



Program Evaluation Models in Technical and Vocational Education and Training: A Systematic Literature Review of Scopus-Indexed Studies (2019–2026)

Muhammad 'Ariqsyah^{1*}, Ambiyar¹, Fahmi Rizal¹

¹ Universitas Negeri Padang, Indonesia

ARTICLE INFO

Article history:

Received: July, 2026

Received in revised from: July, 2026

Accepted: September, 2026

Available online: Sept, 28, 2026

Keywords: TVET; program evaluation; CIPP model; Kirkpatrick model; systematic literature review;

ABSTRACT

Program evaluation is a strategic instrument for quality assurance, accountability, and continuous improvement in Technical and Vocational Education and Training (TVET). Yet TVET evaluation practices remain fragmented across paradigms, models, indicators, and stakeholder configurations. This systematic literature review synthesizes Scopus-indexed empirical and conceptual studies published from 2019 to 2026 on TVET program evaluation. Following PRISMA 2020, 400 Scopus records retrieved through Publish or Perish were screened; 46 relevant records were identified and 33 core TVET studies were retained for thematic synthesis. Keyword co-occurrence mapping was conducted in VOSviewer at two occurrence thresholds. The findings show that high-visibility TVET literature is dominated by vocational education, systematic reviews, technology-enhanced learning, and employability, while standard program-evaluation models such as CIPP, Kirkpatrick, Stake, and Proves remain under-represented. Most studies emphasize reaction and learning outcomes, whereas behavior-level impact, results-level impact, context evaluation, and utilization of findings remain limited. The review recommends context-sensitive, mixed-method, longitudinal, and stakeholder-inclusive evaluation designs integrated with accreditation and SDG indicators.

1. Introduction

Technical and Vocational Education and Training (TVET) is positioned by UNESCO and the International Labor Organization (ILO) as a critical lever for human capital formation, decent work, and inclusive economic development. The World Economic Forum's Future of Jobs Report 2025 projects 170 million new jobs and 92 million displaced jobs, representing a structural labor market shift of 22% of the 1.2 billion formal jobs, with employers expecting 39% of key skills to change by 2030 (World Economic Forum, 2025). IBM research further estimates that workplace skills now have a half-life of approximately five years, with technical skills as short as 2.5 years (Shaji George et al., 2025). These figures explain why credible program evaluation is no longer optional for TVET; it has become structurally imperative.

* Corresponding author.

E-mail address: ariqsyah280300@student.unp.ac.id

<https://doi.org/10.56806/jh.v7i3.456>

In the Indonesian context, Presidential Instruction No. 9/2016 and Presidential Regulation No. 68/2022 on TVET revitalization, along with the Independent Learning and Independent Campus (MBKM) policy, have strengthened the demand for systematic evaluation. The Central Statistics Agency (BPS) reported an Open Unemployment Rate (TPT) of 8.63% for Vocational High School (SMK) graduates in August 2025, the highest among all education levels (BPS, 2025). Without a robust evaluation infrastructure, public investment in TVET risks degenerating into ceremonial reporting rather than evidence-based improvement (Suharno et al., 2020).

Theoretically, program evaluation has evolved into a pluralistic field encompassing post-positivist, constructivist, transformative, and pragmatic paradigms. These paradigms are operationalized through Stufflebeam's Context-Input-Process-Product (CIPP) model, Kirkpatrick's Four Levels (Reaction, Learning, Behavior, Results), Stake's Countenance and Responsiveness evaluation, Provus's Discrepancy model, Scriven's goal-free evaluation, Patton's utilization-focused evaluation, Chen's theory-based evaluation, and logic models (Frye & Hemmer, 2012; Kirkpatrick & Kirkpatrick, 2016; Patton, 2008; Stufflebeam, 2000; Stufflebeam & Coryn, 2014). Alkin's evaluation tree, structured around methods, valuing, and use, remains a central heuristic for navigating this conceptual diversity.

Despite this rich theoretical landscape, three persistent gaps characterize the TVET evaluation literature. First, single-model case studies abound, but cross-contextual syntheses comparing models, contexts, and outcomes are scarce. Second, process and product evaluations dominate the literature, while context evaluation and the utilization of evaluation findings are under-theorized. Third, stakeholder participation is generally limited to students and teachers, while industry and alumni engagement remain largely formal (Abdurrahman et al., 2022; Wahjusaputri & Bunyamin, 2022).

This Systematic Literature Review (SLR) addresses this gap by synthesizing Scopus-indexed studies published between 2019 and 2026. The 2019–2026 period covers post-PRISMA-2020 reporting standards, COVID-19 educational disruption, the industry 4.0/5.0 push, the launch of MBKM, and SDG integration. Exclusive sourcing from Scopus ensures editorial quality while also covering dominant TVET publications, including the Journal of Technical Education and Training, Journal of Vocational Education & Training, Empirical Research in Vocational Education and Training, International Journal of Training Research, International Journal of Evaluation and Research in Education, Jurnal Pendidikan Vokasi, Cakrawala Pendidikan, TEM Journal, Sustainability, and Heliyon.

The objectives of this review are fourfold: (1) to map the evaluation models, paradigms, and objectives adopted in TVET evaluation studies; (2) to characterize the evaluation designs, data collection strategies, indicators, and stakeholder configurations used; (3) to synthesize the strengths, limitations, and emerging issues identified across the studies; and (4) to generate a research agenda for TVET evaluation. The contributions of this review are theoretical (specification of a generic evaluation model for the TVET context), methodological (triangulation of the PRISMA 2020 thematic synthesis with VOSviewer bibliometric mapping), practical (provision of a comparative basis for evaluators, accreditation assessors, and policymakers), and pedagogical (learning resources for PTK lectures and doctoral dissertation research).

Program evaluation has evolved through four main paradigmatic phases. The post-positivist tradition emphasizes measurement, objective indicators, and experimental or quasi-experimental designs. The constructivist tradition emphasizes diverse stakeholder perspectives, naturalistic inquiry, and meaning-making. The transformative paradigm centers on equity, participation, and social justice, which are particularly relevant for marginalized TVET learners. The pragmatic paradigm prioritizes mixed methods, decision-making relevance, and practical utility (Frye & Hemmer, 2012; Mertens et al., 2025).

