**International Journal of Education and Vocational Studies**

Vol. 1, No. 2, November 2026, pp. 50-68
ISSN: 3124-9132 (Media Online)
DOI: https://doi.org/10.62712/ijevs.v1i2.8


# Generative AI in Southeast Asian TVET: A Systematic Review of Learning Outcomes, Ethics and Research Directions

**Vion Age Tricahyo[1*], Haeriani H[2], Husni Mubarak[3]**

[1]Fakultas ilmu pendidikan, Perpustakaan dan Sains Informasi, Universitas Negeri Jakarta, Jakarta, Indonesia
[2]Fakultas Teknik, Pendidikan Teknologi Informasi, Universitas Negeri Gorontalo, Gorontalo, Indonesia
[3]Fakultas Ekonomi dan Bisnis, Bisnis Digital, Universitas Negeri Makassar, Makassar, Indonesia
Author's Email: [1*]vionage@unj.ac.id, [2]haerianih@ung.ac.id, [3]husni.mubarak@unm.ac.id

**Article Info**

***Article history:***

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

***Keywords:***

Academic integrity;
ChatGPT;
Generative artificial intelligence;
Southeast Asia;
Technical and vocational education and training.

**ABSTRACT**

Background: Generative artificial intelligence tools such as ChatGPT entered technical and vocational education and training faster than evidence on their effects. Objective: This review synthesized empirical evidence on the applications, learning outcomes and ethical challenges of generative artificial intelligence in Southeast Asian vocational education. Method: Following PRISMA 2020, ten search strings were run in OpenAlex and Semantic Scholar for 2022–2026, which yielded 254 unique records, of which 25 primary studies met the eligibility criteria. Findings: The studies came from Indonesia (17), Malaysia (5) and Thailand (3), and 18 were published in 2026. Fourteen concerned English language learning. Four quasi-experimental studies reported gains in writing, speaking, writing self-efficacy and AI literacy, while most other evidence was self-reported, and no study measured hands-on technical competence. Academic integrity (9 studies), overreliance (7) and inaccurate output (5) were the most frequent risks, and privacy and unequal access were also reported. Benefits depended on teacher mediation, task design, AI literacy and institutional guidance. Implications: Institutions should favors guided over unrestricted use, and research should prioritize controlled, longitudinal studies with practical assessment. Contribution: The review offers the first region-specific evidence map and a framework for responsible integration.

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

***Corresponding Author:***

Vion Age Tricahyo,
Fakultas ilmu pendidikan, Perpustakaan dan Sains Informasi, Universitas Negeri Jakarta
Jakarta, Indonesia
Email: vionage@unj.ac.id

## 1. INTRODUCTION

Generative artificial intelligence (GenAI) refers to systems, most prominently large language models, that produce text, code, images and feedback in response to natural-language prompts. Since the public release of ChatGPT in late 2022, such tools have spread through education faster than any previous technology, offering personalized explanation, immediate feedback and support for writing, but also raising concerns about inaccurate output, bias, privacy and the erosion of independent work (Kasneci et al., 2023; Samala et al., 2024). Syntheses of the first wave of studies report mostly positive effects on self-regulated learning and critical thinking, qualified by warnings about overreliance (Sardi et al., 2025).

Technical and vocational education and training (TVET) is a distinctive setting for these tools. Its purpose is to develop occupational competence, an integration of knowledge, skills and attitudes demonstrated in work performance (Weigel et al., 2007), and its credibility rests on assessments that resemble real professional tasks (Gulikers et al., 2004). GenAI may support parts of this competence, such as technical communication, documentation and problem diagnosis, but much of vocational learning is procedural and physical. Whether a tool that generates language helps learners to weld, wire, program, nurse or cook is an empirical question, not an assumption that can be carried over from general education.

In Southeast Asia, TVET is central to national development strategies and is delivered through diverse systems: vocational upper secondary schools (*Sekolah Menengah Kejuruan*, SMK) and state polytechnics in Indonesia, polytechnics and community colleges in Malaysia, and public and private vocational colleges in Thailand. Access to devices, connectivity and teacher professional development varies widely within and between these systems, which may shape who benefits from GenAI and who is exposed to its risks. Teachers and institutions have also had to respond to GenAI before national guidance was in place.

The existing reviews leave this setting largely unexamined. A systematic review of 38 studies addressed GenAI and self-regulated learning across education levels (Sardi et al., 2025). A meta-analysis of ten Scopus-indexed experiments in TVET, vocational and engineering education estimated a moderate pooled effect of ChatGPT-assisted instruction, with substantial heterogeneity (Hidayat et al., 2026). A scoping review mapped how vocational teachers' GenAI competence has been defined and measured (Mubarok et al., 2026), and a bibliometric analysis found AI research in vocational education concentrated in health fields and weakly internationalised (Dinata et al., 2025). None of these reviews focused on Southeast Asia, none drew systematically on the regional journals where much of the evidence is published, and none examined learning outcomes, ethical challenges and implementation conditions together. Their scope also leaves open whether reported gains extend beyond language and academic tasks to the practical competences that define TVET.

This review therefore aimed to synthesise the empirical evidence on GenAI in Southeast Asian TVET, evaluate its methodological strength, and translate it into a framework and research agenda. It addressed six research questions. RQ1: How are empirical studies distributed geographically, temporally and methodologically? RQ2: How is GenAI applied in TVET teaching, learning and administration? RQ3: What learning outcomes are reported, and how strong is the evidence? RQ4: What ethical challenges and institutional barriers are reported? RQ5: Which implementation factors are associated with benefits and with risk mitigation? RQ6: What research gaps remain, and what agenda follows from them? The review contributes the first region-specific evidence map, an integrated framework for responsible GenAI integration in TVET, and a prioritised agenda for research on practical skills.

## 2. LITERATURE REVIEW

### 2.1. Generative AI in education

Large language models generate fluent text by predicting likely continuations of a prompt. In education this enables on-demand explanation, drafting and revision support, conversational practice and the rapid production of teaching materials (Kasneci et al., 2023). A review of 34 studies on ChatGPT identified personalised and adaptive learning, instant feedback and improved accessibility as the main advantages, and lack of emotional intelligence, overreliance and privacy as the main drawbacks (Samala et al., 2024). The same capability that makes GenAI useful also creates a design problem: the tool can perform the very task the learner is meant to learn. Research on cognitive offloading shows that people readily shift mental work to external aids, which can improve immediate performance while reducing what is retained once the aid is removed (Risko & Gilbert, 2016). Whether GenAI acts as a scaffold or as a substitute is therefore a central question for any assessment of its learning value.

### 2.2. Competence-based TVET and authentic assessment

Vocational education is organised around competence: the ability to perform occupational tasks to a defined standard in real or realistic conditions (Weigel et al., 2007). Assessment in such programmes is meant to be authentic, that is, aligned with the tasks, physical and social context, results and criteria of professional practice (Gulikers et al., 2004). This has two implications for GenAI research. First, outcomes such as written fluency or self-reported confidence are, at most, partial indicators of vocational competence. Second, evidence that GenAI improves performance on language tasks cannot be generalised to practical skills unless the practical performance itself is assessed.

### 2.3. Ethical risks of generative AI

The ethical risks most often discussed are academic integrity, the accuracy of generated content, overreliance and deskilling, privacy and data governance, and unequal access (Kasneci et al., 2023; Samala et al., 2024). In TVET these risks have a particular weight because certificates attest competence to employers and, in fields such as health, maritime work and electrical installation, to the public. A qualification earned with work produced by a tool may certify a competence the graduate does not have.

### 2.4. Technology acceptance

Much research on educational technology draws on the Unified Theory of Acceptance and Use of Technology (UTAUT), which explains behavioural intention through performance expectancy, effort expectancy, social influence and facilitating conditions (Venkatesh et al., 2003). Studies of GenAI often extend the model with trust, perceived risk and AI literacy. Acceptance research does not measure learning, but it reveals how teachers and students weigh benefits against risks and is therefore relevant to the ethical and institutional questions of this review.

### 2.5. Initial conceptual framework

The review was organised around an initial framework in which GenAI use involves a tool, a user (student, teacher or administrator) and a setting (level, programme and country). Use is mediated by pedagogical and institutional conditions and produces two kinds of consequence: learning outcomes, measured or perceived, and ethical challenges. RQ1 and RQ2 map the inputs and applications, RQ3 and RQ4 the two kinds of consequence, and RQ5 the mediating conditions. The framework is refined in light of the findings in Section 4.10. As a review, the study tested no hypotheses.

## 3. RESEARCH METHOD

### 3.1. Research design

The study was a systematic literature review reported in line with the PRISMA 2020 statement (Page et al., 2021). A systematic design was chosen because the evidence is new, scattered across national and international journals, and methodologically heterogeneous, so a transparent and reproducible procedure was needed to locate it and judge its strength. The review protocol was not registered in advance.

### 3.2. Information sources

The search was run in the OpenAlex and Semantic Scholar indices through a scholarly search interface that covers open-access and subscription literature. These sources were chosen because they index many regional journals that are not covered by Scopus or Web of Science. Scopus, Web of Science, ERIC and Google Scholar were not searched. Table 1 therefore distinguishes the strings actually executed from the Boolean strings proposed for replication in Scopus and Web of Science.

### 3.3. Search strategy

The search was run on 30 October 2026. Ten search strings combined three conceptual groups: GenAI (generative AI, ChatGPT, chatbot), TVET (vocational education, TVET, polytechnic, SMK, vocational college) and geography (Southeast Asian countries). Because the interface ranks results by semantic relevance across titles, abstracts and indexed metadata, field-restricted Boolean syntax could not be applied. Publication years were limited to 2022 onwards. The first page of up to 25 records was retrieved for every string, and a second page for the two broadest strings, which gave 293 records (Table 1). Across publication years, the unique records rose from 4 in 2023 to 21 in 2024, 93 in 2025 and 136 in 2026, and none were from 2021 or 2022.

