

Evaluative AI Literacy, Learner Control, and Cognitive Offloading in GenAI-Supported Higher Education: A Source-Seeded Evidence Map

Chen Wang^{1*}

¹*Jiangxi Fenglin College of Foreign Economy and Trade, Jiujiang, Jiangxi, China*

Accepted

2026-09-25

Keywords

Evaluative AI literacy; learner control; cognitive offloading; generative AI; self-regulated learning; higher education

Corresponding Author

Chen Wang

Copyright 2026 by author(s)

This work is licensed under the

CC BY 4.0



<https://doi.org/10.70693/itphss.v3i5.678>

Abstract

We examined evaluative AI literacy, learner control, and cognitive offloading in higher education by comparing 13 full-text reports in this source-seeded evidence map. These reports were drawn from a 54-record working bibliography; four abstract/preview records were used solely as background. **No core report empirically tests the full hypothesized chain: evaluative AI literacy → learner control → cognitive offloading.** The course-relative offloading distinctions accompany two hypothetical assignments in which work assisted by generative artificial intelligence (GenAI), reasons for decisions, and learning without GenAI would be assessed separately. These proposals remain unvalidated. English-only selection, incomplete search records, and no formal risk-of-bias appraisal prevent comprehensive coverage and field-wide effect estimates.

1. Introduction

Fan et al. (2025) provide a comparison in which ChatGPT advice was associated with better essay scores but no corresponding significant advantage in knowledge or transfer. The independent learning tests therefore did not show the same pattern as the essays. An essay-only assessment would leave this discrepancy undetected. What the student can explain or apply independently of the assistance used to produce the work must be established through a separate task.

We examine the curated set for evidence linking students' capacity to evaluate AI assistance with informed decisions about its use and the work they delegate. None of the 13 core reports examines this whole sequence. The comparison serves to clarify concepts and identify questions that require further study. It provides no basis for identifying what is missing from all published research or deriving an overall estimate of GenAI's effects.

Consider a student who spots an error but does not investigate it, and another who checks

references while leaving the argument to AI. These hypothetical cases call for more than a literacy score. They motivate our comparison of what literacy measures assess, which control decisions are recorded, and what those decisions reveal about delegated work.

2. Literature Review

2.1. Evaluative competence and the exercise of control

Evaluation appears in both Long and Magerko’s (2020) competency framework and B. Wang et al.’s (2023) scale, which separates it from awareness, use, and ethics. In MAILS, Carolus et al. (2023) distinguish AI competencies from self-efficacy and self-management. Using these distinctions does not imply that a description of a competence, a questionnaire report, and a test demonstration provide equivalent evidence.

Evaluative AI literacy means judging the reliability, relevance, and limits of assistance. It draws on disciplinary knowledge: recognizing that citations may be fabricated differs from deciding whether a genuine source supports a theoretical claim. Confidence and frequent use may accompany this competence, but neither demonstrates it.

Control can mean having a choice, feeling that one has a choice, or acting on it. We call these available, perceived, and enacted control. Task and interface permissions define the available choices; requests, checks, revisions, acceptance, and decisions to stop show how they are used. This extends the focus on pacing and navigation in hypermedia research (Scheiter & Gerjets, 2007), without proposing stages or a validated scale. Self-regulated learning (SRL) connects these actions to goals, monitoring, and strategy adjustment (Zimmerman, 2002). Lan and Zhou (2025) review broader AI support for SRL; Xia et al. (2026) focus on GenAI.

2.2. Work allocation and the learning objective

Risko and Gilbert (2016) describe cognitive offloading as using external resources or actions to reduce internal processing demands. In coursework, the relevant question is which demands the learning objective requires students to retain. We use cognitive responsibility for participation in, and accountability for, that reasoning and practice. Constructive alignment links this criterion to course objectives (Biggs, 1996); ICAP (interactive, constructive, active, passive) distinguishes visible activity from constructive engagement (Chi & Wylie, 2014). These are conceptual starting points, not validation of the offloading categories proposed below.

