



ESTIMATING THE EFFECT OF CHATGPT-ASSISTED INSTRUCTION ON LEARNING OUTCOMES IN TVET, VOCATIONAL EDUCATION, AND ENGINEERING EDUCATION: A META-ANALYSIS

Hendra Hidayat	Universitas Negeri Padang, Padang, Indonesia	hendra.hidayat@ft.unp.ac.id
Anggarda Paramita Muji	Universitas Negeri Padang, Padang, Indonesia	anggarda_pm@unp.ac.id
Fitrika Kumala Dewi*	Universitas Negeri Padang, Padang, Indonesia	fitrikakumala@student.unp.ac.id

* Corresponding author

ABSTRACT

Aim/Purpose	The aim of this study is to synthesize empirical evidence from Scopus-indexed studies and estimate the pooled effect of ChatGPT-assisted instruction on learning outcomes across the included TVET, vocational education, and engineering education contexts.
Background	The rapid proliferation of ChatGPT has attracted growing interest in its potential to facilitate learning in TVET, vocational education, and engineering education. However, empirical findings remain inconsistent. Some studies report large learning gains, and others indicate small or weak effects.
Methodology	We conducted a PRISMA-guided screening of Scopus-indexed, English-language, open-access publications from 2022 to 2026. Ten experimental or quasi-experimental studies with sufficient statistical information were used for effect-size calculation. We estimated the pooled effect size in JASP using a random-effects model. Complementary narrative and keyword analyses were also performed to contextualize the statistical findings by identifying the assessed learning domains, research keywords, opportunities, concerns, and future implementation strategies.

Accepting Editor Stamatis Papadakis | Received: June 6, 2026 | Revised: July 27, August 14, August 28, 2026 | Accepted: August 31, 2026.

Cite as: 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, Article 35. <https://doi.org/10.28945/5877>

(CC BY-NC 4.0) This article is licensed to you under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/). When you copy and redistribute this paper in full or in part, you need to provide proper attribution to it to ensure that others can later locate this work (and to ensure that others do not accuse you of plagiarism). You may (and we encourage you to) adapt, remix, transform, and build upon the material for any non-commercial purposes. This license does not permit you to use this material for commercial purposes.

Contribution	The specific contribution of this study lies in combining a quantitative estimate of the average effect with contextual evidence concerning implementation across the educational settings represented by the included studies.
Findings	Across the 10 studies in this bounded corpus, ChatGPT-assisted instruction showed a statistically significant, moderate positive pooled effect across the cognitive, affective, and psychomotor outcomes included in the primary analysis (Hedges' $g = 0.739$, 95% CI = 0.469–1.010, $p < .001$). Residual heterogeneity was substantial ($I^2 = 83.12\%$), indicating considerable variation across studies. The positive direction and moderate magnitude remained stable across the assumed within-study correlations and in the analysis restricted to directly assessed outcomes.
Recommendations for Practitioners	The findings support cautious use of ChatGPT as a supplementary learning tool. The meta-analysis did not establish any instructional condition or implementation strategy as universally effective.
Recommendations for Researchers	Researchers should extend primary evidence to underrepresented educational contexts and report complete statistical and methodological information required for accurate effect-size calculation.
Impact on Society	This study synthesizes evidence on ChatGPT-assisted learning outcomes across the included educational contexts, supporting more informed decisions about its responsible use without implying that the findings represent all TVET, vocational education, or engineering education settings.
Future Research	Future research needs to broaden database coverage, include more types of documents when appropriate, and include non-English and closed-access publications when available. It should also explore the moderating effects of instructional design, prompt training, teacher guidance, assessment strategies, and students' prior knowledge on ChatGPT-assisted learning.
Keywords	ChatGPT, artificial intelligence (AI), learning outcomes, TVET, vocational education, engineering education

INTRODUCTION

Sustainable Development Goal (SDG) 4 emphasizes inclusive and equitable access to quality education and learning opportunities (Abera, 2023; Amoros Molina et al., 2023). To uphold the principle of leaving no one behind (Mensah et al., 2022), education systems need to implement strategies to improve access, enhance the quality of learning, and equip learners with the skills to meet changing social and technological demands. In this context, educational technology is a strategic part of teaching, learning, and skill development, not just an adjunct to instruction.

The role of technology in education has shifted from merely transmitting knowledge to facilitating more interactive, student-centered, and flexible learning environments (Tang, 2023; Yuliana & Hidayat, 2017). Teachers were usually seen as the main source of knowledge and students as passive recipients. Current approaches need to give learners more ownership, enable them to work in teams, encourage independent research, and help them use digital resources effectively. The change is more relevant to Technical and Vocational Education and Training (TVET), where students should be able to apply knowledge to solve real-life technical and vocational problems rather than just understand the principles.

Recent developments in generative artificial intelligence, such as ChatGPT, have energized the debate about the role of AI in education. ChatGPT is a large conversational language model that can generate explanations, feedback, examples, and responses through natural language interaction. Its adoption has led educational stakeholders to consider pedagogical opportunities as well as regulatory issues. Several countries and international organizations have started to formulate policy frameworks and ethical principles for AI in education. Examples include Australia's framework for generative AI in schools (Bower et al., 2025) and Singapore's AIED Ethics Framework (Kundu & Bej, 2025; Tuomi, 2023). Indonesia issued the Ministerial Decree of Communication and Digital Affairs No. 117 of 2026, which provides guidelines for the use and learning of digital technology and artificial intelligence in formal, nonformal, and informal education. General ethical principles for the use of AI were previously established in Circular Letter No. 9 of 2023.

Previous review-level studies have examined ChatGPT in education from different perspectives. These reviews vary in their educational scope, including study designs and outcomes, search coverage, and use of meta-analysis. Table 1 compares the most relevant systematic reviews and meta-analyses to clarify the specific position of the present study.

Table 1. Comparison of previous reviews

Study	Scope	Populations, designs, and outcomes	Search period, databases, and pooled effect	Main difference
(İpek et al., 2023)	Various educational fields	40 articles and theses; populations and designs were not systematically classified; learning outcomes were not quantitatively synthesized	Google Scholar, ScienceDirect, ERIC, Wiley, SpringerLink, Sage, Taylor & Francis, MDPI, and JSTOR; December 2022–February 15, 2023; no pooled effect	Broad review of ChatGPT applications and concerns, without pooled learning outcomes or a vocational and engineering focus
(Heung & Chiu, 2025)	K–12 and higher education across disciplines	17 intervention studies involving 1,735 students; 11 non-randomized and 6 randomized studies; measured behavioral, cognitive, and emotional engagement	Web of Science, Scopus, ProQuest, and ERIC, 2021–August 24, 2024; pooled effects calculated	Focused on engagement rather than domain-specific learning outcomes
(Deng et al., 2025)	Multiple educational levels and disciplines	69 experimental or quasi-experimental studies; 62 were meta-analyzed for academic, affective, cognitive, and psychological outcomes	EBSCOhost, IEEE Xplore, PsycINFO, Scopus, and Web of Science; December 2022–August 31, 2024; pooled effects calculated	Provided a broad cross-disciplinary synthesis rather than a focused vocational and engineering analysis
(Baig & Yadegaridehkordi, 2024)	Higher education	57 articles involving students and other higher-education stakeholders; 28 quantitative, 10 qualitative, 2 mixed-methods, and 17 method-unspecified studies; learning outcomes were not required	Emerald, ERIC, MDPI, SAGE, Elsevier, SpringerLink, Frontiers, PLOS ONE, Wiley, Taylor & Francis, and Google Scholar; January 2023–mid-January 2024; no pooled effect	Examined adoption, applications, benefits, and challenges without pooling comparative learning outcomes

Table 1 summarizes the main focus of the previous reviews of ChatGPT in education. The reviews have mainly focused on teaching practices, student engagement, learning outcomes, adoption, benefits, and challenges. However, the scope of the reviews was generally broad. Most of the included studies belonged to different disciplines or examined higher education without a specific disciplinary

focus. TVET, vocational education, and engineering education are distinct but related areas. For this review, TVET is a broad umbrella term for education and skills development in occupational fields. Vocational education constitutes a central component of this framework. Engineering education is treated as a distinct yet overlapping domain because it develops applied technical and professional competencies; however, it is not automatically classified as TVET. Accordingly, the review retains all three terms as its predefined substantive scope while avoiding their interchangeable use.

Within this defined scope, previous reviews have given limited separate attention to TVET, vocational education, and engineering education. Deng et al. (2025) investigated the effects of ChatGPT-assisted learning on academic and psychological outcomes, synthesizing evidence from technical and non-technical disciplines. Therefore, the effectiveness and implementation of ChatGPT-assisted instruction in these educational domains have not been explored adequately. To fill this domain-specific gap, this study synthesizes the experimental or quasi-experimental evidence on the effects of ChatGPT-assisted instruction on academic learning outcomes in TVET, vocational education, and engineering education. In addition to estimating the pooled effect, it explores the learning domains, implementation opportunities, concerns, and future strategies reported in the included studies. The specific contribution of this study lies in combining quantitative evidence of instructional effectiveness with contextual evidence concerning implementation across the educational contexts represented by the included studies.