Stufflebeam's Context-Input-Process-Product (CIPP) model is a decision-oriented evaluation framework that supports both formative and summative objectives. Recent applications in TVET include evaluations of teaching factory programs (Wahjusaputri & Bunyamin, 2022), competency tests in vocational high schools (Abdurrahman et al., 2022), Field Work Practices (PKL/internships), and the Independent Curriculum. CIPP's strength lies in its comprehensive scope; its limitations include the temptation to apply it superficially as a checklist rather than as a theoretically grounded evaluation strategy.

Kirkpatrick's model operationalizes training evaluation through four hierarchical levels: Reaction, Learning, Behavior, and Results, often expanded with Phillips' Return on Investment (ROI) calculation as a fifth level (Kirkpatrick & Kirkpatrick, 2016). Alsalamah and Callinan (2021) provide a systematic adaptation for evaluating principal training. Persistent limitations of Kirkpatrick's framework include its linearity, weak empirical capture at Levels 3 (Behavior) and 4 (Results), neglect of unintended outcomes, and limited attention to context (Reio Jr. et al., 2017).

Stake's Countenance and Responsive evaluation model emphasizes descriptive metrics and responsiveness to stakeholders. Provus's Discrepancy Model identifies gaps between standards and performance and was applied by Abdurrahman et al. (2022) to automotive competency tests. Scriven's goal-free evaluation intentionally ignores stated objectives to capture the full impact of the program. The logic model traces inputs–activities–outputs–outcomes–impact and is widely used in donor-funded TVET projects. Chen's theory-driven evaluation emphasizes program theory and mechanisms of change.

Mixed-methods case studies are the modal (most frequently used) design in TVET evaluation, while quasi-experimental designs remain rare. Common indicators include competency attainment, employment rates, waiting times to first job, graduate employability, alignment with the Indonesian National Qualifications Framework (KKNI), and SDGs 4 (Quality Education), 8 (Decent Work and Economic Growth), and 9 (Industry, Innovation, and Infrastructure).

TVET evaluations involve a diverse set of stakeholders: students, teachers, principals, industry partners, alumni, parents, accreditation assessors, and regulators. Participatory and democratic evaluation approaches enhance the utility of findings, but most reviewed studies still rely on convenience sampling and report low levels of industry involvement (Semali, 2024).

Five emerging issues deserve attention: (a) digital and AI-based evaluation, including learning analytics; (b) green skills and SDG-oriented evaluation; (c) evaluation of online and hybrid TVET delivery in the post-COVID period; (d) integration of program evaluation with project management standards (e.g., PMBOK); and (e) meta-evaluation of evaluation studies themselves.

Existing systematic reviews of CIPP and Kirkpatrick have focused on medical education, nursing, and principal training. A specific synthesis, Scopus-based, 2019–2026, focusing on TVET, comparing evaluation models, methods, and emerging issues has not been published. This study fills this gap.

Based on the identified gaps in the literature, this review addresses three Research Questions (RQs):

RQ1: Which program evaluation models, paradigms, and theoretical frameworks dominate Scopus-indexed TVET evaluation studies published between 2019 and 2026, and how have their relative shares evolved over time?

RQ2: What evaluation objectives, designs, data collection strategies, indicators, and stakeholder configurations are most frequently adopted in TVET evaluations, and how do these vary by region and program type (vocational high schools, polytechnics, training programs, teaching factories, MBKM)?

RQ3: What strengths, limitations, and emerging issues including utilization of reported findings, accreditation alignment, and SDG integration, and what research agenda do they suggest for TVET evaluation in Indonesia and comparable contexts?

2. Methodology

2.1 Research Design

This study employed a Systematic Literature Review (SLR) combined with bibliometric mapping, following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guidelines (Page et al., 2021). The 2020 PRISMA statement provides up-to-date reporting guidelines covering the identification, screening, and inclusion phases and is recognized as the gold standard for transparent and replicable systematic review reporting.

2.2 Database and Rationale

This review exclusively utilizes Scopus, the world's largest curated abstract and citation database for peer-reviewed literature. Scopus was chosen over Web of Science and Google Scholar because (a) it provides broader coverage of TVET-specific publications, including those published in Asia and Europe; (b) its bibliographic metadata is structured for direct processing by VOSviewer; and (c) Scopus indexing provides a quality threshold consistent with the expectations of doctoral-level research.

2.3 Search Strategy

The literature search was conducted using Publish or Perish (PoP) version 8 connected to the Scopus API. The search was performed on the title, abstract, and author keywords. The main Boolean search string is presented below:

(TITLE-ABS-KEY ("vocational education" OR "technical and vocational education" OR "TVET" OR "vocational training" OR "vocational high school" OR "polytechnic") AND TITLE-ABS-KEY ("program evaluation" OR "training evaluation" OR "educational evaluation" OR "evaluation model" OR "CIPP" OR "Kirkpatrick" OR "Stake" OR "Provus" OR "discrepancy model" OR "logic model")) AND PUBYEAR > 2018 AND PUBYEAR < 2027 AND (LIMIT-TO (DOCTYPE , "ar") OR LIMIT-TO (DOCTYPE , "re") OR LIMIT-TO (DOCTYPE , "cp")) AND (LIMIT-TO (LANGUAGE , "English"))

2.4 Inclusion and Exclusion Criteria

Table 1. Inclusion and Exclusion Criteria

No	Criteria	Inclusion	Exclusion
1	Period	2019–2026	Before 2019 or after 2026
2	Basis data	Scopus indexed publications	Not indexed by Scopus; predatory journal
3	Document type	Articles, reviews, conference papers in journals	Books, theses, editorials, retracted articles
4	Language	English	Non-English without verified translation
5	Topic	Evaluation of programs/training in the context of TVET, vocational, polytechnics	Outside TVET (e.g. medical education, K-12 general education, corporate only)
6	Method	Empirical, conceptual, or review with explicit methods	Opinion writing, news, blogs
7	End	Clear research questions, methods, and results	Insufficient methodological detail

2.5 PRISMA 2020 Selection Process

Article selection followed PRISMA 2020 in three sequential phases: Identification, Screening, and Inclusion. The full-text eligibility step is reported as 'Reports assessed for eligibility' in the Screening phase, consistent with Page et al. (2021).

Identification: 400 records were identified from Scopus via Publish or Perish (PoP) version 8. No duplicates were found in the PoP export, so all 400 records proceeded to screening.

