. Not permitted: AI-written synthesis | Argument map and reflection; unresolved conflict returns to Verification or is stated as a limitation |
| Creation | Produce an analytical product, disclose AI use, and defend two randomly chosen claims in writing or orally | Permitted: declared language editing. Not permitted: AI-generated analysis | Product, disclosure, defence; inconsistent defence leads to revision and re-defence |

Minimum verification protocol. (1) List the discrete factual and inferential claims in the AI output. (2) For each key claim, locate at least two independent sources through a library database or the open web, not through an AI-supplied link alone. (3) Open the source and confirm that it exists and says what is claimed, since fabricated citations are a known failure. (4) Label the claim supported, contradicted, or unverifiable. (5) Contradictory sources are retained and reported, not dropped. An unverifiable key claim cannot be used as a premise and triggers a return path. A key claim is one on which the conclusion depends.

Scaffold fading. Support has three levels: full (templates, sentence starters, exemplars), partial (templates only), and minimal (checklist only). A student moves down one level after scoring Competent or above on all dimensions in two consecutive cycles and moves back up after a Basic score on Verification or Integration. Lecturers decide by rubric scores, not impression. Students who

need language, accessibility, or connectivity support may keep full scaffolds without penalty, and no student is moved below minimal support. Novices therefore receive more structure and more proficient students less, consistent with the expertise reversal effect.

### *Draft rubric and instrument blueprint*

Table 7 gives the draft rubric. Each of the five dimensions is scored 1–3 by two raters, and the weighted total is the sum of weight × (level ÷ 3), giving a maximum of 100. The weights are Formulation 15, Augmentation 15, Verification 30, Integration 25, and Creation 15, reflecting the greater weight of verification and integration in the evidence matrix. Provisional benchmarks are 80 or more for advanced, 60–79 for competent, and below 60 for basic performance. Raters will be calibrated on six benchmark artefacts in two rounds. Descriptors use observable anchors and are subject to expert review before use (Jonsson & Svingby, 2007).

**Table 7.** Draft process-sensitive rubric (provisional)

| Dimension (weight) | Basic (1) | Competent (2) | Advanced (3) |
|---|---|---|---|
| Formulation (15) | Claim stated; premises absent | Claim with at least two premises and listed assumptions | Premises and hidden assumptions stated; scope limits noted |
| Augmentation (15) | Prompts seek answers only | At least three prompts seek objections or alternatives | Iterative prompts test assumptions and compare perspectives |
| Verification (30) | Fewer than half of key claims checked | At least 80% of key claims checked against two independent sources; each labelled | All key claims checked; errors, bias, and fabricated sources identified and reported |
| Integration (25) | AI text appended without comparison | Each premise accepted, revised, or rejected with a reason | Argument map with explicit inference links; rejected premises justified |
| Creation (15) | Product largely reproduces sources; disclosure missing | Original synthesis not stated in any single source; disclosure complete | Steps correctly labelled deductive or inductive; defence consistent with product |

Instrument blueprint. The unaided reasoning test is planned as 40 multiple-choice items per form, 20 analytical (assumption identification, evidence evaluation, argument structure) and 20 logical (deduction, inference strength, fallacy detection), set in general-education topics and written to match the skills named in Facione (1990) and Ennis (1993). [Author to confirm whether an existing instrument will be adapted and its licence, since this blueprint assumes newly written items.] Construct validation will use expert review, item analysis, reliability, and inter-subscale correlations. A correlation above about .85 between subscales will be taken as evidence of a single factor. Distinguishing valid deduction from plausible inference is built into the items by asking students to judge necessity versus probability.

## *Discussion*

This study developed the Symbiotic Critical Human–AI Learning Model (SC-AIM) and its five-stage FAVIC framework Formulation, Augmentation, Verification, Integration, and Creation to provide a structured approach to analytical and logical reasoning in generative artificial intelligence (GenAI)-

supported higher education. The principal outcomes consist of an evidence-to-design matrix, operational specifications for each learning stage, a verification protocol, scaffold-fading criteria, a provisional assessment rubric, and a prospective evaluation design. These outputs address the study's development objectives by translating relevant empirical evidence and theoretical propositions into a coherent instructional framework. Importantly, the findings represent conceptual design products rather than observed improvements in students' cognitive abilities. Although the Analysis and Design phases were completed and draft Development artefacts were produced, expert validation, classroom implementation, and empirical evaluation remain pending. Consequently, the framework's effectiveness, feasibility, and generalisability have not yet been established.

