JETAL: JOURNAL OF ENGLISH TEACHING & APPLIED LINGUISTICS

VOLUME (7) Number (1) Page (55-71) E-ISSN: 2714-9811

# The Implementation of Context, Input, Process, Product Evaluation Model for Assessing Educational Programs

***Enung Nurhasanah[1], I Gede Ratnaya[2], I Gusti Lanang Agung Parwata[3], I Dewa Ayu Made Budhyana[4]***

[1234]Ganesha University of Education

[2]Email: enung@student.undiksha.ac.id

## ABSTRACT

The objective of this study is to undertake a critical examination of the implementation of the CIPP (Context, Input, Process, Product) evaluation model in junior high school educational programs. This examination will be facilitated through a systematic review of 51 studies published between 2015 and 2024. Utilizing a systematic literature review (SLR) approach, this research explores the application contexts, theoretical frameworks, scholarly debates, as well as the practical implications and limitations of using the CIPP model across various educational sectors. The present study utilizes a primary instrument in the form of a data categorization checklist based on the four components of the CIPP model. The data were collected through systematic access to digital journal databases, such as Scopus, with articles filtered based on specific keywords according to inclusion criteria. The selection of articles was conducted in accordance with the PRISMA flow, encompassing the phases of identification, screening, eligibility assessment, and final inclusion. Thematic analysis was employed to code and interpret the content of the selected studies, with a focus on the type of program, evaluation approach, CIPP dimensions analyzed, and key findings. The findings indicate that the CIPP model has been extensively adopted due to its comprehensive and adaptable structure, which effectively assesses the planning, execution, and outcomes of educational programs. The integration of authentic assessment and competency-based learning is a common practice, as it aligns with specific contextual needs. However, the review also identifies key limitations, including the predominance of quantitative methods, uneven geographic distribution of studies, and limited exploration of technology integration.

**Keywords**: *CIPP model; educational evaluation; SLR; junior high school education; program assessment; systematic review;*

## Introduction

Introduction (without subsection, 2-3 pages) includes background, objectives, methods, and literature reviews/theoretical construct (if needed) of the research. The introduction section ends with an emphasis on items to be discussed. The introduction consists of a background of the study explaining the actual phenomenon that has been investigated, supported by references and previous studies that have been done individually or in a group or team. The author must also explain the existence of this research compared to the previous studies. Introduction consists of problem(s) (one problem that is becoming the focus of the study is even better), Novelty of the study, theory used to solve the problem(s). All sources that are cited or paraphrased should be listed in the references list. An introduction does not allow a subchapter.

In the era of globalization and rapid technological advancement, education systems are increasingly required to be accountable, transparent, and outcome-oriented. Educational program evaluations are no longer confined to measuring learning outcomes alone; they must also consider the entire instructional process, from planning to implementation (Farrell, 2020). According to UNESCO (2023), approximately 70% of member countries have identified educational evaluation reform as a top priority, particularly in primary and secondary education, to improve learning quality. This aligns with the growing demand to ensure that curricula,

instructional delivery, and resource allocation function effectively in developing competent and adaptive graduates. Among the emerging approaches in educational evaluation, context-based and comprehensive models such as the CIPP model Context, Input, Process, Product developed by Stufflebeam, are gaining increased attention.

Over the past five years, there has been a significant rise in studies examining the effectiveness and implementation of evaluation models. This surge is largely driven by the increasing need to align educational systems with both local and global demands. Research by.(Li & Hu, 2022) highlights that model-based evaluations like CIPP help bridge the gap between theoretical learning principles and practical educational delivery by tailoring interventions to institutional contexts. Similarly, Jang et al. (2024) demonstrated that the CIPP model is effective for evaluating training programs with diverse participant backgrounds, including migrant workers who face linguistic and cultural barriers. These findings affirm the model's flexibility and relevance across varied educational settings. Nevertheless, despite its growing application in multiple sectors, a systematic analysis focusing specifically on its implementation in junior high school programs remains notably absent.

The central issue underlying this study is the lack of comprehensive documentation and critical analysis concerning the application of the CIPP model at the junior high school level. This stage of education represents a vital transitional period in students' cognitive, emotional, and social development. Unfortunately, many educational programs at this level are still evaluated in a fragmented manner, often focusing solely on outcomes while neglecting the quality of the processes, the relevance of inputs, and contextual appropriateness (Al-Shanawani, 2019). Molope and Oduaran (2019) emphasize that in many developing countries, educational program evaluations frequently overlook critical elements of program planning, such as resource adequacy and implementation strategies. Consequently, educational policies are often not fully informed by valid and academically sound data (Molope & Oduaran, 2020).

Several prior studies have attempted to address these challenges by employing the CIPP evaluation model. For instance, (Suhonen & Sutinen, 2022). applied the model in assessing an international e-learning course in pharmaceutical cold chain management, demonstrating its ability to capture the multidimensionality of educational programs (Heard-Lauréote & Buckley, 2025). further developed the use of the CIPP model in evaluating Continuing Professional Learning (CPL) programs, integrating it with learning cultures theory and Carl Rogers' theory of trust. Their study underscored the importance of blending evaluative and psychological dimensions to create a more humanistic and context-sensitive approach. However, these studies did not specifically examine the model's application in formal junior high school education, which possesses unique curricular structures, learning objectives, and student characteristics.