Screening (title/abstract): 400 records were screened based on title/abstract against the inclusion criteria. A total of 115 records were excluded because they were published before 2019, and 239 records were excluded because they were outside the scope of TVET/vocational/evaluation, leaving 46 relevant records.

Reports assessed for eligibility: 46 relevant records were assessed for eligibility. The relevance audit excluded 13 records that initially passed the keyword filter because they only matched generic terms (e.g., 'evaluation model', 'logic model', 'discrepancy model') without a vocational anchor, leaving 33 core TVET studies.

Inclusion: 33 core studies were included in the final qualitative thematic synthesis. Keyword co-occurrence mapping (VOSviewer) was run on the curated corpus (n = 400) to map macro-intellectual structures.

2.6 PRISMA 2020 Flowchart

The selection process is summarized in Figure 1 (PRISMA 2020 Flowchart). The diagram consists of three phases:

Identification phase: Records identified from Scopus via Publish or Perish (n = 400); duplicates removed before screening (n = 0).

Screening phase: Records screened (n = 400); excluded for publication before 2019 (n = 115) and outside the scope of TVET (n = 239); reports assessed for eligibility (n = 46); excluded after relevance audit (n = 13).

Inclusion phase: TVET core studies included in the thematic synthesis (n = 33); VOSviewer keyword co-occurrence mapping was run on the curated corpus (n = 400).

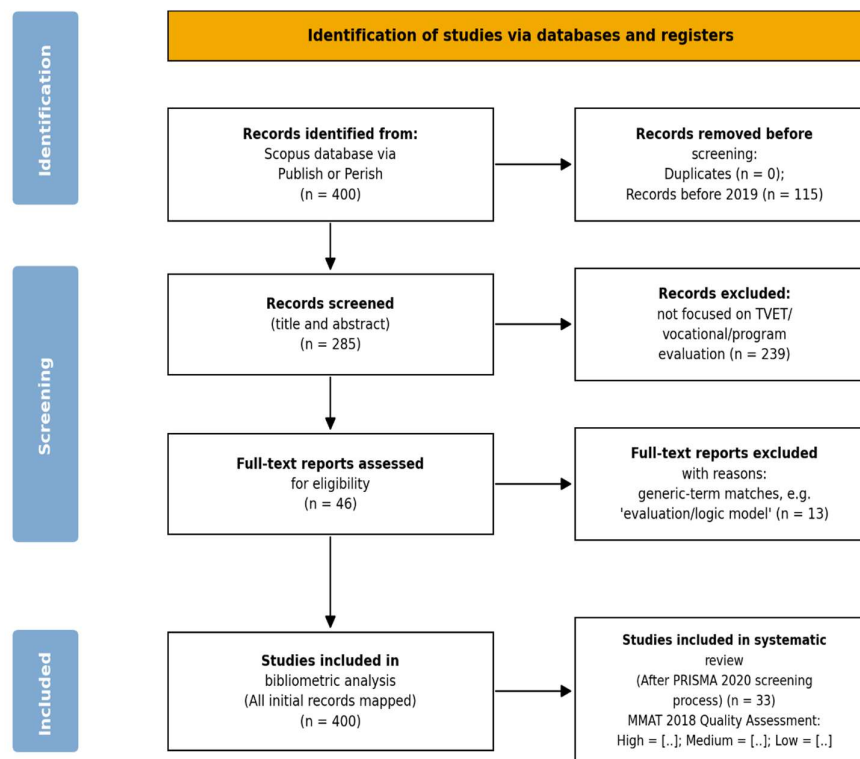


Fig. 1. Study Selection Process Flowchart (PRISMA 2020)

2.7 Data Extraction and Thematic Analysis

A standardized data extraction form captured: bibliographic data, country/region, TVET sub-context (vocational high school, polytechnic, training program, teaching factory, MBKM), evaluation paradigm, evaluation model used, evaluation objectives, design, data collection methods, indicators, stakeholder configuration, sample, key findings, and limitations. Thematic analysis followed Braun and Clarke's (2006) six-phase procedure with both deductive (theory-based) and inductive (data-based) coding. Twenty percent of included studies were double-coded by two reviewers, and inter-rater reliability was calculated using Cohen's κ , with $\kappa \geq 0.80$ set as the acceptance threshold.

2.8 Bibliometric Analysis Using VOSviewer

Bibliometric analyses were performed in VOSviewer version 1.6.20 (Eck & Waltman, 2009; van Eck & Waltman, 2023) using RIS and CSV files exported from Publish or Perish. Three complementary network analyses were performed:

2.8.1 Co-occurrence of Author Keywords

Author keywords were used as the unit of analysis with full counting. The minimum occurrence threshold was set at 5, which is expected to retain approximately 60–80 keywords meeting the threshold from the entire corpus. Anticipated thematic clusters include: (a) CIPP / vocational evaluation; (b) Kirkpatrick / training evaluation; (c) competency / assessment / accreditation; (d) Teaching Factory / MBKM / Project-Based Learning; and (e) employability / tracer study / SDGs.

2.8.2 Co-authorship Analysis

Co-authorship is analyzed at the author and country levels. Thresholds are set at a minimum of 2 documents per author and 5 documents per country. Anticipated leading countries include Indonesia, Malaysia, China, Germany, Australia, Nigeria, and Saudi Arabia.

2.8.3 Citation and Bibliographic Coupling

Citation analysis was conducted at the document and source level with a minimum citation threshold of 5. Bibliographic coupling identified clusters of documents that share common references, revealing intellectual lineages in TVET evaluation research.

2.8.4 Visualization Output

For each analysis, three visualizations were generated: (a) a network visualization, with clusters distinguished by color; (b) an overlay visualization, with nodes colored by year to reveal temporal evolution; and (c) a density visualization, presented as a heat map of conceptual hotspots.

2.9 Quality Assessment

The methodological quality of empirical studies was assessed using the Mixed Methods Appraisal Tool (MMAT) 2018, which provides separate criteria for qualitative, descriptive quantitative, randomized quantitative, non-randomized quantitative, and mixed methods studies. Systematic reviews were assessed using AMSTAR-2. Studies scoring below 60% on the MMAT were excluded from the synthesis.

2.10 Ethical Considerations and Protocol Registration

This review exclusively utilized publicly available metadata and full text from peer-reviewed publications; no primary data involving human subjects was collected. Registration of the review protocol on PROSPERO is recommended prior to data extraction to enhance transparency and reproducibility.