LITERATURE REVIEW

CHATGPT IN TVET, VOCATIONAL EDUCATION, AND ENGINEERING EDUCATION

Chat Generative Pre-trained Transformer (ChatGPT) has attracted worldwide attention since its public release in late 2022 because of its ability to generate human-like responses across a wide range of topics and languages (Athanassopoulos et al., 2023; Jo, 2024). It has a conversational interface that allows users to ask for explanations, examples, summaries, feedback, and help solving problems using natural-language prompts. The rapid adoption of generative AI is evident in public and educational discourse surrounding the model's release; for example, it gained more than 100 million users within two months of its release (Ooi et al., 2025; Teubner et al., 2023). The technical performance is based on natural language processing and large language modeling, which allow the system to generate context-sensitive responses to user input (Le et al., 2024).

ChatGPT can be relevant across TVET, vocational education, and engineering education in applied fields such as computer engineering, industrial engineering, automotive engineering, healthcare, and tourism (Basnet & Pyakurel, 2024; Fadillah et al., 2023; Lukhele & Laseinde, 2024; Poschauko et al., 2024; Verian, 2024; Younis, 2024). Students may use ChatGPT to clarify technical concepts, obtain examples and feedback, test their understanding, and support programming tasks such as debugging. Whether these uses improve learning remains an empirical question addressed by the present review.

TVET, vocational education, and engineering education share an applied orientation that provides a meaningful context for integrating ChatGPT. Learners in these settings are often expected to connect conceptual knowledge with practical activities such as troubleshooting, experimentation, debugging, design, and problem-solving (Hadi, 2021; Yuliana & Hidayat, 2017). ChatGPT has been used to teach programming languages such as Python, Java, and C++ by explaining syntax, identifying errors, providing code examples, and supporting learning (da Silva et al., 2024). Some coding tasks have been reported to achieve high accuracy (M. Li & Krishnamachari, 2024), but these results should be interpreted cautiously because performance depends on task complexity, prompt quality, programming language, and learner expertise. Some similar applications have been reported in the fields of electrical and electronic engineering, automotive education, healthcare, and tourism, where ChatGPT can help in explanations, documentation, simulation of scenarios, and generation of digital content

(Chen & Chen, 2025; Horne, 2025; Kernberg et al., 2024; Murtaza et al., 2025). These examples signal broad potential, yet also underscore the need for empirical synthesis to determine whether ChatGPT consistently improves learning outcomes across these educational contexts.

THEORETICAL PERSPECTIVES

Constructivism, experiential learning, and self-regulated learning offer interpretive perspectives on ChatGPT-assisted learning in applied educational contexts. Constructivism emphasizes knowledge construction through interaction, prior experience, and reflection (Horton, 2001; Mishra, 2023; Yildiz, 2025). Experiential learning emphasizes the cycle of experience, reflection, conceptualization, and experimentation (Kolb, 2014), whereas self-regulated learning emphasizes learners' planning, monitoring, and evaluation of their own learning (Bandura, 1991; Zimmerman, 1990). In this review, these perspectives were not used to formulate hypotheses or research questions, guide the coding of the included studies, or analyze variation in effect sizes. Accordingly, they are presented as interpretive perspectives rather than as an analytical framework. Because the included interventions did not consistently measure these processes, no claim is made that the corresponding mechanisms produced the observed pooled effect.

METHOD

RESEARCH QUESTIONS

Against this background, the review examines the available evidence through four research questions addressing the pooled effect, assessed outcomes, keyword patterns, and reported implementation opportunities, concerns, and future strategies.

- RQ1:** What is the pooled effect of ChatGPT-assisted instruction on learning outcomes across the included educational contexts?
- RQ2:** What aspects are targeted for assessment in the implementation of ChatGPT?
- RQ3:** What is the distribution pattern of keywords in publications on ChatGPT interventions in TVET, vocational education, and engineering education?
- RQ4:** What are the opportunities, concerns, and future strategies for implementing ChatGPT?

ARTICLE SELECTION PROCESS

All the records were extracted from the Scopus database via the Scopus platform. Scopus was selected because its multidisciplinary coverage aligned with the intersection of education, vocational education, engineering education, and educational technology examined in this review. It also provided a consistent source of bibliographic metadata for the meta-analysis and VOSviewer keyword analysis. Accordingly, the review was designed as a bounded Scopus-based evidence synthesis rather than an exhaustive synthesis of all studies indexed across multiple databases.

The search was conducted on April 8, 2026, through the Scopus platform using the Advanced Search Function in the article title, abstract, and keyword fields: TITLE-ABS-KEY ("ChatGPT" AND ("TVET" OR "vocational education" OR "engineering education")). At the database-search stage, the Scopus document-type filter was applied to retain records classified as Article or Conference Paper. No subject-area restriction was applied. All subsequent eligibility criteria were assessed manually during screening in the sequence shown in the PRISMA flow diagram (Hidayat et al., 2024) (Figure 1). First, publications were limited to 2022–2026. Second, only original experimental or quasi-experimental studies were included. Third, publications had to be written in English. Fourth, eligibility was restricted to open-access publications as an access-based condition defining the evidence corpus, rather than as an indicator of methodological quality or reporting completeness. Finally, domain eligibility was determined from the participants' programs and instructional contexts. Studies were classified as TVET or vocational education if they focused on occupation-oriented competencies and as

engineering education if they involved an engineering program or course. Studies involving programming, computer science, educational technology, tourism, or general higher education were included only when they explicitly met either definition; subject labels or technology use alone were insufficient. Eligible learning outcomes were operationalized as a broad, multidomain category encompassing cognitive, affective, and psychomotor outcomes. The outcomes within these domains were not assumed to represent a single homogeneous construct.

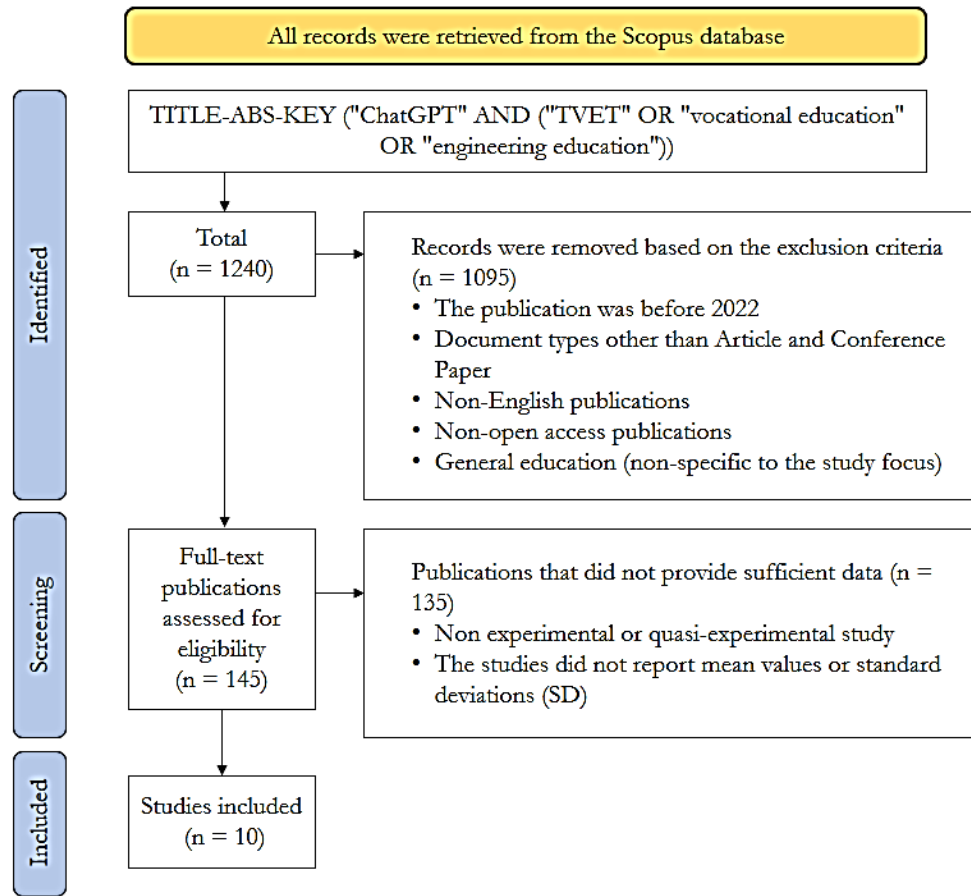


Figure 1. PRISMA flow diagram

All stages of screening were conducted solely by the third author, who repeated the screening and re-checked each eligibility decision against the predefined criteria. Data extraction was conducted by the third author, and the first and third authors subsequently rechecked all extracted values against the source publications.

QUALITY ASSESSMENT

The reported methodological characteristics of the included studies were appraised using the revised JBI Critical Appraisal Checklist for Randomized Controlled Trials and the JBI Critical Appraisal Checklist for Quasi-Experimental Studies (Barker et al., 2023, 2024). Checklist selection was based on the reported study design, unit of allocation, and allocation procedure. Accordingly, Studies 1, 4, 5, and 8 were assessed using the randomized controlled trial checklist, whereas Studies 2, 3, 6, 7, 9, and 10 were assessed using the quasi-experimental checklist. Item-level judgments were recorded for each applicable JBI criterion. The first and third authors independently conducted the appraisals, and disagreements were resolved through discussion and consensus. Table 2 summarizes the methodological characteristics and the JBI checklist applied to each study.