Other studies have explored the adaptation of the CIPP model in secondary education contexts. (Farrell, 2020), for example, evaluated the effectiveness of international market simulations as authentic assessments using the CIPP model, (D. Y. K. Chong et al., 2020) emphasized the model's utility in enhancing clinical competencies among physiotherapy students. Perry et al. (2020) used the CIPP framework to assess self-directed writing tasks in Canadian middle schools, focusing on the development of students' critical thinking and reflective abilities. Although these studies are relevant, they remain sector-specific and do not offer a systematic mapping of how the CIPP model has been consistently applied at the junior high school level. Hence, there remains a need for a comprehensive synthesis of diverse CIPP evaluation practices within this educational tier.

Studies by Tuna and (Basaran et al., 2021) in Turkey and (Ismail et al., 2023) in Iran indicate that the CIPP model is adaptable to different educational systems, even in countries with contrasting sociocultural backgrounds. However, most of these studies focus on a single dimension of the model such as process or product without providing an integrated analysis of all four components. In contrast, research by (Jang et al., 2025) reveals the complexity of implementing the CIPP model within multicultural and multisectoral systems. This underscores the need for a comprehensive literature synthesis that can evaluate the model's full application in similarly complex educational environments, such as junior high schools. As such, this study aims to offer a critical mapping and evidence-based analysis of CIPP implementation within junior secondary education.

A key gap in the literature is the lack of systematic and explicit documentation regarding the use of the CIPP model in junior high school educational programs. Although many studies incorporate individual components of the model, the evaluative processes often remain fragmented and lack standardization. Moreover, there is no

comprehensive review that critically consolidates findings from the past decade (2015–2024) in the context of lower secondary education. This limitation hampers the ability of policymakers and education practitioners to adopt evidence-based approaches in designing and evaluating school programs. Therefore, this research seeks to establish a scientifically grounded foundation for identifying best practices, implementation challenges, and the broader impact of the CIPP model at the junior high school level.

This study aims to critically analyze and systematically synthesize the implementation of the CIPP Evaluation Model in junior high school educational programs from 2015 to 2024**.** Its novelty lies in the application of the Systematic Literature Review (SLR) method, which enables rigorous literature screening based on explicit inclusion criteria and facilitates thematic analysis of each CIPP component. The study will identify common strategies, effectiveness levels, and implementation barriers, providing practical insights for more adaptive and accountable educational policy and program development. The scope of this research is limited to empirical studies published in peer-reviewed academic journals that explicitly reference or apply the CIPP model within formal junior high school education programs. By doing so, this study addresses both scholarly and practical needs in strengthening context-based, data-informed educational evaluation systems.

## Method

This study employed a Systematic Literature Review (SLR) as its primary research design to critically and comprehensively investigate the implementation of the CIPP Evaluation Model within Junior High School Educational Programs over the period from 2015 to 2024**.** The SLR approach was chosen due to its capacity to facilitate a transparent, structured, and replicable process of literature selection, analysis, and synthesis. Rather than involving human participants, the subjects of this study were peer-reviewed academic journal articles that explicitly applied or examined the CIPP model in the context of formal junior secondary education. Inclusion criteria were defined as follows: (1) articles written in English or Indonesian; (2) published between 2015 and 2024; (3) explicitly applying one or more components of the CIPP model context, input, process, or product in junior high school settings; and (4) empirical in nature, thus excluding purely conceptual or theoretical discussions. Exclusion criteria included articles that merely mentioned the CIPP model without practical application, or those applied outside of the junior high school context.

The initial search process yielded 170 articles**.** Following the *removal* of 72 duplicates and 63 articles outside the target time range, 50 articles remained. An additional 9 articles were excluded due to the absence of abstracts or improper indexing, leaving 42 articles for further screening. Among these**,** 24 articles did not meet substantive criteria such as insufficient application depth or lack of relevance to educational programs and 30 articles could not be retrieved in full-text form. Simultaneously, 9 additional articles were obtained from reputable alternative platforms. In *total*, 51 articles were selected for final review: 42 from Scopus and 9 from supplementary sources**.** The primary instrument utilized in this study was a *data categorization checklist*, developed in alignment with the four components of the CIPP framework. The data categorization checklist is a tool used to classify information based on specific categories, such as CIPP components (Context, Input, Process, Product). In the course of the research, the articles were meticulously analyzed, with the relevant sections of each component being marked by the researchers. The information is then systematically recorded in a table or checklist format. Following the categorization of all data, researchers proceed to the analysis of patterns or findings that emerge from each category. This instrument is designed to facilitate the uniformity, organization, and replicability of data analysis processes.

Data collection was carried out by systematically accessing digital journal databases (e.g., Scopus and national academic repositories) and filtering articles based on keyword combinations such as *"CIPP model"*, *"evaluation"*, *"junior high school"*, and *"secondary education"* in accordance with the inclusion criteria. The article selection process adhered strictly to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow, encompassing stages of identification, screening, eligibility review, and final inclusion.

Subsequently, thematic analysis was employed to code and interpret the content of the selected studies, focusing on themes such as program type, evaluation approaches, analyzed CIPP dimensions, and key findings. All coding and interpretations were independently reviewed to ensure reliability and consistency. This methodological framework enables future researchers to replicate the study for different contexts or time periods (Lee et al., 2024; Jang et al., 2024; (Al-Shanawani, 2019); Farrell, 2020; Stocks & Trevitt, 2016; Vesper et al., 2015).