The underlying design proposition is that productive human–AI collaboration requires the deliberate preservation of human cognitive responsibility. SC-AIM assigns problem formulation, evidential evaluation, inferential judgement, and final synthesis to learners, while allowing AI to generate alternative perspectives, counterarguments, and supplementary explanations. This distinction corresponds with Vendrell and Johnston (2026), who argued that the educational effects of large language models depend on instructional design and the preservation of epistemic agency. Similarly, Anders and Speltz (2025) demonstrated the potential of integrating GenAI literacy with self-regulated learning through structured planning, iteration, and evaluation. Nevertheless, SC-AIM differs by specifying a mandatory pre-AI reasoning commitment, explicit verification return rules, and observable assessment artefacts. These differences establish conceptual distinctiveness, not demonstrated superiority over existing instructional frameworks.

Recent international research provides a nuanced empirical foundation for the proposed framework. Bastani et al. (2025) demonstrated that unrestricted GenAI assistance could improve immediate mathematical performance while undermining subsequent unaided performance. Similarly, Fan et al. (2025) identified improvements in AI-assisted writing performance without corresponding advantages in knowledge transfer, accompanied by reduced metacognitive engagement. Stadler et al. (2024) further reported that lower cognitive effort during GenAI-supported scientific inquiry could coincide with weaker reasoning quality. These findings support the decision to prioritise independent formulation, evidence verification, and unaided assessment. However, they originate from different educational populations and tasks, making their applicability to Indonesian university students indirect.

The relationship between AI assistance and cognitive development is not uniformly negative. Daniel et al. (2025), through a systematic literature review, identified substantial opportunities for GenAI-supported academic skill development. Likewise, Cheng et al. (2025) demonstrated that question-asking practices during GenAI-assisted writing were associated with improved task performance. Gu et al. (2026) found that GenAI-supported feedback activities could facilitate feedback literacy and self-regulatory engagement. These results suggest that AI can contribute positively when students actively interrogate outputs and regulate their learning processes. The distinction is crucial: the educational question concerns not merely whether learners use AI, but how cognitive responsibilities are distributed between learners and intelligent systems.

Further evidence reinforces this interpretation. Darvishi et al. (2024) found that students could become dependent on AI assistance rather than internalise the supported behaviours. Meanwhile, Si et al. (2026) demonstrated heterogeneity in relationships between cognitive offloading and critical

thinking among Chinese university students. Guo and Dong (2026) additionally identified relationships among perceived cognitive load, metacognitive monitoring, and overreliance on generative AI feedback. These findings support SC-AIM's emphasis on monitoring the delegation of cognitive operations. Nevertheless, the present framework should not assume that all cognitive offloading is detrimental. Its instructional contribution lies in distinguishing permissible support from delegation of the reasoning processes that constitute the intended learning outcomes.

The Verification and Integration stages are particularly relevant to emerging scholarship on feedback literacy and metacognitive regulation. Yan et al. (2026) found that the relationship between GenAI use and self-reported critical thinking was associated with feedback literacy and reflective thinking. Similarly, Mendoza et al. (2026) identified feedback literacy as a relevant condition shaping the relationship between students' acceptance of ChatGPT and self-regulated learning. Jin et al. (2025), however, observed that favourable perceptions of GenAI-generated feedback did not necessarily translate into sustained engagement. These findings highlight why SC-AIM requires students to examine original sources, justify inferential decisions, and record revisions rather than merely accept AI-generated explanations. Nevertheless, the proposed verification procedure has not yet been tested for accuracy, workload, or effectiveness in improving independent reasoning.