Table 2. Reported methodological characteristics underlying the JBI appraisal

Study	Reported design and allocation	Compared conditions	Reported outcome measurement	Reported analysis	JBI appraisal tool applied
1. (Yilmaz & Karaoglan Yilmaz, 2023)	Pretest-posttest experiment; random allocation reported	ChatGPT-assisted programming (n = 21) vs. no ChatGPT (n = 24)	Computational thinking, programming self-efficacy, and learning-motivation scales	ANCOVA	RCT
2. (H. F. Li, 2023)	Two-class quasi-experiment; one intact class was reportedly randomly assigned to the experimental condition, but the allocation procedure was not described	ChatGPT-FLGA (n = 42) vs. traditional FLGA (n = 39)	Project rubric, self-efficacy, attitude, motivation, and creative-thinking scales	ANCOVA	Quasi-experimental
3. (Sun et al., 2024)	Two-class quasi-experiment; random division stated, but the allocation procedure was not described	ChatGPT-facilitated programming (n = 43) vs. self-directed programming (n = 39)	Programming-performance rubric	t-test	Quasi-experimental
4. (Huesca et al., 2024)	Cluster-randomized experiment; 14 groups randomly designated as intervention (8 groups) or control (6 groups)	ChatGPT-assisted flipped learning (n = 140) vs. video-based flipped learning (n = 114)	Normalized learning gain from a 22-item programming test	Mann-Whitney U	RCT
5. (Wu et al., 2025)	RCT; computer-generated allocation sequence	PA-GPT (n = 31) vs. traditional ChatGPT (n = 30)	Knowledge Construction Test, HOTS Scale, and S-STEM questionnaire	ANCOVA	RCT
6. (El Fathi et al., 2025)	Pretest-posttest quasi-experiment	ChatGPT-assisted instruction (n = 60) vs. traditional instruction (n = 60)	STPFSL-Short conceptual-understanding test	Mann-Whitney U; Wilcoxon signed-rank test	Quasi-experimental
7. (Gu, 2025)	Two-class quasi-experiment; one intact class assigned to each condition, with random assignment stated but not described	Chatbot-assisted learning (n = 31) vs. traditional teaching (n = 31)	Twenty-item HOTS Scale	t-tests; ANCOVA	Quasi-experimental
8. (Baltà-Salvador et al., 2026)	Randomized crossover experiment; condition order randomized	Within-participant comparison of ChatGPT-assisted and unassisted ideation	Idea-based creativity metrics: fluency, flexibility, originality, novelty, and elaboration	Wilcoxon signed-rank test	RCT
9. (Garg et al., 2025)	Three-group quasi-experiment	Internet-only control vs. GenAI without prompt training vs. GenAI with prompt training	Fifteen-item programming and data-analysis test	ANOVA; Tukey HSD	Quasi-experimental
10. (Salem et al., 2025)	Pretest-posttest quasi-experiment; random allocation reported, but allocation procedure not described	ChatGPT-assisted assignments (n = 150) vs. traditional assignments (n = 150)	Tourism-statistics grades	ANOVA	Quasi-experimental

The included studies varied in design, allocation procedures, comparison conditions, outcome measures, and statistical analyses. All 10 studies met the prespecified eligibility criteria and provided comparative data for effect-size and sampling-variance estimation. These reported methodological characteristics were considered when interpreting the pooled estimate.

DATA EXTRACTION AND CALCULATION

Following the quality assessment, the third author extracted the data using a structured Microsoft Excel spreadsheet. Subsequently, the first and third authors rechecked all extracted values against the source publications. The extraction form recorded the authors, publication year, study code, learning outcome, and the sample size, mean, and standard deviation for the experimental and control groups. Studies that did not report sufficient statistical data to calculate effect sizes were excluded, and their authors were not contacted. Using the group-level statistics, the third author calculated Hedges' g and its standard error (SE g) based on the equations and methodological guidance provided by the second author. The calculations were subsequently checked by the first author.

When a study reported multiple eligible learning outcomes, Hedges' g and SE g were calculated separately for each outcome. All eligible outcomes were retained in accordance with the multidomain definition of learning outcomes adopted in this review. They were combined as distinct dimensions of learning evaluated within the same instructional comparison. However, multiple effect sizes derived from the same participant sample were not entered separately into the main meta-analysis because they were statistically dependent. Instead, they were combined into a single study-level composite effect, thereby retaining the study as the unit of analysis. This statistical aggregation did not imply that the combined outcomes represented a single homogeneous construct. Studies reporting only one eligible outcome contributed that effect size directly to the meta-analysis. Hedges' g was selected because it provides a standardized estimate of the difference between the experimental and control groups while correcting for small-sample bias.

Effect-size calculations were selected according to the design of each study. For studies with independent experimental and comparison groups, Hedges' g and its standard error were calculated using Equations (1)–(6). Study 8 used a randomized crossover design in which the same 51 participants completed the AI-assisted and unaided conditions. Its effect size and sampling variance were therefore calculated using a paired standardized mean-difference method with average-variance standardization. Because the within-participant correlation was not reported, $r = .50$ was assumed for the primary analysis, while alternative correlations were examined in the sensitivity analysis.

The calculations were performed as follows:

1. Pooled standard deviation

$$SD_{pooled} = \sqrt{\frac{(n_E - 1)SD_E^2 + (n_C - 1)SD_C^2}{n_E + n_C - 2}} \quad (1)$$

2. Cohen's d

$$d = \frac{M_E - M_C}{SD_{pooled}} \quad (2)$$

3. Correction factor

$$J = 1 - \frac{3}{4(n_E + n_C - 2) - 1} \quad (3)$$

4. Hedges' g

$$g = J \times d \quad (4)$$

5. Variance of d

$$V_d = \frac{n_E + n_C}{n_E n_C} + \frac{d^2}{2(n_E + n_C)} \quad (5)$$

6. Standard error of Hedges' g

$$SE_g = \sqrt{J^2 \times V_d} \quad (6)$$

7. Composite Hedges' g

$$g_c = \frac{1}{m} \sum_{j=1}^m g_j \quad (7)$$

8. Variance of composite Hedges' g

$$V_c = \frac{1}{m^2} (\sum_{j=1}^m SE_{gj}^2 + 2r \sum_{j < k} SE_{gj} SE_{gk}) \quad (8)$$

9. Standard error of composite Hedges' g

$$SE_c = \sqrt{V_c} \quad (9)$$

Studies 1, 2, and 5 reported multiple eligible outcomes measured using the same participant samples. Hedges' g and its standard error were first calculated separately for each outcome using Equations (1)–(6). Within each study, the outcome-level Hedges' g estimates were weighted equally and combined into a composite effect using Equation (7). The composite variance was calculated using Equation (8), which incorporated the sampling variances and pairwise covariances of the outcome-level estimates. Because the correlations between outcome-level effect-size estimates were not reported in the primary studies, a within-study correlation of $r = .50$ was assumed (Julienne et al., 2024), with each pairwise covariance represented by $r SE_{gj} SE_{gk}$. The composite standard error was calculated using Equation (9). The outcome-level values and resulting composite estimates are reported in Table 3 to permit verification of the calculations.

Table 3. Data recapitulation

Author	Study code	Outcome	Experimental group			Control group			g	SE_g
			N	Mean	SD	N	Mean	SD		
(Yilmaz & Karaoglan Yilmaz, 2023)	Study 1a	Computational thinking skills	21	126.73	8.34	24	112.61	15.32	1.10	0.32
	Study 1b	computer programming self-efficacy	21	41.32	5.84	24	33.52	7.64	1.12	0.32
	Study 1c	Learning motivation	21	96.50	7.62	24	85.17	13.84	0.98	0.31
	Composite of Study 1								1.07	0.26
(H. F. Li, 2023)	Study 2a	Project performance	42	88.80	2.73	39	87.27	3.41	0.49	0.22
	Study 2b	Self-efficacy	42	35.05	3.73	39	32.85	4.54	0.53	0.23
	Study 2c	Learning attitudes	42	31.50	2.76	39	30.00	4.21	0.42	0.22
	Study 2d	Learning motivation	42	26.17	2.37	39	24.98	2.09	0.53	0.23
	Study 2e	Creative thinking tendency	42	19.19	3.00	39	17.28	3.96	0.54	0.23
	Composite of Study 2								0.50	0.18

Author	Study code	Outcome	Experimental group			Control group			g	SE g
			N	Mean	SD	N	Mean	SD		
(Sun et al., 2024)	Study 3	Programming skills	43	84.11	19.45	39	78.36	17.59	0.31	0.22
(Huesca et al., 2024)	Study 4	Cognitive skills	140	52.85	24.27	114	41.29	26.84	0.45	0.13
(Wu et al., 2025)	Study 5a	Knowledge construction skills	31	76.50	6.88	30	69.60	7.65	0.94	0.27
	Study 5b	Critical thinking	31	17.30	2.19	30	13.90	1.98	1.61	0.29
	Study 5c	Problem-solving	31	15.90	2.01	30	14.20	2.09	0.82	0.27
	Study 5d	Creativity	31	11.10	1.22	30	10.30	0.96	0.72	0.26
	Study 5e	Attitudes towards science	31	35.90	2.95	30	34.50	2.71	0.49	0.26
	Study 5f	Attitudes towards math	31	36.60	2.86	30	33.10	2.30	1.33	0.28
	Study 5g	Attitudes towards engineering/technology	31	32.20	4.60	30	27.50	4.30	1.04	0.27
Composite of Study 5									0.99	0.21
(El Fathi et al., 2025)	Study 6	Conceptual understanding	60	22.53	3.03	60	18.60	1.90	1.54	0.21
(Gu, 2025)	Study 7	Higher-order thinking skills (HOTS)	31	4.24	0.72	31	3.70	0.55	0.83	0.26
(Baltà-Salvador et al., 2026)	Study 8	Creative thinking	51	7.22	1.08	51	7.06	1.54	0.12	0.14
(Garg et al., 2025)	Study 9	Data analysis	59	6.60	2.69	64	4.28	1.17	1.13	0.19
(Selem et al., 2025)	Study 10	Creative thinking	150	96.12	7.74	150	90.87	8.59	0.64	0.12

Note. N = sample size; Mean = mean score; SD = standard deviation; g = Hedges' g ; SE g = standard error of Hedges' g . For Studies 1, 2, and 5, the composite estimates used in the main analysis were calculated using equal weighting and an assumed within-study correlation of $r = .50$.