**Table 1.** Database search strategy

| No. | Search string | Records |
|---|---|---|
| | *Executed: OpenAlex and Semantic Scholar (relevance-ranked; 30 October 2026; 2022–2026)* | |
| 1 | generative artificial intelligence ChatGPT vocational education Indonesia | 25 |
| 2 | generative artificial intelligence ChatGPT TVET Malaysia | 25 |
| 3 | generative AI ChatGPT technical vocational education Thailand Vietnam Philippines | 24 |
| 4 | generative AI ChatGPT in vocational education TVET students (2 pages)† | 45 |
| 5 | ChatGPT use among polytechnic students Malaysia Indonesia | 25 |
| 6 | artificial intelligence chatbot learning vocational high school SMK students | 25 |
| 7 | generative AI academic integrity ethics vocational college students Southeast Asia | 25 |
| 8 | generative AI technical vocational education Vietnam Thailand Singapore Philippines learning outcomes | 24 |
| 9 | ChatGPT TVET Malaysia students instructors (2 pages) | 50 |
| 10 | vocational teachers generative AI adoption Indonesia polytechnic lecturers | 25 |
| | **Total before removal of duplicates** | **293** |
| | *Proposed for replication (not executed)* | |
| S | Scopus: TITLE-ABS-KEY(("generative AI" OR genAI OR ChatGPT OR "large language model*" OR LLM) AND (TVET OR "technical and vocational" OR "vocational education" OR "technical education" OR polytechnic* OR "vocational school*") AND ("Southeast Asia" OR ASEAN OR Indonesia OR Malaysia OR Thailand OR Vietnam OR Philippines OR Singapore OR Brunei OR Cambodia OR Laos OR Myanmar OR "Timor-Leste")) AND PUBYEAR > 2021 | – |

| No. | Search string | Records |
|---|---|---|
| W | Web of Science: TS=(("generative AI" OR ChatGPT OR "large language model*") AND (TVET OR "vocational education" OR "technical education" OR polytechnic*) AND ("Southeast Asia" OR Indonesia OR Malaysia OR Thailand OR Vietnam OR Philippines OR Singapore OR Brunei OR Cambodia OR Laos OR Myanmar OR "Timor-Leste")) | – |

Note. The first page of string 4 was limited to 2023–2026. Field restriction was not available in the executed searches.

### 3.4. Eligibility criteria

Table 2 sets out the eligibility criteria. Two decisions shaped the evidence base. First, the setting had to be identifiable from the title or abstract, either by country, by a named institution or by a country-specific institution type such as SMK, which excluded some plausibly regional studies. Second, studies of artificial intelligence were included only when the abstract named a generative tool or described AI use that generated text or responses for the user; studies of rule-based chatbots or unspecified AI were excluded.

**Table 2.** Inclusion and exclusion criteria

| Criterion | Inclusion | Exclusion |
|---|---|---|
| Study type | Empirical primary study (any design) | Reviews, bibliometric analyses, conceptual or discussion papers |
| Technology | Named GenAI tool, or AI described as generating text or responses | Rule-based chatbots; AI type not specified |
| Context | TVET: SMK, polytechnic, community or vocational college, vocational higher education, TVET teacher education | Mixed samples with TVET results not separable; general higher education |
| Geography | Southeast Asian country stated or identifiable from a named institution | Other regions; setting not identifiable |
| Outcomes | Learning outcomes (measured or perceived) or ethical challenges | Adoption determinants only, with no risk or trust construct |
| Publication type | Journal article or conference paper | Books, book chapters in repositories, preprints |
| Language | English | Other languages |
| Period | 2022–2026 | Before 2022 |
| Full text | Not required; decisions based on title, abstract and metadata | – |

### 3.5. Study selection

Duplicates were removed by DOI or index identifier, leaving 254 records. These were screened on title and abstract for the joint presence of an artificial intelligence term and a vocational or TVET term, which excluded 134 records. One reviewer assessed the remaining 120 records against the full criteria, using titles, abstracts and bibliographic metadata. Full texts were not retrieved, so no full-text assessment stage took place. Ninety-five records were excluded: 34 were not empirical primary studies, 32 were outside Southeast Asia or did not identify their setting, 20 did not specify generative AI, 4 did not have a TVET-specific sample, 3 reported adoption determinants only and 2 were excluded for other reasons (an instrument-validation pilot and an administrative information chatbot). Twenty-five studies were included (Figure 1).

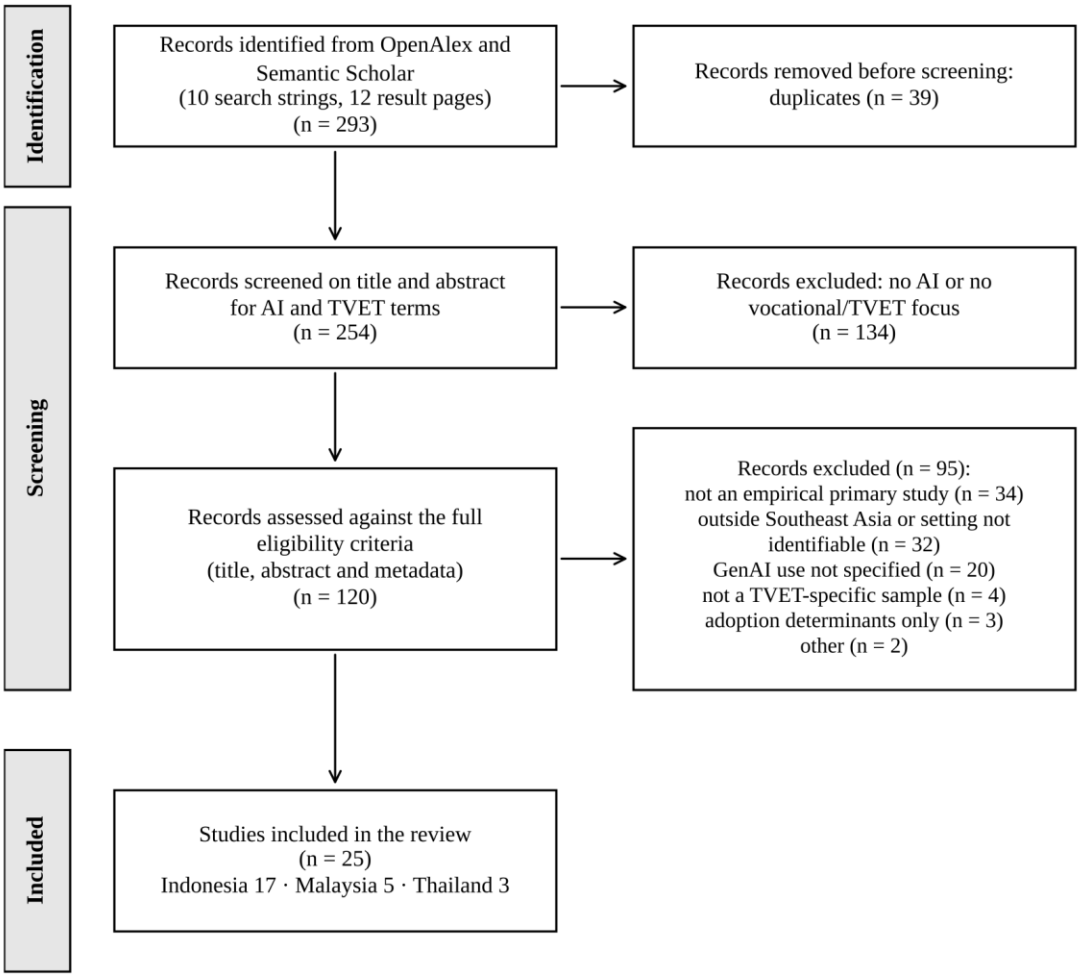



**Figure 1.** PRISMA 2020 flow diagram

### 3.6. Data extraction

A standard extraction form recorded, for each study, the author and year, country, educational context, research aim, design, sample, GenAI tool, learning activity, outcomes, ethical issues, main findings and reported limitations. Author affiliations were taken from article landing pages where these were accessible. Information that was not reported in the abstract was coded as not reported (NR); research aims, learning activities and limitations were often incompletely described and are summarised only where reported. The extracted data are presented in Tables 3 to 5.

### 3.7. Data analysis and synthesis

The analysis combined descriptive analysis and thematic synthesis. Descriptive analysis counted studies by country, year, level, participant group, design, tool, application and outcome. Designs were grouped by strength into comparative (quasi-experimental with a comparison group), single-group pre-post or classroom action research (CAR), survey or mixed methods, and qualitative or programme evaluation. Outcomes were classified by type of evidence, distinguishing measured outcomes from correlational and self-reported ones. Statements about risks and implementation were coded inductively and grouped into themes following thematic synthesis (Thomas & Harden, 2008). Term frequencies in titles and abstracts were counted descriptively to check the thematic coding; no bibliometric network analysis was performed. Because designs and measures were too heterogeneous for pooling, no meta-analysis was attempted.

## 4. RESULTS AND DISCUSSION

### 4.1. Overview of the included studies

The 25 included studies reported data from about 3,400 participants: at least 1,923 students, 1,145 teachers or lecturers and 330 administrators, with sample sizes not reported in three abstracts (Table 3). Seventeen studies involved students only, five teachers or lecturers, two both students and teachers, and one administrators. Thirteen were set in vocational secondary schools, eleven in post-secondary institutions and one did not state the level. Thirteen institutions were named as study sites, and SMK Model Patriot IV Ciawigebang was the only one that appeared in two studies. Author affiliations could be identified for 12

studies; all were from the same country as the study site, and no cross-border collaboration was observed among them.