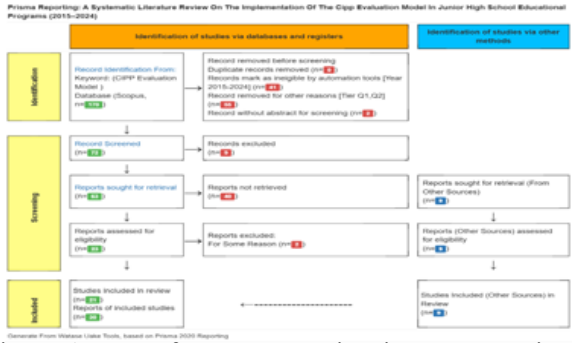

Figure 1. Steps for a Systematic Literature Review

A thematic analysis was subsequently carried out to extract and interpret the core findings from the selected studies. This process employed the Watase Uake System, a structured and adaptable analytical framework developed by Wahyudi (2024), which integrates both inductive and deductive reasoning to ensure coding clarity and consistency throughout the review process. Although not the most recently introduced methodology, the Watase Uake System remains a reliable and effective tool for organizing complex literature reviews. The system delineates the review into six sequential phases: (1) defining search keywords, inclusion criteria, and scope limitations; (2) screening relevant literature; (3) refining article selection and managing exclusions; (4) analyzing titles, abstracts, and keyword patterns; (5) extracting and coding significant data from eligible articles; and (6) conducting classification, network analysis, hypothesis mapping, and data visualization. The structured nature of the system enhances auditability and minimizes subjective bias, making it particularly well-suited for interdisciplinary research topics such as training evaluation across diverse professional fields. Each step of this study from initial identification to final reporting was rigorously aligned with PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines, ensuring that the findings were analytically robust, thematically coherent, and methodologically sound.

## Findings and Discussion

### *Research Output Trends and Contextualized*

Figure 2 illustrates a word cloud generated from keyword metadata extracted from the Scopus database, providing a visual representation of the most dominant themes found in literature related to the application of the CIPP Evaluation Model. The thematic analysis of these keywords reveals several major research directions that reflect the current academic focus on educational program evaluation.

The most frequently occurring overarching themes are centered on training and educational evaluation, with prominent keywords such as *"CIPP model," "evaluation," "junior high school,"* and *"secondary education"* appearing consistently across studies. Supporting subthemes include training effectiveness, outcome evaluation, and curriculum development, indicating a strong research interest in assessing the impact of training and educational initiatives across diverse settings. Notably, many of these contexts are within the healthcare sector, as shown by recurring terms like *"nursing education," "medical education,"* and *"pharmacy."*

These themes are interrelated through a shared emphasis on improving the quality of education and training, with the CIPP model serving as the central evaluative framework. Regional patterns are also evident, with a significant portion of research focusing on localized contexts in Asia (e.g., *Korea, Pakistan, Laos*) and Africa (e.g., *Ethiopia*). This underscores the importance of adapting the CIPP model to align with the unique needs of different educational and healthcare systems.

This trend reflects a broader global priority to enhance training and educational effectiveness, particularly in fields where learning outcomes are closely tied to professional competencies. Looking ahead, future challenges include the integration of digital tools and instructional technology such as *e-learning* and *simulation-based training* as well as addressing disparities in the application of evaluation models across regions. These emerging themes present valuable opportunities for further research, particularly in advancing contextual innovation and evaluative adaptability.


Figure 2. Word Cloud Based on Metadata from Scopus

After extracting data from the 51 eligible articles, the findings are presented in a concise summary, as shown in Table 1. This table outlines the core insights derived from the reviewed studies, encompassing a range of dimensions relevant to the research focus. The summary provides a comprehensive overview of each study, including country of origin, country classification, contextual background, methodological approach, key outcomes, and citation details. By organizing the findings in this way, the table offers a clearer

understanding of how the evaluation model has been applied across varied educational settings and national contexts, thereby enriching the interpretation of its use in diverse environments. The systematic analysis of 51 studies employing the CIPP (Context, Input, Process, Product) evaluation model within secondary and higher education contexts reveals a broad geographical distribution. This widespread adoption underscores a growing global interest in the applicability and utility of the CIPP framework across diverse educational systems. These studies span five continents, with a substantial concentration observed in Asia and Oceania.

From a quantitative perspective, Iran emerges as the most prolific contributor, accounting for nine studies. This prominence reflects the nation's strong institutional commitment to the development and evaluation of educational programs using the CIPP model, particularly in fields such as health education and e-learning. Australia follows with eight studies, most of which center on authentic assessment practices and experiential learning strategies in higher education environments. China ranks third, contributing six studies that primarily address the integration of arts and humanities education, as well as the development of educational resources. Both Canada and South Korea present five studies, with emphases on STEM education reform, case-based learning, and digital safety education initiatives. Turkey contributes four studies, highlighting its focus on curriculum evaluation and the construction of CIPP-based assessment tools. Meanwhile, the United States and the United Kingdom contribute three studies each, often in the context of postgraduate education and professional development programs. Additional countries such as Saudi Arabia, South Africa, Finland, and Singapore each provide between one and two studies, illustrating the global reach and applicability of the CIPP model in varied educational settings.

Notably, one study is identified as multinational, involving collaboration between the World Health Organization and institutions in the United States, Australia, and Switzerland. This study emphasizes the application of the CIPP framework in the global rollout of e-learning initiatives for health professionals.