For Studies 1, 2, and 5, the outcome-level Hedges' g estimates were combined into one equally weighted composite effect per study. The corresponding composite standard error incorporated the sampling variances and pairwise covariances of the outcome-level estimates using an assumed within-study correlation of $r = .50$. Only the resulting composite effect and standard error were included in the meta-analysis.

DATA ANALYSIS

Study-level effect estimates were analyzed in JASP using a random-effects meta-analysis (Bartoš et al., 2025). This model was selected because contextual and methodological diversity made a single common true effect implausible (Durlak & Lipsey, 1991; Moreau & Gamble, 2022). The analysis pro-

duced the pooled effect size, confidence interval, standard error, z value, significance level, and heterogeneity statistics. Sensitivity analyses assessed the robustness of the primary results across assumed within-study correlations of $r = 0, .30, .50, .70$, and $.90$ and in an outcome-restricted analysis that retained directly assessed cognitive, academic, skill, and performance outcomes while excluding affective and other self-reported outcomes. The same random-effects settings were used throughout. Potential small-study effects were explored using funnel-plot inspection, Kendall's rank correlation test, and Rosenthal's fail-safe N . Because only 10 studies were included, these assessments were treated as exploratory, and moderator analysis was not conducted. Keyword co-occurrence analysis was conducted to map the thematic content of the 10 studies included in the meta-analysis. The records were managed in Mendeley and exported as a single RIS file to VOSviewer. The analysis used author keywords, full counting, and default settings, with the minimum occurrence threshold set to 1; therefore, all 41 keywords were included. No manual cleaning or merging of keyword variants was performed. VOSviewer generated three color-coded clusters based on keyword co-occurrence, which were interpreted descriptively according to the keywords grouped within each cluster.

RESULTS AND DISCUSSION

The results and discussion are organized according to the four research questions. The meta-analysis was used to estimate the average effect of ChatGPT-assisted instruction across the included studies. The domain analysis described the assessed learning outcomes, while the keyword analysis mapped the dominant research themes. The narrative synthesis subsequently examined the implementation opportunities, concerns, and future strategies reported in the included studies. These analytical components address distinct research questions and are interpreted together to provide complementary quantitative and contextual evidence.

SUMMARY EFFECT

RQ1: What is the pooled effect of ChatGPT-assisted instruction on learning outcomes across the included educational contexts?

The study-level effects varied in magnitude. A random-effects meta-analysis was therefore used to estimate the average effect while accounting for between-study heterogeneity. Table 4 presents the random-effects model and heterogeneity test.

Table 4. Random-effects model and heterogeneity test

	Q	df	p	τ^2	I^2
Omnibus test of model coefficients	28.751	1	< .001	-	-
Test of residual heterogeneity	49.838	9	< .001	0.153	83.121%

The omnibus test of model coefficients was significant ($Q(1) = 28.751, p < .001$), confirming that the average effect differed from zero. Residual heterogeneity was significant ($Q(9) = 49.838, p < .001, \tau^2 = 0.153, I^2 = 83.121\%$), indicating substantial variation across study-level effects. Accordingly, the pooled effect represents an average estimate rather than uniform effectiveness across all instructional conditions. Because instructional design, teacher guidance, prompting practices, learner characteristics, exposure duration, and assessment type were not tested as moderators, they should be regarded as directions for future research rather than established explanations for the heterogeneity. The pooled random-effects estimate is presented in Table 5.

Table 5 presents a pooled random-effects estimate of 0.739 (95% CI = 0.469–1.010; $z = 5.362; p < .001$). According to Cohen's criteria of 0.2 for a small effect, 0.5 for a medium effect, and 0.8 for a large effect (Cohen, 2013), this represents a statistically significant moderate positive average effect of ChatGPT-assisted instruction relative to the comparison conditions. The estimate summarizes the

standardized effects across the cognitive, affective, and psychomotor outcomes included in the primary analysis. A sensitivity analysis was conducted using only directly assessed cognitive, academic, skill, and performance outcomes. The analysis included Studies 2, 3, 4, 5, 6, 8, 9, and 10, while Studies 1 and 7 were excluded because they did not provide eligible directly assessed outcomes. The analysis yielded a statistically significant pooled estimate of 0.689 (SE = 0.165, 95% CI = 0.365–1.013, $z = 4.167$, $p < .001$). This estimate was 0.05 lower than the primary estimate of 0.739 but remained positive, statistically significant, and moderate in magnitude.

Table 5. Pooled effect size

					95% confidence interval	
	Estimate	Standard error	z	p	Lower	Upper
Pooled effect	0.739	0.138	5.362	< .001	0.469	1.010

Table 6 presents the sensitivity analysis results across the five assumed within-study correlations. The pooled Hedges' g estimates ranged from 0.730 to 0.753, and all 95% confidence intervals remained above zero. The estimated between-study variance (τ^2) ranged from 0.144 to 0.162, while I^2 ranged from 83.12% to 87.05%. Cochran's Q remained statistically significant under all five correlation assumptions ($p < .001$), indicating consistently high between-study heterogeneity. Therefore, the direction, statistical significance, and moderate interpretation of the pooled effect remained stable across the examined within-study correlation assumptions.

Table 6. Sensitivity analysis across assumed within-study correlations

r	Pooled Hedges' g	95% CI	τ^2	Q	df	p	I^2
0	0.753	0.495–1.011	0.144	53.213	9	< .001	86.53%
.30	0.745	0.480–1.010	0.148	47.273	9	< .001	83.34%
.50	0.739	0.469–1.010	0.153	49.838	9	< .001	83.12%
.70	0.735	0.461–1.009	0.157	56.758	9	< .001	84.03%
.90	0.730	0.451–1.009	0.162	82.894	9	< .001	87.05%

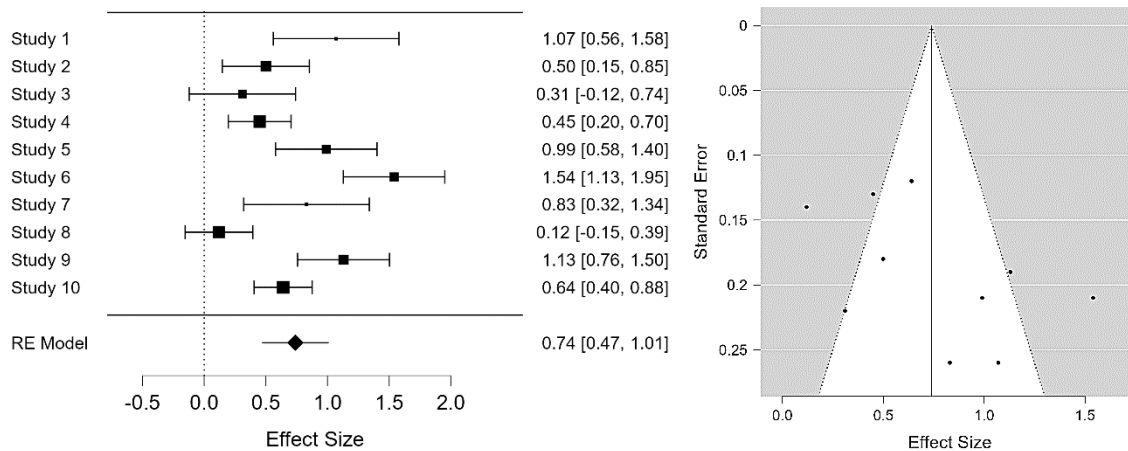


Figure 2. Forest plot and funnel plot

Visual inspection of the funnel plot presented in Figure 2 did not reveal a clear pattern of asymmetry, although the small number of studies made the visual assessment inconclusive. The rank correlation test likewise did not detect statistically significant funnel-plot asymmetry (Kendall's $\tau = 0.250$, $p = 0.321$). Rosenthal's fail-safe N was 557, indicating that, under the assumptions of this method, 557 hypothetical studies averaging a null result would be required to make the combined result statistically non-significant (Rosenthal, 1979). This value exceeded the conventional $5k+10$ criterion ($5 \times 10 + 10 = 60$). However, the fail-safe N does not estimate the actual number of missing studies. Given the inclusion of only 10 studies and the significant residual heterogeneity, these analyses had limited power. Therefore, they detected no clear evidence of funnel-plot asymmetry but could not establish the absence of publication bias or distinguish publication bias from heterogeneity, chance, or other study-level differences.