**Table 3.** Characteristics of the 25 included empirical studies

| Study | Country | Level | Setting and programme | Participants (n) |
|---|---|---|---|---|
| Apoko (2025) | ID | Sec. | SMKs, East Jakarta; EFL | Teachers (37) |
| Azlina & Noh (2025) | MY | Post-sec. | Politeknik Sultan Abdul Halim Mu'adzam Shah; marketing | Students (250) |
| Choeykhunthot et al. (2026) | TH | Post-sec. | Private vocational colleges, nationwide | Teachers (513) |
| Chooppawa et al. (2026a) | TH | Post-sec. | Private vocational colleges; administration | Administrators (330) |
| Chooppawa et al. (2026b) | TH | Post-sec. | Private vocational education institutions | Instructors (544) |
| Halim & Hairuddin (2026) | ID | Post-sec. | Politeknik Negeri Ujung Pandang; civil engineering | Students (50) |
| Indrayani & Anjani (2026) | ID | Sec. | Two public vocational schools, Way Kanan | Students (168) |
| Janah et al. (2026) | ID | Sec. | SMK YPE Sawunggalih Kutoarjo; EFL writing | Students (32) |
| Kardiyasa et al. (2026) | ID | Sec. | Vocational school, Bali; EFL writing | Students (68) |
| Khor et al. (2026) | MY | Post-sec. | Politeknik Kota Bharu; four departments | Students (344) |
| Lestari (2025) | ID | Sec. | SMK Tri Arga 2 | Students (105) |
| Majid et al. (2026) | ID | Sec. | Vocational high schools, West Java | Teachers (7) |
| Manan et al. (2026) | MY | Post-sec. | Temerloh Community College; English | Students (30) |
| Mariadass @ Manickam et al. (2026) | MY | Post-sec. | Polytechnic Ungku Omar | Students (381) |
| Moeis et al. (2024) | ID | Sec. | SMKS Wahyu Makassar | Students (NR) |
| Pebriani et al. (2025) | ID | Sec. | SMK Model Patriot IV Ciawigebang; English | Students (134) |
| Pranoto et al. (2025) | ID | NR | Five vocational departments incl. nursing and IT; ESP | Students (50) |
| Pratama et al. (2026) | ID | Sec. | SMK Methodist 2 Palembang; English | Students (100), teachers (4) |
| Ratnaningsih (2025) | ID | Post-sec. | Maritime polytechnic; maritime English | Students (69) |
| Sari et al. (2026) | ID | Post-sec. | State polytechnic, Ambon; business administration | Students (70) |
| Subramaniam (2026) | MY | Post-sec. | Malaysian polytechnics, nationwide; ESL | Lecturers (40) |
| Sudarmaji et al. (2026) | ID | Sec. | SMK Pembangunan Jaya Yakapi | Students (41) |
| Yahya et al. (2026) | ID | Sec. | SMKN 1 Polewali; pharmacy | Students (31) |
| Yusuf et al. (2026) | ID | Sec. | Vocational high schools; Islamic religious education | Students, teachers (NR) |
| Yuniarti & Pebriani (2026) | ID | Sec. | SMK Model Patriot IV Ciawigebang; English | Students (NR) |

Note. ID = Indonesia; MY = Malaysia; TH = Thailand; Sec. = vocational secondary; Post-sec. = post-secondary; SMK = vocational upper secondary school; EFL/ESL = English as a foreign/second language; ESP = English for specific purposes; NR = not reported in the abstract.

### 4.2. Geographical and temporal distribution

The evidence is recent and concentrated. No eligible study was published in 2022 or 2023; one appeared in 2024, six in 2025 and eighteen in 2026 (Figure 2a). Indonesia accounted for 68% of the studies, Malaysia for 20% and Thailand for 12%, and no eligible study came from Vietnam, the Philippines, Singapore or the other ASEAN member states (Figure 2b). The concentration in Indonesia mirrors the size of its SMK system and the number of its national journals. The absence of studies from Vietnam and the Philippines, both with large TVET systems, is a clear gap, although the English-language search may have missed studies published in national languages.

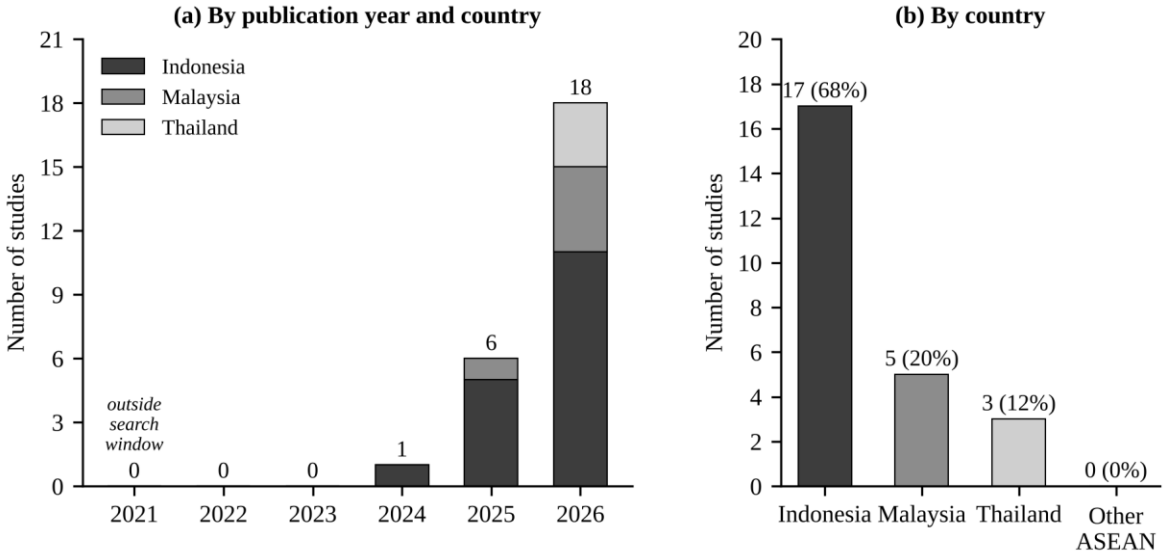



**Figure 2.** Distribution of studies by country and publication year (n = 25)
Note. The search covered 2022 onwards, so 2021 lies outside the search window.

### 4.3. Methodological characteristics

Designs were weighted towards description (Table 4; Figure 3). Ten studies were cross-sectional surveys, four used mixed methods, three were qualitative and one evaluated a training programme. Only four studies, all Indonesian, used a quasi-experimental design with a comparison group; none randomised participants. One used a single-group pre-post design and two used classroom action research. All five Malaysian and all three Thai studies were surveys or mixed-methods studies. Samples were mostly drawn from a single institution by convenience or purposive sampling, and the three Thai studies used partial least squares structural equation modelling on large teacher and administrator samples. Instruments were predominantly researcher-developed questionnaires; only the experimental and action research studies used performance tests or rubrics, and only two reported standardised effect sizes (Indrayani & Anjani, 2026; Yahya et al., 2026). Where reported, interventions lasted four to twelve weeks, and no study reported a delayed post-test.

**Table 4.** Methodological characteristics of the included studies

| Study | Design | Data collection | Analysis |
|---|---|---|---|
| Apoko (2025) | Mixed (sequential explanatory) | Survey; semi-structured interviews | Descriptive statistics; qualitative analysis |
| Azlina & Noh (2025) | Cross-sectional survey | 19-item questionnaire; random sampling | Inferential statistics (SPSS 29) |
| Choeykhunthot et al. (2026) | Cross-sectional survey | Structured questionnaire (extended UTAUT) | PLS-SEM |
| Chooppawa et al. (2026a) | Cross-sectional survey | Questionnaire (UTAUT, trust, privacy, policy fit) | PLS-SEM |
| Chooppawa et al. (2026b) | Cross-sectional survey | Seven-point questionnaire; stratified random sampling | PLS-SEM; PLSpredict |
| Halim & Hairuddin (2026) | Mixed (descriptive survey) | 24-item Likert scale; three open-ended items | Descriptive statistics; thematic analysis |
| Indrayani & Anjani (2026) | Quasi-experimental, non-equivalent groups | Pre/post self-report scales; 8 weeks | ANCOVA |
| Janah et al. (2026) | Classroom action research (2 cycles) | Observation; writing tasks; interviews; reflective journals | Rubric scoring; thematic analysis |
| Kardiyasa et al. (2026) | Quasi-experimental | Writing test; plagiarism-awareness questionnaire | One-way MANOVA |
| Khor et al. (2026) | Cross-sectional survey | Structured questionnaire | Descriptive statistics (SPSS 26) |
| Lestari (2025) | Cross-sectional survey | Online Likert questionnaire | Descriptive percentages |

| Study | Design | Data collection | Analysis |
|---|---|---|---|
| Majid et al. (2026) | Qualitative case study | Interviews; observation; document analysis | Interactive analysis model |
| Manan et al. (2026) | Cross-sectional survey | Questionnaire | Descriptive statistics |
| Mariadass @ Manickam et al. (2026) | Cross-sectional survey | Questionnaire (ABCE framework); convenience sample | Descriptive statistics; correlation |
| Moeis et al. (2024) | Programme evaluation | Participant feedback after training | Descriptive |
| Pebriani et al. (2025) | Cross-sectional survey | 16-item Likert questionnaire; purposive sample | Descriptive statistics |
| Pranoto et al. (2025) | Single-group pre-post | Pre/post tests; usage logs; Likert surveys | Paired t-test; ANOVA; OLS and LASSO |
| Pratama et al. (2026) | Qualitative case study | Interviews; questionnaires; observation | Thematic analysis; triangulation |
| Ratnaningsih (2025) | Classroom action research (2 cycles) | Pre/post speaking tests; questionnaires | Paired t-test |
| Sari et al. (2026) | Quasi-experimental; 12 weeks | 30-item self-efficacy questionnaire | Paired and independent t-tests |
| Subramaniam (2026) | Mixed (exploratory descriptive) | 32-item Likert scale (UTAUT2); open-ended items | Descriptive statistics; thematic analysis |
| Sudarmaji et al. (2026) | Cross-sectional correlational (with SLR) | AI literacy scale; workshop response scale | Spearman correlation |
| Yahya et al. (2026) | Quasi-experimental; 4 weeks | Speaking rubric (Heaton) | Paired and independent t-tests; Cohen's d |
| Yusuf et al. (2026) | Qualitative case study | Interviews; observation; assignment documents | Thematic analysis |
| Yuniarti & Pebriani (2026) | Mixed methods | Questionnaire; open-ended items; interviews | Descriptive and qualitative analysis |