Research conducted in Indonesia provides further justification for developing context-sensitive AI-supported instructional models. Darwin et al. (2024) reported that Indonesian university students recognised GenAI's potential for academic exploration while expressing concerns about limited nuance and dependence on generated information. Hamsiah et al. (2026) similarly identified tensions between improved academic writing efficiency and students' intellectual autonomy, authorship, and cognitive engagement. These findings suggest that accessibility and convenience do not automatically produce deeper reasoning. However, their perceptual and mixed-methods evidence cannot establish causal effects on independent analytical competence. SC-AIM responds conceptually by requiring learners to formulate initial claims and retain responsibility for the resulting arguments.

Additional Indonesian research emphasises the importance of AI literacy and self-regulated learning. Ramadhan and Susanti (2026) identified positive associations involving GenAI use, AI literacy, ethical awareness, and critical-thinking measures among university students. Hanifatunnisa and Septiadi (2026) similarly examined relationships between AI literacy, prompting skills, and critical-thinking disposition. Diana and Islam (2026) investigated GenAI use and self-regulated learning among Indonesian EFL students. These findings are compatible with SC-AIM's emphasis on responsible prompting and reflective evaluation. Nevertheless, differences in research designs, measurement instruments, and disciplinary contexts prevent direct comparisons with the model's anticipated effects on analytical and logical reasoning.

Related educational investigations provide additional contextual insights. Hidayah and Sodikin (2026) reported favourable differences in academic confidence and decision-making quality associated with explainable AI-supported physics learning, although methodological constraints limited causal interpretation. Damanik and Remtula (2026) highlighted the association between metacognitive readiness and lifelong learning orientation. Srinawati and Rahmidani (2026) similarly examined self-regulated learning in relation to perceived AI usefulness and self-reported critical thinking among Indonesian vocational students. Furthermore, Mardianti et al. (2026) emphasised the need to reorient educator competencies toward digitally supported and reflective pedagogical

practices. Collectively, these studies suggest that AI-supported instruction requires more than technological access: it demands pedagogical guidance, reflective competence, and appropriate assessment. Nevertheless, their results cannot be treated as direct validation of SC-AIM.

The findings have interconnected theoretical, pedagogical, and policy implications. Theoretically, SC-AIM operationalises complementary principles from cognitive offloading, metacognitive regulation, scaffolding, and hybrid human–AI learning by specifying which reasoning operations remain under human control and which forms of assistance may be delegated. The framework extends these perspectives through mandatory pre-AI formulation, stage-specific verification requirements, iterative return rules, and explicit scaffold-fading criteria. Pedagogically, lecturers may use the FAVIC sequence to design inquiry-based assignments requiring students to defend initial claims, evaluate alternative explanations, verify sources, and reflect on revisions. These practices may support analytical reasoning, digital literacy, academic integrity, and responsible identity formation as independent learners. However, their effects on engagement and emotional well-being remain untested. Alsaiari et al. (2025) demonstrated that emotionally enriched AI feedback could reduce negative emotional responses without necessarily improving engagement or performance, reinforcing the need to distinguish affective from cognitive outcomes. At the policy level, higher education institutions should establish transparent AI-use guidelines, assessment practices that distinguish assisted from unaided performance, accessible technological alternatives, and privacy safeguards for student-generated data. Such policies should promote human agency and responsible digital participation rather than treating AI use as inherently beneficial or harmful.

The principal contribution of this study lies in translating evidence concerning GenAI-assisted learning into a process-sensitive instructional model designed specifically to preserve analytical and logical reasoning. Although previous frameworks have addressed AI literacy, self-regulated learning, metacognition, and human–AI collaboration, SC-AIM integrates these principles through the FAVIC sequence, which specifies permitted and prohibited AI functions, stage-specific learning artefacts, verification requirements, and iterative correction mechanisms. Particularly distinctive is the requirement that students establish an initial reasoning position before receiving AI assistance and subsequently demonstrate ownership of their conclusions through independent verification and defence. The proposed assessment rubric also assigns greater weight to Verification and Integration, reflecting the framework's emphasis on evidential judgement rather than polished final products. Methodologically, the evidence-to-design matrix distinguishes empirical support, theoretical assumptions, and author-derived design decisions. The study consequently contributes an operational instructional proposition and an auditable development framework, rather than a validated new theory or demonstrated improvement in cognitive performance.

