International Journal of Education and Vocational Studies

Vol. 1, No. 2, November 2026, pp. 69-91
ISSN: 3124-9132 (Media Online)
DOI: https://doi.org/10.62712/ijevs.v1i2.9

Web-Based and Mobile AI Tools in Southeast Asian Higher Education: A Systematic Review and Thematic Synthesis

Supiyandi[1*], Ramlah binti Mailok[2]

[1]Faculty of Computational Science and Digital Intelligence, Information Technology, Universitas Pembangunan Panca Budi, Medan, Indonesia
[2]Faculty of Computing and Meta-Technology, Universiti Pendidikan Sultan Idris, Tanjung Malim, Perak, Malaysia
Author's Email: [1*]supiyandi@dosen.pancabudi.ac.id, [2]mramlah@meta.upsi.edu.my

Article Info

Article history:

Received: August 28, 2026
Accepted: October 15, 2026
Published: November 1, 2026

Keywords:

Artificial intelligence in education;
Generative artificial intelligence;
Higher education;
Southeast Asia;
Systematic review.

ABSTRACT

Web-based and mobile artificial intelligence tools have spread rapidly in Southeast Asian universities, yet evidence on what they achieved in real learning remained fragmented. This systematic review synthesised empirical studies that implemented or evaluated such tools in degree-level higher education in Southeast Asia between 2021 and 2026. Following the PRISMA 2020 guideline, 642 Scopus records were screened, 79 full texts were assessed and 59 studies were included. Study quality was appraised with the Mixed Methods Appraisal Tool, measured outcomes were summarised by effect direction, and findings were integrated through three-stage thematic synthesis. Most studies came from Indonesia, Vietnam and Thailand, concerned language learning, used general-purpose chatbots and relied on non-randomised designs. Of 45 studies that measured learning, 36 reported improvements, mainly in uncontrolled designs or against routine teaching, whereas four of six comparisons with equivalent human support found no difference. Five analytical themes emerged: polished products with uncertain learning, pedagogy over presence, an affective double edge, unequal readiness and comparator-dependent conclusions. The evidence supported these tools as comparable and scalable complements within teacher-mediated, bounded designs rather than as superior replacements. The review proposed an orchestration framework and a research agenda that prioritised active comparators and independent, delayed measures of learning.

This work is licensed under a CC BY-SA License.

Corresponding Author:

Supiyandi,
Faculty of Computational Science and Digital Intelligence, Information Technology, Universitas Pembangunan Panca Budi,
Medan, Indonesia
Email: supiyandi@dosen.pancabudi.ac.id

1. INTRODUCTION

Artificial intelligence in higher education has moved from specialised systems, such as intelligent tutors, automated writing evaluation and learning analytics, to general-purpose tools that any student can open in a browser or on a phone. An early systematic review of AI applications in higher education found a field dominated by technical development and asked, "where are the educators?" (Zawacki-Richter et al., 2019). A later state-of-the-field review confirmed rapid growth in AI research in higher education across assessment, tutoring and student support (Crompton & Burke, 2023). The public release of large language models changed the scale of the question: conversational systems now generate explanations, feedback, examples and drafts on demand, which creates educational opportunities and risks at the same time (Kasneci et al., 2023).

Southeast Asian universities combine rapid digital adoption with large differences between countries, institutions and learners in resources, connectivity and English proficiency. Many students learn through

English or another additional language, and many institutions operate with large classes and limited individual feedback. Web-based and mobile AI tools are therefore attractive as a low-cost source of practice and feedback. The same conditions may also limit how well tools are used: uneven AI literacy, unreliable access and different institutional policies can shape what students gain. Evidence from this region is growing quickly but is dispersed across national and regional journals, disciplines and research traditions.

Three problems make the regional evidence hard to use. First, much research on AI in education measures acceptance, intention or perception rather than what happens when a tool is implemented in teaching. Second, implementation studies vary widely in design, from one-group pre/post comparisons to randomised trials, and positive findings from weak designs are easily read as effectiveness. Third, outcomes are often reported as students' perceptions of improvement, which can diverge from measured learning. Searches of OpenAlex and the web (30 October 2026) identified reviews of AI in higher education globally and of AI in other Southeast Asian sectors, but no systematic review restricted to implementation evidence from Southeast Asian higher education that separates measured from self-reported outcomes; this reflects our searches rather than proof of absence.

This review therefore synthesises empirical studies that implemented or evaluated AI tools in real learning activities in degree-level higher education in Southeast Asia between 2021 and 2026. It combines a structured appraisal of study quality, an effect-direction summary of measured outcomes and a thematic synthesis of reported experiences. The research questions are:

RQ1. What are the contexts, AI applications, delivery modes and research designs of empirical studies on AI-supported learning in Southeast Asian higher education published between 2021 and 2026?

RQ2. What learning outcomes have been reported, and how do findings differ between measured and self-reported outcomes and across study designs and comparator types?

RQ3. Which pedagogical, learner-related and institutional conditions are associated with more or less favourable outcomes?

RQ4. What challenges and risks do students and instructors report when AI is integrated into learning?

RQ5. How methodologically robust is this evidence base, and which gaps should future research address?

2. LITERATURE REVIEW

2.1. Delivery, pedagogy and learning theory

Most of these tools are web-based or mobile services rather than institutional systems. Students reach them outside timetabled contact, often without institutional mediation, and teachers increasingly build them into feedback, practice and project work. This makes the pedagogical arrangement around the tool, not only the tool itself, a central object of study. Established theory points the same way. Feedback improves learning when it answers where the learner is going, how they are going and where to next (Hattie & Timperley, 2007). Learners benefit when they plan, monitor and reflect on their own work (Zimmerman, 2002), and hybrid human–AI arrangements are expected to support rather than replace such regulation (Molenaar, 2022). From a cognitive-load perspective, help that removes productive effort can raise immediate performance while weakening learning (Sweller, 1988).

2.2. The performance–learning gap

Recent experiments outside Southeast Asia illustrate why this matters. In a large field experiment in high school mathematics, unrestricted access to a generative AI tutor improved performance during practice but harmed later performance once access was removed, whereas a version with pedagogical guardrails largely mitigated the harm (Bastani et al., 2025). A laboratory study in higher education found that ChatGPT support improved essay scores without corresponding gains in knowledge or transfer, and described the pattern as "metacognitive laziness" (Fan et al., 2025). Both studies suggest that measured gains in AI-assisted products need not reflect independent learning.

2.3. Conceptual lens of the review

Building on these perspectives, this review reads the evidence through three distinctions. The first separates measured learning outcomes from students' perceptions, because feedback and motivation theories predict that confidence and learning can diverge. The second separates the quality of products made with AI from independent performance without it, which follows from the performance–learning gap described above. The third separates comparisons with routine teaching from comparisons with human support of equal intensity, because added practice and attention can produce gains regardless of the tool. These distinctions shaped data extraction, the effect-direction summary and the analytical themes.

3. RESEARCH METHOD

3.1. Research design

This is a systematic review with a convergent segregated mixed-methods synthesis. The review methodology followed standard systematic-review stages; reporting follows PRISMA 2020 (Page et al., 2021). Because the included studies combine quantitative, qualitative and mixed-methods designs with

heterogeneous outcomes and instruments, meta-analysis was not appropriate. Measured outcomes were therefore summarised by effect direction following the Synthesis Without Meta-analysis (SWiM) guideline (Campbell et al., 2020), and all findings were integrated through thematic synthesis (Thomas & Harden, 2008).

3.2. Review protocol and amendment

The protocol was not prospectively registered. Its elements (objective, eligibility criteria, search, selection, extraction, appraisal and synthesis) were fixed before full-text assessment and are documented in the screening workbook. One amendment was made on 30 October 2026, before the full texts of the affected records were retrieved, to sharpen three criteria: (a) pre-university, foundation and pre-sessional programmes (ISCED levels 3–4) do not count as higher education; (b) multi-country studies are eligible only when Southeast Asian results are reported separately; and (c) predictive learning-analytics studies without a learner-facing AI intervention and a measured learning outcome are excluded (new code E7). Three records whose full texts had already been read were excluded under the amended criteria at the full-text stage.

3.3. Database and search strategy

Scopus was searched as the single bibliographic database because it indexes the regional and international journals in which most Southeast Asian education research is published and supports reproducible field-restricted searching. No other database was searched; this is acknowledged as a limitation. The search combined four concept groups (Table 1).

Table 1. Database search strategy

Element	Specification
Database	Scopus (Elsevier)
Search field	TITLE-ABS-KEY
Concept 1: AI	"artificial intelligence" OR "machine learning" OR "deep learning" OR "generative AI" OR chatbot *OR "large language model*" OR ChatGPT OR "intelligent tutoring" OR "learning analytics"
Concept 2: higher education	universit *OR "higher education" OR undergraduate* OR "college student*" OR "medical student*"
Concept 3: implementation	experiment *OR "quasi-experiment*" OR intervention OR "randomi?ed" OR "pre-test" OR "post-test" OR implement *OR evaluat*
Concept 4: Southeast Asia	Indonesia *OR Malaysia* OR Thai *OR Vietnam* OR Philippin *OR Filipino OR Singapore* OR Brunei OR Cambodia *OR Lao* OR Myanmar OR Timor
Limits	PUBYEAR > 2020 AND PUBYEAR < 2027; DOCTYPE = article; LANGUAGE = English
Export date	30 October 2026 (09:54 UTC)
Records retrieved	642

3.4. Eligibility criteria

Table 2. Inclusion and exclusion criteria

Domain	Inclusion	Exclusion (code)	Rationale
Study type	Primary empirical study (quantitative, qualitative or mixed)	Reviews, bibliometric, conceptual or framework papers, protocols (E1)	The review synthesises primary evidence
Role of AI	AI/ML tool used by learners or teachers in a real learning activity	AI as course content, administrative use, ML only as an analysis method, studies outside education (E2)	Focus is AI as a learning tool
Implementation	A specific AI tool or AI-integrated activity implemented in instruction and evaluated	General acceptance, adoption or perception surveys; students' self-initiated AI use not organised by the course (E3)	Implementation evidence, not attitudes, is the object
Population	Degree-level higher education (ISCED 5–8) students or teachers	School, pre-university, foundation and pre-sessional programmes (E4)	Population consistency
Setting	Conducted in a Southeast Asian country; multi-country only if SEA data separable	Non-SEA studies or inseparable multi-country data (E5)	Regional scope

Domain	Inclusion	Exclusion (code)	Rationale
Outcome	Learner outcome, experience or learning process reported	System or AI performance, usability only (E6); predictive models without a learner-facing intervention (E7)	Learning, not system accuracy, is the object
Publication	Journal article, English, 2021–2026	Other document types or languages	Feasibility and reproducibility

3.5. Study selection

One duplicate was removed, leaving 641 records for screening. Titles and abstracts were assessed against the eligibility criteria using the exclusion codes in Table 2. Abstracts were read in full when the title and methods information were insufficient to determine eligibility. Full texts of records considered eligible or uncertain were sought and assessed against the same criteria. Full texts were obtained from open-access sources or through the author's institutional access; where a DOI was available, it was checked to confirm the record's identity. Screening, data extraction, and appraisal were conducted by one reviewer, with an AI assistant (Claude, Anthropic) used under the author's direction to support the process. The reviewer made and verified all final decisions. No independent second-reviewer screening was conducted; this is a limitation of the review. The selection process is shown in Figure 1.