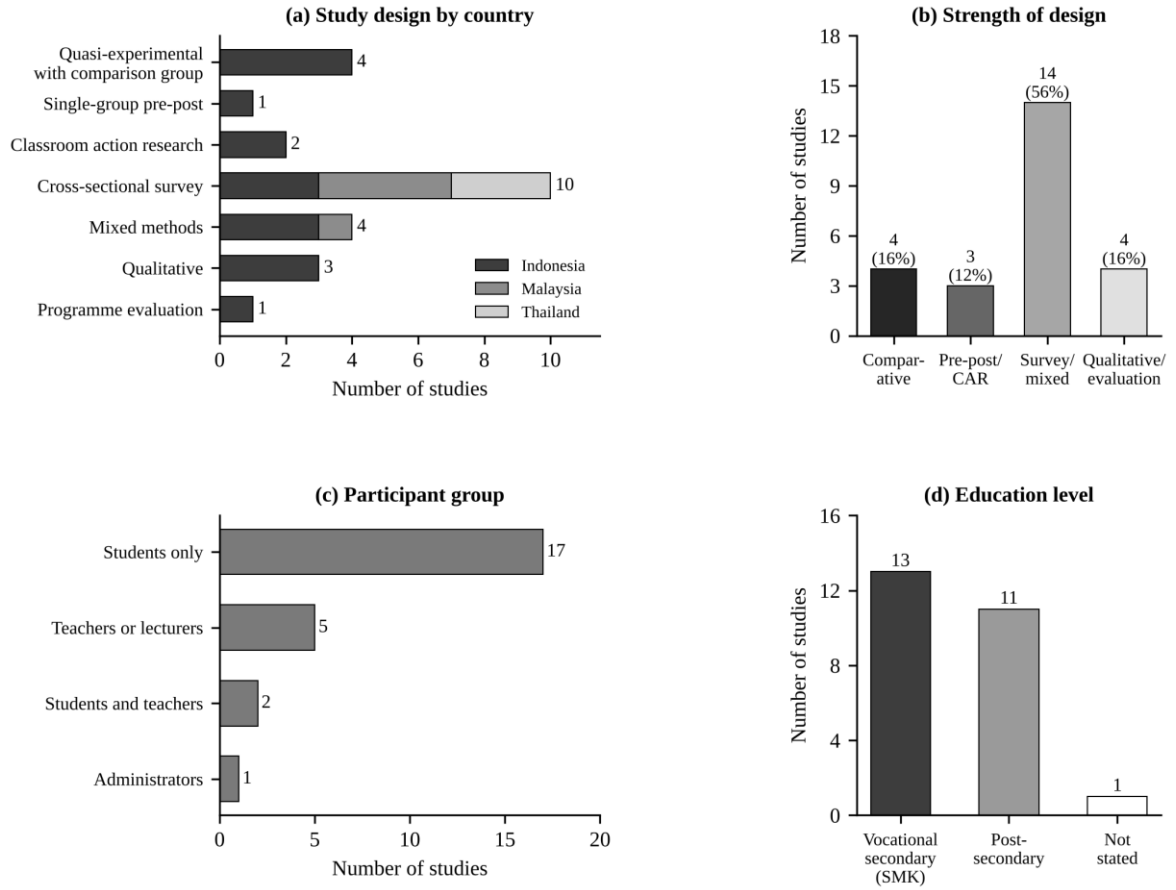



**Figure 3.** Research design and methodological characteristics (n = 25): (a) design by country; (b) strength of design; (c) participant group; (d) education level

### 4.4. GenAI applications in TVET

ChatGPT was named in eleven studies, Gemini in two (one alongside ChatGPT) and Microsoft Copilot in one; five referred to GenAI in general and seven described generative uses of unnamed AI tools or chatbots. Figure 4 maps the applications against user groups. Student-facing applications were dominated by writing support and feedback (eight studies), explanation and tutoring (six), speaking and conversation practice (three) and assignment completion and general study support (four; e.g. Lestari, 2025; Moeis et al., 2024). Teacher-facing applications concerned lesson and material preparation (six studies), with one study reporting teachers' use for programming assistance and media production (Majid et al., 2026) and one examining college administration (Chooppawa et al., 2026a). No study used GenAI for hands-on practice, simulation or practical assessment in a technical programme. Even where the setting was technical, such as pharmacy, maritime studies or civil engineering, GenAI was applied to the English component of the programme (Halim & Hairuddin, 2026; Ratnaningsih, 2025; Yahya et al., 2026).

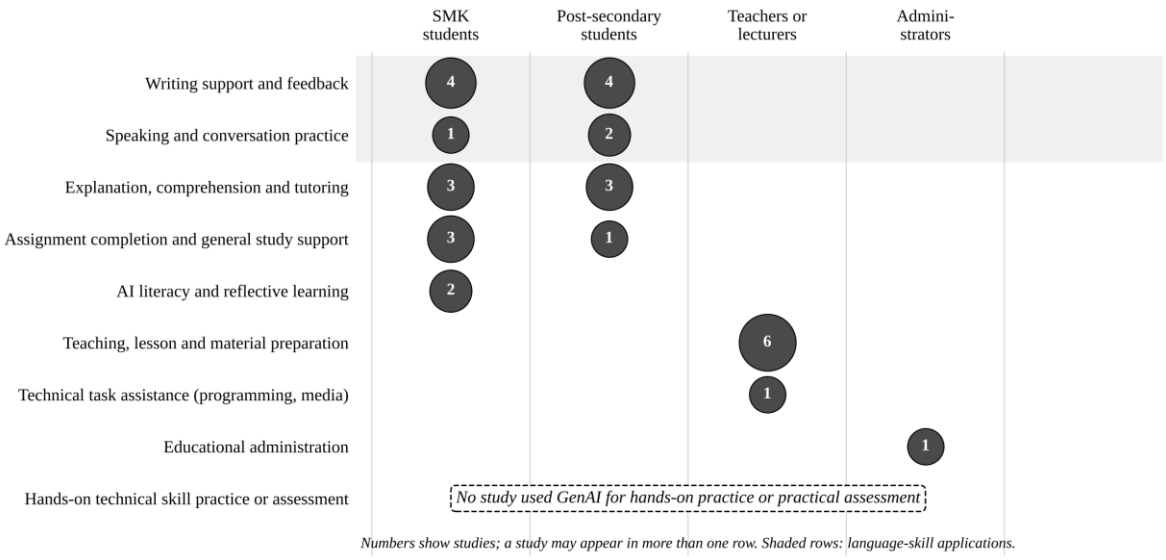



**Figure 4.** Thematic evidence map of GenAI applications in TVET by user group

### 4.5. Learning outcomes

Table 5 and Figure 5 summarise the outcomes by type of evidence. Measured improvements came from seven studies. The four studies with comparison groups all favoured the GenAI condition. In a vocational school in Bali, ChatGPT-assisted process writing combined with explicit integrity guidance improved writing scores (M = 73.76 vs 62.76) and plagiarism awareness relative to conventional instruction (Kardiyasa et al., 2026). SMK pharmacy students who practised patient-counselling role-plays with Gemini outperformed controls in speaking (d = 1.06), with the largest gain in comprehensibility (Yahya et al., 2026). Polytechnic business students receiving ChatGPT-assisted instruction for twelve weeks gained more in business email writing self-efficacy than controls (19.40 vs 6.26), although this outcome was measured by questionnaire (Sari et al., 2026). An eight-week AI literacy and socio-emotional learning intervention improved socio-emotional adjustment and AI literacy (Indrayani & Anjani, 2026). Three single-group studies reported gains in ESP proficiency (Pranoto et al., 2025), speaking (Ratnaningsih, 2025) and surface features of writing (Janah et al., 2026).