Table 1. Construct boundaries and the evidence needed to distinguish them

Construct	Evidence to distinguish it	Insufficient substitute
Evaluative AI literacy	Independent judgments of reliability, relevance, and limits.	Confidence or use frequency.
Learner control	Permitted options; perceived choice; observed decisions.	One undifferentiated score or a sequence of stages.
Self-regulated learning	Goals, monitoring, and justified strategy adjustment.	A reflective prompt being available.
Cognitive offloading	Who performs and checks each target-related operation.	Message counts, word edits, or later test scores.
Independent learning	Understanding, retention, or transfer with assistance conditions stated.	Assisted products or perceived improvement.

Note. Author synthesis; analytic distinctions, not validated measures, coding rules, or causal evidence. A checking event cannot be reused as independent evidence of successive links.

3. Methodology and Procedures

3.1. A custom narrative review variant

Source-seeded evidence map names this custom narrative review; it is not an established review type. Grant and Booth (2009) discuss systematized reviews without defining this variant. This is neither a systematic review nor a scoping review conducted under JBI methodology and reported through PRISMA-ScR (Peters et al., 2020; Tricco et al., 2018). It organizes inherited sources by constructs, measurements, and evidential limits.

Conceptual differences and questions for further study, rather than field coverage or a combined effect estimate, were examined through the bibliography. Scoping reviews also clarify concepts and gaps (Peters et al., 2020); the distinction here concerns how sources were found and examined. This explanation is retrospective. Separate, documented work would be needed for a systematic review; reporting guidance cannot retrospectively supply that work (Page et al., 2021).

3.2. Eligibility and the historical candidate register

The empirical window runs from 1 January 2022 to 22 September 2026, using first public availability. Eligible English-language reports concerned higher-education learners' GenAI use and addressed literacy, control/regulation, offloading, or learning comparisons relevant to delegation. They need not measure all three constructs. Multiple versions count once. Foundational work outside this window supplies concepts and methods, not empirical learner evidence for the comparison.

Earlier drafts supplied 48 entries, followed by five targeted additions (IDs 49–53) and a control report (ID 54). Supplement 1 preserves their identifiers and decisions, including eight incomplete identities. This is a working register, not a database export; individual discovery routes and a separate project as its origin remain unestablished. Figure 1 reconciles the records with the present core.

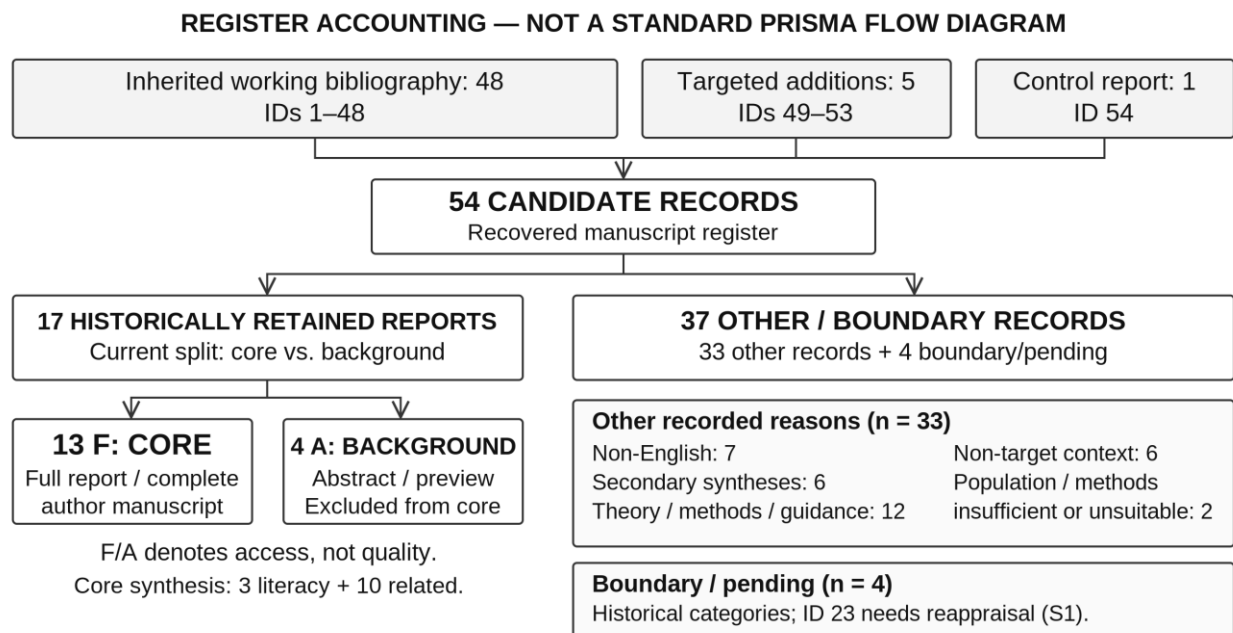


Figure 1. Historical register accounting and the revised 13-report core.