EFFECTS OF CHATGPT

RQ2: What aspects are targeted for assessment in the implementation of ChatGPT?

The learning outcomes reported in the included studies were categorized based on Bloom's three learning domains: cognitive, affective, and psychomotor (Mukherjee & Kittur, 2025). The cognitive domain included knowledge, reasoning, understanding, and problem solving. The affective domain covered attitudes, motivation, perceptions, and self-efficacy. The psychomotor domain covered skill performance and practical competencies. Table 7 identifies the specific outcomes reported by each study, while Figure 3 summarizes their distribution across the three domains. This classification was used to retain the conceptual distinctions among the outcomes included within the broad category of learning outcomes.

Table 7. Profiles of the studies and learning outcomes

Study	Author	Educational context	Experiment group (EG)	Control group (CG)	Learning outcome
Study 1	(Yilmaz & Karaoglan Yilmaz, 2023)	Higher education: programming education	ChatGPT-assisted programming	Non-ChatGPT Programming	Computational thinking skills; self-efficacy; motivation
Study 2	(H. F. Li, 2023)	Preservice teacher education: modern educational technology course	ChatGPT-based flipped learning guiding approach (ChatGPT-FLGA)	Flipped classroom instruction	Project performance; self-efficacy; learning attitudes; intrinsic motivation; creative thinking
Study 3	(Sun et al., 2024)	Higher education: programming course	ChatGPT-facilitated programming (CFP)	Self-directed programming (SDP)	Programming skill
Study 4	(Huesca et al., 2024)	Engineering education: basic programming course	ChatGPT-assisted flipped learning	Video-based flipped learning	Cognitive skills
Study 5	(Wu et al., 2025)	Engineering/STEM education: network embedded systems course	Peer assessment cycle with ChatGPT	Traditional ChatGPT	Knowledge construction skills; critical thinking; problem-solving; creativity; attitudes towards science, math, and engineering/technology

Study	Author	Educational context	Experiment group (EG)	Control group (CG)	Learning outcome
Study 6	(El Fathi et al., 2025)	Engineering education: introductory thermodynamics course	ChatGPT-assisted instruction	Traditional instruction	Conceptual understanding
Study 7	(Gu, 2025)	Applied design education: 3D animation design course	Chatbot-assisted learning model	Traditional teaching model	Higher-order thinking skills (HOTS)
Study 8	(Baltà-Salvador et al., 2026)	Engineering education: creative ideation and design	ChatGPT-assisted learning	Unaided ideation	Creative thinking
Study 9	(Garg et al., 2025)	Engineering education: data analysis and programming	GenAI with prompt training	Internet-only control	Data analysis
Study 10	(Salem et al., 2025)	Tourism and hospitality education: tourism statistics course	ChatGPT apps	Traditional assignments	Creative thinking

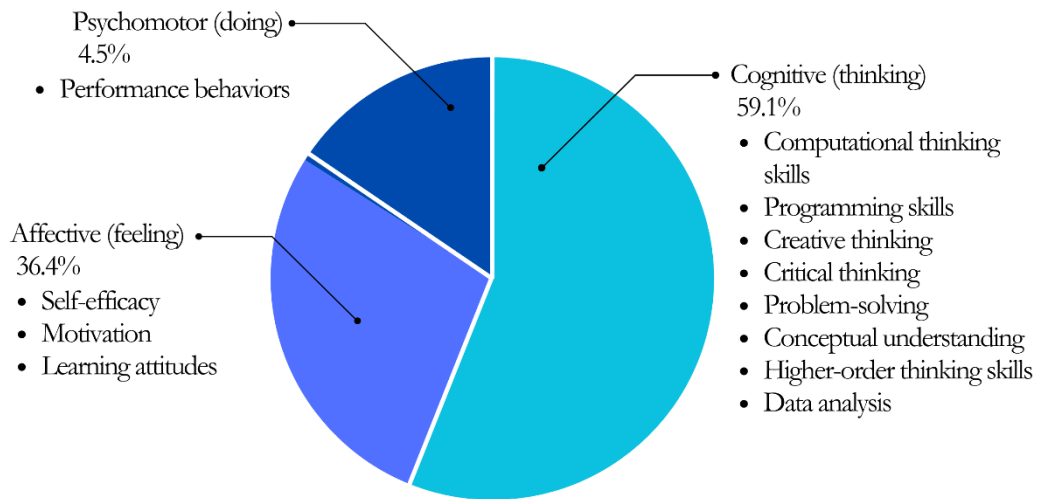


Figure 3. Percentage distribution of reported learning outcomes by domain

The included studies represented varied educational contexts, including programming education, modern educational technology, engineering and STEM education, applied design education, and tourism and hospitality education. These contexts should not be interpreted as equivalent or as representing the full range of TVET. The control-group interventions also varied, although most involved traditional instruction, whereas the experimental groups consistently integrated ChatGPT. Across the 22 learning outcomes reported, 13 outcomes (59.1%) were classified as cognitive, 8 (36.4%) as affective, and 1 (4.5%) as psychomotor. The cognitive outcomes included computational and programming skills, creativity, critical thinking, conceptual understanding, higher-order thinking, data analysis

skills, knowledge construction skills, and problem-solving (Altun & Yildirim, 2023; Amrina et al., 2018; Mangaroska et al., 2022; Ragonis & Shmallo, 2022). The affective outcomes included self-efficacy, motivation, intrinsic motivation, learning attitudes, and attitudes toward science, mathematics, and engineering or technology (Wang et al., 2026). Project performance was classified as psychomotor because it assessed students' applied performance in designing and producing an operable courseware project. Its scoring rubric also covered instructional, scientific, technological, and artistic dimensions. The classification presented in Table 7 and Figure 3 preserves the conceptual distinctions among these outcomes rather than treating them as interchangeable measures of a single construct. These percentages describe the distribution of measured and reported outcomes in the included studies, rather than differences in the magnitude of ChatGPT's effects across the three domains.

From an experiential learning perspective, doing can be understood as active participation in learning tasks, not merely physical activity (Salinas-Navarro et al., 2024). Previous studies have also associated digital interaction with questioning and problem-solving activities (Syahmaidi et al., 2019). However, the included studies did not consistently measure active experimentation or other experiential-learning processes. These references, therefore, provide an interpretive perspective only and are not used to claim that experiential-learning mechanisms produced the observed distribution of outcomes or the pooled effect.

KEYWORD CO-OCCURRENCE NETWORK ANALYSIS

RQ3: What is the distribution pattern of keywords in publications on ChatGPT interventions in TVET, vocational education, and engineering education?

The keyword co-occurrence analysis descriptively mapped relationships among author-provided keywords across the 10 included studies and was not used to explain the pooled intervention effect. The results are presented in Figure 4.

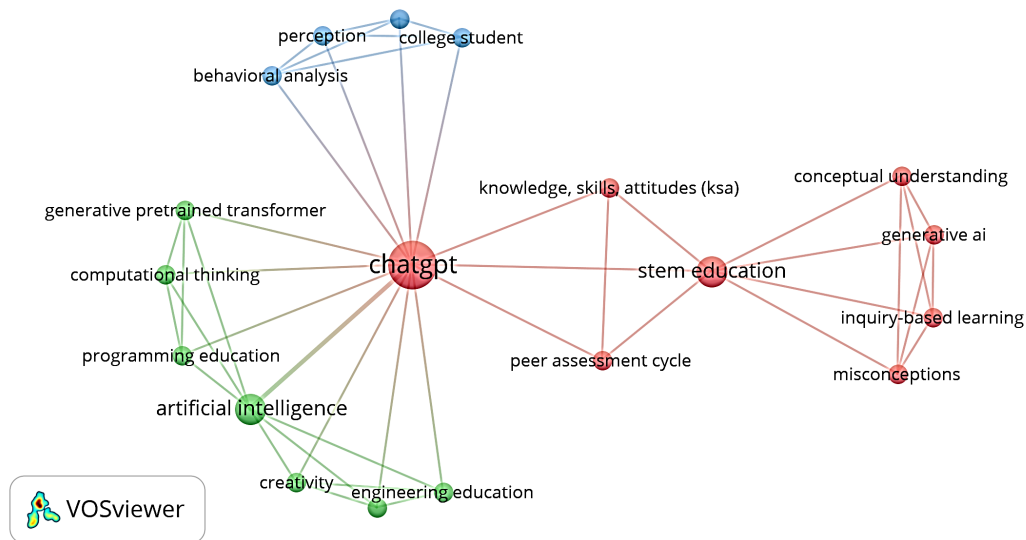


Figure 4. Keyword analysis

Each node represents a keyword, while the connecting lines indicate keyword co-occurrence. Node size is proportional to occurrence frequency (Jiang et al., 2025), and color distinguishes the three clusters. The network comprised three clusters, with eight nodes in red, seven in green, and four in blue. Within this network, “ChatGPT” was the largest and most connected node, while “artificial intelligence” and “STEM education” were also relatively prominent. The red cluster connected

ChatGPT with knowledge, skills, and attitudes (KSAs), STEM education, peer assessment, conceptual understanding, generative AI, inquiry-based learning, and misconceptions. The green cluster covered generative pretrained transformers, computational thinking, programming education, artificial intelligence, creativity, and engineering education, while the blue cluster focused on perception, college students, and behavioral analysis. Spelling variants, abbreviations, and closely related terms were reviewed before the analysis. Related but non-equivalent terms were kept separate, and no thesaurus was used to preserve the keywords provided by the authors. Since the network was based on only 10 records and a minimum occurrence threshold of one, it should be interpreted as a description of keyword relationships within this specific corpus rather than as a stable thematic structure across the broader field.