A regional analysis of the study distribution reveals the following: 1. Asia leads with a total of 27 studies, including contributions from Iran, China, South Korea, Saudi Arabia, Iraq, Singapore, and Hong Kong. This high representation reflects the dynamic educational reforms and evaluation needs in densely populated regions with rapidly evolving education systems. 2. Oceania, represented primarily by Australia, contributes eight studies, showcasing regional leadership in pedagogical innovation and authentic assessment evaluation. 3. North America, encompassing Canada and the United States, offers eight studies in total, often focused on student engagement and simulation-based assessment practices. 4. Europe, particularly the United Kingdom and Finland, contributes four studies, emphasizing reflective assessment design and the development of innovation competencies in higher education. 5. Africa, represented by South Africa, accounts for two studies, primarily highlighting professional and community-based education initiatives. 6. The Middle East, with substantial representation from Iran and Saudi Arabia, contributes over ten studies, indicating a significant regional investment in model-based curriculum evaluation.

This global distribution of studies affirms the CIPP model's versatility and adaptability across a range of educational systems and national policy frameworks. The concentration of research in countries such as Iran, China, and Australia further suggests the presence of policy environments that are receptive to structured evaluation models aimed at enhancing educational quality and accountability.

Moreover, the geographical diversity of the reviewed studies constitutes a major strength of this systematic review, as it allows for comparative analysis across cultural, institutional, and policy contexts. However, the disproportionate representation from certain regions also highlights a critical need to expand CIPP-related research in underrepresented areas such as Latin America, sub-Saharan Africa, and parts of Southeast Asia, to ensure a more comprehensive understanding of the model's global implementation and relevance.

A comprehensive systematic review of 51 studies employing the CIPP (Context, Input, Process, Product) evaluation model in junior high school education and broader educational contexts reveals significant variation in research settings across global regions and education sectors. To better capture this diversity, the reviewed studies have been grouped into seven major thematic domains based on their contextual focus.

Healthcare education emerges as the most prevalent domain. The reviewed studies span evaluations of clinical training, virtual healthcare education, nursing education, and interprofessional health programs. The frequent application of the CIPP model in these contexts reflects a rising demand for quality assurance and evidence-based evaluation in medical and health sciences

education. Notably, studies such as (Molope & Oduaran, 2020), (Yazdani & Moradi, 2017) , and (E. J. M. Chong et al., 2016) highlight the model's usefulness in assessing the effectiveness and responsiveness of health education programs.

Higher education represents the second most prominent category. Research within this domain focuses on authentic assessment methods, innovation competence development, and curriculum design, particularly within applied university settings. These studies often explore the alignment between institutional learning outcomes and labor market expectations. For example, (Sokhanvar et al., 2021), (Keenan & Stewart-Wells, 2021), and (Keenan & Stewart-Wells, 2021) demonstrate how the CIPP model facilitates comprehensive program evaluation to ensure the relevance and rigor of tertiary education curricula. Basic and early childhood education is another key area, encompassing studies that evaluate curriculum design and implementation in early learning and primary school settings. These studies apply the CIPP model to assess the contextual relevance of educational programs, availability and adequacy of learning resources, and pedagogical engagement at foundational education levels. For instance, (Gandomkar, 2018) and Perry et al. (2020) used the model to uncover implementation challenges and opportunities for curriculum improvement in early education.

Occupational safety and professional training also feature prominently among the reviewed studies. This category includes research on educational programs designed for specific workforce needs, such as migrant worker safety training and community-based professional development. The CIPP model has proven effective in evaluating both structural components and learning outcomes of these applied training programs. Studies by Jang et al. (2024) and Molope & Oduaran (2019) demonstrate how the model supports regulatory compliance and workforce capability development.

Community development and youth training programs form another domain of interest. These studies explore educational initiatives targeting marginalized and underserved populations. They emphasize how contextual variables such as economic status, cultural diversity, and infrastructure limitations influence program execution and learner outcomes. As highlighted by Molope & Oduaran (2019) and (Hakan & Seval, 2011) the CIPP model is instrumental in identifying both enabling and inhibiting factors within these challenging educational environments. Entrepreneurship and innovation education, though fewer in number, comprise a significant emerging context. Studies in this domain investigate how entrepreneurial thinking and innovation can be nurtured through higher education programs, particularly in China and Finland. The CIPP model has been applied to assess creativity development, goal-setting strategies, and systems thinking. Examples such as (Warju, 2016) and Keinänen et al. (2018) illustrate the model's relevance in fostering forward-looking educational approaches aligned with economic transformation.

Finally, several studies apply the CIPP model to specific academic disciplines, including English as a Foreign Language (EFL), mathematics, and arts education. These studies reveal the model’s adaptability to domain-specific pedagogical challenges, such as promoting learner autonomy, improving subject-specific proficiency, and integrating creative methodologies. Work by Ismail et al. (2023) and Dong & Shen (2024) underscores how CIPP facilitates evaluation tailored to distinct curricular and instructional demands.