Several limitations must be acknowledged. First, the study employed a targeted narrative literature review rather than a systematic review, creating potential selection bias and limiting the completeness of its evidence base. Second, the Analysis phase did not include direct needs assessment involving Indonesian students or lecturers, while expert review, prototype testing, implementation, and evaluation have not yet been conducted. Third, the proposed instructional procedures may impose additional cognitive, administrative, and assessment demands, particularly in open and distance learning environments characterised by heterogeneous digital literacy, connectivity, and learning autonomy. Furthermore, the draft rubric and reasoning instruments have not undergone psychometric validation, and the framework's applicability across disciplines remains

uncertain. Future research should begin with independent expert validation and usability-oriented prototype testing before proceeding to classroom-based pilot evaluation. Subsequent studies should compare SC-AIM with existing AI-supported instructional approaches using supervised unaided reasoning assessments, delayed retention measures, transfer tasks, and process-level evidence. Longitudinal and multisite investigations should also examine learner agency, feedback literacy, engagement, equity, and emotional well-being. Such investigations would establish whether the framework can transform human–AI interaction from a mechanism of cognitive substitution into a sustainable form of intellectually accountable collaboration.

## Conclusion

This article produced a conceptual model-development contribution: the SC-AIM model with a restricted definition of symbiosis, the FAVIC specification with return rules, a minimum verification protocol and fading criteria, a draft process-sensitive rubric and instrument blueprint, and an evidence-to-design matrix that states the warrant of each source. Analysis and Design were completed, Development reached the level of draft artefacts, and Implementation and Evaluation are proposed. The contribution is bounded. No expert-validation, pilot, or learner-outcome data are reported, so effects on analytical and logical reasoning, feasibility, acceptability, and generalizability remain untested. The model is intended for pilot use in Indonesian higher education after expert judgement and prototype testing.

## References


Aiken, L. R. (1985). Three coefficients for analyzing the reliability and validity of ratings. Educational and Psychological Measurement, 45(1), 131–142. https://doi.org/10.1177/0013164485451012

Alsaiari, O., Baghaei, N., Lahza, H., Lodge, J. M., Boden, M., & Khosravi, H. (2025). Emotionally enriched AI-generated feedback: Supporting student well-being without compromising learning. *Computers & Education, 239*, Article 105363. https://doi.org/10.1016/j.compedu.2025.105363

Anders, A. D., & Speltz, E. D. (2025). Developing generative AI literacies through self-regulated learning: A human-centered approach. *Computers and Education: Artificial Intelligence, 9*, Article 100482. https://doi.org/10.1016/j.caeai.2025.100482

Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, Ö., & Mariman, R. (2025). Generative AI without guardrails can harm learning: Evidence from high school mathematics. Proceedings of the National Academy of Sciences, 122(26), Article e2422633122. https://doi.org/10.1073/pnas.2422633122

Branch, R. M. (2009). Instructional design: The ADDIE approach. Springer. https://doi.org/10.1007/978-0-387-09506-6

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

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, Article 38. https://doi.org/10.1186/s41239-023-00408-3

Cheng, Y., Fan, Y., Li, X., Chen, G., Gašević, D., & Swiecki, Z. (2025). Asking generative artificial intelligence the right questions improves writing performance. *Computers and Education: Artificial Intelligence, 8*, Article 100374. https://doi.org/10.1016/j.caeai.2025.100374

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

Creswell, J. W., & Plano Clark, V. L. (2018). Designing and conducting mixed methods research (3rd ed.). SAGE Publications.