Note. **Historical record accounting, not a PRISMA flow diagram.** Numbers summarize recorded decisions, not database yields or systematic screening stages. Four A records remain outside the core. Reason groups are retrospective; ID 23 requires eligibility reappraisal (Supplement 1). Missing search and deduplication totals have not been reconstructed.

3.3. Full-text access and the core/background boundary

F denotes a complete article or author manuscript with accessible methods and results. A denotes abstracts, summaries, previews, or fragments leaving those sections incomplete. A DOI alone is insufficient. The 13 F/four A designations are inherited access labels, not quality ratings. Earlier revisions checked bibliographic records and selected passages without repeating every report’s extraction.

The earlier count of 17 included both access groups. Here, 13 F reports enter Table 2 and Sections 4.1–4.3; IDs 6, 26, 30, and 53 appear only as discussion background. Supplement 1 explains the corrected reason recorded for Zhu et al. (2026) and the updated publication details for L. Zhang and Chen (2026). Updating those records did not amount to a new screening exercise.

3.4. Narrative comparison and the limits of appraisal

Studies were compared according to their measures, permitted assistance, delegated work, and assessment conditions and timing. **No effect sizes were combined, and no formal quality or risk-of-bias ratings, certainty grades, or proposed offloading codes were applied.** The descriptive discussion of these concerns appears in Section 3.5. The source on which this set’s direct perceived-control comparison depends is Qin et al.’s (2025) exploratory preprint. Removing it leaves 12 reports without that comparison in this form. This is a check of source dependence, not statistical sensitivity.

3.5. Common methodological weaknesses across core included reports

The questionnaires in Fitriani et al. (2026) and Liu et al. (2025) record students’ reports without observing their decisions during a task. Qin et al. (2025) examine the reported sense of control, leaving unanswered whether students checked answers or delegated work appropriately. In Abbas et al. (2024), survey responses were collected at different times; cognitive skill was not repeatedly tested.

The learning assessment in Fan et al. (2025) used independent post-tests. Products, prompts, or group discussions alone cannot provide equivalent evidence, and retention is unknown where later testing is absent. Table 2 identifies other interpretive limitations: reliance on volunteers, differing interfaces, brief observation periods, and several instructional changes introduced together. These observations are not formal bias ratings or an overall certainty grade.

4. Results and Discussion

Table 2 contains three literacy-focused reports and ten concerning regulation, offloading, or learning comparisons. Their retained accounts contain no joint test of evaluative literacy, learner control, and cognitive offloading. Separate findings cannot become a tested chain when measures, tasks, and samples differ. The four A records contribute background only, outside this synthesis.

Table 2. The 13-report full-text core: contributions and limits

Report / role / access	Evidence contributed	Interpretive limit
Jin et al. (2025) Literacy · F	20-item GLAT; objective GenAI knowledge and prediction of supported performance.	Neither isolated evaluative judgment nor an instructional effect on unaided retention.
Tzirides et al. (2024) Literacy · F	AI review, peer feedback, and image-generation reflection in graduate courses.	Volunteers and bundled activities; perceived literacy development.

Report / role / access	Evidence contributed	Interpretive limit
Fitriani et al. (2026) Literacy · F	Survey associations among literacy, dependency, offloading, and perceived ability.	Self-reported ability; the offloading significance claim depends on a testing approach not fully specified.
Qin et al. (2025) Related · F; preprint	Higher perceived control in an instructional-agent comparison.	New scale, differing interfaces, and post-randomization exclusions.
Fan et al. (2025) Related · F	Advice-only ChatGPT support improved essay revision; no significant advantage in knowledge gain or transfer.	Not unrestricted essay outsourcing; no evidence of lasting skill decline.
Liu et al. (2025) Related · F	Distinguishes task-processing strategies from regulation of AI applications.	Reported strategies are not observed decisions.
Edwards et al. (2025) Related · F	Prototype for socially shared regulation in collaborative work.	Prompts supplied do not establish effective shared regulation.
Y. Zhang et al. (2027) [†] Related · F	Distribution of regulatory activity in human–AI collaboration.	Group processes do not establish individual retention.
Kestin et al. (2025) Related · F	Immediate learning advantage for structured AI tutoring over in-class active learning.	Tutoring design and setting change together.
Urban et al. (2024) Related · F	Improved assisted creative problem solving; no difference in absolute self-evaluation accuracy.	Not delayed independent creativity or improved calibration.
Misiejuk et al. (2026) Related · F	281 prompts from 122 submissions; prompting related to submission quality.	Units are prompts/products, not independent skill-change scores.
Elshall and Badir (2025) Related · F	Environmental data-science assessment design.	A planned take-home critical-thinking component was not implemented.
Abbas et al. (2024) Related · F	Time-separated university surveys of GenAI use and reported consequences.	Memory complaints are not repeated objective memory tests.