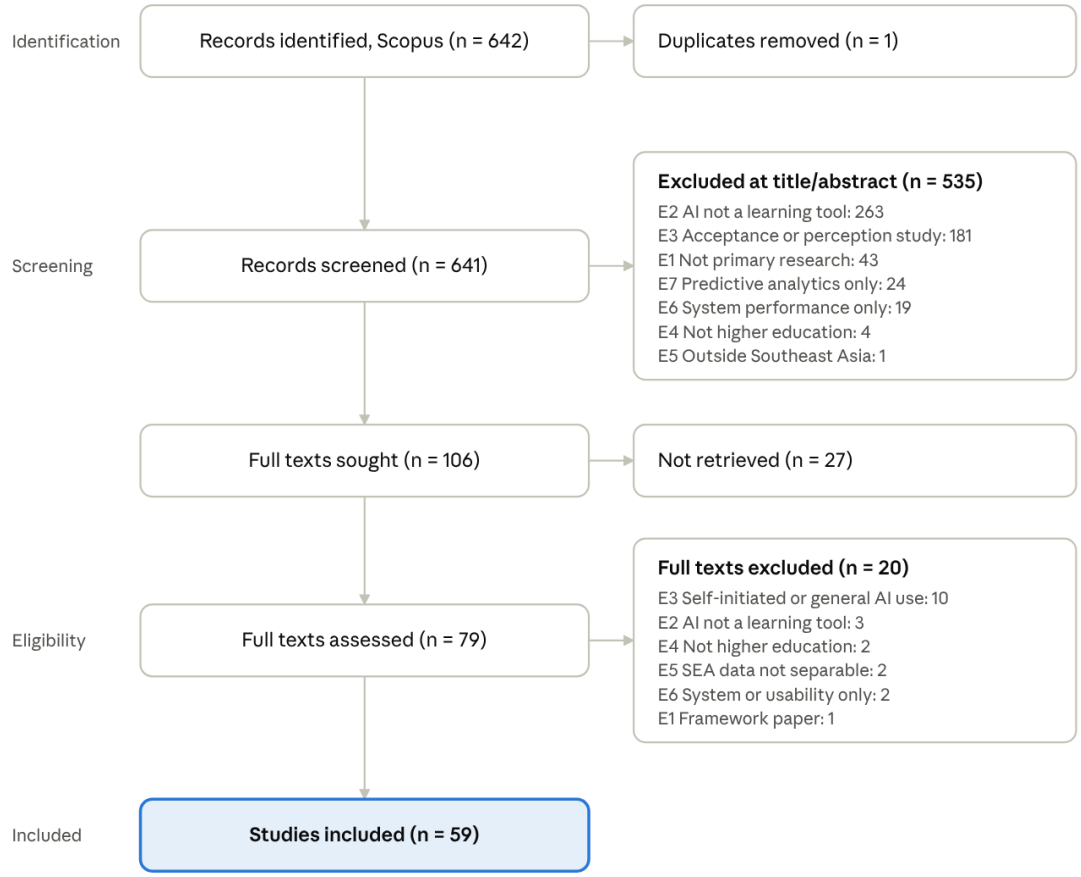


Figure 1. PRISMA 2020 flow diagram. Most exclusions at title and abstract concerned AI not used as a learning tool or general acceptance studies; 27 full texts could not be retrieved.

3.6. Data extraction

A standardized matrix recorded, for each study: citation and DOI; country, institution and discipline; objective; design; sample; AI tool or intervention; comparator; duration; outcomes and instruments; measured results; self-reported results; contextual and implementation factors; limitations stated by the authors; theory; recommendations; and notes by the review authors. Information not reported was recorded as "NR" and never inferred. Delivery mode (web-based general-purpose tool, purpose-built web platform or LMS integration, mobile or messaging application, desktop software, not reported) was coded from the full texts after the scope was finalized.

3.7. Quality appraisal

Methodological quality was appraised with the Mixed Methods Appraisal Tool (MMAT) version 2018 (Hong et al., 2018), which covers qualitative, randomized, non-randomized, quantitative descriptive and mixed-methods designs with two screening questions and five design-specific criteria. Each criterion was rated Yes, No or Can't tell with a written justification. Following MMAT guidance, no overall score was calculated; instead, ratings were used to interpret findings, especially for criteria on confounding and outcome measurement. Weaknesses of reporting (for example, inconsistent degrees of freedom) were recorded separately from weaknesses of design.

3.8. Synthesis

Quantitative strand. For each study that measured a learning outcome, the direction of the main measured result was coded as improvement or advantage, no difference or parity, mixed, or worse. Results were stratified by design tier (randomized; controlled non-randomized; uncontrolled) and by comparator type (no comparator; business-as-usual or non-equivalent comparator; active human-support comparator offering the same function, such as teacher feedback or standardized patients; AI-versus-AI comparison of design variants). Direction counts describe the distribution of findings and are not interpreted as votes on effectiveness.

Thematic synthesis. Following Thomas and Harden (2008), findings and contextual information from each study were coded line by line into descriptive codes (Stage 1), grouped into descriptive themes (Stage 2) and interpreted into analytical themes that address the research questions (Stage 3). Measured outcomes, self-reported outcomes and review-author interpretations were kept distinct throughout, and every code–study link is traceable to the extracted evidence in the synthesis workbook. Code frequencies indicate how many studies reported a finding, not its importance.

4. RESULTS AND DISCUSSION

4.1. Study selection

Of 642 records, one was a duplicate and 535 were excluded at title and abstract: 263 did not use AI as a learning tool (E2), 181 were general acceptance or perception studies (E3), 43 were not primary research (E1), 24 were predictive analytics without a learner-facing intervention (E7), 19 reported only system performance (E6), four were outside higher education (E4) and one was outside Southeast Asia (E5). Of 106 records sought, 27 full texts could not be retrieved. Of 79 assessed, 20 were excluded: 10 for self-initiated or general AI use (E3), three for AI not used as a learning tool (E2), two each for population (E4), setting (E5) and system or usability outcomes only (E6), and one framework paper (E1). Fifty-nine studies were included (Figure 1).

4.2. Characteristics of included studies (RQ1)

The 59 studies came from six of the eleven Southeast Asian countries: Indonesia (20), Vietnam (15), Thailand (14), Malaysia (6), the Philippines (3) and Singapore (1). No eligible study came from Brunei, Cambodia, Laos, Myanmar or Timor-Leste. The evidence is recent: 32 studies were published in 2026, 20 in 2025, five in 2024 and two in 2021, with none in 2022–2023 (Figure 2). Language education dominated (35 studies: 22 on writing, six on speaking, seven on other language skills), followed by education and teacher education (6), STEM and computing (5), health professions (5), business and entrepreneurship (3) and four other fields. Table 3 lists every study.

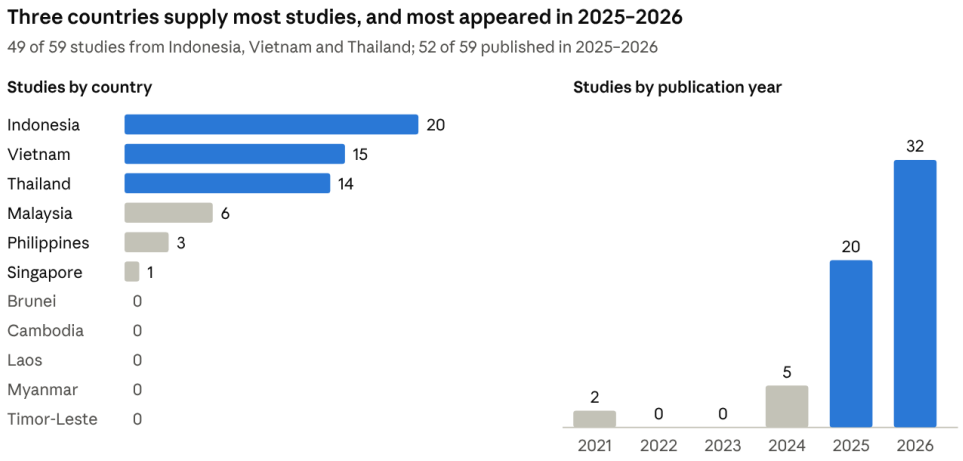


Figure 2. Geographical and temporal distribution of the 59 included studies. Five Southeast Asian countries contributed no eligible study, and no study was published in 2022–2023.

General-purpose generative AI, mostly ChatGPT and occasionally Gemini, was the tool in 32 studies; 10 used purpose-built or configured AI systems (for example AI simulated patients, an agentic English coach and a retrieval-augmented support chatbot), six used automated writing evaluation or writing assistants, and 11 combined several tools or did not name them. By delivery mode, 32 studies used web-based general-purpose tools, 12 used purpose-built web platforms or LMS integration, four used mobile or messaging applications, three involved desktop software in whole or part, and eight did not report how the tool was accessed.

Table 3. Characteristics of the included empirical studies (sorted by country and author). Measured: ▲ improvement or advantage, ◆ no difference or parity, ◐ mixed, — no measured learning outcome. MMAT: QUAL qualitative, RCT randomised, NRS non-randomised, DESC quantitative descriptive, MM mixed methods; counts of Yes / Can't tell / No across the five design criteria.

No	Study	Country	Discipline	AI tool (delivery)	Design	n	Measured	MMAT (Y/CT/N)
454	Andewi et al. (2025)	Indonesia	STEM/computing	Asynchronous ChatGPT help-seeking (web)	Quasi-experiment (controlled)	60	▲	MM 4/1/0
98	Apriani et al. (2025)	Indonesia	Language: writing	ChatGPT-3.5 writing support (web)	Quasi-experiment (controlled)	2 classes	▲	MM 2/2/1
457	Arifin et al. (2026)	Indonesia	Teacher education	MoralQ app (AI chatbot) (mobile)	One-group pre/post	138	▲	NRS 2/1/2
409	Darmawansah et al. (2024)	Indonesia	Language: writing	ChatGPT argument maps (web)	Quasi-experiment (controlled)	67	▲	NRS 3/2/0
492	Dwipayana et al. (2026)	Indonesia	Social sciences/humanities	DISCERN-AI multi-tool model (desktop, partly)	Quasi-experiment (controlled)	120	▲	MM 3/2/0
330	Fadilah (2026)	Indonesia	Language: writing	Constrained ChatGPT in peer talk (web)	One-group pre/post	22	▲	MM 3/1/1
378	Fardian et al. (2025)	Indonesia	STEM/computing	ChatGPT alone / with instructor (web)	Quasi-experiment (controlled)	≈30 per group (target)	◐	MM 2/2/1
238	Khusniyah et al. (2026)	Indonesia	Language: writing	ChatGPT feedback vs peer review (web)	Quasi-experiment (controlled)	41	◆	NRS 3/2/0
426	Mandasari et al. (2026)	Indonesia	Language: writing	AI in collaborative writing (web)	Qualitative	7	—	QUAL 5/0/0
57	Margareth et al. (2025)	Indonesia	Language: translation	ML-enabled CAT tool (desktop)	Quasi-experiment (controlled)	50	▲	NRS 1/3/1
281	Marzuki et al. (2026)	Indonesia	Language: writing	NLP writing assistants (NR)	One-group pre/post	100	▲	MM 4/0/1
601	Mediyawati et al. (2021)	Indonesia	Language: Indonesian (journalism)	U-Tapis language filter (NR)	One-group pre/post	30	▲	NRS 0/4/1
536	Oktarin et al. (2024)	Indonesia	Language: writing	ChatGPT writing support (web)	Post-only controlled	50	▲	MM 2/2/1
26	Puteri et al. (2025)	Indonesia	Language: writing	ChatGPT + teacher feedback (web)	Quasi-experiment (controlled)	35	◆	NRS 3/2/0
210	Rizal et al. (2026)	Indonesia	Social sciences/civic	Health-information chatbot programme (NR)	Cross-sectional (SEM)	210	—	MM 3/2/0
298	Sumakul (2025)	Indonesia	Language: EFL general	AI apps incl. ELSA Speak (mobile)	Qualitative	21 + 2 teachers	—	QUAL 3/2/0