3. Results

This section reports the results of a synthesis of 33 TVET core studies that passed the relevance audit as well as the results of a bibliometric mapping of keyword co-occurrence conducted in VOSviewer on the curated corpus. The presentation is structured in three layers: (1) corpus characteristics, (2) intellectual structure based on two VOSviewer maps at different minimum occurrence thresholds, and (3) a thematic synthesis that answers the three research questions.

3.1 Characteristics of the Corpus

Thirty-three core studies were distributed between 2019 and 2026, with a focus on the pandemic and post-pandemic periods: 2019 (n = 5); 2020 (n = 13); 2021 (n = 4); 2022 (n = 5); 2023 (n = 3); 2024 (n = 3). Peak productivity occurred in 2020 (n = 13), coinciding with the disruption of learning caused by COVID-19, which prompted a surge in studies on online learning in vocational settings. In terms of document type, the majority were journal articles, supplemented by a number of reviews and conference papers. The most frequently cited sources were the Journal of Vocational Education and Training, the Indonesian Journal of Science and Technology, the International Journal of Educational Development, and the International Journal of Evaluation and Research in Education, confirming the representation of both mainstream and Indonesian TVET publications.

Table 2. Ten TVET Studies with the Highest Citations in the Core Corpus

No	Author (Year)	Citation	Short Title	Source
1	Antonietti dkk. (2022)	236	Can teachers' digital competence influence technology acceptance in vocational education?	Computers in Human Behavior
2	Dahalan et al. (2024)	191	Gamification and Game Based Learning for Vocational Education and Training: A Systematic Litera	Education and Information Technologies
3	Cattaneo dkk. (2022)	171	How digitalised are vocational teachers? Assessing digital competence in vocational education a	Computers and Education
4	Suharno et al. (2020)	170	Vocational education in Indonesia: History, development, opportunities, and challenges	Children and Youth Services Review
5	Beer and Mulder (2020)	145	The effects of technological developments on work and their implications for continuous vocatio	Frontiers in Psychology
6	Li and Pilz (2023)	144	International transfer of vocational education and training: a literature review	Journal of Vocational Education and Tr
7	Chiang DKK (2022)	144	Augmented reality in vocational training: A systematic review of research and applications	Computers in Human Behavior
8	Jabarullah and Iqbal Hussain (2019)	142	The effectiveness of problem-based learning in technical and vocational education in Malaysia	Education and Training
9	Syauqi et al. (2020)	131	Students' perceptions toward vocational education on online learning during the COVID-19 pandem	International Journal of Evaluation an
10	Zeng dkk. (2019)	124	Prevalence and correlates of depression, anxiety and symptoms of stress in vocational college n	Psychology Health and Medicine

The most frequently cited studies focused on the digital competencies of vocational teachers (Antonietti et al., 2022; Cattaneo et al., 2022), gamification and augmented reality in vocational training (Chiang et al., 2022; Dahalan et al., 2024), and studies of TVET systems and policies (Li & Pilz, 2023; McGrath & Yamada, 2023; Suharno et al., 2020). This pattern suggests that the corpus is more focused on learning innovation and policy than on the explicit application of program evaluation models.

3.2 Intellectual Structure: Keyword Co-occurrence Analysis (VOSviewer)

Keyword co-occurrence mapping was performed with two minimum occurrence thresholds to test the stability of the cluster structure. Keywords were derived from the title field and standardized through the thesaurus file (metadata export does not include author keywords). Figure 2 presents the map at a minimum threshold of 5 occurrences, while Figure 3 presents the map at a minimum threshold of 2 occurrences.

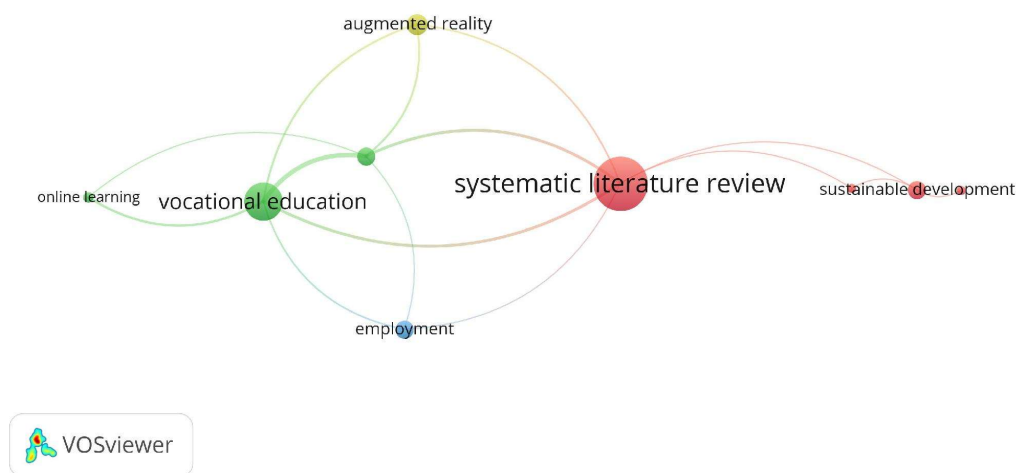


Fig. 2. Keyword Co-Occurrence Network Map (minimum occurrence threshold = 5)

At a more stringent threshold (Figure 2), the structure simplifies to a robust core. The two largest nodes are “systematic literature review” and “vocational education,” indicating that systematic reviews have become a common methodological approach while simultaneously affirming vocational education as a thematic axis. Four color-distinct clusters are formed: (i) a vocational education cluster linked to “online learning”; (ii) a methodological-conceptual cluster linking “systematic literature review” to “sustainable development”; (iii) an “augmented reality” cluster; and (iv) an “employment” cluster. The strongest link connects “vocational education” to “systematic literature review” and “online learning,” indicating that the most mature body of literature lies at the intersection of vocational education, online learning, and literature synthesis.