**Table 5.** GenAI applications and learning outcomes in the included studies

| Study | Tool | Application | Outcome | Evidence | Reported result |
|---|---|---|---|---|---|
| Apoko (2025) | GenAI | EFL lesson preparation and instruction | Teachers' perceptions | Self-report | Positive; access and training needs |
| Azlina & Noh (2025) | ChatGPT | General study support | Attitudes towards use | Self-report | Usefulness, ease and social influence significant; credibility not |
| Choeykhunthot et al. (2026) | GenAI | Teaching and professional use | Intention; digital competency | Self-report | Trust and risk predicted intention; intention did not predict competency |
| Chooppawa et al. (2026a) | GenAI | College administration | Intention; use | Self-report | Trust, privacy concern and policy fit significant |
| Chooppawa et al. (2026b) | GenAI | Innovative pedagogy | Intention; innovative pedagogy | Self-report | Trust and social influence strongest |
| Halim & Hairuddin (2026) | AI tools | Vocabulary, grammar, ideas, presentations | Perceived communication gains | Self-report | Positive (M = 4.12) with concerns |
| Indrayani & Anjani (2026) | AI-supported | Reflective prompting; AI literacy | Socio-emotional adjustment; AI literacy | Comparison group | Positive (partial $\eta^2$ = .10–.21) |

| Study | Tool | Application | Outcome | Evidence | Reported result |
|---|---|---|---|---|---|
| Janah et al. (2026) | ChatGPT | Feedback in process writing | Writing; argumentation | Pre-post (CAR) | Surface gains; limited argumentation gains |
| Kardiyasa et al. (2026) | ChatGPT | Process writing with integrity guidance | Writing; plagiarism awareness | Comparison group | Positive (73.76 vs 62.76) |
| Khor et al. (2026) | ChatGPT and others | Feedback and clarification | Perceived comprehension | Self-report | Positive; writing highest |
| Lestari (2025) | AI tools | Completing assignments | Frequency of use | Self-report | 52.4% often used AI for assignments |
| Majid et al. (2026) | AI tools | Media production; programming assistance | Teachers' perceptions | Qualitative | Positive; integrity and dependency concerns |
| Manan et al. (2026) | ChatGPT | Scriptwriting | Motivation; engagement | Self-report | Positive; dependency concerns |
| Mariadass @ Manickam et al. (2026) | ChatGPT | Explaining difficult concepts | Problem solving; ethical awareness | Correlational | r = .729 with self-reported problem solving |
| Moeis et al. (2024) | Copilot | Academic task support (training) | Perceived usefulness | Self-report | Training rated useful |
| Pebriani et al. (2025) | ChatGPT and others | Writing correction; explanations | Perceived usefulness; trust | Self-report | 74.7% positive; 39.5% fully trusted output |
| Pranoto et al. (2025) | ChatGPT, Gemini | ESP practice | ESP test scores | Pre-post | Mean gain 24.42; baseline strongest predictor |
| Pratama et al. (2026) | AI chatbots | Reading and writing feedback | Autonomy; engagement | Qualitative | Positive; dependency and integrity concerns |
| Ratnaningsih (2025) | ChatGPT | Speaking practice | Speaking test scores; confidence | Pre-post (CAR) | Significant gains |
| Sari et al. (2026) | ChatGPT | Business email writing | Writing self-efficacy | Comparison group | Larger gain (19.40 vs 6.26) |
| Subramaniam (2026) | ChatGPT | Lesson and material preparation | Lecturers' acceptance | Self-report | High acceptance; integrity concerns |
| Sudarmaji et al. (2026) | GenAI | AI literacy workshop | AI literacy | Correlational | rho approx. .60 with workshop response |
| Yahya et al. (2026) | Gemini | Patient-counselling role-play | Speaking scores | Comparison group | Positive (d = 1.06) |
| Yusuf et al. (2026) | AI tools | Understanding material; composing answers | Integrity practices | Qualitative | Integrity shifted, mediated by teachers |
| Yuniarti & Pebriani (2026) | AI tools | Comprehension, vocabulary, feedback | Perceived benefits and risks | Self-report | Positive with concerns |

The remaining evidence was self-reported or qualitative. Students reported that GenAI helped them understand material, expand vocabulary, correct grammar and generate ideas (Halim & Hairuddin, 2026; Pebriani et al., 2025), with the largest perceived gains in writing (Khor et al., 2026). Self-reported problem solving rose with frequency of ChatGPT use (r = .729; Mariadass @ Manickam et al., 2026), and community college students reported higher motivation and engagement in scriptwriting (Manan et al., 2026). These results agree with the tested ones in direction, but they cannot separate learning from enthusiasm, and correlations between use and self-rated skill may simply reflect that more confident students use the tools more.

Three qualifications apply to all outcome evidence. First, higher-order skills improved less than surface skills: in the CAR study, vocabulary, grammar and organisation improved over two cycles while

argumentation and critical thinking changed little (Janah et al., 2026). Second, baseline competence was the strongest predictor of post-test performance, which suggests that GenAI may amplify existing differences (Pranoto et al., 2025). Third, no included study measured technical or practical competence, the core outcome of TVET (Figure 5).

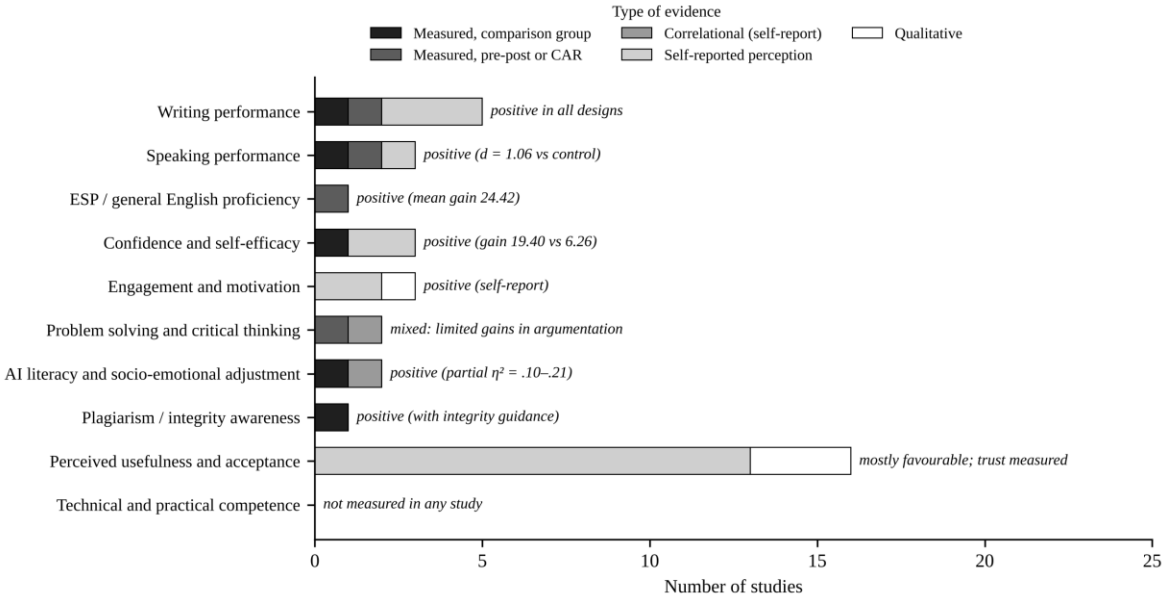



**Figure 5.** Learning outcomes and evidence of effectiveness by outcome domain (n = 25)
Note. A study may contribute to more than one outcome domain.

### 4.6. Ethical challenges and institutional barriers

Thematic coding identified nine themes, which Table 6 and Figure 6 organise at the learner, tool and system levels. At the learner level, academic integrity was the most frequent concern (nine studies). Half of the SMK students in one survey often used AI to complete assignments (Lestari, 2025), teachers described integrity violations as a major obstacle to integration (Majid et al., 2026; Subramaniam, 2026), and students in religious education classes showed an ambivalent mix of efficiency-seeking and honesty norms (Yusuf et al., 2026). Overreliance was reported in seven studies, usually together with concern about weaker independent thinking (Janah et al., 2026; Pratama et al., 2026; Yuniarti & Pebriani, 2026).

At the tool level, doubts about the accuracy of generated output appeared in five studies. Only 39.5% of SMK students fully trusted AI corrections of their writing (Pebriani et al., 2025), yet perceived credibility did not predict polytechnic students' attitudes towards ChatGPT (Azlina & Noh, 2025), which suggests that students keep using tools whose output they do not fully trust. Privacy concerns appeared in four studies, and Thai administrators' intention to use GenAI depended on trust, privacy concern and institutional policy fit rather than on expected performance (Chooppawa et al., 2026a). At the system level, unequal access to devices and connectivity, differences in digital competence, gaps in teachers' prompting skills and the absence of formal guidelines were each reported in four or five studies, and they turned ethical risks into everyday practice problems (Apoko, 2025; Khor et al., 2026; Majid et al., 2026).

**Table 6.** Ethical challenges and implementation barriers

| Theme | Level | n | Illustrative evidence | Studies |
|---|---|---|---|---|
| Academic integrity and plagiarism | Learner | 9 | AI used to complete assignments; ambivalent honesty norms | Kardiyasa et al. (2026); Lestari (2025); Majid et al. (2026); Yusuf et al. (2026) |
| Overreliance and dependency | Learner | 7 | Verbatim reliance on generated responses | Janah et al. (2026); Manan et al. (2026); Pratama et al. (2026) |
| Limited critical thinking | Learner | 5 | Surface gains without argumentation gains | Janah et al. (2026); Subramaniam (2026); Yuniarti & Pebriani (2026) |
| Accuracy and credibility of output | Tool | 5 | 39.5% fully trusted AI corrections | Azlina & Noh (2025); Halim & Hairuddin (2026); Pebriani et al. (2025) |
| Privacy and data governance | Tool | 4 | Privacy concern shaped administrators' intention | Chooppawa et al. (2026a); Halim & Hairuddin (2026); Sudarmaji et al. (2026) |

| Theme | Level | n | Illustrative evidence | Studies |
|---|---|---|---|---|
| Trust and perceived risk | Tool | 3 | Perceived risk positively related to intention | Choeykhunthot et al. (2026); Chooppawa et al. (2026a, 2026b) |
| Unequal access and digital divides | System | 5 | Uneven infrastructure and baseline ability | Apoko (2025); Khor et al. (2026); Pranoto et al. (2025) |
| Prompting skill and teacher readiness | System | 5 | Limited prompt engineering; need for training | Janah et al. (2026); Khor et al. (2026); Majid et al. (2026) |
| Absent or weak guidelines | System | 4 | No formal rules on AI use in assignments | Lestari (2025); Majid et al. (2026); Mariadass @ Manickam et al. (2026) |

Note. n = number of studies; a study may report more than one theme.