No	Study	Country	Discipline	AI tool (delivery)	Design	n	Measured	MMAT (Y/CT/N)
185	Susiawati et al. (2026)	Indonesia	Language: Arabic	AI Arabic learning tools (NR)	Qualitative	78	—	QUAL 2/3/0
75	Suwardika et al. (2026)	Indonesia	Digital literacy	ChatGPT in flipped design thinking (LMS)	Crossover (within-subjects)	26	—	MM 3/2/0
456	Syafei & Nuraeningsih (2025)	Indonesia	Language: writing	ChatGPT, QuillBot, Grammarly et al. (web)	One-group pre/post	66	▲	MM 0/4/1
402	Wikanengsih et al. (2025)	Indonesia	Counselling education	ChatGPT + QuillBot in PBL (web)	Post-only controlled	160	▲	MM 2/2/1
45	Chuan et al. (2026)	Malaysia	STEM/computing	ChatGPT within IEEC framework (web)	One-group pre/post	60	▲	MM 4/0/1
322	Dahri et al. (2025)	Malaysia	Teacher education	Mobile ChatGPT after-class review (mobile)	Quasi-experiment (controlled)	32	▲	NRS 3/2/0
369	Kailin & Saeed (2026)	Malaysia	Language: writing	ChatGPT proposal feedback (web)	Qualitative	4	—	QUAL 5/0/0
444	Me et al. (2025)	Malaysia	Design	Multiple design AI tools (desktop, partly)	Post-only controlled	38	◐	MM 2/2/1
77	Rath (2025)	Malaysia	Health professions	ChatGPT terminology support (web)	One-group pre/post	35	▲	MM 4/0/1
292	Yusof (2025)	Malaysia	Teacher education	ChatGPT retrieval practice (web)	Historical cohort	128	▲	NRS 3/1/1
431	Niez et al. (2026)	Philippines	Language: ESL general	Agentic AI coach vs reactive chat (web platform)	Quasi-experiment (controlled)	128	▲	MM 4/1/0
377	Ravelo et al. (2026)	Philippines	Language: writing	ChatGPT feedback cycles (web)	Repeated measures (one group)	175	▲	NRS 2/2/1
219	Silor & Silor (2026)	Philippines	Mixed courses	AI feedback in course modules (web platform)	Quasi-experiment (controlled)	210	▲	MM 2/2/1
570	Kong et al. (2021)	Singapore	Health professions	Virtual Integrated Patient (web platform)	Post-only survey	106	—	DESC 2/3/0
221	Chinwong et al. (2025)	Thailand	Health professions	ChatGPT simulated smoker (web)	Post-only survey	145	—	NRS 2/1/2
366	Jitpaisarnwattana & Saville (2026)	Thailand	Language: writing	Write & Improve AWE (web platform)	One-group pre/post	53	▲	NRS 3/1/1
345	Jitpaisarnwattana (2026)	Thailand	Language: writing	ChatGPT IELTS feedback (web)	One-group pre/post	52	◐	MM 4/0/1
294	Kanoksilapatham & Takrudkaew (2025)	Thailand	Language: speaking	ChatGPT conversation scripts (web)	One-group pre/post	93	▲	MM 3/1/1
274	Kawinkoonlasate (2025)	Thailand	Language: writing	Gemini vs ChatGPT feedback (web)	Quasi-experiment (controlled)	80	▲	MM 3/1/1
397	Sangpom & Sangpom (2026)	Thailand	Business	ChatGPT decision support in PBL (web)	One-group pre/post	32	▲	MM 4/0/1

No	Study	Country	Discipline	AI tool (delivery)	Design	n	Measured	MMAT (Y/CT/N)
18	Shen et al. (2026)	Thailand	Business/entrepreneurship	GPT-4 prompt co-adaptation coach (web platform)	RCT	120	—	RCT 4/0/1
477	Sindhvananda et al. (2026)	Thailand	Health professions	AI simulated patient (web platform)	One-group pre/post	25	▲	MM 4/1/0
287	Sinlapaninman & Yonwilad (2025)	Thailand	Graduate research	Dialogflow coaching chatbot (LINE, mobile)	Post-only survey	30	—	MM 2/2/1
318	Smithsarakarn (2026)	Thailand	Language: writing	ChatGPT feedback + reflection (web)	Quasi-experiment (controlled)	60	◆	MM 3/2/0
357	Wangdi & Rofiah (2026)	Thailand	Language: EFL general	GenAI-integrated EFL class (web)	Qualitative	22	—	QUAL 4/1/0
629	Wiboolyasarin et al. (2024)	Thailand	Language: writing (Thai L2)	Wiki + ChatGPT feedback (web platform)	Quasi-experiment (controlled)	39	▲	NRS 2/3/0
181	Wiboolyasarin et al. (2026)	Thailand	Language: writing (Thai L2)	ChatGPT feedback timing (web)	Quasi-experiment (controlled)	64	▲	NRS 2/3/0
236	Wuttikamonchai (2025)	Thailand	STEM/computing	AI in blended PjBL model (NR)	One-group pre/post	108	▲	NRS 1/3/1
53	Dang (2026)	Vietnam	Language: speaking	ChatGPT-aided mind maps (web)	One-group pre/post	30	▲	NRS 3/1/1
157	Dang et al. (2026)	Vietnam	Health professions	AI virtual standardised patients (web platform)	RCT	67	◆	RCT 1/3/1
9	Duc (2026)	Vietnam	Language: reading	AI-assisted reading projects (NR)	Post-only survey	130	—	MM 1/3/1
519	Duong & Suppasetseree (2024)	Vietnam	Language: speaking	AI voice chatbot (web platform)	One-group pre/post	30	▲	MM 1/3/1
585	Furze et al. (2024)	Vietnam	Institution-wide	AI Assessment Scale with GenAI (web)	Historical cohort	all submissions	▲	NRS 2/2/1
78	Nguyen & Tran (2026)	Vietnam	Language: writing	Grammarly, QuillBot (web)	One-group pre/post	63	◐	MM 4/1/0
213	Pham (2026)	Vietnam	Language: writing (ESP)	ChatGPT-4o essay feedback (web)	Post-only survey	55	—	DESC 2/3/0
308	Phung et al. (2026)	Vietnam	STEM/computing	TutorAI + RAG chatbot (web platform)	Historical cohort	≈2,500	▲	NRS 1/3/1
173	The Nguyen et al. (2026)	Vietnam	Language: speaking	AI speaking tools in CLT (NR)	Quasi-experiment (controlled)	70	▲	NRS 2/2/1
482	The et al. (2026)	Vietnam	Language: speaking	AI speaking tools (NR)	Quasi-experiment (controlled)	80	▲	NRS 2/2/1
445	Thuan & Hanh (2025)	Vietnam	Language: pronunciation	MS Teams Reading Progress (web platform)	Post-only survey	123	—	DESC 2/2/1
332	Tran (2025)	Vietnam	Language: writing	Gemini feedback order (web)	Descriptive comparison	14	▲	DESC 2/3/0
223	Tran et al. (2026)	Vietnam	Entrepreneurship	GenAI reflective prompts (web)	Case study (mixed)	25	▲	MM 3/1/1
439	Tuong & Tran (2025)	Vietnam	Language: writing	Grammarly, Write & Improve, Slick	One-group pre/post	200	◐	MM 2/3/0

No	Study	Country	Discipline	AI tool (delivery)	Design	n	Measured	MMAT (Y/CT/N)
				Write (web)				
131	Vo & Ho (2026)	Vietnam	Language: writing	ChatGPT grammar feedback (web)	Quasi-experiment (controlled)	90	▲	NRS 1/4/0

4.3. Methodological characteristics (RQ1, RQ5)

Non-randomised designs dominated (Table 4, Figure 3). Eighteen studies were controlled quasi-experiments, 17 used a one-group pre/post design, six were post-only surveys, three compared with a previous cohort, three compared groups after the intervention only, and only two were randomised (Dang et al., 2026; Shen et al., 2026). Samples were small: 22 studies had 50 participants or fewer, and the largest learner samples came from institutional implementations. Interventions ranged from a single task (Pham, 2026) to a year-long institutional pilot (Furze et al., 2024); most lasted between two and sixteen weeks, and we identified no study with a delayed post-test of retention.

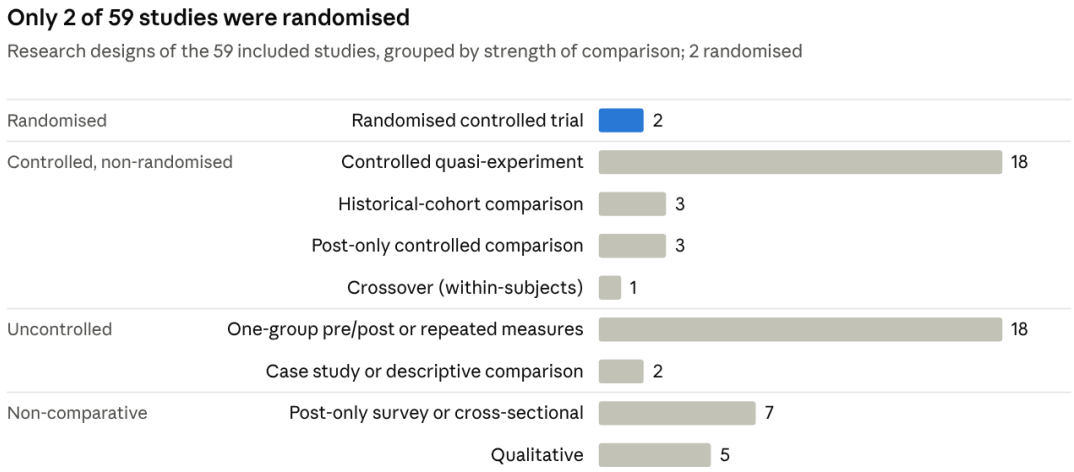


Figure 3. Methodological characteristics of the included studies by design and strength of comparison.