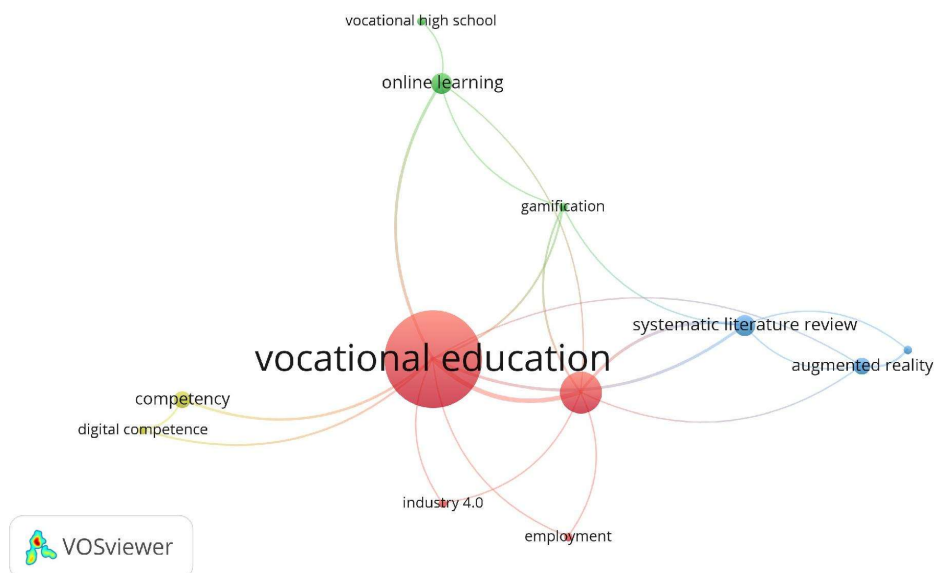


Fig. 3. Keyword Co-Occurrence Network Map (minimum occurrence threshold = 2)

At a looser threshold (Figure 3), more terms pass, resulting in a richer cluster structure, and “vocational education” emerges as the central node with the highest total link strength. Four clusters are identified: (1) the red cluster of vocational education, industry 4.0, and employment represents the axis of vocational education, industrial transformation, and employment outcomes; (2) the green cluster of online learning, vocational high school, and gamification represents learning innovations at the vocational secondary level; (3) the blue cluster of systematic literature review and augmented reality represents immersive technologies and their methodological synthesis; and (4) the yellow cluster of competency and digital competence represents the competency agenda. A comparison of the two maps reveals a stable core (vocational education, online learning, augmented reality, systematic literature review) and a threshold-sensitive periphery (competency, gamification, vocational high school, industry 4.0).

The most striking bibliometric finding is its apparent absence: not a single keyword node referring to the standard program evaluation models CIPP, Kirkpatrick, Stake, or Provus reached the threshold for occurrence. This indicates that, in the high-visibility TVET literature from 2019–2026, model-specific program evaluation has not yet become a dominant discourse and remains fragmented as single case studies. This finding provides the empirical basis for the gap argument raised in this review.

3.3 Thematic Synthesis

A close reading of the 33 core studies yielded five themes aligned with the VOSviewer clusters, plus one peripheral theme.

Theme 1: Systems, policies, and strengthening the position of vocational education. A group of studies frame TVET at the systems and policy levels: the history and challenges of vocational education in Indonesia (Suharno et al., 2020), the international transfer of TVET systems (Li & Pilz, 2023), skills trends for development (McGrath et al., 2020; McGrath & Yamada, 2023), strengthening the standing of vocational education (Billett, 2020), the contribution of vocational institutions in manufacturing areas (Lund & Karlsen, 2020), vocational governance reform in China (Ling et al., 2023), and Indonesia's vision of competitiveness and human capital (Indrawati & Kuncoro, 2021). This theme emphasizes the urgency of evaluation at the policy level, which is the least operationalized using formal evaluation models.

Theme 2: Online learning and COVID-19 disruption. The 2020 surge was driven by studies of online learning: distance learning in West Java vocational schools (Mulyanti et al., 2020), student perceptions of online learning (Syauqi et al., 2020), development of an LMS-based e-learning system (Wagino et al., 2023, 2026), online instructional design based on self-regulated learning (Sangsawang, 2020), and analysis of practicum content through expert systems (Ana, 2020). These studies primarily measured reactions and perceptions (equivalent to Kirkpatrick's Levels 1–2) but rarely explored behavioral impacts and outcomes (Levels 3–4).

Theme 3: Immersive technologies, gamification, and digital competencies. The next largest cluster relates to learning technologies: a systematic review of gamification (Dahalan et al., 2024), gamification in blended e-learning (Jayalath & Esichaikul, 2022), a systematic review of augmented reality in vocational training (Chiang et al., 2022), a motivational AR application framework (Acosta et al., 2019), mobile-based AR (Radosavljevic et al., 2020), simulators and VR for welder training (Lavrentieva et al., 2020), and vocational teachers' digital competencies and technology acceptance (Antonietti et al., 2022; Cattaneo et al., 2022). This theme is the most methodologically rich, but almost all of them evaluate learning interventions, not comprehensive TVET programs.

Theme 4: Competence, 21st-Century Skills, and Work Readiness. Studies in this theme highlight the challenges of implementing 21st-century skills (Mutohhari et al., 2021), the effectiveness of

problem-based learning in TVET in Malaysia (Jabarullah & Iqbal Hussain, 2019), and the adoption of technology for vocational education through PLS-SEM (Chatterjee et al., 2021; Wagino et al., 2024). Recurring indicators include competency achievement and work readiness, which can potentially be mapped to the Product dimension in CIPP and the Level of Learning in Kirkpatrick.

Theme 5: Employability, Industry 4.0, and Sustainability. This theme links TVET to employment outcomes and industrial transformation: the impact of TVET on adult skills and employment (Choi et al., 2019), the comparison of vocational versus general education on employment and earnings (Korber & Oesch, 2019), the implications of technological developments for continuous VET (Beer & Mulder, 2020), the impact of the fourth industrial revolution on vocational skills (Spöttl & Windelband, 2021), and the use of Industry 4.0 technologies in Nigerian TVET (George et al., 2024). The appearance of “sustainable development” on the map suggests a link to the SDG agenda, although explicit measurement of SDG indicators remains rare.

Peripheral theme: Well-being of vocational students. A small number of studies have addressed psychosocial dimensions of students, such as symptoms of depression and anxiety in vocational nursing students (Zeng et al., 2019) and bullying in vocational schools (Xu et al., 2022). While peripheral to program evaluation, this theme highlights a relevant but overlooked dimension of context (Context in CIPP).