Note. All 13 records are F (complete report/author manuscript in the retained access ledger); F is not a quality rating. “Related” reports are part of the core, unlike the four A background records omitted here. A report is not necessarily an independent cohort. Entries cannot establish a joint causal chain. [†]Y. Zhang et al.: citation year 2027; first online 27 June 2026, within the cutoff. Source: listed reports and retained extraction account.

4.1. Literacy measures do not establish informed action

Although considered together here, Jin et al.’s (2025) knowledge test, the coursework in

Tzirides et al. (2024), and the questionnaires in Fitriani et al. (2026) address different questions. None measures evaluative competence separately and follows it through control decisions and delegated work in one task. These reports therefore leave unresolved whether accurate evaluators also check, revise, or reject assistance under comparable conditions.

4.2. Feeling in control and exercising judgment

Qin et al. (2025) report students' experience of control; Liu et al. (2025) examine the strategies they say they use. To judge enacted control, we need to know what help a student requested, what was checked, and why an answer was accepted. Interface options do not answer those questions or establish whether the decision to delegate was sound.

Regulation within groups is the subject of Edwards et al. (2025) and Y. Zhang et al. (2027; first online 27 June 2026). One member could supply the reasoning in a group answer, leaving the others' understanding unclear. Pairing records of each member's participation with individual post-tests without GenAI would enable future research to examine each student's delegated work and learning, rather than infer them from the group product.

4.3. Assisted performance and independent learning

The comparisons are not uniformly negative. Kestin et al. (2025) report an immediate learning advantage with structured tutoring; Urban et al. (2024) find better assisted creative problem solving. Fan et al. (2025) show why revision gains need separate learning measures. Different interventions and assessments do not estimate one common effect. Efklides' (2006) metacognitive account likewise separates experienced ease from knowledge and control.

Independent skill change remains unmeasured in Misiejuk et al. (2026); the reported relationship concerns prompting and submission quality. Elshall and Badir (2025) did not implement the assessment component, so its effect remains unestablished. Interpreting the memory complaints in Abbas et al. (2024) requires appropriate baseline and follow-up evidence to distinguish less practice, failure to learn, and loss of an existing skill.

4.4. Untested relationships and course-relative offloading

Figure 2 presents questions suggested by the comparison, leaving direction, timing, and absent associations open. **This map has tested neither a causal pathway nor mediation.** SRL provides theoretical context, not an estimated moderator. The lower panels identify separate evidence about assisted work and independent learning, rather than outcomes predicted by a fitted model.