Table 4. Methodological characteristics and key MMAT judgements

Characteristic	Studies (n)	Key appraisal finding
Controlled quasi-experiment	18	Intact classes; baseline adjustment (ANCOVA) in a minority; allocation often confounded with instructor or class
One-group pre/post or repeated measures	18	No comparator; testing, maturation and practice effects not separable from the intervention
Historical-cohort or post-only controlled	6	No baseline or a different cohort; concurrent changes (policy, curriculum) not controlled
Randomised controlled trial	2	One cluster-randomised with four clusters; one individually randomised with self-reported outcomes
Post-only survey / cross-sectional	7	Perception outcomes only; mostly unvalidated short instruments
Qualitative or case study	6	Generally coherent analysis; small purposive samples
Crossover / descriptive comparison	2	Carry-over between conditions; very small samples
MMAT non-randomised (n = 20): confounders accounted for	4 Yes / 3 Can't tell / 13 No	Confounding was the most frequent weakness
MMAT non-randomised (n = 20): appropriate outcome measurement	3 Yes / 15 Can't tell / 2 No	Rater reliability and blinding rarely reported
MMAT mixed methods (n = 28): each	0 Yes / 10 Can't tell /	Quantitative strands usually uncontrolled

Characteristic	Studies (n)	Key appraisal finding
component meets its quality criteria	18 No	

Reporting problems recurred alongside design limits: degrees of freedom inconsistent with sample size (Vo & Ho, 2026; Wikanengsih et al., 2025; Mediyawati et al., 2021), an effect size inconsistent with its test statistic (Dahri et al., 2025), and contradictions between abstract and results (Tuong & Tran, 2025). In several studies it was unclear whether students could use AI during the post-test (Apriani et al., 2025; Syafei & Nuraeningsih, 2025), and in two the outcome was scored by the same AI system that delivered the intervention or rehearsed with it (Arifin et al., 2026; Kanoksilapatham & Takrudkaew, 2025).

4.4. Research applications of AI (RQ1)

Seven pedagogical roles emerged from the studies (Table 5). AI most often served as a feedback provider (22 studies), mainly on writing, followed by tutor or coach (14), project and content scaffold (10), simulated patient or conversation partner (7), collaborative partner in peer work (3), institutional or assessment system (2) and generator of retrieval-practice questions (1).

Table 5. Main research applications and reported empirical outcomes

Pedagogical role of AI	Studies (n)	Typical tools	Measured outcomes ▲/◆/◐ (not measured)	Main reported outcomes
Feedback provider	22	ChatGPT, Gemini; AWE and writing assistants	13/3/3 (3)	Gains mainly in grammar, vocabulary and mechanics; coherence, task achievement and lexical range unchanged or mixed; parity with teacher or peer feedback in controlled comparisons
Tutor or coach	14	ChatGPT; purpose-built coaches	7/0/1 (6)	Achievement, self-efficacy, proficiency and self-regulation gains; ChatGPT alone weakest in mathematics; agentic scaffolding outperformed reactive chat
Project and content scaffold	10	Multiple tools in PBL and design	7/0/1 (2)	Higher project, skill and critical-thinking scores; creativity higher but technical execution unchanged
Simulated patient or practice partner	7	AI simulated patients; speaking chatbots	4/1/0 (2)	Clinical history-taking parity with standardised patients (RCT) or within-group gains; speaking gains in uncontrolled or unadjusted designs
Collaborative partner	3	ChatGPT in peer work	2/0/0 (1)	Stronger argument structure and group efficacy; negative emotions dominant in collaborative writing
Institutional or assessment system	2	AI Assessment Scale; TutorAI and RAG chatbot	2/0/0 (0)	Lower AI-related misconduct and higher pass rates against historical baselines
Retrieval-practice generator	1	ChatGPT	1/0/0 (0)	Higher final-exam scores than a previous cohort

4.5. Reported outcomes and effect direction (RQ2)

Forty-five studies measured a learning outcome; 14 reported only perceptions or experiences. Of the 45, 36 reported improvement or an advantage for the AI condition, four found no difference or parity, five were mixed and none reported a worse result for the AI condition as a whole. The direction of results depended strongly on design and comparator (Figure 4):

1. **No comparator (20 studies).** Seventeen reported gains and three were mixed. Within-group effects were often large, for example d = 1.87 for writing with NLP assistants (Marzuki et al., 2026) and d = 1.93 for OSCE history-taking with an AI simulated patient (Sindhvananda et al., 2026).

2. **Business-as-usual or non-equivalent comparator (16 studies).** Fourteen favoured AI and two were mixed. Many had post-only comparisons (Oktarin et al., 2024; Wikanengsih et al., 2025), historical controls (Yusof, 2025; Furze et al., 2024; Phung et al., 2026), unadjusted baseline differences (The Nguyen et al., 2026; The et al., 2026) or allocation confounded with ability (Margareth et al., 2025).
3. **Active human-support comparator (6 studies).** Four found parity: AI virtual versus standardised patients in a randomised trial (Dang et al., 2026), ChatGPT plus teacher feedback versus teacher feedback (Puteri et al., 2025), ChatGPT versus lecturer-guided peer review (Khusniyah et al., 2026) and ChatGPT versus teacher-led feedback (Smithsarakarn, 2026). The two positive results came from a two-week intervention with 32 students (Dahri et al., 2025) and a comparison in which ChatGPT was available at any time while the lecturer was not (Andewi et al., 2025).
4. **AI versus AI (3 studies with measured outcomes, plus one randomised trial with self-reported outcomes).** Design variants mattered: an agentic coach outperformed a reactive chat interface (Niez et al., 2026; partial $\eta^2 = .19$), immediate and combined feedback outperformed delayed feedback (Wiboolyasarin et al., 2026), adaptive prompt co-adaptation outperformed static and adaptive prompting on self-reported mindset outcomes (Shen et al., 2026), and the better tool depended on learner proficiency (Kawinkoonlasate, 2025).

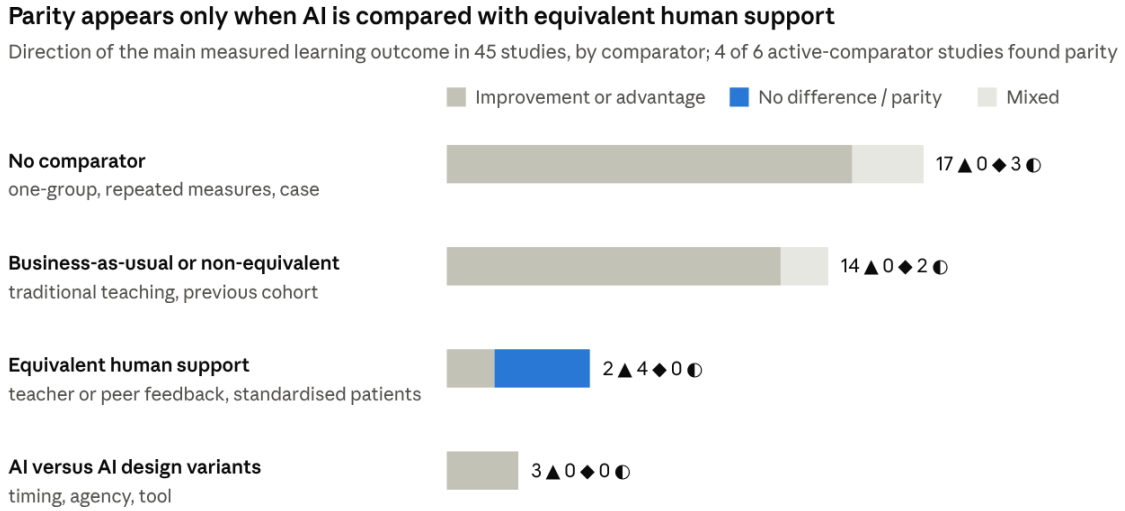


Figure 4. Evidence map of measured learning outcomes by comparator type (SWiM effect direction). Counts describe direction only, not magnitude.

The most direct test of mediation came from a three-arm study in linear algebra: students taught only through ChatGPT scored lowest (53.25), conventional instruction scored 68.40 and ChatGPT combined with an instructor scored highest (87.58) (Fardian et al., 2025).

Across outcome types, measured gains clustered in surface features. Writing studies found improvements in grammatical accuracy and vocabulary but no change in coherence and a decline in lexical range (Tuong & Tran, 2025), limited change in thesis development and disciplinary argument (Nguyen & Tran, 2026), and no gain in task achievement (Jitpaisarnwattana, 2026). A design study found higher creativity scores but no difference in technical execution (Me et al., 2025). The clearest exception is constrained use: when ChatGPT served only to prompt idea exploration in peer talk, warrants, backing and rebuttals improved while surface features changed little (Fadilah, 2026). Self-reported outcomes were almost uniformly positive, including confidence or self-efficacy (12 studies) and engagement or motivation (15 studies).

4.6. Thematic synthesis (RQ2–RQ4)

Stage 1 produced 24 descriptive codes linked to 190 code–study pairs (Table 6). Stage 2 grouped them into six descriptive themes, and Stage 3 produced five analytical themes (Table 7, Figure 5).

Table 6. Initial coding framework and descriptive themes

Code	Descriptive code	Evidence type	Studies (n)	Descriptive theme
C01	Immediate, on-demand availability of feedback and	Self-reported	10	DT1

Code	Descriptive code	Evidence type	Studies (n)	Descriptive theme
	support			
C02	Gains concentrated in surface language features (grammar, vocabulary, mechanics)	Measured	11	DT1
C04	Lower anxiety; non-judgemental practice	Self-reported	7	DT1
C05	Increased confidence or self-efficacy	Self-reported	12	DT1
C06	Increased engagement or motivation	Self-reported	15	DT1
C07	Idea generation and content scaffolding	Self-reported	7	DT1
C03	Limited or no gains in higher-order skills	Measured	7	DT2
C08	Scaffolding of self-regulation, reflection or revision cycles	Measured	6	DT2
C23	AI integrated into peer collaboration	Design feature	5	DT2
C09	Personalisation or adaptivity of AI support	Measured	3	DT3
C17	Human (teacher) mediation decisive; hybrid arrangements	Context	13	DT3
C18	AI embedded in an explicit pedagogical model	Design feature	16	DT3
C19	Bounded AI use that preserves learner cognitive work	Design feature	4	DT3
C21	Design of AI support (timing, agency, adaptive prompting) changes outcomes	Measured	5	DT3
C15	Technical, connectivity or access constraints	Context	7	DT4
C16	Gaps in AI literacy or prompting skill	Context	5	DT4
C20	Learner proficiency or usage intensity moderates benefit	Measured	5	DT4
C22	Institutional context and resources shape implementation	Context	7	DT4
C10	Over-reliance, dependency or cognitive offloading	Self-reported	12	DT5
C11	Inaccurate output and verification burden	Self-reported	10	DT5
C12	Generic or impersonal feedback; continued wish for human feedback	Self-reported	7	DT5
C13	Authorship, voice and academic-integrity concerns	Self-reported	9	DT5
C14	Negative emotions during AI-supported work	Self-reported	3	DT5
C25	AI simulated patients or partners for clinical and communicative practice	Design feature	4	DT6

DT1 = gains cluster in surface accuracy, confidence and engagement; DT2 = higher-order gains are uneven; DT3 = human mediation and pedagogical structure condition outcomes; DT4 = learner and institutional moderators; DT5 = epistemic, ethical and affective risks; DT6 = simulated practice partners in professional education.