3.4 Discussion and Answers to Research Questions

RQ1 (dominant models and paradigms). Contrary to initial expectations, the 2019–2026 high-perception corpus does not place standard program evaluation models as a central discourse; none of the CIPP, Kirkpatrick, Stake, or Provus nodes reached the threshold on either map. Vocational education as the object, systematic reviews as the method, and learning technologies and employability as the issues are dominant. Thus, the answer to RQ1 is corrective: model-based program evaluation in TVET remains fragmented and under-indexed, representing a gap rather than a strength in the mainstream literature.

RQ2 (design, indicators, stakeholders). The most frequently encountered designs were perception surveys, mixed-methods case studies, and systematic reviews. Recurring indicators included competencies (including digital competencies), employability, employment and earnings, and technology acceptance. The stakeholder configuration was dominated by students and teachers; industry and alumni engagement was largely implicit. The geographic context centered on Indonesia/vocational high schools and a few European (e.g., Switzerland) and African contexts, in line with the employment and Industry 4.0 clusters in Figure 3.

RQ3 (strengths, limitations, agenda). The strengths of the literature lie in its methodological diversity and integration of learning technologies. The main limitations are the lack of theory- and model-based evaluations, weak evaluation of context and utilization of findings, and shallow capture of behavioral/outcome impacts (Kirkpatrick Levels 3–4). The proposed research agenda: (a) application of CIPP and Kirkpatrick in a theory-based manner, not as a checklist; (b) a hybrid CIPP–Kirkpatrick design to bridge program and training evaluation; (c) a rigorous tracer study to close the Level 4 gap; (d) substantive industry and alumni engagement; and (e) explicit integration of SDG indicators and green skills.

3.5 Limitations

Several limitations should be noted for a balanced interpretation. First, the metadata is sourced from a Publish or Perish export that does not include abstracts or author keywords; therefore, keywords for co-occurrence analysis were derived from titles and standardized through a thesaurus, so the map reflects title terminology, not author indexing. Second, the automated search also

included records from adjacent domains; 13 records that initially passed keyword screening were excluded through a relevance audit because they only matched generic terms (e.g., “evaluation model,” “logic model”), so the thematic synthesis relied on 33 core studies. Third, the RIS export from Publish or Perish does not include a list of cited references, so citation analysis and bibliographic coupling are suboptimal; it is recommended that the map be regenerated from a direct Scopus export for validation. Fourth, the reported map is generated from a broader curated corpus; replication on a more rigorous core corpus is recommended for final publication.

4. Conclusions

This systematic literature review maps and synthesizes the Scopus-indexed TVET program evaluation literature for the 2019–2026 period using a combination of PRISMA 2020 and VOSviewer bibliometric mapping. Of the 400 records identified, 33 core studies were thematically synthesized. Three main conclusions can be drawn. First (RQ1), contrary to initial expectations, standard program evaluation models—CIPP, Kirkpatrick, Stake, or Provus—do not appear as keyword nodes in either map; the dominant discourse is vocational education as the object, systematic reviews as the method, and learning technologies (augmented reality, gamification, online learning) and employability as the issues. Thus, model-based program evaluation in TVET remains fragmented and has not yet become mainstream—a gap, not a strength. Second (RQ2), common designs are perception surveys, mixed-methods case studies, and systematic reviews, with recurring indicators of competency attainment (including digital competencies), work readiness, and employment and earnings. Stakeholder configurations are still dominated by students and teachers, while industry and alumni involvement is largely implicit. Third (RQ3), the strengths of the literature lie in methodological diversity and integration of learning technologies, while limitations include weak theory- and model-based evaluations, shallow capture of behavioral impacts and outcomes (Kirkpatrick's Levels 3–4), and minimal context evaluation and utilization of findings.

4.1 Implications and Recommendations

Theoretically, these findings underscore the need to specify a context-sensitive evaluation model for TVET and reposition model-based evaluation as a priority agenda, not a supplement. Methodologically, triangulation of PRISMA 2020 with VOSviewer mapping proved useful for reading intellectual structures and conducting in-depth synthesis; for further research, it is recommended to use metadata export directly from Scopus to optimize citation analysis and bibliographic coupling. Practically, the review results provide a comparative basis for evaluators, accreditation assessors, and vocational program managers in selecting and implementing appropriate evaluation models. Policy-wise, TVET program evaluation needs to be integrated with accreditation regimes (BAN-PT, IABEE, AUN-QA, and the like) and Sustainable Development Goal indicators (SDGs 4, 8, and 9), so that evaluation does not cease to be a ceremonial reporting but becomes an evidence-based improvement instrument.

4.2 Further Research Agenda

Based on the identified gaps, five research directions are recommended: (1) the application of CIPP and Kirkpatrick in a theory-based manner—not as a checklist—with impact measurement at Levels 3–4; (2) the development of a hybrid CIPP–Kirkpatrick design to bridge program evaluation and training evaluation; (3) the implementation of a rigorous, longitudinal tracer study to close the gap in evaluating outcomes and graduate utilization; (4) substantive industry and alumni engagement through a participatory and democratic evaluation approach; and (5) the explicit integration of green skills and SDG indicators, including exploration of digital and artificial intelligence-based evaluation.

With this agenda, TVET program evaluation is expected to move from the dominance of single case studies to theory-based, context-sensitive, and utilization-oriented evaluation practices, in line with the demands of skills transformation in the industry 4.0/5.0 era.