**Table 1. Distribution of Studies**

| No | Author | Country | Context | Research Method | Result |
|---|---|---|---|---|---|
| 1 | Kurucz et al., 2025 | Canada | Educational reform in a southern Ontario school district, focusing on STEM education and the UN Sustainable Development Goals | Case study | Program strengths included integrated learning and project-based approaches; challenges included traditional expectations and workload issues |
| 2 | Gerayllo et al., 2025 | Iran | Virtual education implementation during COVID-19 at an Iranian medical university | Cross-sectional survey | Moderate quality across CIPP components; highest agreement: course alignment with syllabus (context), suitable class hours (input), reduced professor commitment (process), low student participation (product); significant differences between faculty/students for input construct and among students by age/education/faculty/marital status |
| 3 | Dong and Shen, 2024 | China | Universities in Xinjiang, China focusing on integration of literature and art education under `Five Education` | Survey and modeling | Teaching process management is the most significant contributor to integration; teachers` cognitive, practical, and monitoring abilities significantly affect teaching ability |

| | | | | | |
|---|---|---|---|---|---|
| 4 | Huang et al., 2024 | China | Evaluation of teaching quality for ASD course in application-oriented higher education in China | Mixed-method (qualitative criteria with quantitative weighting and analysis) | ASD course evaluated as excellent; team and content strong; but needs environmental improvements and enhanced student intuition |
| 5 | Salinas-Navarro et al., 2024 | United Kingdom, Mexico | Higher education, generative artificial intelligence, experiential learning, authentic assessment, constructive alignment. | Thing ethnography | GenAI tools can act as agents-to-support learning experiences, agents-to-think-about ILOs, agents-to-teach-and-learn-with, agents-to-assess-learning-with, and agents-to-learn-with for authentic assessment. |
| 6 | Jang et al., 2024 | South Korea | High industrial accident rates among migrant construction workers in Korea due to language barriers, inadequate safety education, and cultural differences; Need for innovative solutions amid labor shortages and regulatory changes (Serious Accidents Punish | Modified Delphi study | 70 validated items across 4 modules and 15 criteria for DHSE framework; Expert consensus achieved (mean ≥3.50, consensus ≥0.75, convergence ≤0.50, CV<0.25); DHSE addresses communication barriers and enhances engagement |
| 7 | Lee et al., 2024 | South Korea | Pharmacology education for first-year medical students at Kangwon National University, South Korea, focusing on integrating basic science with clinical practice through blended CBL. | Program evaluation | High student satisfaction (mean 4.17/5) with CBL integration; highest satisfaction in process evaluation (4.31) and faculty feedback (4.50); improved clinical knowledge application and critical thinking noted. |
| 8 | Ismail et al., 2023 | Iran | Conducted with 57 intermediate-level male EFL learners at Melal English Language Institute in Ahvaz, Iran; quasi-experimental design comparing Experimental Group (authentic assessment) and Control Group (non-authentic assessment). | Quasi-experimental | 1. Experimental Group significantly outperformed Control Group in self-regulated learning, autonomy, and self-efficacy post-tests. 2. Experimental Group held positive attitudes toward authentic assessment. 3. Authentic assessment enhances psychological variables and language skill development. |
| 9 | Anderson et al., 2022 | Australia | Undergraduate health promotion subject at La Trobe University, Australia (2019–2021). | Experiential and authentic assessment design with community-based research elements | Students developed a wide range of competencies aligned with IUHPE standards through authentic, experiential learning assessments. Student feedback highlighted appreciation for real-world application and community engagement. |
| 10 | Tuna and Ba?dal, 2021 | Turkey | Tourism undergraduate programs in four Turkish universities | Survey | Students generally satisfied with programs; physical environment needs improvement; Gastronomy students more satisfied than Tourism Management students |
| 11 | Sokhanvar et al., 2021 | Multiple, dominated by Australia | Higher education settings using authentic assessments between 2010 and 2019. | Systematic literature review | Authentic assessment enhances engagement, satisfaction, self-confidence, communication, collaboration, critical thinking, and reflective skills; bridges knowledge and practice. |
| 12 | Way, 2021 | Australia | Postgraduate Occupational Health and Safety Management course at an Australian University, fully online | Mixed Methods (dominantly qualitative) | Online mimetic simulation supported seven out of eight authentic assessment elements; increased student engagement, interaction, and mastery of outcomes; emotion recognition enhanced learning behaviors |
| 13 | Keenan and Stewart-Wells, 2021 | United States | Doctoral program in education at a Midwestern Christian university in the USA, focusing on literacy leadership outcomes | Case study | Authentic assessments: 1) Develop scholarly identity and professional skills; 2) Enhance retention (93.8%); 3) Integrate faith/service values; 4) Foster community impact through real-world projects |
| 14 | Farrell, 2020 | Canada | Undergraduate international marketing course using | Content Analysis | Simulation was found to be authentic in reflection, real-world applicability, and complexity; but weak in |