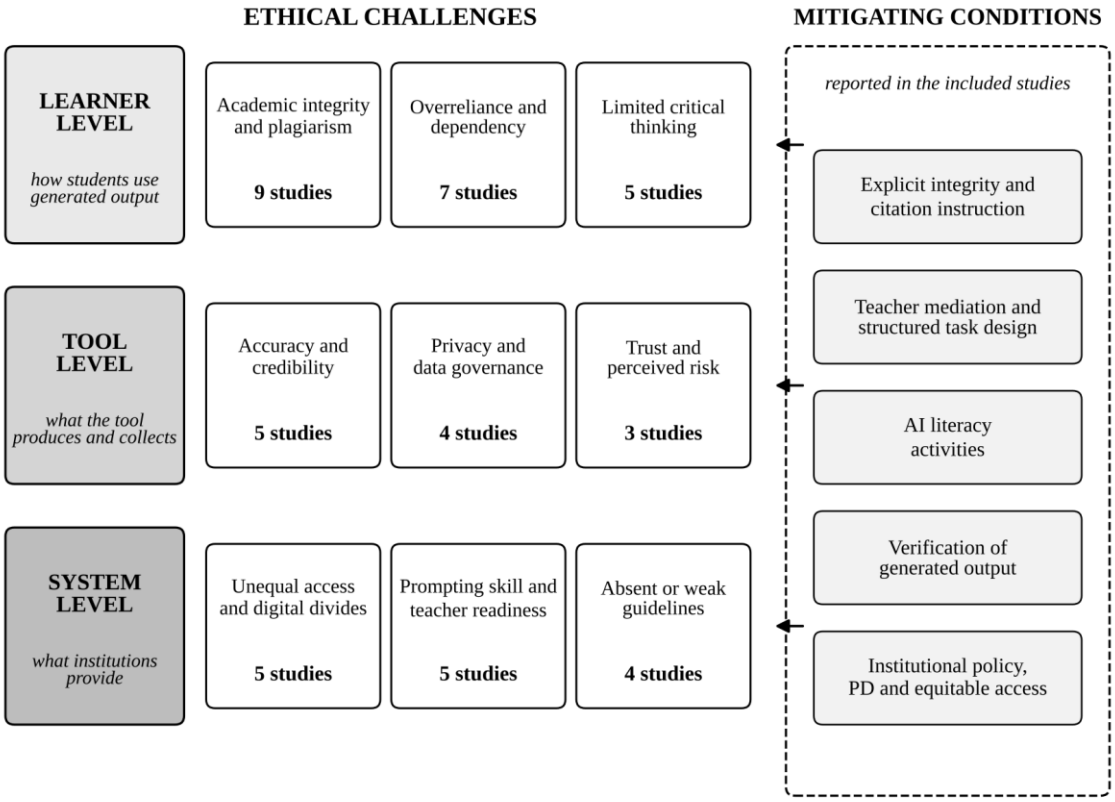



**Figure 6.** Ethical challenges of GenAI integration in TVET and the mitigating conditions reported in the included studies

### 4.7. Implementation factors

The studies converged on four implementation factors. The first was teacher meditation and task design. Benefits appeared when GenAI was built into structured, workplace-relevant tasks, such as patient-counselling role-plays (Yahya et al., 2026) or business email writing (Sari et al., 2026), rather than left to open-ended use. Teachers who positioned AI as a verification tool and kept human judgement central (Majid et al., 2026), or who used authentic assessment and reflective dialogue (Yusuf et al., 2026), reported more responsible student use. The second was explicit integrity instruction. In the Bali study, guidance on paraphrasing, citation and ethical AI use accompanied ChatGPT-assisted writing, and students improved both writing and plagiarism awareness (Kardiyasa et al., 2026), which shows that learning gains and integrity need not be traded against each other.

The third factor was AI literacy. Structured AI literacy activities improved students' AI literacy and perceived support (Indrayani & Anjani, 2026), while a study combining a literature synthesis with an SMK survey concluded that most young people use AI tools regularly but fewer than a quarter can explain how they work or govern user data (Sudarmaji et al., 2026). The fourth was institutional guidance. Clear policies, professional development and equitable access appeared repeatedly as prerequisites for safe adoption (Apoko, 2025; Chooppawa et al., 2026a; Mariadass @ Manickam et al., 2026). The adoption studies added an important nuance: among Thai vocational teachers, trust and effort expectancy predicted intention to use GenAI, perceived AI risk was positively related to intention, and intention did not predict digital competency (Choeykhunthot et al., 2026).

### 4.8. Interpretation and comparison with international research

Taken together, the findings show a positive but narrow evidence base. The direction of the effects agrees with international syntheses: most studies in a cross-level review reported positive effects of GenAI on self-regulated learning and critical thinking (Sardi et al., 2025), and a meta-analysis of ten Scopus-indexed experiments in TVET, vocational and engineering education found a moderate pooled effect of ChatGPT-assisted instruction (Hedges' $g = 0.739$) with substantial heterogeneity ($I^2 = 83\%$) (Hidayat et al., 2026). The regional evidence adds three points. It is dominated by English language outcomes; it is concentrated in national journals that international syntheses are likely to miss; and it shows the same heterogeneity, with gains depending on how the tool was embedded in teaching. The regional studies also confirm the warning that overreliance may limit higher-order learning (Sardi et al., 2025): surface writing features improved more consistently than argumentation.

### 4.9. Language learning versus practical vocational training

The most important finding concerns what has not been studied. Fourteen of the 25 studies concerned English language learning, and no study assessed hands-on technical competence. This matters because language tasks are the domain in which a language model is most likely to help. Writing and speaking are mediated by language, feedback on them can be generated directly, and practice can happen in conversation with the tool. Practical vocational skills differ in kind. They involve perception, manual dexterity, safety judgement and the use of equipment, and they are learned through deliberate practice in authentic conditions and assessed by observing performance (Gulikers et al., 2004; Weigel et al., 2007).

Positive language outcomes therefore cannot be generalised to technical skill acquisition. GenAI may still support practical training indirectly, for example through explanation of procedures, troubleshooting guidance, generation of practice scenarios, documentation and reflection, but each of these mechanisms needs to be tested against performance on practical tasks. The cognitive offloading literature also suggests a specific risk for procedural learning: if learners rely on generated instructions during practice, immediate performance may improve while independent competence does not (Risko & Gilbert, 2016). Competence-based TVET, with its emphasis on authentic performance assessment, provides the right criterion for such research, but no included study used it.

### 4.10. Theoretical implications and integrated framework

Two theoretical implications follow. First, the evidence supports a mediation view of GenAI in TVET: outcomes depended less on the tool than on the conditions of use. The same conditions that produced learning gains, namely teacher mediation, structured tasks, integrity instruction and AI literacy, also mitigated the ethical risks. Second, the adoption studies challenge the usual emphasis of UTAUT on performance expectancy (Venkatesh et al., 2003). In vocational institutions, trust and social influence were the strongest predictors of intention (Chooppawa et al., 2026b), and trust, privacy concern and policy fit mattered while performance expectancy did not (Chooppawa et al., 2026a); teachers who were aware of the risks were more rather than less willing to use GenAI (Choeykhunthot et al., 2026). Acceptance in these settings appears to be governed as much by institutional legitimacy as by individual cost-benefit judgements.

Figure 7 integrates these findings into a framework for responsible GenAI integration in TVET. Inputs (the tool, the user and the setting) act through five implementation conditions: teacher mediation, task design linked to workplace tasks, integrity and citation instruction, AI literacy and prompting skill, and institutional policy, professional development and access. These conditions enable learning outcomes and mitigate ethical challenges, and the challenges that do arise feed back into institutional policy and teacher guidance. The framework refines the initial model in Section 2.5 by specifying the mediating conditions. Its outcome side rests mainly on language outcomes, and its link to technical competence is a hypothesis for testing rather than a finding.