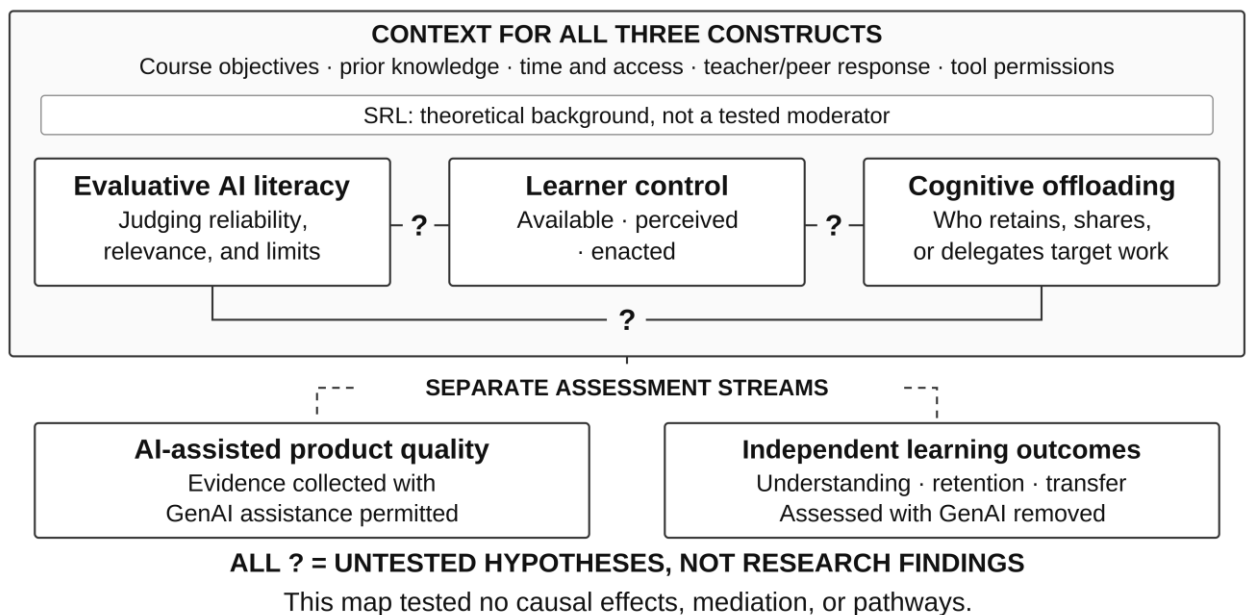


Figure 2. Untested conceptual relationships and separate assessment streams.

Note. Author proposal, not a fitted model. **Every question-mark connector represents an untested hypothesis; no mediation effect has been tested.** The lines are non-directional and specify neither timing nor a required intermediary. Control aspects are not stages. Dashed branches organize assessment rather than estimated effects; SRL provides theoretical context.

Supportive, substitutive, and mixed offloading are unvalidated, course-relative heuristics. Supportive offloading retains the student's target reasoning or reconstruction. Substitutive offloading transfers required judgment or practice to AI without that work being reconstructed by the learner. Mixed offloading describes changes in allocation across operations or stages. The terms refer to how a task is done, not to student types or demonstrated learning benefits.

None of the 13 core reports was coded with these heuristics. Empirical use would first require a manual specifying task-stage units, course objectives, decision rules, examples, and treatment of missing evidence. Pilot coding should establish whether independent coders can apply the rules consistently and whether the distinctions capture the intended processes. Learning scores should be concealed during classification so that a favorable outcome does not determine the label.

Zhu et al. (2026) distinguish dependent offloading, which transfers core thinking to AI, from autonomous offloading, which retains the user's reasoning. Their three-wave survey of students and early-career workers relates these potentially coexisting modes to agency, motivation, and perceived outcomes. The dimensions do not label fixed learner types. This outside-core study supplies an empirical comparator, not validation of the present course-relative heuristics.

A student in a hypothetical task could select a model and check its assumptions, leaving the arithmetic to AI. This could count as supportive offloading when model selection is the objective, but as substitutive offloading when the calculation itself must be learned. Checking one stage while leaving another unexamined could be mixed; missing evidence stays unclassified. Independent learning measures are required to examine whether this course-relative distinction adds to Zhu et al.'s (2026) survey dimensions. Jiang and Chen (2026) remain a boundary record awaiting extraction.

4.5. Hypothetical assessment illustrations—not validated interventions

The following examples keep the assisted submission, the student's reasons for decisions, and a task without GenAI as separate sources of evidence. **Both assignments are hypothetical. Neither has been implemented or evaluated for this map, and both need piloting.**

The writing example would use two supplied articles for a 600-word argument, with students identifying AI contributions to the essay and explaining two decisions through reference to source passages. A ten-minute task without GenAI would subsequently assess a new claim–source pairing. In this hypothetical design, separate scores would be recorded for the essay, the decision justifications, and the independent judgments.

A 500-word interpretation would form the hypothetical methods assignment. It would contain a rejected explanation, a plausible confounder, and evidence that could change the conclusion. In a subsequent case without GenAI, students would critique a causal claim while retaining equivalent source access and necessary accessibility support. Brief decision notes could replace full chat records, with research use kept separate from grading.