| | | | | | |
|---|---|---|---|---|---|
| | | | Country Manager simulation at a Canadian business school | | communication, collaboration, and knowledge transfer |
| 15 | Perry et al., 2020 | Canada | Grade 3 classrooms in a mid-size school district in Western Canada | Participatory action research | High SRL classrooms showed better student engagement, metacognitive strategy use, higher quality writing, and more meaningful self-assessments compared to low SRL emphasis classrooms |
| 16 | Chong et al., 2020 | Hong Kong, China | Year-2 undergraduate physiotherapy students at a Hong Kong university; six-week exercise prescription/instruction simulation involving paired role-playing (therapist/client), LMS-based submissions, and multi-stage rubric feedback | Mixed methods | Students engaged in cyclical self-regulated learning (planning-execution-reflection). Rubrics guided planning/feedback but limited creativity. Discussion forum access significantly correlated with scores ($\rho$=0.59) despite low perceived usefulness. LMS facilitated progress tracking. |
| 17 | Ajjawi et al., 2019 | Australia | Australian higher education; WIL placements in health, education, psychology, and law; Tightly-coupled (accredited) vs loosely-coupled (non-accredited) placements | Qualitative interpretative approach | Three key misalignments identified: 1) Between assessment tasks and future professional identities 2) Between workplace activities and assessment methods 3) Between university and industry roles/practices. Performance-based assessment perceived as more authentic than written reports. |
| 18 | Molope and Oduaran, 2019 | South Africa | Post-1994 South Africa; North West Province (42% poverty, mining/agriculture-based economy); Government-led community development; CDPs employed across departments (Social Development, Health, etc.); 135 practitioners in North West DSD | Program evaluation | Lack of alignment between CDPs` needs (policy formulation, project management) and program activities; Centralized planning silences practitioner input; Budget constraints and year-end fund dumping; Inadequate facilitator expertise; Inconsistent participant selection causing tensions; Workplace-based CPD valued for practical learning |
| 19 | Jopp, 2019 | Australia | Higher Education in Australia; Swinburne University; Special Interest Tourism unit | Case study | Increased student engagement and creativity; reduced plagiarism; positive student/industry feedback; challenges in technology support |
| 20 | Al-Shanawani, 2019 | Saudi Arabia | Kindergartens in Saudi Arabia using self-learning curriculum since 1976, lacking evaluation and modernization | Mixed-methods | Curriculum moderately aligns with context, inputs, and processes but lacks in technology integration, emotional needs, and comprehensive evaluation; outputs support basic skills but weak in socioemotional intelligence. |
| 21 | KeinÃ¤nen et al., 2018 | Finland | Finnish higher education institutions (universities of applied sciences) with students from courses using constructivist and student-centered approaches. | Quantitative survey with factor analysis | The five-factor model (creative problem-solving, systems thinking, goal orientation, teamwork, networking) outperformed the original three-factor model, showing acceptable reliability and validity. Innovation competence combines individual and social factors, requiring revised pedagogical practices. |
| 22 | Wright and Veness, 2017 | Australia | Australian higher education archaeology programs; skills gap between academia and cultural heritage industry; case study at Australian National University | Case study | Positive student engagement; development of transferable skills; industry-aligned competencies; improved critical thinking; some student discomfort with immersive methods |
| 23 | Chong et al., 2016 | Singapore | Nursing students in a post-secondary institution in Singapore during their clinical practice at a tertiary hospital | Quasi-experimental study | Authentic assessment pedagogy significantly improved cognitive, affective, and critical thinking domains of nursing students; psychomotor domain showed improvement but was not statistically significant |
| 24 | Stocks and Trevitt, 2016 | UK and Australia | Continuing Professional Learning (CPL) programmed in the UK and Australia | Conceptual and case-based exploration | Trust is central to enabling authentic reflection in CPL portfolios; participants struggle when trust is lacking and when institutional cultures emphasize performativity; formative and developmental aspects of assessment often clash with summative pressures |
| 25 | Kearney et al., 2015 | Australia | Higher education; first-year Bachelor of Education students; mathematics unit; University of Notre Dame Australia, Sydney | Case study (quantitative) | Students could self/peer assess reasonably (83% marks within 5% variance); final mark very close to lecturer mark (mean diff 0.01); higher achievers under-marked, lower over-marked; ASPAL training improved confidence and engagement |
| 26 | Vesper et al., 2015 | Multinational (USA, | Development of an e-learning course on pharmaceutical cold | Design-based | Proactive risk identification (e.g., IT access barriers, cultural collaboration issues) enabled mitigation |

| | | | | | |
|---|---|---|---|---|---|
| | | Australia, Switzerland) | chain management by the World Health Organization (WHO) for global vaccine professionals | research | strategies like DVD distribution for restricted regions, improving course accessibility and reducing implementation failures |
| 27 | Sudario et al. 2022 | USA | Medical education during the COVID-19 pandemic, focusing on pandemic preparedness and clinical knowledge. | Pilot study | Students showed high satisfaction, knowledge attainment, and confidence in PPE use but lower confidence in ventilator management. The course was |
| 28 | Quinton, et al. 2022 | United Kingdom | MST4Life program within a West Midlands (UK) housing service for homeless youth | Process Evaluation | Learner enjoyment emerged as a significant predictor of learning outcomes, with transfer intention serving as a mediating factor in this relationship. While engagement was positively associated with favorable program responses, it did not show a direct correlation with actual learning achievements |
| 29 | Shallal et al. 2023 | irak | Online faculty development training in bioethics for multidisciplinary educators at Hawler Medical University | Quasi-experimental program evaluation | Significant improvement in knowledge among clinical faculty, high satisfaction with course content, and moderate evidence of behavioral application in teaching practices |
| 30 | Wang 2024 | Cina | Applied universities in China, focusing on the integration of innovation entrepreneurship education with professional education | Mixed-methods | The study constructed an evaluation index system with 4 primaries indicators, 9 secondary indicators, and 23 tertiary indicators, with the learning layer having the highest weight (0.5579). |
| 31 | Depasquale et al. 2024 | United Kingdom | interprofessional education (IPE) in undergraduate pharmacy programmed in the United Kingdom | Cross-sectional survey | Campus-based IPE is common and varies widely; limited planned practice-based IPE; students benefit from opportunistic IPE but need more structured opportunities |
| 32 | Golestani et al. 2024 | Iran | Higher education in Iran focusing on traffic safety education among medical students | Interventional Study | The traffic safety course significantly improved student reactions, knowledge, behavior, and results, with a favorable cost- effectiveness ratio especially at the results level |
| 33 | Mirmoghtadaie et al. 2024 | Iran | E-learning on sexual health for female students at Shahid Beheshti University of Medical Sciences, Iran | Quasi-experimental intervention | Participants` knowledge scores improved significantly after the course (from 15.34 to 16.88, $P < 0.001$); overall satisfaction rate was 88%, with highest scores for course duration (94.4%) and content delivery |
| 34 | Rucks et al. 2024 | USA | Evaluation of the Evaluate project supported by NSF`s | Reflective evaluation of an ECB initiative using an adapted CIIP framework | adapted model enabled multidimensional evaluation of ECB; findings influenced redesign of coaching service, improved understanding of training effects, supported funding decisions, and confirmed increased application and improved evaluation quality |
| 35 | Gerayllo et al. 2025 | Iran | Using the CIPP Model to elicit perceptions of health professions faculty and students about virtual learning | Quasi-experimental | The CIPP model revealed faculty-student perceptions on virtual learning, showing mixed satisfaction, tech/resource gaps, engagement issues, and concerns on skill retention |