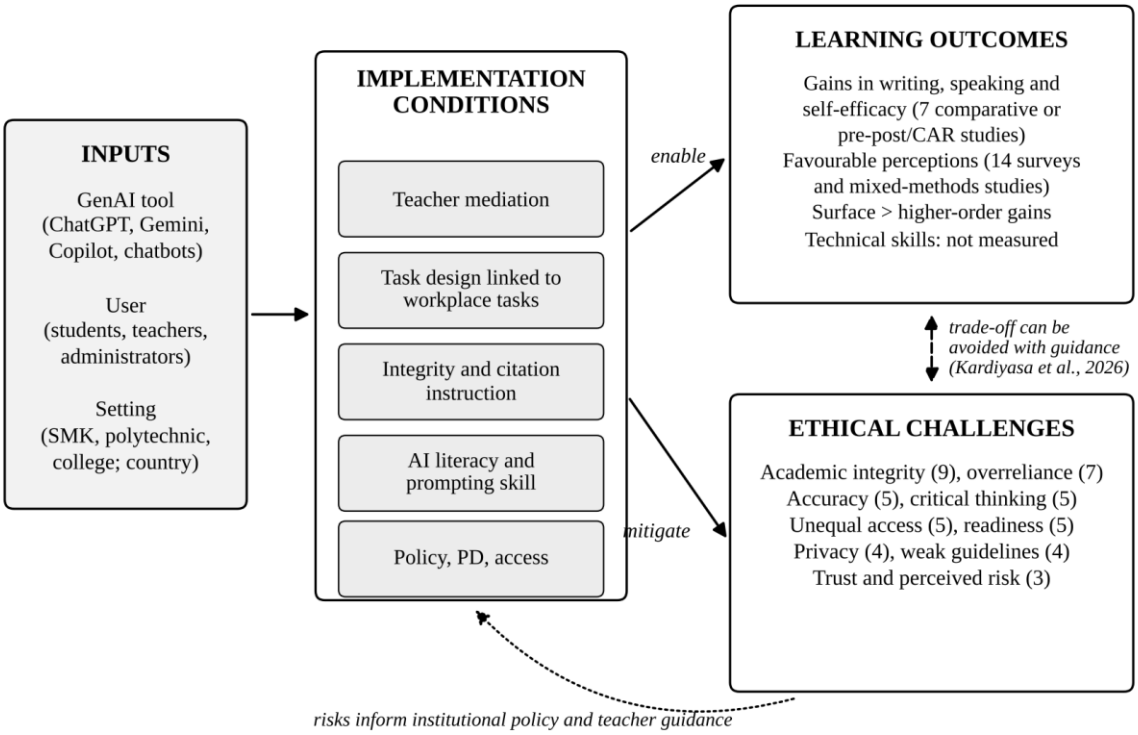



**Figure 7.** Integrated conceptual framework of GenAI in Southeast Asian TVET

### 4.11. Methodological weaknesses in the evidence

The evidence base has five weaknesses. Only 16% of the studies used a comparison group, none randomised participants, and the comparative studies used intact classes with samples of 31 to 168. Most outcomes were self-reported, and several studies measured perceptions or intentions rather than learning. Instruments were largely researcher-developed, and their validity was rarely reported in the abstracts. Interventions were short, and no study tested retention after GenAI support was withdrawn. Finally, publication bias is likely: almost all studies reported favorable results, many were small single-site studies by teachers of their own classes, and null or negative findings may be less likely to reach publication.

### 4.12. Implications for practice and policy

For TVET educators, the evidence supports guided rather than prohibited or unrestricted use. GenAI should be embedded in workplace-relevant tasks, combined with explicit instruction in verification, paraphrasing and citation, and assessed in ways that capture process as well as product, such as drafts, oral defence or observed performance. For institutional policymakers, clear rules on acceptable use, data protection and assessment, together with professional development and equitable access to devices and connectivity, are prerequisites, because governance factors shaped adoption more than expected usefulness. For curriculum developers, AI literacy, including how GenAI works, how to verify its output and how it handles data, should be integrated into vocational curricula alongside occupational standards, and assessment of practical competence should remain independent of AI assistance. For AI developers, the findings point to a need for tools that support vocational practice, such as procedure-aware feedback, transparent sourcing and privacy-protective defaults suited to school settings.

### 4.13. Future research agenda

Table 7 and Figure 8 set out the research agenda. The highest priority is controlled studies of GenAI in technical skill training, such as programming, electrical and mechanical work, hospitality and health, with performance measured by practical assessment. Longitudinal designs with delayed post-tests are needed to test whether gains persist once AI support is withdrawn, which is the core of the overreliance concern. Randomised or matched comparison groups would strengthen causal claims about language and self-efficacy outcomes. Evidence is needed from Vietnam, the Philippines and the other ASEAN member states, including national-language sources. The gap between intention and competence among teachers calls for intervention studies on professional development. Finally, studies should report sample sizes, instrument validity and effect sizes in full, so that future syntheses can pool them.

**Table 7.** Research gaps and future research directions

| Gap | Evidence from this review | Future direction | Suggested design |
|---|---|---|---|
| Technical and practical competence | No study measured hands-on performance | GenAI in programming, electrical, mechanical, hospitality and health training | Controlled trials with practical assessment |
| Durability of learning | Interventions of 4–12 weeks where reported; no delayed tests | Whether gains persist once AI support is withdrawn | Longitudinal studies; delayed post-tests |
| Causal strength | 4 of 25 studies used a comparison group | Stronger evidence on language and self-efficacy outcomes | Randomised or matched comparison groups |
| Geographical coverage | Three countries; English-language records | Vietnam, the Philippines and other ASEAN states | Multi-country studies; national-language sources |
| Teacher competence | Intention did not predict digital competency | Professional development for GenAI and integrity | Intervention studies with teachers |
| Reporting quality | Sample sizes and instrument validity often unreported | Complete reporting of methods and effect sizes | Validated instruments; reporting standards |

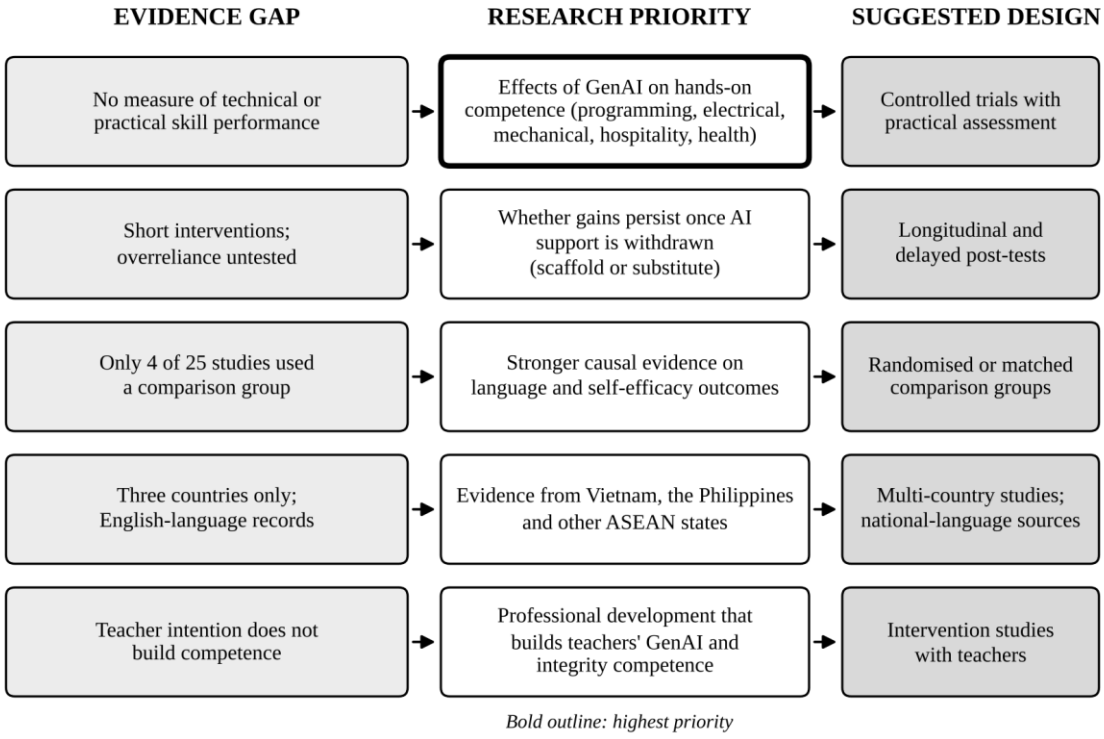



**Figure 8.** Future research agenda for GenAI in Southeast Asian TVET

### 4.14. Limitations of this review

This review has several limitations. Screening and data extraction were carried out by a single reviewer on the basis of titles, abstracts and metadata, because full texts were not retrieved; some study characteristics are therefore missing, and judgements about eligibility could not be checked against full reports. Studies whose setting was not stated in the abstract were excluded, which may have removed eligible regional studies. The search used the OpenAlex and Semantic Scholar indices, retrieved the top-ranked records for each string rather than every match, and covered English-language records only; searches of Scopus, Web of Science and national databases such as SINTA and MyCite with the strings proposed in Table 1 may identify more studies. No formal risk-of-bias tool was applied beyond coding by design. Finally, literature is growing quickly, and the findings describe the evidence available in October 2026.

## 5. CONCLUSION

This review identified 25 empirical studies on GenAI in Southeast Asian TVET, almost all published in 2025 and 2026 and concentrated in Indonesia, Malaysia and Thailand. GenAI was applied mainly to writing

support, explanation, speaking practice and teachers' lesson preparation. The strongest evidence came from four quasi-experimental studies that reported gains in writing, speaking, writing self-efficacy and AI literacy; most other evidence was self-reported, and no study measured hands-on technical competence. The most frequent ethical challenges were academic integrity, overreliance and inaccurate output, followed by privacy and unequal access. Benefits and risk mitigation depended on teacher mediation, task design, AI literacy and institutional guidance, and one controlled study showed that integrity instruction and learning gains can be achieved together. The review is limited by abstract-based screening by a single reviewer and by the databases searched. Its contribution is a region-specific evidence map, an integrated framework for responsible GenAI integration and a prioritized research agenda. The most urgent need is for controlled, longitudinal studies of GenAI in practical skill training, assessed through authentic performance, across a wider range of Southeast Asian countries.

## 6. ACKNOWLEDGEMENTS

The authors have no acknowledgements to declare.