Damanik, F. H. S., & Remtula, J. N. (2026). Metacognitive readiness in digital learning and its role in lifelong learning pathways. *Digital Learning, Social Science, and Life-course Studies, 2*(1), 57–71. https://doi.org/10.70211/disolife.v1i2.349

Daniel, K., Msambwa, M. M., & Zhang, W. (2025). Can generative AI revolutionise academic skills development in higher education? A systematic literature review. *European Journal of Education, 60*(1), Article e70036. https://doi.org/10.1111/ejed.70036

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

Darwin, Rusdin, D., Mukminatien, N., Suryati, N., Laksmi, E. D., & Marzuki. (2024). Critical thinking in the AI era: An exploration of EFL students' perceptions, benefits, and limitations. Cogent Education, 11(1), Article 2290342. https://doi.org/10.1080/2331186X.2023.2290342

Diana, S. N., & Islam, S. (2026). Generative AI use, self-regulated learning, and critical thinking in EFL writing: A correlational survey of Indonesian university students. *Jurnal Pendidikan Terapan, 4*(4), 1–14. https://doi.org/10.61255/jupiter.v4i4.1479

Ennis, R. H. (1993). Critical thinking assessment. Theory Into Practice, 32(3), 179–186. https://doi.org/10.1080/00405849309543594

Facione, P. A. (1990). Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction (ERIC Document No. ED315423). California Academic Press. https://eric.ed.gov/?id=ED315423

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

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. American Psychologist, 34(10), 906–911. https://doi.org/10.1037/0003-066X.34.10.906

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

Gu, J., Chen, J., & Yan, Z. (2026). Fostering feedback literacy by scaffolding self-regulated feedback: A comparative study of GenAI and human peers. *Frontiers in Education, 10*, Article 1736446. https://doi.org/10.3389/feduc.2025.1736446

Guo, Y., & Dong, P. (2026). University students' overreliance on generative AI feedback in the learning process: The effects of perceived cognitive load, cognitive offloading, and metacognitive monitoring. *Frontiers in Psychology, 17*, Article 1998976. https://doi.org/10.3389/fpsyg.2026.1998976

Hamsiah, A., Rizal, A., & Malik, A. R. (2026). ChatGPT use in higher education writing and its relationship with critical thinking autonomy struggle and authorship. *Discover Education, 5*, Article 580. https://doi.org/10.1007/s44217-026-01627-2

Hanifatunnisa, F. A., & Septiadi, J. (2026). The simultaneous influence of AI literacy and prompting skills on students' critical thinking disposition. *Journal of Educational Sciences, 10*(5), 112–125. https://doi.org/10.31258/jes.10.5.p.112-125

Hedges, L. V. (2007). Effect sizes in cluster-randomized designs. Journal of Educational and Behavioral Statistics, 32(4), 341–370. https://doi.org/10.3102/1076998606298043

Hidayah, N., & Sodikin, S. (2026). Explainable AI, academic confidence, and decision-making quality in physics learning. *Digital Learning, Social Science, and Life-course Studies, 2*(1), 28–45. https://doi.org/10.70211/disolife.v2i1.777

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), Article 248. https://doi.org/10.1145/3571730

Jin, F. J.-Y., Maheshi, B., Lai, W., Li, Y., Gasevic, D., Chen, G., Charwat, N., Chan, P. W. K., Martinez-Maldonado, R., Gašević, D., & Tsai, Y.-S. (2025). Students' perceptions of generative AI-powered learning analytics in the feedback process: A feedback literacy perspective. *Journal of Learning Analytics, 12*(1), 152–168. https://doi.org/10.18608/jla.2025.8609

Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. Educational Research Review, 2(2), 130–144. https://doi.org/10.1016/j.edurev.2007.05.002

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

Kapur, M. (2016). Examining productive failure, productive success, unproductive failure, and unproductive success in learning. Educational Psychologist, 51(2), 289–299. https://doi.org/10.1080/00461520.2016.1155457

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. Journal of Chiropractic Medicine, 15(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Lee, H.-P., Sarkar, A., Tankelevitch, L., Drosos, I., Rintel, S., Banks, R., & Wilson, N. (2025). The impact of generative AI on critical thinking: Self-reported reductions in cognitive effort and confidence effects from a survey of knowledge workers. In Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (Article 1121, pp. 1–22). Association for Computing Machinery. https://doi.org/10.1145/3706598.3713778

Licklider, J. C. R. (1960). Man-computer symbiosis. IRE Transactions on Human Factors in Electronics, HFE-1(1), 4–11. https://doi.org/10.1109/THFE2.1960.4503259

Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (pp. 1–16). Association for Computing Machinery. https://doi.org/10.1145/3313831.3376727