References

- Abdurrahman, Parmin, & Muryanto, S. (2022). Evaluation on the automotive skill competency test through 'discontinuity' model and the competency test management of vocational education school in Central Java, Indonesia. *Heliyon*, 8(2). <https://doi.org/10.1016/j.heliyon.2022.e08872>
- Acosta, J. L. B., Navarro, S. M. B., Gesa, R. F., & Kinshuk, K. (2019). Framework for designing motivational augmented reality applications in vocational education and training. *Australasian Journal of Educational Technology*, 35(3). <https://doi.org/10.14742/ajet.4182>
- Alsalamah, A., & Callinan, C. (2021). Adaptation of Kirkpatrick's Four-Level Model of Training Criteria to Evaluate Training Programmes for Head Teachers. *Education Sciences*, 11(3), 116. <https://doi.org/10.3390/educsci11030116>
- Ana, A. (2020). Trends in Expert System Development: A Practicum Content Analysis in Vocational Education for Over Grow Pandemic Learning Problems. *Indonesian Journal of Science and Technology*, 5(2), 246–260. <https://doi.org/10.17509/ijost.v5i2.24616>
- Antonietti, C., Cattaneo, A., & Amenduni, F. (2022). Can teachers' digital competence influence technology acceptance in vocational education? *Computers in Human Behavior*, 132, 107266. <https://doi.org/10.1016/j.chb.2022.107266>
- Central Bureau of Statistics of Indonesia. (2025). Condition of the Labor Force in Indonesia August 2025. <https://www.bps.go.id/id/publication/2025/12/19/42a75ee61332755586fdfcdd/keadaan-angkatan-kerja-di-indonesia-agustus-2025.html>
- Beer, P., & Mulder, R. H. (2020). The Effects of Technological Developments on Work and Their Implications for Continuous Vocational Education and Training: A Systematic Review. *Frontiers in Psychology*, 11. <https://doi.org/10.3389/fpsyg.2020.00918>
- Billett, S. (2020). Perspectives on enhancing the standing of vocational education and the occupations it serves. *Journal of Vocational Education & Training*, 72(2), 161–169. <https://doi.org/10.1080/13636820.2020.1749483>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Cattaneo, A. A. P., Antonietti, C., & Rauseo, M. (2022). How digitalised are vocational teachers? Assessing digital competence in vocational education and looking at its underlying factors. *Computers & Education*, 176, 104358. <https://doi.org/10.1016/j.compedu.2021.104358>
- Chatterjee, S., Bhattacharjee, K. K., Tsai, C.-W., & Agrawal, A. K. (2021). Impact of peer influence and government support for successful adoption of technology for vocational education: A quantitative study using PLS-SEM technique. *Quality & Quantity*, 55(6), 2041–2064. <https://doi.org/10.1007/s11135-021-01100-2>
- Chiang, F.-K., Shang, X., & Qiao, L. (2022). Augmented reality in vocational training: A systematic review of research and applications. *Computers in Human Behavior*, 129, 107125. <https://doi.org/10.1016/j.chb.2021.107125>
- Choi, S. J., Jeong, J. C., & Kim, S. N. (2019). Impact of vocational education and training on adult skills and employment: An applied multilevel analysis. *International Journal of Educational Development*, 66, 129–138. <https://doi.org/10.1016/j.ijedudev.2018.09.007>
- Dahalan, F., Alias, N., & Shaharom, M. S. N. (2024). Gamification and Game Based Learning for Vocational Education and Training: A Systematic Literature Review. *Education and Information Technologies*, 29(2), 1279–1317. <https://doi.org/10.1007/s10639-022-11548-w>
- Eck, N. van, & Waltman, L. (2009). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, 84(2), 523–538. <https://doi.org/10.1007/s11192-009-0146-3>
- Frye, A. W., & Hemmer, P. A. (2012). Program evaluation models and related theories: AMEE Guide No. 67. *Medical Teacher*, 34(5), e288–e299. <https://doi.org/10.3109/0142159X.2012.668637>
- George, W. K., Silas, E. I., Pandey, D., Pandey, B. K., George, W. K., Silas, E. I., Pandey, D., & Pandey, B. K. (2024). Utilization of Industry 4.0 Technologies in Nigerian Technical and Vocational Education: A Conundrum for Educators [Chapter]. IGI Global Scientific Publishing. (utilization-of-industry-40-technologies-in-nigerian-technical-and-vocational-education). <https://Services.igi->

- Global.Com/Resolvedoi/Resolve.Asp?Doi=10.4018/978-1-6684-9962-7.Ch014. <https://doi.org/10.4018/978-1-6684-9962-7.ch014>
- Indrawati, S. M., & Kuncoro, A. (2021). Improving Competitiveness Through Vocational and Higher Education: Indonesia's Vision For Human Capital Development In 2019–2024. *Bulletin of Indonesian Economic Studies*, 57(1), 29–59. <https://doi.org/10.1080/00074918.2021.1909692>
- Jabarullah, N. H., & Iqbal Hussain, H. (2019). The effectiveness of problem-based learning in technical and vocational education in Malaysia. *Education + Training*, 61(5), 552–567. <https://doi.org/10.1108/ET-06-2018-0129>
- Jayalath, J., & Esichaikul, V. (2022). Gamification to Enhance Motivation and Engagement in Blended eLearning for Technical and Vocational Education and Training. *Technology, Knowledge and Learning*, 27(1), 91–118. <https://doi.org/10.1007/s10758-020-09466-2>
- Kirkpatrick, J. D., & Kirkpatrick, W. K. (2016). Kirkpatrick's four levels of training evaluation. Association for Talent Development.
- Korber, M., & Oesch, D. (2019). Vocational versus general education: Employment and earnings over the life course in Switzerland. *Advances in Life Course Research*, 40, 1–13. <https://doi.org/10.1016/j.alcr.2019.03.003>
- Lavrentieva, O. O., Arkhypov, I. O., Kuchma, O. I., & Uchitel, A. D. (2020). Use of simulators together with virtual and augmented reality in the system of welders' vocational training: Past, present, and future. *Augmented Reality in Education : Proceedings of the 2nd International Workshop (AREdu 2019)*, Kryvyi Rih, Ukraine, March 22, 2019, (2547), 201–216. <http://ceur-ws.org/Vol-2547/paper15.pdf>
- Li, J., & Pilz, M. (2023). International transfer of vocational education and training: A literature review. *Journal of Vocational Education & Training*, 75(2), 185–218. <https://doi.org/10.1080/13636820.2020.1847566>
- Ling, Y., Chung, S. J., & Wang, L. (2023). Research on the reform of management system of higher vocational education in China based on personality standard. *Current Psychology*, 42(2), 1225–1237. <https://doi.org/10.1007/s12144-021-01480-6>
- Lund, H. B., & Karlsen, A. (2020). The importance of vocational education institutions in manufacturing regions: Adding content to a broad definition of regional innovation systems. *Industry and Innovation*, 27(6), 660–679. <https://doi.org/10.1080/13662716.2019.1616534>
- McGrath, S., Ramsarup, P., Zeelen, J., Wedekind, V., Allais, S., Lotz-Sisitka, H., Monk, D., Openjuru, G., & Russon, J.-A. (2020). Vocational education and training for African development: A literature review. *Journal of Vocational Education & Training*, 72(4), 465–487. <https://doi.org/10.1080/13636820.2019.1679969>
- McGrath, S., & Yamada, S. (2023). Skills for development and vocational education and training: Current and emergent trends. *International Journal of Educational Development*, 102, 102853. <https://doi.org/10.1016/j.ijedudev.2023.102853>
- Mertens, D. M., Hall, J. N., & Wilson, A. T. (2025). *Program evaluation theory and practice*. Guilford Publications.
- Mulyanti, B., Purnama, W., & Pawinanto, R. E. (2020). Distance Learning in Vocational High Schools during the COVID-19 Pandemic in West Java Province, Indonesia. *Indonesian Journal of Science and Technology*, 5(2), 271–282. <https://doi.org/10.17509/ijost.v5i2.24640>
- Mutohahari, F., Sutiman, S., Nurtanto, M., Kholifah, N., & Samsudin, A. (2021). Difficulties in implementing 21st century skills competence in vocational education learning. *International Journal of Evaluation and Research in Education (IJERE)*, 10(4), 1229–1236. <https://doi.org/10.11591/ijere.v10i4.22028>
- 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>
- Patton, M. Q. (2008). *Utilization-focused evaluation*. Sage publications.
- Radosavljevic, S., Radosavljevic, V., & Grgurovic, B. (2020). The potential of implementing augmented reality into vocational higher education through mobile learning. *Interactive Learning Environments*, 28(4), 404–418. <https://doi.org/10.1080/10494820.2018.1528286>
- ReioJr., T. G., Rocco, T. S., Smith, D. H., & Chang, E. (2017). A Critique of Kirkpatrick's Evaluation Model. *New Horizons in Adult Education and Human Resource Development*, 29(2), 35–53. <https://doi.org/10.1002/nha3.20178>