The background sources describe writing (Fyfe, 2023; Anders & Dux Speltz, 2025), metacognitive support (Xu et al., 2025), and virtual-reality feedback (W.-S. Wang et al., 2025). Their limited-access accounts do not validate either assignment. Pilots would establish the marking workload, student completion time, accessibility, and feasibility within class time,

providing practical evaluation before testing effectiveness. The proposal concerns better assessment, not discouraging appropriate GenAI use.

4.6. Separating measures before testing relationships

Inspecting each item is necessary when reviewing a literacy scale. Lee and Kan (2026) include self-regulated use within literacy; L. Zhang and Chen (2026) study the constructs alongside one another. When a predictor includes the behavior designated as the outcome, the concern raised by these boundary sources is that shared content may account for part of the apparent relationship. High reliability or separate factors do not remove that overlap. Deleting items therefore requires the revised scale to undergo validation.

One possible assessment would ask students to examine the same reliable and flawed AI outputs against supplied evidence before using AI for their own work. They would judge relevance and limitations and explain corrections without generating more answers. Domain knowledge would be tested separately, with scoring blinded. Later tasks would record available choices, perceived control, decisions made, and who did each operation. Those work records, rather than reused evaluation scores or message counts, would measure offloading.

Parallel tasks without GenAI would be administered immediately and after a delay to assess independent learning. The measures must undergo content review and piloting and show evidence that they assess distinct constructs before the comparisons proceed. Questionnaires may provide context, but cannot replace performance or delegation records. Only when these requirements are met and the allowed assistance is specified could studies compare evaluation instruction or control scaffolds.

Following learners across comparable tasks would connect repeated decision records to later independent performance. A standardized manual, sensitive to course objectives, would keep offloading codes consistent across coders and occasions. This design could examine the chain beyond one questionnaire. Sequence alone cannot establish causal mediation; instructional experiments would also have to address confounding of intermediate choices.

4.7. Limits of coverage and reproducibility

This evidence map should not be read as a scoping review. Earlier manuscript arguments shaped the seed, which could then reinforce the emerging framework—a risk of confirmation bias inherent in this design. English-only eligibility omits non-English empirical work. The January 2022–22 September 2026 window is also a snapshot: new studies, models, permissions, and teaching practices could alter the comparison. Apparent gaps concern this curated set of 13 core reports, not the entire field.

We did not conduct a systematic search of grey or unpublished literature. Nor does the inclusion of one preprint identify omitted sources or permit estimation of the effect of omissions. **With queries, search returns, and provenance missing, reconstruction of the original search is not possible; the register supports count checks only.** There was no complete multi-database search, preregistration, or independent duplicate screening/extraction. Eight unresolved source identities, retrospective eligibility decisions, and selection by access remain limitations despite removing the four A records from the core.

A future review should register questions and decisions, preserve dated searches and raw exports, link duplicate versions, and retain both reviewers' screening, extraction, and appraisal records. This could produce a reproducible new evidence set, but cannot supply the missing original history. Supplement 1, S6 describes these prospective steps.

5. Conclusion and Suggestion

The full evaluative literacy–control–offloading chain remains untested in these 13 reports. The comparison identifies the need for different measures of knowledge, perceived choice, delegated work, and assisted output when studying the chain. The course-relative distinctions and hypothetical assignments are proposals, not established mechanisms or estimates for the field. Separate assessment of assisted work, reasons for decisions, and performance without GenAI would allow these proposals to be tested without presuming either benefits or harms.

Declarations

Funding. This research received no specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

AI assistance. Generative AI tools assisted source discovery and organization, language revision, and figure formatting. The author retains responsibility for source verification, interpretation, and the final conclusions.

Data availability. No new participant data were collected. Supplement 1 preserves the 54-record working register, the 13-core/four-background distinction, and version notes. The earlier verification record documents the bibliographic checks made before this language revision. Original complete search returns and extraction archives remain unavailable, preventing full reproduction of the earlier retrieval.