OPPORTUNITIES, CONCERNS, AND FUTURE STRATEGIES

RQ4: What are the opportunities, concerns, and future strategies for implementing ChatGPT?

The included studies identified both opportunities and concerns associated with ChatGPT-assisted instruction. Table 8 synthesizes these study-level findings and the future directions recommended by the respective authors. Table 8 shows that the included studies reported opportunities related to programming support, human–computer interaction, personalized and interactive learning, peer assessment, and learning assessment. They also identified concerns regarding inaccurate or biased outputs, plagiarism, over-reliance, reduced learner autonomy and creativity, privacy and infrastructure constraints, and subscription costs. These findings summarize the opportunities and concerns reported in the primary studies; their prevalence and causal effects were not examined in this review. The included studies proposed prompt training, verification of AI-generated responses, explicit guidance for responsible use, integration with peer- or problem-based learning, and continued evaluation. The primary studies reported these strategies but were not compared or tested as moderators in this meta-analysis. They should therefore be treated as proposals for future evaluation rather than as proven conditions for effectiveness.

Table 8. Opportunities, concerns, and future strategies

Study	Authors	Opportunities	Concerns	Future strategies
Study 1	(Yilmaz & Karaoglan Yilmaz, 2023)	AI supports programming learning while also enhancing student motivation.	Some students find it difficult to collaborate with their peers, especially when working on group projects.	Developing digital literacy skills to train students in writing effective prompts. Well-structured and well-directed prompts produce more accurate and relevant responses from ChatGPT.
Study 2	(H. F. Li, 2023)	Support human-computer conversations.	ChatGPT is feared to potentially reduce students' creative thinking abilities.	• Teachers should utilize ChatGPT to enhance students' problem-solving skills. • Course designers and academic staff can leverage ChatGPT to implement student-centered pedagogical transformation. • Course designers should master the use of ChatGPT to foster students' creative thinking abilities.
Study 3	(Sun et al., 2024)	ChatGPT assists in computer programming by providing guidance, answering	Students may potentially take answers from ChatGPT and present them as their own work.	Teachers should be aware of both the potential benefits and drawbacks associated with integrating ChatGPT or other AI language

Study	Authors	Opportunities	Concerns	Future strategies
		questions, and generating code snippets.		models into their computer programming courses from a pedagogical perspective.
Study 4	(Huesca et al., 2024)	ChatGPT provides a more active, interactive, and personalized approach to the learning process. This technology can also enhance the effectiveness of existing pedagogical strategies.	• Lecturers are concerned that students' critical thinking skills may be weakened due to dependence on this tool. • ChatGPT may encourage individuals with limited experience to overestimate their abilities, leading them to mistakenly believe they are more competent than they actually are.	• The long-term impact must be continuously evaluated, as it may affect students' future careers. • Educators should consistently consider ethical aspects and the risks of over-reliance on AI tools in learning. • A variety of assessment techniques and learning activities (which may or may not involve the use of ChatGPT) should be integrated into the learning process.
Study 5	(Wu et al., 2025)	ChatGPT can act as a virtual peer in peer assessment activities and personalized learning.	• Ethical issues, algorithmic bias, and their impact on student–teacher relationships • Students' excessive dependence on AI • Lack of instructional guidance from teachers	• It is recommended to conduct periodic evaluations of ChatGPT, as it may influence students' competencies and career choices. • The implementation of ChatGPT in learning should be integrated with approaches such as problem-based learning or flipped classrooms.
Study 6	(El Fathi et al., 2025)	ChatGPT can be used as a supplementary pedagogical tool.	• Inherent AI biases • Data privacy • Infrastructural limitations	• Implementing the CILP Framework learning strategy • Encouraging the use of well-directed prompts • Evaluating AI-generated responses • Promoting reflective learning • Encouraging ethical and strategic use of AI • Integrating AI interaction with peer learning
Study 7	(Gu, 2025)	AI chatbots support learning in solving 3D problems.	The long-term impact of chatbot use is still unclear.	To examine the relationship between the duration of chatbot use and learning effectiveness, as well as to explore its impact on motivation, engagement, self-efficacy, and learning performance.
Study 8	(Baltà-Salvador et al., 2026)	Support human-computer interaction.	• Creative limitation • Risk of plagiarism	• Precision in requests to get relevant responses • Use AI just for initial inspiration

Study	Authors	Opportunities	Concerns	Future strategies
			• Cognitive conditioning • Loss of autonomy • Homogeneity • Intellectual laziness • Unreliable information • Lack of empathy • Individual responsibility	• Learn to interact with AI • Request relevant information instead of direct ideas • Start with individual unaided idea • Provide clear context
Study 9	(Garg et al., 2025)	Helps in learning assessment.	There are limitations in the quality and accuracy of the generated content.	• Training in the use of effective prompt engineering • Verifying every response generated by ChatGPT • Enhancing understanding of code logic and data analysis techniques
Study 10	(Selem et al., 2025)	Personalized learning experiences.	• Hindrance to creativity • Data privacy issues • Psychological impacts • Subscription costs	• Building AI literacy • Teachers need to equip students with the skills required to use AI as a supplementary resource • Teachers should develop clear guidelines for using ChatGPT • ChatGPT should be used not only individually but also in group settings • Stakeholders should establish strong privacy policies and provide financial and psychological support mechanisms

CONCLUSION

Within this bounded corpus of 10 English-language, open-access studies indexed in Scopus, the pooled Hedges' g of 0.739 indicates the average standardized difference across the eligible learning outcomes between ChatGPT-supported instruction and the comparison conditions. The pooled effect was statistically significant and moderate in size, and this interpretation remained stable across the examined within-study correlation assumptions. The moderate interpretation also remained unchanged when the analysis was restricted to directly assessed cognitive, academic, skill, and performance outcomes. Significant between-study heterogeneity indicated that the magnitude of the effect varied across the included studies. The evidence was primarily from higher education students and concentrated mainly in programming, computer science, and engineering (STEM) courses, with a smaller body of evidence from educational technology, applied design, and tourism and hospitality education. The evidence was dominated by cognitive outcomes, with fewer affective outcomes and only one psychomotor outcome; accordingly, the pooled estimate represents an average across these conceptually diverse outcomes rather than equivalent effects within each domain. Hands-on and trade-based programs, workplace or apprenticeship learning, and non-higher-education vocational learner groups were poorly represented or absent. Therefore, the pooled estimate should not be generalized to all TVET, vocational education, or engineering education settings. Given the small and narrowly composed evidence base, it remains unclear whether the observed effect varies according to

instructional design, teacher guidance, prompting practices, learner characteristics, prior knowledge, exposure duration, assessment type, and learning context.

LIMITATION

This review was restricted to English-language, open-access empirical journal articles and conference papers indexed in Scopus from 2022 to 2026. Consequently, eligible studies indexed exclusively in other databases or published behind subscription barriers may have been omitted. These restrictions may have affected the geographical, institutional, disciplinary, and educational-system representation of the evidence and may also have increased or decreased the pooled effect estimate, although the direction and magnitude of this possible influence cannot be determined. Therefore, the findings represent the 10 studies included in this bounded evidence corpus rather than the complete international literature on ChatGPT-assisted learning.

FUTURE WORK

Future research should expand the search strategy to multiple databases, include a broader range of document types when methodologically appropriate, and consider non-English and closed-access publications when available. More primary studies are needed across a broader range of TVET, vocational education, and engineering education contexts beyond the programming- and computer science-dominated evidence base, particularly for psychomotor skills, workplace performance, long-term retention, and professional competence. As the evidence base grows, future meta-analyses should test moderators of the between-study variation identified as unresolved in this review.