| | | | | | |
|---|---|---|---|---|---|
| | | | | intervention | |
| 36 | Dong, et al, 2023 | Tiongkok | Literature and art integration education A Study on the Integration of Teaching Resources in Colleges and Universities in the Context of Five Educational Programs | Quantitative survey using CIPP to evaluate teaching resource | Teaching resource integration was rated good overall, with strengths in teaching process and faculty management, but facilities lagged. Faculty's cognitive, practical, and monitoring abilities significantly influenced teaching effectiveness. |
| 37 | Huang et al, 2024 | Cina | Literature and art integration education A Study on the Integration of Teaching Resources in Colleges and Universities in the Context of Five Educational Programs | Quantitative survey using CIPP framework, stratified sampling, Likert-scale questionnaire, and SEM analysis. | Teaching resource integration scored strongly in process management and faculty leadership, while facilities lagged. High faculty cognitive, practical, and monitoring abilities significantly enhanced integration effectiveness, supporting the "five-education" framework across diverse programs. |
| 38 | Lee et al, 2024 | Korea | Evaluation of a basic-clinical integrated pharmacology case-based learning program from a student perspective using the CIPP model | Mixed-method study using CIPP framework; data collected via student questionnaires and focus group interviews, analyzed through thematic and descriptive analysis. | Students expressed strong satisfaction with program context and resources (Input) and praised the relevance of case-based methodology (Process). Learning outcomes (Product) were positively reported, though some desired more clinical exposure and feedback |
| 39 | Tuna et al, 2021 | Turkey | Curriculum evaluation of tourism undergraduate programs in Turkey A CIPP model-based framework | This study used a qualitative case study design based on the CIPP model, collecting data through interviews, document analysis, and stakeholder feedback from tourism undergraduate programs | The CIPP evaluation revealed strong alignment in **Context** and **Input**—curriculum goals and resources were appropriate. **Process** showed engaging field experiences. However, **Product** outcomes indicated inconsistent skill integration and insufficient stakeholder feedback, suggesting need for enhanced industry collaboration |
| 40 | Sokhanvar et al, 2021 | Australia | Advantages of authentic assessment for improving the learning experience and employability skills of higher education students A systematic literature review | SLR | Authentic assessments consistently improved student engagement, real-world skill acquisition, and reflective learning. They enhanced employability skills—such as critical thinking and communication—while also increasing student satisfaction. Educators reported stronger alignment between learning objectives and industry expectations. |
| 41 | Hakan et al, 2020 | Turki | CIPP evaluation model scale development, reliability and validity | The study employed a quantitative design | The study successfully developed a CIPP-based evaluation scale with strong construct validity and high internal consistency. Exploratory and confirmatory factor analyses confirmed the scale's four-factor structure, ensuring reliability for educational |
| 42 | AbdiShah shahani et | Iran | The Evaluation of Reproductive Health PhD | Qualitative evaluation | The evaluation revealed strong alignment in curriculum goals and expert faculty (Context and |