## 7. DECLARATIONS

Funding: This research received no specific grant from any funding agency in the public, commercial or non-profit sectors.
Conflict of interest: The authors declare no conflict of interest.
Author contributions: Vion Age Tricahyo: Conceptualization, methodology, investigation, data curation, formal analysis, visualization, writing – original draft. Haeriani H: Validation, writing – review and editing. Husni Mubarak: Supervision, validation, writing – review and editing. All authors reviewed and approved the final manuscript.
Data availability: The search strings are reported in Table 1, and the extracted data and thematic synthesis are presented in Tables 3-6. The list of retrieved records and screening decisions is available from the corresponding author upon reasonable request.
Ethical approval: Not applicable. This review analyzed published studies and did not involve the recruitment of human participants or the collection of new data from human participants.
Use of generative AI: ChatGPT (OpenAI) was used to assist with drafting and editing this declarations section. The authors reviewed and edited the generated text and took responsibility for the final wording. Any other use of generative AI in preparing this manuscript should also be disclosed here**.**

## 8. REFERENCES

Apoko, T. W. (2025). Generative artificial intelligence in English instruction: Indonesian EFL vocational high school teachers' perspectives. *Journal of Languages and Language Teaching*, *13*(3), 1432. https://doi.org/10.33394/jollt.v13i3.14190

Azlina, H., & Noh, N. S. (2025). Factors influencing the attitudes towards using ChatGPT for learning among marketing students at Politeknik Sultan Abdul Halim Mu'adzam Shah. *Journal of STEM and Education*, [volume(issue), pages]. https://openalex.org/W7219950743

Choeykhunthot, A., Limpinan, P., & Jantakoon, T. (2026). Extending UTAUT for generative AI adoption among vocational teachers: The roles of trust, AI literacy, and risk awareness. *Higher Education Studies*, *16*(3), 565. https://doi.org/10.5539/hes.v16n3p565

Chooppawa, P., Limpinan, P., & Jantakoon, T. (2026a). Intention to use generative AI for vocational college administration. *World Journal of Education*, *16*(3), 157. https://doi.org/10.5430/wje.v16n3p157

Chooppawa, P., Limpinan, P., & Jantakoon, T. (2026b). Acceptance of generative AI for supporting innovative learning in vocational education. *World Journal of Education*, *16*(3), 30. https://doi.org/10.5430/wje.v16n3p30

Dinata, C., Milansari, I. L., Triyono, M. B., & Pratama, G. N. I. P. (2025). The potential of artificial intelligence in vocational education research and development: A bibliometric study. *Jurnal Pendidikan Vokasi*, *15*(1), 59–74. https://doi.org/10.21831/jpv.v15i1.77873

Gulikers, J. T. M., Bastiaens, T. J., & Kirschner, P. A. (2004). A five-dimensional framework for authentic assessment. *Educational Technology Research and Development*, *52*(3), 67–86. https://doi.org/10.1007/BF02504676

Halim, S., & Hairuddin, N. H. (2026). Vocational higher education students' perceptions on the use of artificial intelligence tools in learning English communication skills. *IDEAS: Journal on English Language Teaching and Learning, Linguistics and Literature*, [volume(issue), pages]. https://openalex.org/W7221274276

Hidayat, H., Muji, A. P., & Dewi, F. K. (2026). Estimating the effect of ChatGPT-assisted instruction on learning outcomes in TVET, vocational education, and engineering education: A meta-analysis. *Journal*

*of Information Technology Education: Research*, *25*, 35. https://doi.org/10.28945/5877

Indrayani, S., & Anjani, R. N. (2026). Enhancing vocational students' socio-emotional adjustment through artificial intelligence supported learning. *AI and Developmental Insights in Education*, *2*(1), 144–158. https://doi.org/10.58524/aidie.v2i1.157

Janah, R., Tusino, T., & Dewi, P. (2026). Exploring the role of ChatGPT in developing writing skills and critical thinking among vocational high school students. *AL-ISHLAH: Jurnal Pendidikan*, *18*(1). https://doi.org/10.35445/alishlah.v18i1.8580

Kardiyasa, I. G., Santosa, M. H., & Kusuma, I. P. I. (2026). The effect of ChatGPT on vocational students' writing and plagiarism awareness. *Jurnal Ilmiah Global Education*, *7*(3), 2330–2334. https://doi.org/10.55681/jige.v7i3.6900

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

Khor, M. S. C., Che' Lah, C. F., & Hamzah, R. H. (2026). AI usage patterns and effects on English language comprehension among students at Politeknik Kota Bharu. *International Journal on e-Learning and Higher Education*, *21*(1), 1–20. https://doi.org/10.24191/ijelhe.v21n1.2111

Lestari, T. P. (2025). Analyzing the impact of artificial intelligence on student learning: A case study of SMK Tri Arga 2. *Journal of Applied Informatics and Computing*, *9*(1), 221–226. https://doi.org/10.30871/jaic.v9i1.8902

Majid, N. W. A., Barri, M. A. F., Sari, A. P., Sodikin, R. A., Azman, M. N. A., & Prestoza, M. J. R. (2026). Vocational teachers' insights on artificial intelligence and computational thinking integration. *Jurnal Penelitian dan Pengkajian Ilmu Pendidikan: e-Saintika*, *10*(1), 1–15. https://doi.org/10.36312/e-saintika.v10i1.3660

Manan, N. A., Radzuan, N. R. M., & Masrom, U. K. (2026). Understanding English writing motivation among TVET learners: The role of ChatGPT in scriptwriting task. In *Atlantis Advances in Applied Sciences* (pp. 127–140). Atlantis Press. https://doi.org/10.2991/978-94-6239-745-3_10

Mariadass @ Manickam, J., Chyng, L. Y., & Gobal, N. R. (2026). Exploring ChatGPT utilization and ethical awareness in TVET education through the ABCE framework. *International Journal of Software Engineering and Computer Systems*, *12*(1), 73–84. https://doi.org/10.15282/ijsecs.12.1.2026.7.0153

Moeis, D., Usman, N., Faisal, M., Harmin, A., Mulyadi, I., & Thamrin, M. (2024). Microsoft Copilot training for monitoring student learning: A case study vocational high school Makassar - Indonesia. *I-Com: Indonesian Community Journal*, *4*(3), 1911–1922. https://doi.org/10.33379/icom.v4i3.5134

Mubarok, H., Rostika, Y., Komaro, M., & Saripudin. (2026). Defining, operationalizing, and measuring generative AI competence among vocational teachers: A scoping review. *International Journal of Educational Practice and Policy*, *4*(3), 37–55. https://doi.org/10.66314/ijepp.v4i3.1182

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

Pebriani, S. R., Suryana, Y., & Yuniarti. (2025). Students' perceptions on the use of artificial intelligence (AI) as English learning tools at SMK Model Patriot IV Ciawigebang. *Darussalam English Journal (DEJ)*, *5*(2), 105–119. https://doi.org/10.30739/dej.v5i2.4008

Pranoto, D. C. Y., Rufi'i, R., Sabariah, S., & Bandono, A. (2025). Harnessing generative AI for ESP: A cross-disciplinary vocational education framework with predictive modeling evidence from Indonesia. *International Journal of Artificial Intelligence Research*, *9*(1.1). https://doi.org/10.29099/ijair.v9i1.1.1569

Pratama, V., Tahrun, T., & Yukaman, H. (2026). Students' and teachers' perspectives on the use of AI chatbot in English language teaching at the vocational high school level in Methodist 2 Palembang. *Esteem Journal of English Education Study Programme*, *9*(1), 930–943. https://doi.org/10.31851/b0j1dv33

Ratnaningsih, D. (2025). Enhancing students' speaking skills using ChatGPT application in ESP maritime English learning. *JEES (Journal of English Educators Society)*, *10*(2), 186–197. https://doi.org/10.21070/jees.v10i2.1964

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

Samala, A. D., Zhai, X., Aoki, K., Bojic, L., & Zikic, S. (2024). An in-depth review of ChatGPT's pros and cons for learning and teaching in education. *International Journal of Interactive Mobile Technologies (iJIM)*, *18*(2), 96–117. https://doi.org/10.3991/ijim.v18i02.46509

Sardi, J., Darmansyah, Candra, O., Yuliana, D. F., Habibullah, Yanto, D. T. P., & Eliza, F. (2025). How generative AI influences students' self-regulated learning and critical thinking skills? A systematic

review. *International Journal of Engineering Pedagogy (iJEP)*, *15*(1), 94–108. https://doi.org/10.3991/ijep.v15i1.53379

Sari, H. F. A., Pattipeilohy, V., & Putra, H. (2026). Can ChatGPT strengthen students' confidence in business email writing? A quantitative study of EFL learners' self-efficacy. *IDEAS: Journal on English Language Teaching and Learning, Linguistics and Literature*, [volume(issue), pages]. https://openalex.org/W7221275193

Subramaniam, R. (2026). Lecturers' perceptions of ChatGPT use in Malaysian polytechnic ESL classrooms: An exploratory descriptive study. *International Journal of Humanities Philosophy and Language*, *9*(34), 32. https://doi.org/10.35631/ijhpl.934003

Sudarmaji, E., Herlan, Yatim, M. R., Yunas, S., Prakoso, R., Kaniati, R., Ramadh, O. T., & Juwanto, S. N. (2026). AI literacy and cognitive evolution in Generation Z: A synthesis of empirical evidence and workshop-based insights. *Journal of Social Science and Human Research Studies*, *2*(4). https://doi.org/10.65150/ep-jsshrs/v2e4/2026-05

Thomas, J., & Harden, A. (2008). Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Medical Research Methodology*, *8*(1), 45. https://doi.org/10.1186/1471-2288-8-45

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, *27*(3), 425–478. https://doi.org/10.2307/30036540

Weigel, T., Mulder, M., & Collins, K. (2007). The concept of competence in the development of vocational education and training in selected EU member states. *Journal of Vocational Education & Training*, *59*(1), 53–66. https://doi.org/10.1080/13636820601145549

Yahya, S. M., Munir, & Ariyani, A. (2026). Improving SMK students' speaking skills with AI-based learning in vocational contexts: A quasi-experimental study. *IDEAS: Journal on English Language Teaching and Learning, Linguistics and Literature*, [volume(issue), pages]. https://openalex.org/W7221211561

Yuniarti, Y., & Pebriani, S. R. (2026). Vocational high school students' perceptions of using AI for learning English: Benefits and challenges. *Journal of English Educational Study (JEES)*, *9*(1), 36–45. https://doi.org/10.31932/jees.v9i1.5448

Yusuf, W. F., Imelda, N. F., & Kholiq, A. (2026). Integration of artificial intelligence in Islamic religious education learning in vocational schools. *MUMTAZ: Jurnal Pendidikan Agama Islam*, *6*(1), 173–195. https://doi.org/10.69552/mumtaz.v6i1.3753