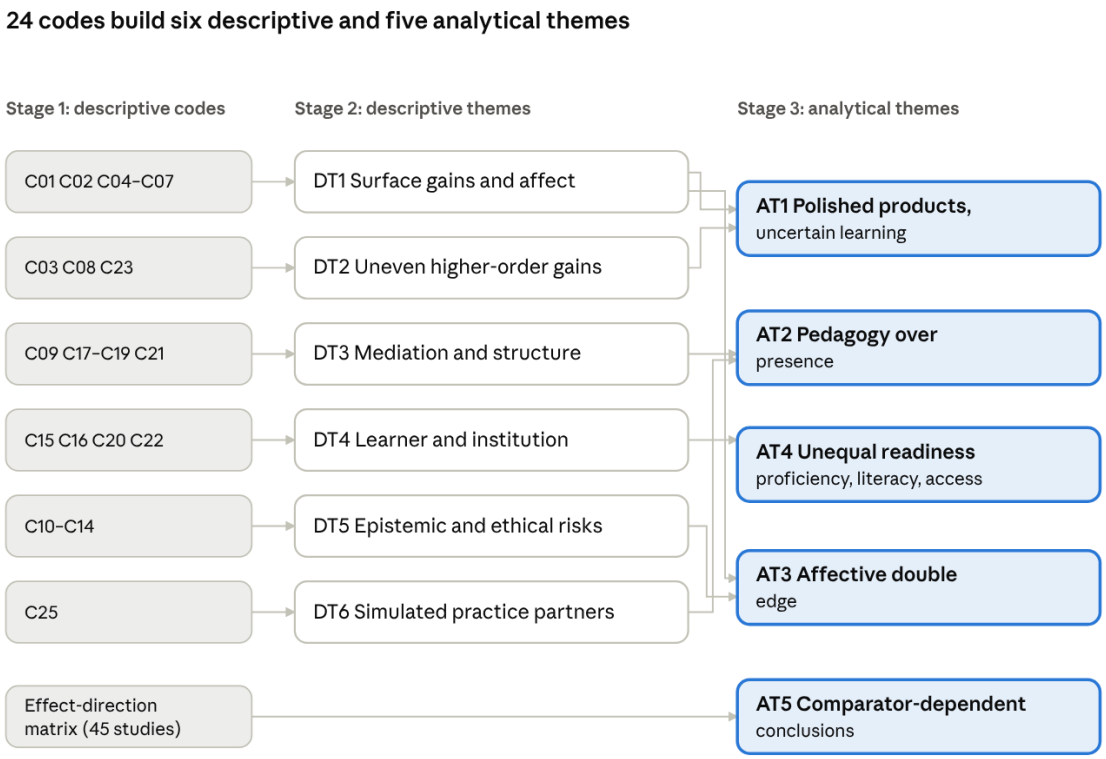


Figure 5. Thematic coding and synthesis framework (Thomas & Harden, 2008). Arrows show which descriptive themes informed each analytical theme; AT5 draws on the effect-direction matrix.

Table 7. Descriptive and analytical themes

Analytical theme	Descriptive themes (codes)	Key empirical evidence	Supporting studies (examples)	Analytical interpretation
AT1 Polished products, uncertain learning	DT1, DT2 (C02, C03, C05, C06)	Surface gains in 11 studies; no or limited higher-order gains in 7; self-reports positive	Tuong & Tran (2025); Nguyen & Tran (2026); Jitpaisarnwattana (2026); Me et al. (2025)	AI reliably improves what learners produce and how they feel about it; evidence of deeper, independent competence is thin
AT2 Pedagogy over presence	DT3, DT6 (C17, C18, C19, C21, C09)	ChatGPT alone lowest, with instructor highest; agentic > reactive; timing matters; parity with equivalent human support	Fardian et al. (2025); Niez et al. (2026); Wiboolyasarin et al. (2026); Dang et al. (2026); Fadilah (2026)	Outcomes depend on how AI is orchestrated rather than whether it is present
AT3 The affective double edge	DT1, DT5 (C04, C10–C14)	Lower anxiety in 7 studies; frustration, distrust and impersonal feedback in others	The Nguyen et al. (2026); Kanoksilapatham & Takrudkaew (2025); Wangdi & Rofiah (2026); Mandasari et al. (2026); Kailin & Saeed (2026)	AI lowers the social cost of practice but adds emotional and epistemic work learners must regulate
AT4 Unequal readiness	DT4 (C15, C16, C20, C22)	Benefits vary with proficiency, AI literacy, usage and institutional resources	Kawinkoonlasate (2025); Jitpaisarnwattana & Saville (2026); Tuong & Tran (2025); Chinwong et al. (2025)	Without attention to readiness, AI may widen existing gaps
AT5 Comparator-	Effect-	Advantage against	Dang et al. (2026);	Evidence supports AI as

Analytical theme	Descriptive themes (codes)	Key empirical evidence	Supporting studies (examples)	Analytical interpretation
dependent conclusions	direction matrix	business-as-usual; parity in 4 of 6 active-human comparisons	Puteri et al. (2025); Khusniyah et al. (2026); Smithsarakarn (2026)	a comparable, scalable complement, not a superior replacement

AT1 — Polished products, uncertain learning. The most consistent measured gains concerned grammar, vocabulary and mechanics, together with self-reported confidence and engagement. Gains in coherence, argumentation, task achievement and technical execution were absent or inconsistent in the studies that measured them separately. Several studies measured AI-assisted products or rehearsed performances, so their outcomes cannot be read as independent learning.

AT2 — Pedagogy over presence. Studies that varied how AI was used, rather than whether, produced the most informative results. Teacher mediation, explicit pedagogical models (16 studies), bounded use that left reasoning to learners, and the design of AI support (timing, agency, adaptivity) were associated with better outcomes. Where AI replaced a human function of equal intensity, outcomes were usually equivalent.

AT3 — The affective double edge. Students described AI as non-judgemental and available, which lowered anxiety about speaking, asking for help and making mistakes. The same studies and others reported over-reliance (12 studies), inaccurate output and the burden of verifying it (10), impersonal feedback (7), integrity and voice concerns (9) and, in qualitative studies, frustration, confusion and anxiety that learners had to regulate.

AT4 — Unequal readiness. Benefits differed by proficiency (larger for lower-band or intermediate learners in some studies, for higher-proficiency learners with ChatGPT in another), by usage intensity, by institution type (urban versus rural, public versus private) and by AI literacy, which many students lacked.

AT5 — Comparator-dependent conclusions. The strongest comparisons point to parity, the weakest to large gains. This pattern is itself a finding about the evidence base and shapes how all other themes should be read.

4.7. Challenges, barriers and implementation factors (RQ4)

Table 8 separates challenges reported by participants from limitations acknowledged by the original authors and interpretations made by the review authors; the most frequently coded challenges are shown in Figure 6.

Table 8. Challenges, barriers and implementation factors

Category	Challenge or factor	Source	Studies (n)	Examples
Epistemic	Over-reliance, dependency, cognitive offloading	Participants	12	Andewi et al. (2025); Tuong & Tran (2025); Me et al. (2025)
Epistemic	Inaccurate output; verification burden	Participants	10	Fardian et al. (2025); Rath (2025); Kailin & Saeed (2026)
Relational	Generic, mechanical or impersonal feedback; wish for human feedback	Participants	7	Pham (2026); Jitpaisarnwattana (2026); Sindhvananda et al. (2026)
Ethical	Authorship, voice and academic integrity	Participants; institutional data	9	Tuong & Tran (2025); Furze et al. (2024)
Affective	Frustration, confusion, anxiety	Participants	3	Wangdi & Rofiah (2026); Mandasari et al. (2026); Kailin & Saeed (2026)
Infrastructure	Technical problems, connectivity, device access	Participants; authors	7	Chinwong et al. (2025): 88.3% reported technical issues; Kawinkoonlasate (2025)
Capability	Limited AI literacy and prompting skill	Participants; authors	5	Chinwong et al. (2025): 58.6%; Andewi et al. (2025)
Learner	Proficiency and usage intensity moderate benefit	Measured	5	Kawinkoonlasate (2025); Jitpaisarnwattana & Saville (2026)
Institutional	Resources, institution type, policy	Authors; review authors	7	Tuong & Tran (2025); Furze et al. (2024); Phung et al. (2026)

Category	Challenge or factor	Source	Studies (n)	Examples
Methodological	No control, confounded allocation, unclear AI access at post-test, AI-scored outcomes	Review authors	most studies	See Table 4

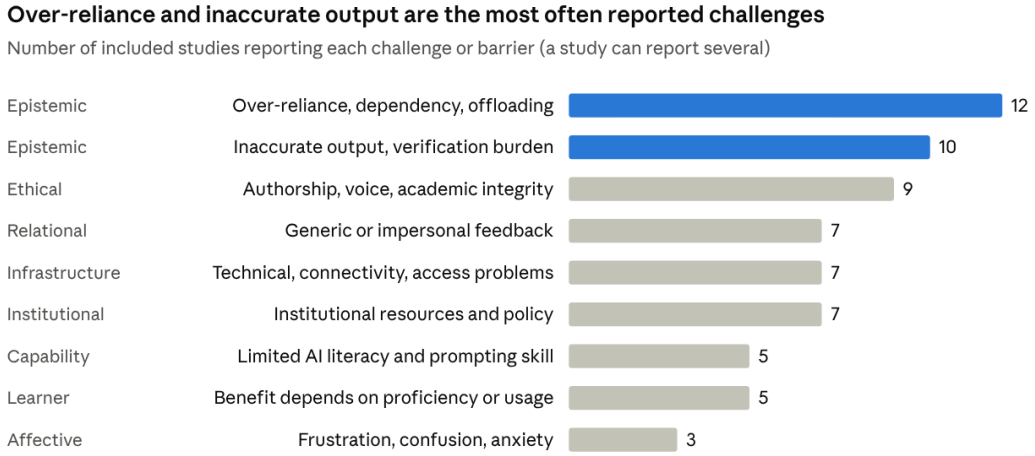


Figure 6. Main challenges and implementation barriers by number of studies reporting them. Frequencies show how often a challenge was reported, not its importance.

4.8. Interpreting the evidence

The review shows a young, fast-growing and positive evidence base whose positivity is concentrated in its weakest designs. Its central contribution is to show that, in Southeast Asian higher education, what AI achieves depends on how it is orchestrated and against what it is compared.

Three patterns dominate. First, language learning, and writing in particular, accounts for 35 of 59 studies. This reflects the high stakes of English proficiency for graduates in the region, the heavy feedback load of large writing classes, and the fit between text-generating tools and text-based tasks. Second, almost all tools were reached through the web or a mobile application, mostly general-purpose chatbots used without institutional configuration; purpose-built systems appeared mainly in medicine and in a few well-resourced projects. Third, 52 of 59 studies appeared in 2025–2026 and most used classroom-based one-group or intact-class designs. The evidence therefore mirrors the speed of adoption: teachers evaluated what they introduced in their own classes, with the designs available to them.

Three distinctions are needed to read the outcomes correctly. The first is between measured and perceived outcomes. Self-reports of confidence, engagement and usefulness were almost uniformly positive, but perceptions were also collected in studies where measured gains were small, mixed or absent, so positive perceptions cannot stand in for learning. The second is between products and learning. Many studies scored work produced with AI or performances rehearsed with it, and several did not state whether AI was available during the post-test. This mirrors evidence from outside the region that AI support can raise performance during use while weakening independent performance afterwards (Bastani et al., 2025) and can improve essay quality without deeper knowledge gains (Fan et al., 2025). The surface-level profile of gains in the included writing studies is consistent with that interpretation.

The third distinction is between comparators. Against business-as-usual teaching, AI conditions usually added practice, feedback and attention, so an advantage is expected regardless of the tool. Against an equivalent human function, the best-designed comparisons found parity, including the only randomised trial with measured outcomes (Dang et al., 2026). Read together, the evidence supports a modest but practically important claim: AI can deliver feedback and simulated practice of comparable value at scale, which matters where teacher time and standardised patients are scarce.