REFERENCES

- Abera, H. G. (2023). The role of education in achieving the sustainable development goals (SDGs): A global evidence based research article. *International Journal of Social Science and Education Research Studies*, 3(1), 67–81. <https://doi.org/10.55677/ijssers/V03I1Y2023-09>
- Altun, E., & Yildirim, N. (2023). What does critical thinking mean? Examination of pre-service teachers' cognitive structures and definitions for critical thinking. *Thinking Skills and Creativity*, 49, 101367. <https://doi.org/10.1016/j.tsc.2023.101367>
- Amoros Molina, A., Hellden, D., Alfven, T., Niemi, M., Leander, K., Nordenstedt, H., Rehn, C., Ndejjo, R., Wanyenze, R., & Biermann, O. (2023). Integrating the United Nations sustainable development goals into higher education globally: a scoping review. *Global Health Action*, 16(1), Article 2190649. <https://doi.org/10.1080/16549716.2023.2190649>
- Amrina, Z., Desfitri, R., Zuzano, F., Wahyuni, Y., Hidayat, H., & Alfino, J. (2018). Developing instruments to measure students' logical, critical, and creative thinking competences for Bung Hatta University students. *International Journal of Engineering and Technology*, 7(4), 128–131. <https://doi.org/10.14419/ijet.v7i4.9.20633>
- Athanassopoulos, S., Manoli, P., Gouvi, M., Lavidas, K., & Komis, V. (2023). The use of ChatGPT as a learning tool to improve foreign language writing in a multilingual and multicultural classroom. *Advances in Mobile Learning Educational Research*, 3(2), 818–824. <https://doi.org/10.25082/AMLER.2023.02.009>
- Baig, M. I., & Yadegaridehkordi, E. (2024). ChatGPT in the higher education: A systematic literature review and research challenges. *International Journal of Educational Research*, 127, 102411. <https://doi.org/10.1016/j.ijer.2024.102411>
- Baltà-Salvador, R., Brasó-Vives, E., & Peña, M. (2026). Evaluating AI-assisted creative ideation: A crossover study in higher education. *Thinking Skills and Creativity*, 59, 101958. <https://doi.org/10.1016/j.tsc.2025.101958>
- Bandura, A. (1991). Social cognitive theory of self-regulation. *Organizational Behavior and Human Decision Processes*, 50(2), 248–287. [https://doi.org/10.1016/0749-5978\(91\)90022-L](https://doi.org/10.1016/0749-5978(91)90022-L)
- Barker, T. H., Habibi, N., Aromataris, E., Stone, J. C., Leonardi-Bee, J., Sears, K., Hasanoff, S., Klugar, M., Tufanaru, C., & Moola, S. (2024). The revised JBI critical appraisal tool for the assessment of risk of bias

- for quasi-experimental studies. *JBIE Evidence Synthesis*, 22(3), 378–388. <https://doi.org/10.11124/JBIES-23-00268>
- Barker, T. H., Stone, J. C., Sears, K., Klugar, M., Tufanaru, C., Leonardi-Bee, J., Aromataris, E., & Munn, Z. (2023). The revised JBI critical appraisal tool for the assessment of risk of bias for randomized controlled trials. *JBIE Evidence Synthesis*, 21(3), 494–506. <https://doi.org/10.11124/JBIES-22-00430>
- Bartoš, F., Wagenmakers, E.-J., & Viechtbauer, W. (2025). *Meta-analysis with JASP, Part I: Classical approaches*. PsyArXiv. <https://doi.org/10.48550/arXiv.2509.09845>
- Basnet, R., & Pyakurel, C. K. (2024). Enhancing performance of health assistants through technical and vocational education and training for better healthcare access. *Journal of Technical and Vocational Education and Training*, 18(1), 89–101. <https://doi.org/10.3126/tvet.v18i1.62747>
- Bower, M., Henderson, M., Slade, C., Southgate, E., Gulson, K., & Lodge, J. (2025). What generative Artificial Intelligence priorities and challenges do senior Australian educational policy makers identify (and why)? *Australian Educational Researcher*, 52(3), 2069–2094. <https://doi.org/10.1007/s13384-025-00801-z>
- Chen, W., & Chen, Z. (2025). Digital tourism: Exploring the ChatGPT–metaverse synergy. *Tourism Review*, 1–20. <https://doi.org/10.1108/TR-04-2025-0406>
- Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Routledge. <https://doi.org/10.4324/9780203771587>
- da Silva, C. A. G., Ramos, F. N., de Moraes, R. V., & dos Santos, E. L. (2024). ChatGPT: Challenges and benefits in software programming for higher education. *Sustainability*, 16(3), 1245. <https://doi.org/10.3390/su16031245>
- Deng, R., Jiang, M., Yu, X., Lu, Y., & Liu, S. (2025). Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. *Computers and Education*, 227, Article 105224. <https://doi.org/10.1016/j.compedu.2024.105224>
- Durlak, J. A., & Lipsey, M. W. (1991). A practitioner’s guide to meta-analysis. *American Journal of Community Psychology*, 19(3), 291–332. <https://doi.org/10.1007/BF00938026>
- El Fathi, T., Saad, A., Larhzil, H., Lamri, D., & Al Ibrahim, E. M. (2025). Integrating generative AI into STEM education: enhancing conceptual understanding, addressing misconceptions, and assessing student acceptance. *Disciplinary and Interdisciplinary Science Education Research*, 7, Article 6. <https://doi.org/10.1186/s43031-025-00125-z>
- Fadillah, R., Ganefri, Yulastri, A., & Hidayat, H. (2023). Development of mobile learning based on digital entrepreneurs using raspberry Pi on TVET. *International Journal on Advanced Science, Engineering and Information Technology*, 13(6), 2231–2239. <https://doi.org/10.18517/ijascit.13.6.18608>
- Garg, A., Nisumba Soodhani, K., & Rajendran, R. (2025). Enhancing data analysis and programming skills through structured prompt training: The impact of generative AI in engineering education. *Computers and Education: Artificial Intelligence*, 8, 100380. <https://doi.org/10.1016/j.caeai.2025.100380>
- Gu, Y. (2025). The impact of Chatbot-assisted Learning Model on Higher-Order Thinking Skills (HOTS) in a 3D animation design course. *Proceedings of 2nd International Conference on Informatics Education and Computer Technology Applications* (pp. 122–127). Association for Computing Machinery. <https://doi.org/10.1145/3732801.3732826>
- Hadi, B. (2021). Fenomena learning loss pada pendidikan sekolah menengah kejuruan di Indonesia [The phenomenon of learning loss in vocational secondary education in Indonesia]. *Edudikara: Jurnal Pendidikan Dan Pembelajaran*, 6(4), 290–296. <https://doi.org/10.32585/edudikara.v6i4.262>
- Heung, Y. M. E., & Chiu, T. K. F. (2025). How ChatGPT impacts student engagement from a systematic review and meta-analysis study. *Computers and Education: Artificial Intelligence*, 8, 100361. <https://doi.org/10.1016/j.caeai.2025.100361>
- Hidayat, H., Zuhendra, Z., Efrizon, E., Delianti, V. I., Dewi, F. K., Isa, M. R. M., Harmanto, D., & Tanucan, J. C. M. (2024). Computational thinking skills in engineering education: Enhancing academic achievement through innovations, challenges, and opportunities. *TEM Journal*, 13(4), 3454–3467. <https://doi.org/10.18421/TEM134-78>

- Horne, C. (2025). Leveraging AI chatbots to enhance student understanding of electric circuits. *Proceedings of the IEEE Global Engineering Education Conference, London, United Kingdom*, 1–3. <https://doi.org/10.1109/EDUCON62633.2025.11016290>
- Horton, B. (2001). 'I hear and I forget, I see and I remember, I do and I understand' – Putting learning models into practice. *Planet*, 3(1), 12–14. <https://doi.org/10.11120/plan.2001.00030012>
- Huesca, G., Martínez-Treviño, Y., Molina-Espinosa, J. M., Sanromán-Calleros, A. R., Martínez-Román, R., Cendejas-Castro, E. A., & Bustos, R. (2024). Effectiveness of using ChatGPT as a tool to strengthen benefits of the flipped learning strategy. *Education Sciences*, 14(6), 660. <https://doi.org/10.3390/educsci14060660>
- İpek, Z. H., Gözümlü, A. İ. C., Papadakis, S., & Kallogiannakis, M. (2023). Educational applications of the ChatGPT AI system: A systematic review research. *Educational Process: International Journal*, 12(3), 26–55. <https://doi.org/10.22521/edupij.2023.123.2>
- Jiang, L., Li, J., Wider, W., Tanucan, J. C. M., Lobo, J., Fauzi, M. A., Hidayat, H., & Zou, R. (2025). A bibliometric insight into immersive technologies for cultural heritage preservation. *npj Heritage Science*, 13, Article 126. <https://doi.org/10.1038/s40494-025-01704-z>
- Jo, H. (2024). Subscription intentions for ChatGPT plus: A look at user satisfaction and self-efficacy. *Marketing Intelligence & Planning*, 42(6), 1052–1073. <https://doi.org/10.1108/MIP-08-2023-0411>
- Julienne, A., Verbecque, E., & Besnard, S. (2024). Normative data for instrumented posturography: a systematic review and meta-analysis. *Frontiers in Human Neuroscience*, 18, 1498107. <https://doi.org/10.3389/fnhum.2024.1498107>
- Kernberg, A., Gold, J. A., & Mohan, V. (2024). Using ChatGPT-4 to create structured medical notes from audio recordings of physician-patient encounters: Comparative study. *Journal of Medical Internet Research*, 26, e54419. <https://doi.org/10.2196/54419>
- Kolb, D. A. (2014). *Experiential learning: Experience as the source of learning and development*. FT Press.
- Kundu, A., & Bej, T. (2025). AI policies in school education: a comparative study on China, Singapore, Finland, and the US. *Journal of Science and Technology Policy Management*. <https://doi.org/10.1108/JSTPM-06-2024-0218>
- Le, K. D. R., Tay, S. B. P., Choy, K. T., Verjans, J., Sasanelli, N., & Kong, J. C. H. (2024). Applications of natural language processing tools in the surgical journey. *Frontiers in Surgery*, 11, 1403540. <https://doi.org/10.3389/fsurg.2024.1403540>
- Li, H. F. (2023). Effects of a ChatGPT-based flipped learning guiding approach on learners' courseware project performances and perceptions. *Australasian Journal of Educational Technology*, 39(5), 40–58. <https://doi.org/10.14742/ajet.8923>
- Li, M., & Krishnamachari, B. (2024). Evaluating ChatGPT-3.5 efficiency in solving coding problems of different complexity levels: An empirical analysis. *PsyArXiv*. <https://doi.org/10.48550/arXiv.2411.07529>
- Lukhele, B. N., & Laseinde, O. T. (2024). Exploring mechanical engineering equipment at TVET colleges in South Africa, towards integrating virtual and cyber-physical learning. *Procedia Computer Science*, 232, 2690–2700. <https://doi.org/10.1016/j.procs.2024.02.086>
- Mangaroska, K., Sharma, K., Gašević, D., & Giannakos, M. (2022). Exploring students' cognitive and affective states during problem solving through multimodal data: Lessons learned from a programming activity. *Journal of Computer Assisted Learning*, 38(1), 40–59. <https://doi.org/10.1111/jcal.12590>
- Mensah, J., Mensah, A., & Mensah, A. N. (2022). Understanding and promoting the 'leaving no one behind' ambition regarding the sustainable development agenda: A review. *Visegrad Journal on Bioeconomy and Sustainable Development*, 11(1), 6–15. <https://doi.org/10.2478/vjbsd-2022-0002>
- Mishra, N. R. (2023). Constructivist approach to learning: An analysis of pedagogical models of social constructivist learning theory. *Journal of Research and Development*, 6(1), 22–29. <https://doi.org/10.3126/jrdn.v6i01.55227>