Mardianti, D., Kusimawati, I., & Arifin, B. (2026). Reorienting teacher and lecturer competencies in the era of deep learning and generative artificial intelligence in the transformation of 21st-century learning. *JPI: Jurnal Pustaka Indonesia, 6*(1), 414–432. https://doi.org/10.62159/jpi.v6i1.2257

Marton, F., & Säljö, R. (1976). On qualitative differences in learning: I Outcome and process. British Journal of Educational Psychology, 46(1), 4–11. https://doi.org/10.1111/j.2044-8279.1976.tb02980.x

Mendoza, N. B., Xiong, Y., & Yan, Z. (2026). Making sense of AI feedback: How students' feedback literacy moderates the link between ChatGPT acceptance and self-regulated learning. *Educational Psychology*. Advance online publication. https://doi.org/10.1080/01443410.2026.2641521

Molenaar, I. (2022). Towards hybrid human-AI learning technologies. European Journal of Education, 57(4), 632–645. https://doi.org/10.1111/ejed.12527

Molenda, M. (2003). In search of the elusive ADDIE model. Performance Improvement, 42(5), 34–37. https://doi.org/10.1002/pfi.4930420508

Mosqueira-Rey, E., Hernández-Pereira, E., Alonso-Ríos, D., Bobes-Bascarán, J., & Fernández-Leal, Á. (2023). Human-in-the-loop machine learning: A state of the art. Artificial Intelligence Review, 56(4), 3005–3054. https://doi.org/10.1007/s10462-022-10246-w

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, Article 100041. https://doi.org/10.1016/j.caeai.2021.100041

Ramadhan, G., & Susanti, A. (2026). The impact of generative AI use, AI literacy, and AI usage ethics on critical thinking skills among students at Universitas Negeri Semarang. *Improvement: Jurnal Ilmiah untuk Peningkatan Mutu Manajemen Pendidikan, 13*(1), 63–77. https://doi.org/10.21009/improvement.v13i1.68359

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

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). Experimental and quasi-experimental designs for generalized causal inference. Houghton Mifflin.

Si, S., Qi, Y., Xu, J., & Qi, X. (2026). When thinking is outsourced: Cognitive offloading and the heterogeneity of critical thinking among Chinese university students using generative artificial intelligence. *Journal of Intelligence, 14*(7), Article 116. https://doi.org/10.3390/jintelligence14070116

Srinawati, F. O., & Rahmidani, R. (2026). Perceived usefulness, perceived ease of use, and self-reported critical thinking among Indonesian vocational students: The indirect role of self-regulated learning. *JPI: Jurnal Pustaka Indonesia, 6*(2), 494–514. https://doi.org/10.62159/jpi.v6i2.2634

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

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

Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. Research in Science Education, 48(6), 1273–1296. https://doi.org/10.1007/s11165-016-9602-2

Tankelevitch, L., Kewenig, V., Simkute, A., Scott, A. E., Sarkar, A., Sellen, A., & Rintel, S. (2024). The metacognitive

demands and opportunities of generative AI. In Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (Article 680, pp. 1–24). Association for Computing Machinery. https://doi.org/10.1145/3613904.3642902

UNESCO. (2023). Guidance for generative AI in education and research. https://doi.org/10.54675/EWZM9535

Vendrell, M., & Johnston, S.-K. (2026). Scaffolding critical thinking with generative AI: Design principles for integrating large language models in higher education. *Computers and Education: Artificial Intelligence, 10*, Article 100572. https://doi.org/10.1016/j.caeai.2026.100572

Wang, J., & Fan, W. (2026). Retraction note: The effect of ChatGPT on students' learning performance, learning perception, and higher-order thinking: Insights from a meta-analysis. Humanities and Social Sciences Communications, 13, Article 528. https://doi.org/10.1057/s41599-026-07310-z

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

Yan, Z., Fang, Z., & Zou, W. (2026). The association between generative AI use and university students' critical thinking: A moderated mediation model of GenAI feedback literacy and reflective thinking. *Frontiers in Psychology, 17*, Article 1939286. https://doi.org/10.3389/fpsyg.2026.1939286

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, Article 39. https://doi.org/10.1186/s41239-019-0171-0

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

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