4.9. Mechanisms, barriers and contradictions

The studies point to four mechanisms. **Teacher mediation**: AI alone produced the weakest mathematics outcomes and AI with an instructor the strongest (Fardian et al., 2025), and teachers frequently remained the source of judgement, motivation and contextual feedback. **Bounded use**: when AI was limited to prompting ideas within peer reasoning, higher-order argument improved (Fadilah, 2026), whereas open use often improved surface features only. This is consistent with cognitive-load arguments that help which removes productive effort can reduce learning (Sweller, 1988). **Feedback design**: immediate and combined feedback outperformed delayed feedback (Wiboolyasarin et al., 2026), and feedback embedded in revision and

reflection cycles echoes the feed-up, feed-back and feed-forward functions of effective feedback (Hattie & Timperley, 2007). **Regulation support**: an agentic coach that scaffolded goal setting, monitoring and reflection outperformed a reactive chat interface (Niez et al., 2026), as self-regulated learning theory would predict (Zimmerman, 2002).

The challenges are not side effects but part of what determines outcomes. Over-reliance and inaccurate output are two faces of the same problem: learners who cannot evaluate AI output either accept it or must spend effort verifying it. Limited AI literacy and prompting skill, reported for example by 58.6% of pharmacy students (Chinwong et al., 2025), and frequent technical problems (88.3% in the same study) limit both use and benefit. Concerns about voice, authorship and integrity were widespread, and one institutional study suggests that explicit assessment design can reduce AI-related misconduct (Furze et al., 2024), although its historical design cannot separate the framework from deterrence and detection effects.

Several findings conflict, and we do not harmonise them. Proficiency moderated benefits in opposite directions: lower-band learners gained most from automated writing evaluation (Jitpaisarnwattana & Saville, 2026), whereas ChatGPT feedback favoured higher-proficiency learners and Gemini lower-proficiency learners in another study (Kawinkoonlasate, 2025). Differences in tool, task and instrument plausibly explain this, but no study tested the interaction directly. Affective findings also diverge: surveys and quasi-experiments reported lower anxiety, while qualitative studies of collaborative and revision tasks found negative emotions to be dominant (Mandasari et al., 2026; Wangdi & Rofiah, 2026). The difference may reflect task demands as much as method: individual, low-stakes practice versus graded, collaborative writing. Finally, one study's abstract and regression disagree on whether private-university students gained more or less (Tuong & Tran, 2025), and two studies from the same team report near-identical designs and results at different universities (The Nguyen et al., 2026; The et al., 2026); both warrant caution.

4.10. Theoretical, institutional and policy implications

The evidence extends rather than replaces existing theory. It supports hybrid human–AI regulation, in which AI handles parts of monitoring and feedback while teachers retain judgement (Molenaar, 2022), and it suggests that the unit of analysis should be the orchestrated activity, not the tool. It qualifies motivational accounts: AI raised self-reported motivation and confidence, consistent with satisfying needs for competence and autonomy (Ryan & Deci, 2000), but such gains did not reliably translate into deeper performance. It adds an affective dimension often missing from technology studies: AI changes the emotional appraisal of learning tasks, lowering threat in practice while creating new sources of frustration and doubt, which fits control-value and emotion-regulation perspectives (Pekrun, 2006; Gross, 1998). Finally, it challenges the use of acceptance or perception as a proxy for effectiveness, a habit that shaped much of the excluded literature.

Recommendations should be proportional to evidence that is mostly non-randomised. Institutions can reasonably (a) use AI to extend feedback and simulated practice where human capacity is scarce, since the strongest comparisons show parity; (b) design activities that bound AI use and keep reasoning with learners; (c) assess learning without AI access at some point, so that AI-assisted products are not mistaken for competence; (d) build AI literacy and verification skills before or alongside use; (e) address access and connectivity gaps between institutions; and (f) evaluate new AI initiatives with an active comparator and a delayed measure rather than one-group pre/post designs.

4.11. Integrated analytical framework

The synthesis suggests an Orchestrated AI-Supported Learning (OASL) framework in which outcomes follow from how an AI application is orchestrated, filtered through learner and institutional readiness and through learners' affective and epistemic regulation (Figure 7). The framework was derived from the analytical themes; it has not been empirically validated.

Its five components are: (1) **AI application**, defined by tool type, pedagogical role and delivery mode; (2) **pedagogical orchestration**, comprising teacher mediation, bounded use, feedback design and embedding in an explicit model; (3) **readiness**, comprising proficiency, AI literacy, access and institutional resources and policy; (4) **regulation processes**, comprising self-regulation, verification and emotion regulation; and (5) **outcomes**, separated into product quality, independent learning, and perceptions and affect. Comparator and design act as an evidential lens on every observed relationship.

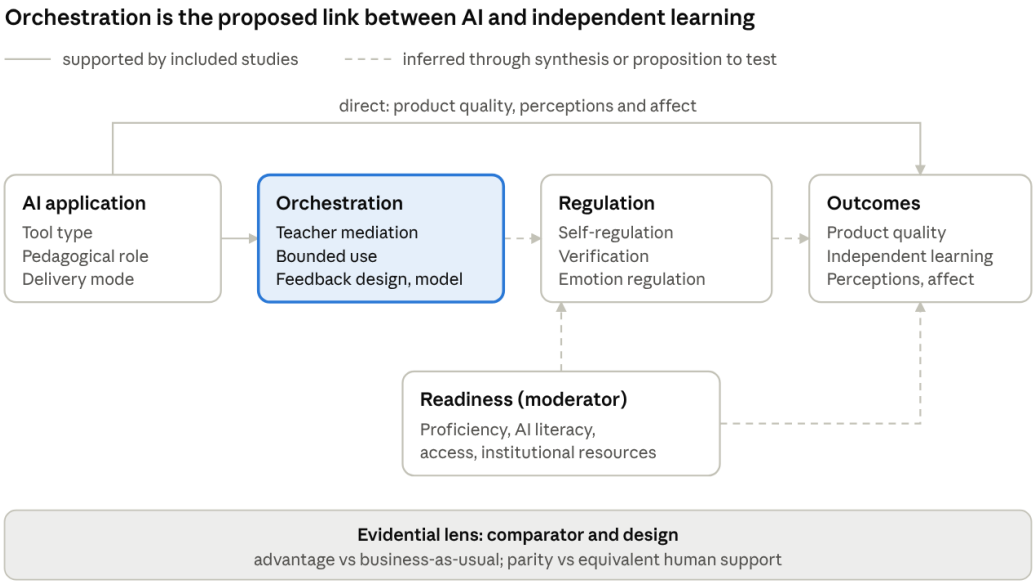


Figure 7. The Orchestrated AI-Supported Learning (OASL) framework. Solid arrows mark relationships supported by the included studies; dashed arrows mark relationships inferred through synthesis or proposed for testing (Table 9).

Table 9. Status of relationships in the framework

Relationship	Status	Basis
AI application → product quality and surface accuracy	Supported by measured evidence (mostly non-randomised)	Surface gains in 11 studies; within-group gains in uncontrolled designs
AI application → perceptions and affect (confidence, engagement, lower anxiety)	Supported by self-reported evidence	12–15 studies; qualitative accounts
Orchestration (mediation, bounded use, feedback design) → independent learning	Inferred through synthesis; partly supported by comparative studies	Fardian et al. (2025); Niez et al. (2026); Wiboolyasarin et al. (2026); Fadilah (2026)
AI ≈ equivalent human support for the same function	Supported by the strongest available comparisons	Dang et al. (2026) RCT; Puteri et al. (2025); Khusniyah et al. (2026); Smithsarakarn (2026)
Readiness (proficiency, AI literacy, access) moderates benefit	Inferred; inconsistent direction	Kawinkoonlasate (2025); Jitpaisarnwattana & Saville (2026); Chinwong et al. (2025)
Over-reliance and unregulated use → weaker independent learning	Theoretical proposition for testing	Consistent with Bastani et al. (2025) and Fan et al. (2025); no SEA study measured it directly
Emotion and verification regulation mediate the effect of AI on learning	Theoretical proposition for testing	Wangdi & Rofiah (2026); Mandasari et al. (2026); Kailin & Saeed (2026)

4.12. Research gaps and future research agenda

The most urgent gap is not more studies but stronger ones: active-comparator designs with independent, delayed measures of learning. Table 10 links each gap to the evidence that revealed it; Figure 8 orders the priorities.

Table 10. Research gaps and future research agenda

Gap	Supporting evidence	Recommended research question	Suggested methodology	Potential contribution
Methodological: few randomised or active-comparator designs	2 RCTs; 36 of 45 measured studies uncontrolled or business-as-usual	Does AI-supported feedback or practice outperform equivalent human support of the same intensity?	Cluster-randomised or matched designs with active human comparators; pre-registration	Separates added effort from added value
Measurement:	Unclear AI access at	Do AI-assisted gains	AI-free post-tests,	Distinguishes

Gap	Supporting evidence	Recommended research question	Suggested methodology	Potential contribution
products and rehearsed tasks measured instead of independent learning	post-test; AI-scored outcomes; surface-only gains	persist when learners work without AI?	delayed retention and transfer tasks; blinded, double-rated scoring	performance from learning
Longitudinal	No delayed post-test identified; interventions mostly 2–16 weeks	How do benefits and dependency evolve across a semester or programme?	Longitudinal cohort and interrupted time-series designs	Tests the over-reliance proposition
Mechanism and orchestration	Only 3 AI-versus-AI studies; one three-arm mediation study	Which features (bounded use, timing, agency, teacher mediation) drive learning?	Factorial and component-analysis experiments	Builds design principles
Geographical	No studies from Brunei, Cambodia, Laos, Myanmar, Timor-Leste; one from Singapore	How do outcomes differ in low-resource and multilingual systems?	Multi-site comparative studies; regional consortia	Regional representativeness
Population and discipline	35 of 59 studies in language learning; few in STEM, health and social sciences	Do patterns hold for disciplinary reasoning, clinical skills and quantitative work?	Discipline-specific trials with domain measures	Generalisability beyond language
Readiness and equity	Inconsistent proficiency moderation; access and AI-literacy barriers	Who benefits, who is left behind, and can AI-literacy training close gaps?	Moderator analyses with pre-registered hypotheses; equity-focused designs	Equitable implementation
Affective and epistemic regulation	Qualitative reports of frustration, distrust and verification burden	Do emotion regulation and verification skills mediate AI effects?	Mixed methods with process data (logs, think-aloud) and validated affect measures	Tests framework propositions
Institutional and policy	One institution-wide pilot with historical comparison	Which assessment and governance frameworks sustain integrity and learning?	Multi-institution policy evaluations; difference-in-differences	Evidence for institutional policy
Reporting quality	Inconsistent statistics and incomplete instrument reporting	—	Adoption of reporting checklists (e.g., CONSORT-style items for educational trials)	Usable, synthesisable evidence

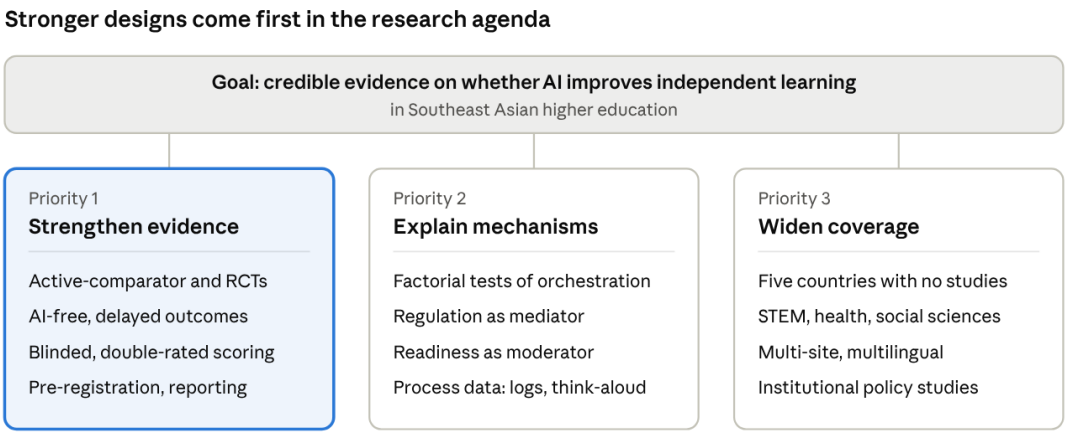


Figure 8. Future research agenda, ordered by priority. Stronger designs come first because every other question depends on credible measures of independent learning.