- Moreau, D., & Gamble, B. (2022). Conducting a meta-analysis in the age of open science: Tools, tips, and practical recommendations. *Psychological Methods*, 27(3), 426–432. <https://doi.org/10.1037/met0000351>
- Mukherjee, S., & Kittur, J. (2025). Learning through cognitive, affective, and psychomotor domains: Understanding undergraduate engineering students' perspectives in the United States. *Engineering Education Review*, 3(3). <https://doi.org/10.54844/eer.2025.0916>
- Murtaza, M., Cheng, C. T., Fard, M., & Zeleznikow, J. (2025). Transforming driver education: A comparative analysis of LLM-augmented training and conventional instruction for autonomous vehicle technologies. *International Journal of Artificial Intelligence in Education*, 35(2), 736–773. <https://doi.org/10.1007/s40593-024-00407-z>
- Ooi, K. B., Tan, G. W. H., Al-Emran, M., Al-Sharafi, M. A., Capatina, A., Chakraborty, A., Dwivedi, Y. K., Huang, T. L., Kar, A. K., Lee, V. H., Loh, X. M., Micu, A., Mikalef, P., Mogaji, E., Pandey, N., Raman, R., Rana, N. P., Sarker, P., Sharma, A., ... Wong, L. W. (2025). The potential of generative artificial intelligence across disciplines: Perspectives and future directions. *Journal of Computer Information Systems*, 65(1), 76–107. <https://doi.org/10.1080/08874417.2023.2261010>
- Poschauko, V. C., Kreuzer, E., Hirz, M., & Pacher, C. (2024). Engineering education goes lifelong learning: Modularized technical vocational education and training program for the automotive sector. *Procedia Computer Science*, 232, 1799–1808. <https://doi.org/10.1016/j.procs.2024.02.002>
- Ragonis, N., & Shmallo, R. (2022). The application of higher-order cognitive thinking skills to promote students' understanding of the use of static in object-oriented programming. *Informatics in Education*, 21(2), 331–352. <https://doi.org/10.15388/infedu.2022.10>
- Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, 86(3), 638–641. <https://doi.org/10.1037/0033-2909.86.3.638>
- Salinas-Navarro, D. E., Vilalta-Perdomo, E., Michel-Villarreal, R., & Montesinos, L. (2024). Using generative artificial intelligence tools to explain and enhance experiential learning for authentic assessment. *Education Sciences*, 14(1), 83. <https://doi.org/10.3390/educsci14010083>
- Selem, K. M., Saghier, E. G., Khreis, S. H. A., & Tan, C. C. (2025). ChatGPT-related risk patterns and students' creative thinking toward tourism statistics course: Pretest and posttest quasi-experimentation. *Journal of Hospitality and Tourism Education*, 37(3), 313–328. <https://doi.org/10.1080/10963758.2025.2456638>
- Sun, D., Boudouaia, A., Zhu, C., & Li, Y. (2024). Would ChatGPT-facilitated programming mode impact college students' programming behaviors, performances, and perceptions? An empirical study. *International Journal of Educational Technology in Higher Education*, 21, Article 14. <https://doi.org/10.1186/s41239-024-00446-5>
- Syahmaidi, E., Hidayat, H., Hartanto, S., & Rahmadani, A. F. (2019). Needs analysis of designing online computer-assisted training to improve pedagogical competencies in engineering education. *International Journal of Engineering and Advanced Technology*, 8(6), 4619–4624. <https://doi.org/10.35940/ijeat.F8905.088619>
- Tang, K. H. D. (2023). Student-centered approach in teaching and learning: What does it really mean? *Acta Pedagogica Asiana*, 2(2), 72–83. <https://doi.org/10.53623/apga.v2i2.218>
- Teubner, T., Flath, C. M., Weinhardt, C., van der Aalst, W., & Hinz, O. (2023). Welcome to the era of ChatGPT. *Business & Information Systems Engineering*, 65, 95–101. <https://doi.org/10.1007/s12599-023-00795-x>
- Tuomi, I. (2023). A Framework for Socio-Developmental Ethics in Educational AI. *Proceedings of the Annual Hawaii International Conference on System Sciences*, 6208–6217. <https://doi.org/10.24251/hicss.2023.752>
- Verian, M. T. (2024). Revolutionizing TVET education: Tailored methods for tourism and hospitality skills. *2024 International Conference on TVET Excellence & Development, Melaka, Malaysia*, 146–150. <https://doi.org/10.1109/ICTeD62334.2024.10844664>
- Wang, Z., Zou, D., Zhang, R., Lee, L.-K., Xie, H., & Wang, F. L. (2026). ChatGPT-enhanced self-regulated learning in programming education: impacts on motivation, self-efficacy, and learning outcomes. *Interactive Learning Environments*, 34(5), 3041–3066. <https://doi.org/10.1080/10494820.2025.2559919>

- Wu, T. T., Lee, H. Y., Chen, P. H., Lin, C. J., & Huang, Y. M. (2025). Integrating peer assessment cycle into ChatGPT for STEM education: A randomised controlled trial on knowledge, skills, and attitudes enhancement. *Journal of Computer Assisted Learning*, 41(1), e13085. <https://doi.org/10.1111/jcal.13085>
- Yildiz, T. (2025). From constructivism to cultural cognition: A comparative analysis of Piaget, Vygotsky, and Tomasello's Theories of Cognitive Development. *Humanitas – Uluslararası Sosyal Bilimler Dergisi*, 13(25), 411–429. <https://doi.org/10.20304/humanitas.1601228>
- Yilmaz, R., & Karaoglan Yilmaz, F. G. (2023). The effect of generative artificial intelligence (AI)-based tool use on students' computational thinking skills, programming self-efficacy and motivation. *Computers and Education: Artificial Intelligence*, 4, 100147. <https://doi.org/10.1016/j.caeai.2023.100147>
- Younis, B. (2024). Developing and validating the contextual technology andragogy/pedagogy entrepreneurship work content knowledge model: A framework for vocational education. *IEEE Transactions on Education*, 68(1), 57–66. <https://doi.org/10.1109/TE.2024.3449636>
- Yuliana, Y., & Hidayat, H. (2017). How to implement technology science for entrepreneurship by using product-based learning approach and participatory action learning system in higher education? *Advanced Science Letters*, 23(11), 10918–10921. <https://doi.org/10.1166/asl.2017.10186>
- Zimmerman, B. J. (1990). Self-regulated learning and academic achievement: An overview. *Educational Psychologist*, 25(1), 3–17. https://doi.org/10.1207/s15326985ep2501_2

AUTHORS



Hendra Hidayat is a doctor in technology and vocational education. He is a lecturer in the Department of Electronics Engineering Education at the Engineering Faculty of Universitas Negeri Padang, Padang, Indonesia. His current expertise is in electronics learning technology, electronics engineering, and technopreneurship. His research interests are augmented reality for electronics engineering education, gamification, integrated mobile application-based learning, the design of prototype electronics engineering learning tools, and digital entrepreneurship with artificial intelligence and mobile apps.



Anggarda Paramita Muji is a lecturer at the Department of Early Childhood Teacher Education, Faculty of Education, Universitas Negeri Padang, Padang, Indonesia. Her expertise includes early childhood education and pedagogy of learning. Her research interests include early childhood learning and development, teacher education and teaching readiness, entrepreneurship education for young children, curriculum and program evaluation, project-based learning, and technology-enhanced learning.



Fitrika Kumala Dewi is a doctoral student in the Technological and Vocational Education program at the Faculty of Engineering, Universitas Negeri Padang, Padang, Indonesia. Her research focuses on technology in electronics engineering education, mobile technology, and augmented reality.