References

- Abbas, M., Jam, F. A., & Khan, T. I. (2024). Is it harmful or helpful? Examining the causes and consequences of generative AI usage among university students. *International Journal of Educational Technology in Higher Education*, 21, Article 10. <https://doi.org/10.1186/s41239-024-00444-7>
- Anders, A. D., & Dux Speltz, E. (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>
- Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education*, 32(3), 347–364. <https://doi.org/10.1007/BF00138871>
- Carolus, A., Koch, M. J., Straka, S., Latoschik, M. E., & Wienrich, C. (2023). MAILS—Meta AI literacy scale: Development and testing of an AI literacy questionnaire based on well-founded competency models and psychological change- and meta-competencies. *Computers in Human Behavior: Artificial Humans*, 1(2), Article 100014. <https://doi.org/10.1016/j.chbah.2023.100014>
- Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, 49(4), 219–243. <https://doi.org/10.1080/00461520.2014.965823>
- Edwards, J., Nguyen, A., Lämsä, J., Sobocinski, M., Whitehead, R., Dang, B., Roberts, A.-S., & Järvelä, S. (2025). Human-AI collaboration: Designing artificial agents to facilitate socially shared regulation among learners. *British Journal of Educational Technology*, 56(2), 712–733. <https://doi.org/10.1111/bjet.13534>
- Efklides, A. (2006). Metacognition and affect: What can metacognitive experiences tell us about the learning process? *Educational Research Review*, 1(1), 3–14. <https://doi.org/10.1016/j.edurev.2005.11.001>
- Elshall, A. S., & Badir, A. (2025). Balancing AI-assisted learning and traditional assessment: The FACT assessment in environmental data science education. *Frontiers in Education*, 10, Article 1596462. <https://doi.org/10.3389/feduc.2025.1596462>

- 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>
- Fitriani, A. N., Risaldy, R., Rauf, A., & Afidatunisa, S. (2026). Academic dependency, AI literacy, and cognitive offloading predict students' cognitive ability in generative AI learning. *Artificial Intelligence in Lifelong and Life-Course Education*, 1(2), 53–66. <https://doi.org/10.66053/ailce.v1i2.18>
- Fyfe, P. (2023). How to cheat on your final paper: Assigning AI for student writing. *AI & SOCIETY*, 38, 1395–1405. <https://doi.org/10.1007/s00146-022-01397-z>
- Grant, M. J., & Booth, A. (2009). A typology of reviews: An analysis of 14 review types and associated methodologies. *Health Information & Libraries Journal*, 26(2), 91–108. <https://doi.org/10.1111/j.1471-1842.2009.00848.x>
- Jiang, Y., & Chen, H. (2026). AI-based cognitive offloading by students in higher education in China. *Computers in Human Behavior: Artificial Humans*, 8, Article 100327. <https://doi.org/10.1016/j.chbah.2026.100327>
- Jin, Y., Martinez-Maldonado, R., Gašević, D., & Yan, L. (2025). GLAT: The generative AI literacy assessment test. *Computers and Education: Artificial Intelligence*, 9, Article 100436. <https://doi.org/10.1016/j.caeai.2025.100436>
- Kestin, G., Miller, K., Klales, A., Milbourne, T., & Ponti, G. (2025). AI tutoring outperforms in-class active learning: An RCT introducing a novel research-based design in an authentic educational setting. *Scientific Reports*, 15, Article 17458. <https://doi.org/10.1038/s41598-025-97652-6>
- Lan, M., & Zhou, X. (2025). A qualitative systematic review on AI empowered self-regulated learning in higher education. *npj Science of Learning*, 10, Article 21. <https://doi.org/10.1038/s41539-025-00319-0>
- Lee, Y. E., & Kan, J. S. (2026). AI literacy and self-perceived cognitive learning outcomes among university students in AI-integrated courses: Associations with instructor feedback and AI use indicators. *Education Sciences*, 16(9), Article 1384. <https://doi.org/10.3390/educsci16091384>
- Liu, X., Xiao, Y., & Li, D. (2025). Assessing strategic use of artificial intelligence in self-regulated learning: Instrument development and evidence from Chinese university students. *International Journal of Educational Technology in Higher Education*, 22, Article 69. <https://doi.org/10.1186/s41239-025-00567-5>
- 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>
- Misiejuk, K., López-Pernas, S., Kaliisa, R., & Saqr, M. (2026). Cognitive offloading in student–AI collaboration: A longitudinal analysis of prompting strategies. *Computers in Human Behavior Reports*, 22, Article 101130. <https://doi.org/10.1016/j.chbr.2026.101130>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., . . . Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, Article n71. <https://doi.org/10.1136/bmj.n71>
- Peters, M. D. J., Marnie, C., Tricco, A. C., Pollock, D., Munn, Z., Alexander, L., McInerney, P., Godfrey, C. M., & Khalil, H. (2021). Updated methodological guidance for the conduct of

scoping reviews. *JBIE Evidence Synthesis*, 18(10), 2119–2126. <https://doi.org/10.11124/JBIES-20-00167>