5. CONCLUSION

Across 59 studies from six Southeast Asian countries, web-based and mobile AI tools in higher education were associated with better surface-level performance, higher confidence and engagement, and comparable outcomes to equivalent human support, but not with clearly demonstrated gains in deeper, independent learning. **RQ1.** The evidence is recent (52 studies in 2025–2026), concentrated in Indonesia, Vietnam and Thailand, dominated by language learning and by general-purpose chatbots accessed through the web, and built mostly on non-randomised designs. **RQ2.** Of 45 studies measuring learning, 36 reported improvements, but advantages appeared mainly without a comparator or against business-as-usual teaching; four of six comparisons with equivalent human support found parity, and measured gains were concentrated in surface features while self-reports were almost uniformly positive. **RQ3.** Teacher mediation, bounded use, feedback design and embedding in explicit pedagogical models were associated with more favourable outcomes, while proficiency, AI literacy, access and institutional resources moderated benefit. **RQ4.** Students and instructors reported over-reliance, inaccurate output, impersonal feedback, integrity concerns, negative emotions, technical problems and limited AI literacy. **RQ5.** The evidence base is weak in design and reporting, with confounding, unclear outcome measurement and missing delayed measures as the main problems. Theoretically, the review supports hybrid human–AI regulation and shifts attention from the tool to the orchestrated activity. Practically, it supports using AI to extend feedback and practice under bounded, teacher-mediated designs, with assessment that also measures learning without AI. **Limitations.** This review searched one database, so relevant studies indexed elsewhere may be missing. Twenty-seven of 106 sought full texts could not be retrieved, including several with strong designs. Screening, extraction, appraisal and coding were performed by a single reviewer with AI assistance; second-reviewer verification is required before submission. The protocol was not registered and was amended once before the affected full texts were retrieved. English-language restriction may exclude studies published in national languages. Effect direction does not capture magnitude, and the analytical framework has not been tested.

6. ACKNOWLEDGEMENTS

Acknowledgements: The authors have no acknowledgements to declare. Funding: This research received no external funding. Conflict of interest: The authors declare that they have no competing interests. Data availability: The screening workbook, extraction matrix with quality appraisal, and thematic synthesis workbook are available in Scopus at publisher. Use of AI-assisted technologies: The authors used Claude (Anthropic) to assist with record screening, data extraction, quality appraisal, thematic coding, figure preparation, and language editing. The authors reviewed and verified all AI-assisted content and remain fully responsible for the final manuscript.

7. REFERENCES

Andewi, W., Waziana, W., Wibisono, D., Putra, K. A., Hastomo, T., & Oktarin, I. B. (2025). From prompting to proficiency: A mixed-methods analysis of prompting with ChatGPT versus lecturer interaction in an EFL classroom. *Journal of Studies in the English Language, 20*(2), 210–238. https://doi.org/10.64731/jsel.v20i2.282318

Apriani, E., Hamidah Daulay, S., Aprilia, F., Gafur Marzuki, A., Warsah, I., Supardan, D., & Muthmainnah. (2025). A mixed-method study on the effectiveness of using ChatGPT in academic writing and students' perceived experiences. *Journal of Language and Education, 11*(1), 26–45. https://doi.org/10.17323/jle.2025.17913

Arifin, S., Setiyadi, D., Kurniawan, W., & Ekawati, S. (2026). Fostering pre-service teachers' empathy through AI-enhanced MoralQ application: An experimental study. *Journal of Interdisciplinary Studies in Education, 15*(7). https://doi.org/10.32674/zjkyyp67

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

Campbell, M., McKenzie, J. E., Sowden, A., Katikireddi, S. V., Brennan, S. E., Ellis, S., Hartmann-Boyce, J., Ryan, R., Shepperd, S., Thomas, J., Welch, V., & Thomson, H. (2020). Synthesis without meta-analysis (SWiM) in systematic reviews: Reporting guideline. *BMJ, 368*, l6890. https://doi.org/10.1136/bmj.l6890

Chinwong, D., Penthinapong, T., & Chinwong, S. (2025). Integrating ChatGPT for smoking cessation counseling practice in pharmacy education: A single group quasi-experimental study. *Tobacco Induced Diseases, 23*, Article 182. https://doi.org/10.18332/tid/211706

Chuan, O. Y., Singh, U. G., & Alwi, S. S. E. (2026). IEEC: A framework for AI literacy in programming. *Asia Pacific Journal of Educators and Education, 41*(1), 337–367. https://doi.org/10.21315/apjee2026.41.1.15

Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education: The state of the field.

International Journal of Educational Technology in Higher Education, 20, 22. https://doi.org/10.1186/s41239-023-00392-8

Dahri, N. A., Al-Rahmi, W. M., Alhashmi, K. A., & Bashir, F. (2025). Enhancing mobile learning with AI-powered chatbots: Investigating ChatGPT's impact on student engagement and academic performance. *International Journal of Interactive Mobile Technologies, 19*(11), 17–38. https://doi.org/10.3991/ijim.v19i11.54643

Dang, H. N. N., Luong, T. V., Vo, H. T. H., Nguyen, L. T. K., Tran, T. N., Pham, H. T. A., Truong, H. T., Tran, H., Tran, T. T., Hoang, T. A., Doan, T. C., Huynh, Q., & Marwick, T. H. (2026). Artificial intelligence–powered virtual standardized patients in teaching history-taking skills to medical students: A randomized controlled trial. *BMC Medical Education, 26*, Article 984. https://doi.org/10.1186/s12909-026-09305-5

Dang, T. H. T. (2026). Enhancing English speaking skills of Vietnamese university students through ChatGPT-aided mind maps. *Journal of Language Teaching and Research, 17*(3), 999–1007. https://doi.org/10.17507/jltr.1703.24

Darmawansah, D., Rachman, D., & Febiyani, F. (2024). Collaborative argument map-based ChatGPT to enhance EFL students' argumentation complexity and group efficacy. *Taiwan International ESP Journal, 15*(2), 97–124. https://doi.org/10.6706/TIESPJ.202412_15(2).0004

Duc, D. H. (2026). Operationalizing SAMR Redefinition in EFL reading: AI as a mediating tool for literacy innovation. *Digital Technologies Research and Applications, 5*(1), 66–82. https://doi.org/10.54963/dtra.v5i1.1900

Duong, T., & Suppasetseree, S. (2024). The effects of an artificial intelligence voice chatbot on improving Vietnamese undergraduate students' English speaking skills. *International Journal of Learning, Teaching and Educational Research, 23*(3), 293–321. https://doi.org/10.26803/ijlter.23.3.15

Dwipayana, I. K. A., Sutama, I. M., Rasna, I. W., & Artika, I. W. (2026). AI-assisted multimodal discourse analysis learning model to enhance critical literacy skills. *Suranaree Journal of Social Science, 20*(2), Article e284991. https://doi.org/10.55766/sjss284991

Fadilah, E. (2026). Artificial intelligence-mediated peer collaboration in EFL argumentative writing: Exploring learners' inferential reasoning development. *Technology in Language Teaching and Learning, 8*, Article 103944. https://doi.org/10.29140/tltl.2026.103944

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

Fardian, D., Suryadi, D., Prabawanto, S., & Jupri, A. (2025). Integrating Chat-GPT in the classroom: A study on linear algebra learning in higher education. *International Journal of Information and Education Technology, 15*(4), 732–751. https://doi.org/10.18178/ijiet.2025.15.4.2279

Furze, L., Perkins, M., Roe, J., & MacVaugh, J. (2024). The AI Assessment Scale (AIAS) in action: A pilot implementation of GenAI-supported assessment. *Australasian Journal of Educational Technology, 40*(4), 38–55. https://doi.org/10.14742/ajet.9434

Gross, J. J. (1998). The emerging field of emotion regulation: An integrative review. *Review of General Psychology, 2*(3), 271–299. https://doi.org/10.1037/1089-2680.2.3.271

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112. https://doi.org/10.3102/003465430298487

Hong, Q. N., Fàbregues, S., Bartlett, G., Boardman, F., Cargo, M., Dagenais, P., Gagnon, M.-P., Griffiths, F., Nicolau, B., O'Cathain, A., Rousseau, M.-C., Vedel, I., & Pluye, P. (2018). The Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers. *Education for Information, 34*(4), 285–291. https://doi.org/10.3233/EFI-180221

Jitpaisarnwattana, N. (2026). Generative AI as an IELTS tutor: The effects of AI-generated feedback on EFL students' writing. *Language Education and Assessment, 9*, Article 104046. https://doi.org/10.29140/lea.2026.104046

Jitpaisarnwattana, N., & Saville, N. (2026). Using automated writing evaluation to improve EFL students' IELTS writing: Effects across proficiency levels and learning behaviours. *LEARN Journal: Language Education and Acquisition Research Network, 19*(2), 265–296. https://doi.org/10.70730/XYWK3308

Kailin, Z., & Saeed, M. A. (2026). Chinese EFL learners' engagement with ChatGPT feedback on academic writing: A case study in Malaysia. *Computers and Composition, 79*, Article 102976. https://doi.org/10.1016/j.compcom.2025.102976

Kanoksilapatham, B., & Takrudkaew, T. (2025). Integrating ChatGPT to enhance university students' communication skills: Pedagogical considerations. *World Journal of English Language, 15*(6), 173–183. https://doi.org/10.5430/wjel.v15n6p173

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., … Kasneci, G. (2023). ChatGPT for good? On

opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Kawinkoonlasate, P. (2025). A comparative study of Google Gemini and ChatGPT in enhancing English language learning for EFL learners: A case study of the English research writing course. *Pedagogical Research, 10*(4), Article em0251. https://doi.org/10.29333/pr/17670

Khusniyah, N. L., Hakim, L., & Muslimin, A. I. (2026). Comparing AI-based and peer-based feedback in teaching the CaRS model: A quasi-experimental study on postgraduate academic writing. *International Journal of Information and Education Technology, 16*(3), 687–694. https://doi.org/10.18178/ijiet.2026.16.3.2541

Kong, J. S. M., Teo, B. S., Lee, Y. J., Pabba, A. B., Lee, E. J. D., & Sng, J. C. G. (2021). Virtual integrated patient: An AI supplementary tool for second-year medical students. *The Asia Pacific Scholar, 6*(3), 87–90. https://doi.org/10.29060/TAPS.2021-6-3/SC2394

Mandasari, B., Basthomi, Y., Anugerahwati, M., & Wulyani, A. N. (2026). EFL freshmen's emotions and emotion regulation strategies in AI-supported collaborative argumentative essay writing. *Applied Research on English Language, 15*(1), 1–24. https://doi.org/10.22108/are.2025.146800.2621