- Sangsawang, T. (2020). An Instructional Design for Online Learning in Vocational Education according to a Self-Regulated Learning Framework for Problem Solving during the CoViD-19 Crisis. *Indonesian Journal of Science and Technology*, 5(2), 283–298. <https://doi.org/10.17509/ijost.v5i2.24702>
- Semali, L. (2024). Leveraging multistakeholder partnerships in technical, vocational education, and training. *Vocation, Technology & Education*, 1(1), e3. <https://doi.org/10.54844/vte.2024.0558>
- Shaji George, A., Baskar, T., & Siranchuk, N. (2025). Navigating the AI Disruption: Strategic Frameworks for Workforce Transformation and Innovation in the Age of Intelligent Automation. *Partners Universal International Innovation Journal*, 3(3), 83–95. <https://puuij.com/index.php/research/article/view/189>
- Spöttl, G., & Windelband, L. (2021). The 4th industrial revolution – its impact on vocational skills. *Journal of Education and Work*, 34(1), 29–52. <https://doi.org/10.1080/13639080.2020.1858230>
- Stufflebeam, D. L. (2000). The CIPP Model for Evaluation. Dalam D. L. Stufflebeam, G. F. Madaus, & T. Kellaghan (Ed.), *Evaluation Models: Viewpoints on Educational and Human Services Evaluation* (hlm. 279–317). Springer Netherlands. https://doi.org/10.1007/0-306-47559-6_16
- Stufflebeam, D. L., & Coryn, C. L. (2014). *Evaluation theory, models, and applications*. John Wiley & Sons.
- Suharno, Pambudi, N. A., & Harjanto, B. (2020). Vocational education in Indonesia: History, development, opportunities, and challenges. *Children and Youth Services Review*, 115, 105092. <https://doi.org/10.1016/j.childyouth.2020.105092>
- Syauqi, K., Munadi, S., & Triyono, M. B. (2020). Students' perceptions toward vocational education on online learning during the COVID-19 pandemic. *International Journal of Evaluation and Research in Education (IJERE)*, 9(4), 881–886. <https://doi.org/10.11591/ijere.v9i4.20766>
- van Eck, N. J., & Waltman, L. (2023). *VOSviewer Manual*. http://www.vosviewer.com/documentation/Manual_VOSviewer_1.6.1.pdf
- Wagino, W., Habibullah, H., Nasution, I., Giatman, M., Koto, R. D., Samala, A. D., Criollo-C, S., & Polin, K. C. (2024). Enhancing Academic Achievement and Critical Thinking through e-Learning: Exploring Motivation, Creativity, Participation, and Gadget Utilization in Higher Education. *International Journal of Information and Education Technology*, 14(12), 1679–1687. <https://doi.org/10.18178/ijiet.2024.14.12.2198>
- Wagino, W., Maksum, H., Purwanto, W., Krismadinata, K., Suhendar, S., & Koto, R. D. (2023). Exploring the Full Potential of Collaborative Learning and E-Learning Environments in Universities: A Systematic Review. *TEM Journal*, 1772–1785. <https://doi.org/10.18421/TEM123-60>
- Wagino, W., Tymofiiv, V., Polin, K., Milana, M., Astuti, M., & Koto, R. D. (2026). Longitudinal Analysis of Sustainable Impact of e-Learning Integration on Academic Capabilities, Critical Thinking Skills, and Digital Behavior of Automotive Engineering Students. *International Journal of Information and Education Technology*, 16(1), 259–277. <https://doi.org/10.18178/ijiet.2026.16.1.2500>
- Wahjusaputri, S., & Bunyamin, B. (2022). Development of teaching factory competency-based for vocational secondary education in Central Java, Indonesia. *International Journal of Evaluation and Research in Education (IJERE)*, 11(1), 353–360. <https://doi.org/10.11591/ijere.v11i1.21709>
- World Economic Forum. (2025). *The Future of Jobs Report 2025*. World Economic Forum. <https://www.weforum.org/publications/the-future-of-jobs-report-2025/>
- Xu, S., Ren, J., Li, F., Wang, L., & Wang, S. (2022). School Bullying Among Vocational School Students in China: Prevalence and Associations With Personal, Relational, and School Factors. *Journal of Interpersonal Violence*, 37(1–2), NP104–NP124. <https://doi.org/10.1177/0886260520907360>
- Zeng, Y., Wang, G., Xie, C., Hu, X., & Reinhardt, J. D. (2019). Prevalence and correlates of depression, anxiety and symptoms of stress in vocational college nursing students from Sichuan, China: A cross-sectional study. *Psychology, Health & Medicine*, 24(7), 798–811. <https://doi.org/10.1080/13548506.2019.1574358>