Qin, F., Hao, Z., Yu, J., Liu, Z., & Zhang, Y. (2025). *AI instructional agent improves student's perceived learner control and learning outcome: Empirical evidence from a randomized controlled trial* [Preprint]. arXiv. <https://doi.org/10.48550/arXiv.2505.22526>

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>

Scheiter, K., & Gerjets, P. (2007). Learner control in hypermedia environments. *Educational Psychology Review*, 19(3), 285–307. <https://doi.org/10.1007/s10648-007-9046-3>

Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., Moher, D., Peters, M. D. J., Horsley, T., Weeks, L., Hempel, S., Akl, E. A., Chang, C., McGowan, J., Stewart, L., Hartling, L., Aldcroft, A., Wilson, M. G., Garritty, C., . . . Straus, S. E. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. *Annals of Internal Medicine*, 169(7), 467–473. <https://doi.org/10.7326/M18-0850>

Tzirides, A. O., Zapata, G., Kastania, N. P., Saini, A. K., Castro, V., Ismael, S. A., You, Y., Santos, T. A. D., Searsmith, D., O'Brien, C., Cope, B., & Kalantzis, M. (2024). Combining human and artificial intelligence for enhanced AI literacy in higher education. *Computers and Education Open*, 6, Article 100184. <https://doi.org/10.1016/j.caeo.2024.100184>

Urban, M., Děchtěrenko, F., Lukavský, J., Hrabalová, V., Svacha, F., Brom, C., & Urban, K. (2024). ChatGPT improves creative problem-solving performance in university students: An experimental study. *Computers & Education*, 215, Article 105031. <https://doi.org/10.1016/j.compedu.2024.105031>

Wang, B., Rau, P.-L. P., & Yuan, T. (2023). Measuring user competence in using artificial intelligence: Validity and reliability of artificial intelligence literacy scale. *Behaviour & Information Technology*, 42(9), 1324–1337. <https://doi.org/10.1080/0144929X.2022.2072768>

Wang, W.-S., Lin, C.-J., Lee, H.-Y., Huang, Y.-M., & Wu, T.-T. (2025). Enhancing self-regulated learning and higher-order thinking skills in virtual reality: The impact of ChatGPT-integrated feedback aids. *Education and Information Technologies*, 30, 19419–19445. <https://doi.org/10.1007/s10639-025-13557-x>

Xia, Q., Liu, Q., Tlili, A., & Chiu, T. K. F. (2025). A systematic literature review on designing self-regulated learning using generative artificial intelligence and its future research directions. *Computers & Education*, 240, Article 105465. <https://doi.org/10.1016/j.compedu.2025.105465>

Xu, X., Qiao, L., Cheng, N., Liu, H., & Zhao, W. (2025). Enhancing self-regulated learning and learning experience in generative AI environments: The critical role of metacognitive support. *British Journal of Educational Technology*, 56(5), 1842–1863. <https://doi.org/10.1111/bjet.13599>

Zhang, L., & Chen, S. C. (2026). Towards human-AI collaborative learning: Synergizing self-regulation and artificial-intelligence literacy. *Education and Information Technologies*, 31, 1879–1907. <https://doi.org/10.1007/s10639-025-13880-3>

Zhang, Y., Meng, X., Feng, S., & Lin, J. (2027). Human–AI collaboration reconfigures group regulation from socially shared to hybrid co-regulation. In E. G. Blanchard, G. Chen, M. Chi, & S. Isotani (Eds.), *Artificial intelligence in education* (Lecture Notes in Computer Science, Vol. 16585, pp. 318–326). Springer. https://doi.org/10.1007/978-3-032-29770-9_35
First published online 27 June 2026.

Zhu, Q., Li, X., Dong, Y., Chang, P., & Fan, M. (2026). Not all cognitive offloading is equal: Distinguishing dependent and autonomous offloading to generative AI. *Frontiers in Psychology*, 17, Article 1878629. <https://doi.org/10.3389/fpsyg.2026.1878629>

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