Margareth, A., Sukardjo, M., & Situmorang, R. (2025). The evaluation and comparison of translation technologies on the learning outcomes of legal text translation studies. *LEARN Journal: Language Education and Acquisition Research Network, 18*(1), 568–593. https://doi.org/10.70730/DFYU5629

Marzuki, A. G., Huriyah, H., Anita, A., Triassanti, R., Kencanawati, D., & Santiana, S. (2026). Enhancing Indonesian EFL students' academic writing proficiency through NLP-based writing assistants: A mixed methods study. *Southeastern Philippines Journal of Research and Development, 31*(2), 90–104. https://doi.org/10.53899/spjrd.v31i2.1300

Me, R. C., Kamil, M. J. M., Razali, A. F., Li, J., Abidin, S. Z., & Ramli, S. H. (2025). Artificial intelligence (AI)-aided collaborative design in industrial design education for final year projects (FYP): Improving workflow and innovation. *Academic Quarter,* (31), 28–46. https://doi.org/10.54337/academicquarter.i31.11269

Mediyawati, N., Young, J. C., & Nusantara, S. B. (2021). U-Tapis: Automatic spelling filter as an effort to improve Indonesian language competencies of journalistic students. *Cakrawala Pendidikan, 40*(2), 402–412. https://doi.org/10.21831/cp.v40i2.34546

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

Nguyen, P. B. T., & Tran, N. B. C. (2026). AI-supported tools for English academic writing and publishing: A case study of Vietnamese biotechnology undergraduates. *Arab World English Journal, 17*(1), 437–468. https://doi.org/10.24093/awej/vol17no1.27

Niez, R. A., Peñaranda, J. O., Zipagan, M. N., Sabonsolin, J. C. M., Jaca, C. A. L., Tancinco, N. P., Magno, J. M., Gelizon, M. J. B., & Cerera, J. P. (2026). Agentic AI as process scaffolding in undergraduate ESL instruction: Effects on proficiency, engagement, and self-regulated language learning. *Theory and Practice in Language Studies, 16*(10), 3346–3357. https://doi.org/10.17507/tpls.1610.04

Oktarin, I. B., Saputri, M. E. E., Magdalena, B., Hastomo, T., & Maximilian, A. (2024). Leveraging ChatGPT to enhance students' writing skills, engagement, and feedback literacy. *Edelweiss Applied Science and Technology, 8*(4), 2306–2319. https://doi.org/10.55214/25768484.v8i4.1600

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*, n71. https://doi.org/10.1136/bmj.n71

Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review, 18*(4), 315–341. https://doi.org/10.1007/s10648-006-9029-9

Pham, Q. N. (2026). Affordances and constraints of ChatGPT feedback in ESP writing: Vietnamese medical students' perspectives. *The Asian Journal of Applied Linguistics, 10*(3), Article 1336. https://doi.org/10.25442/hku.31827469

Phung, T.-N., Do, D.-C., Nguyen, T.-T., Nguyen, V.-S., Nguyen, T.-V., & Le, D.-N. (2026). An integrated framework for outcome based education and AI supported blended learning in curriculum redesign and intelligent training management. *Discover Computing, 29*, Article 196. https://doi.org/10.1007/s10791-026-10088-y

Puteri, C. G., Cahyono, B. Y., Widiati, U., & Suryati, N. (2025). Using ChatGPT as a writing assistant: A study on essay quality development among Indonesian EFL university students. *Indonesian Journal of Applied Linguistics, 15*(2), 332–346. https://doi.org/10.17509/3ebmy158

Rath, A. (2025). Leveraging ChatGPT to support terminology learning in oral anatomy: A mixed-methods study among linguistically diverse dental students. *BMC Medical Education, 25*, Article 1425. https://doi.org/10.1186/s12909-025-07968-0

Ravelo, V. M., Ancho, I. V., Mecha, J. P. F., & Albakri, I. S. M. A. (2026). Leveraging artificial intelligence as a Vygotskian tool to enhance Filipino university students' language learning and critical thinking skills. *The Palawan Scientist, 18*(1), 192–206. https://doi.org/10.69721/TPS.J.2026.18.1.20

Rizal, A., Kadir, F., Permana, I., Veddayana, C., Suhendra, A., Damara, M. K., & Akhiruddin. (2026). Artificial intelligence for health citizenship education: Bridging civic competence and physical health literacy in universities. *Journal of Intelligent Decision Making and Information Science, 3*(6s), 560–575. https://doi.org/10.59543/jidmis.v3.1414

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*(1), 68–78. https://doi.org/10.1037/0003-066X.55.1.68

Sangpom, W., & Sangpom, N. (2026). Artificial intelligence as an educational decision support system in problem-based learning: Critical thinking, decision-making, and human–AI collaboration in higher education. *Journal of Intelligent Decision Making and Information Science, 3*(5s), 142–167. https://doi.org/10.59543/jidmis.v3.1074

Shen, Q., Pituksung, S., Atsawamaitree, L., & Pituksung, P. (2026). Reciprocal prompt co-adaptation with GPT-4 for entrepreneurship education. *International Journal of Systematic Innovation, 10*(4), Article 026110022. https://doi.org/10.6977/IJoSI.202608_10(4).0004

Silor, A. C., & Silor, F. S. C. (2026). AI-driven learning tools and gender equity: Rethinking inclusive digital pedagogies in higher education. *International Journal of Learning, Teaching and Educational Research, 25*(2), 111–129. https://doi.org/10.26803/ijlter.25.2.6

Sindhvananda, K., Ruengwattanachot, K., Leksuwankun, S., Pongpitakmetha, T., Hiransuthikul, A., Surawattanawong, T., Anukoolwittaya, P., & Panjasriprakarn, P. (2026). AI-powered simulated patients with automated feedback for enhancing headache history-taking skills: A convergent mixed-methods study. *BMC Medical Education, 26*, Article 1151. https://doi.org/10.1186/s12909-026-09511-1

Sinlapaninman, U., & Yonwilad, W. (2025). AI-enhanced research coaching for graduate education: A mixed-methods evaluation of learning outcomes and educational design effectiveness. *Educational Process: International Journal, 19*, Article e2025535. https://doi.org/10.22521/edupij.2025.19.535

Smithsarakarn, P. (2026). AI-assisted feedback and EFL writing development: Evidence from Thai university students in an English for Business course. *World Journal of English Language, 16*(6), 323–332. https://doi.org/10.5430/wjel.v16n6p323

Sumakul, D. T. Y. G. (2025). The interplay between L2 motivation and artificial intelligence: What and how. *CALL-EJ, 26*(3), 107–128. https://doi.org/10.54855/callej.252636

Susiawati, I., Mardani, D., Raswan, Syarifah, Matin, I., & Thamrin, M. H. (2026). Exploring AI-assisted Arabic learning for linguistic intelligence development among Indonesian university students. *Indonesian Journal of Applied Linguistics, 16*(1), 81–91. https://doi.org/10.17509/8c995612

Suwardika, G., Sopandi, A. T., Indrawan, I. P. O., & Masakazu, K. (2026). Flipped design thinking with generative AI for digital literacy in ODL. *Asian Association of Open Universities Journal, 21*(1), 48–64. https://doi.org/10.1108/AAOUJ-11-2025-0206

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

Syafei, M., & Nuraeningsih. (2025). Leveraging artificial intelligence in writing: ELT students' perspectives and experiences. *International Journal of Learning, Teaching and Educational Research, 24*(5), 416–430. https://doi.org/10.26803/ijlter.24.5.22

The, N. H., Du, T. T., & Hoang, T. T. (2026). Exploring the impact of artificial intelligence on English majors' speaking skills: A case study at a Vietnamese university. *Arab World English Journal*, Special Issue on Artificial Intelligence (3), 166–179. https://doi.org/10.24093/awej/AI3.11

The Nguyen, H., Du Tran, T. D., Huong Nguyen, T. K. H., & Xuan Nguyen, T. X. (2026). The impact of AI-integrated CLT on students' oral communication skills: A case study at Thu Dau Mot University, Vietnam. *Multidisciplinary Science Journal, 8*(10), Article e2026626. https://doi.org/10.31893/multiscience.2026626

Thomas, J., & Harden, A. (2008). Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Medical Research Methodology, 8*, 45. https://doi.org/10.1186/1471-2288-8-45

Thuan, P. D., & Hanh, N. T. H. (2025). Students' perspectives on application of AI-powered technology in learning English pronunciation: A case of Reading Progress. *International Journal of Evaluation and Research in Education, 14*(1), 731–738. https://doi.org/10.11591/ijere.v14i1.31381

Tran, K. L., Nguyen, L. T., & Pham, M. K. (2026). Enhancing student success through AI-assisted reflective learning in entrepreneurship education: An empirical study. *Student Success, 17*(3), 120–130. https://doi.org/10.63608/ssj.4067

Tran, T. T. T. (2025). Enhancing EFL writing revision practices: The impact of AI- and teacher-generated feedback and their sequences. *Education Sciences, 15*(2), Article 232. https://doi.org/10.3390/educsci15020232

Tuong, T.-P.-L., & Tran, T.-T. (2025). The differential impact of AI tools among EFL university learners: A process writing approach. *International Journal of Learning, Teaching and Educational Research, 24*(5), 452–471. https://doi.org/10.26803/ijlter.24.5.24

Vo, T. L., & Ho, T. M. P. (2026). A quasi-experimental study on enhancing EFL learners' grammar precision in writing skills under supported-ChatGPT approach at a university in a remote region of Vietnam. *Arab World English Journal*, Special Issue on Artificial Intelligence (3), 361–375. https://doi.org/10.24093/awej/A3.23

Wangdi, T., & Rofiah, N. L. (2026). Generative AI-induced emotions in EFL classrooms: Effects and students' regulation strategies. *Journal of Education Culture and Society, 17*(1), 609–629. https://doi.org/10.15503/jecs2026.1.609.629

Wiboolyasarin, W., Chomchuen, F., Kamonsawad, R., Boonprawes, S., Wiboolyasarin, K., & Jinowat, N. (2026). Timing is everything: Investigating the impact of immediate, delayed, and combined AI-generated feedback on L2 writing proficiency. *Social Sciences & Humanities Open, 13*, Article 102790. https://doi.org/10.1016/j.ssaho.2026.102790

Wiboolyasarin, W., Wiboolyasarin, K., Suwanwihok, K., Jinowat, N., & Muenjanchoey, R. (2024). Synergizing collaborative writing and AI feedback: An investigation into enhancing L2 writing proficiency in wiki-based environments. *Computers and Education: Artificial Intelligence, 6*, Article 100228. https://doi.org/10.1016/j.caeai.2024.100228

Wikanengsih, Puspita, R. D., Wardani, D. S., Nurhayati, E., & Moid, I. B. (2025). Enhancing academic writing in higher education: The impact of ChatGPT and QuillBot in AI-driven project-based learning. *Journal of Information Technology Education: Innovations in Practice, 24*, Article 025. https://doi.org/10.28945/5646

Wuttikamonchai, O. (2025). A blended and project-based learning management model using artificial intelligence to enhance Thai undergraduate student digital media creation skills. *Edelweiss Applied Science and Technology, 9*(4), 2692–2705. https://doi.org/10.55214/25768484.v9i4.6640

Yusof, I. J. (2025). ChatGPT-assisted retrieval practice and exam scores: Does it work? *Journal of Information Technology Education: Research, 24*, Article 8. https://doi.org/10.28945/5474

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

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