| | | | | | |
|---|---|---|---|---|---|
| | al, 2025 | | Program in Iran A CIPP Model Approach | using interviews and document analysis via CIPP | Input). However, gaps were found in interdisciplinary collaboration and student research support. Graduates expressed moderate satisfaction, highlighting needs for improved mentorship and practical learning opportunities. |
| 43 | Al-Shanawani et al, 2029 | Arab Saudi | Evaluation of Self-Learning Curriculum for Kindergarten Using Stufflebeam s CIPP Model | Qualitative case study using interviews, observations, and CIPP framework analysis. | The evaluation found the curriculum contextually appropriate and resource-supported (Context, Input). However, inconsistencies were observed in instructional delivery (Process). Outcome analysis showed developmental benefits but highlighted the need for ongoing monitoring, teacher training, and parent engagement to ensure effectiveness. |
| 44 | Molope et al, 2019 | Aprika Selatan | Evaluation of the community development practitioner's professional development programmed CIPP model application | Mixed-methods approach using surveys, interviews, and document analysis with CIPP. | The program was contextually relevant and aligned with community needs. Resources and content (Input) were adequate. Process evaluations revealed strong facilitation but limited peer collaboration. Product outcomes showed improved practitioner skills, though impact on community empowerment remained moderately demonstrated |
| 45 | Jopp et al, 2019 | Australia | A case study of a technology enhanced learning initiative that supports authentic assessment | Qualitative case study using observations, interviews, and document analysis | The initiative effectively fostered student engagement through real-world tasks and digital collaboration tools. Participants reported enhanced critical thinking and practical skills. However, challenges included inconsistent technology access and varying instructor readiness, impacting assessment consistency and learning outcomes. |
| 46 | Jang, et al, 220 | Korea selatan | Validation of a Digital Human-Based Safety Education Framework for Migrant Construction Workers in Korea Using the CIPP Model and a Modified Delphi Study | Modified Delphi method and CIPP-based evaluation with expert panel input | The validated framework addressed language, cultural, and technological barriers effectively. Experts confirmed contextual relevance (Context), resource adequacy (Input), interactive delivery (Process), and learning improvements (Product). Recommendations included ongoing refinement to enhance adaptability across diverse migrant worker populations. |
| 47 | Farrell et al,2020 | canada | Do international marketing simulations provide an authentic assessment of learning A student perspective | Qualitative survey and focus group analysis from student perspectives | Students perceived simulations as authentic and engaging, enhancing practical marketing skills and strategic thinking. While most found it aligned with real-world challenges, some noted gaps in team dynamics and assessment fairness, suggesting areas for instructional refinement and support. |
| 48 | Perry et al, 2020 | Canada | Collaborating with teachers to design and implement assessments for self-regulated learning in the context of authentic classroom writing tasks | Design-based research using teacher collaboration, classroom observations, and interviews. | The study found that teacher collaboration in designing assessments enhanced students' self-regulated learning strategies. Authentic writing tasks promoted metacognitive awareness, goal-setting, and reflective practices. However, implementation varied, highlighting the need for sustained professional development and contextualized assessment support. |
| 49 | Keinänen et al, 2019 | Firlandia | How to measure student's innovation competences in higher education Evaluation of an assessment tool in authentic learning environments | Mixed-method approach using surveys, observations, and statistical analysis. | Assessment tool effectively captured innovation competences in authentic learning settings. It supported evaluation of creativity, collaboration, and problem-solving. However, challenges in standardization and context adaptation were identified, indicating the need for refinement to enhance reliability and cross-disciplinary applicability. |
| 50 | AbdiShahshahani,2015 | Iran | The Evaluation of Reproductive Health PhD Program in Iran A CIPP | Qualitative case study using | Evaluation showed alignment between program goals and national health needs. While input and faculty quality were satisfactory, challenges emerged in |

| | | | | |
|---|---|---|---|---|
| | | Model Approach | interviews and document analysis. | research infrastructure and interdisciplinary collaboration. Graduates expressed moderate satisfaction, suggesting improvements in mentorship, curriculum integration, and practical training |
| 51 | Stocks et Inggris al, 2016 | The place of trust in Continuing Professional Learning programmed supporting authentic reflection in portfolio assessment | Qualitative study using narrative inquiry, interviews, and portfolio content analysis. | Between learners and facilitators was essential for meaningful reflection in portfolio assessment. Programs fostering openness and relational support enhanced authenticity, while lack of trust hindered honest self-assessment, limiting the effectiveness of professional learning outcomes. |

The review of 51 studies employing the CIPP evaluation model across various educational settings reveals a notable diversity in methodological approaches. Three primary methodologies were most prevalent: Mixed Methods**,** Qualitative**,** and Quantitative, accompanied by several complementary strategies such as Quasi-Experimental**,** Case Study**,** Survey**,** and Program Evaluation**.**

The Mixed Methods approach was the most frequently employed, appearing in eight studies**.** This methodology integrates both quantitative and qualitative techniques to produce a more comprehensive analysis of educational program effectiveness. It is particularly suited to evaluating complex programs that require data triangulation from multiple sources, as demonstrated in studies involving interprofessional training, e-learning integration, and clinical simulation-based instruction. The Qualitative approach was equally prominent, also used in eight studies. These investigations commonly utilized interviews, document analysis, classroom observations, and narrative inquiry. Qualitative methods were instrumental in capturing participants' perceptions, experiences, and responses to curriculum implementation, training effectiveness, and policy execution. The approach enables in-depth exploration of contextual variables, particularly in relation to reflective practices, collaborative skills, and trust-building in professional learning environments.

The Quantitative approach was found in five studies, often employing Likert-scale surveys, factor analyses, and both descriptive and inferential statistics. This method is well-suited for measuring program impact using numerical indicators such as score improvement, learner satisfaction, content effectiveness, and resource efficiency. Some studies also focused on validating CIPP-based assessment instruments through quantitative means.

In addition to the three dominant approaches, five studies utilized a quasi-experimental design, typically structured around non-randomized control and experimental groups. These designs aimed to assess the effectiveness of specific educational interventions such as authentic assessment strategies, health training programs, and digital learning platforms through pre-test and post-test measurements.

A further six studies fell under the Other category, encompassing methods such as Design-Based Research**,** Delphi Studies**,** Content Analysis**,** Participatory Action Research**,** and Ethnography**.** These approaches contributed significantly to the development of evaluation frameworks, the construction of new indicators, and the examination of technological applications in cross-cultural educational contexts.

The distribution of these methodological approaches highlights the adaptability of the CIPP evaluation model to diverse research contexts, ranging from local initiatives to international collaborations, and from primary to secondary data sources. The dominance of mixed methods reflects a growing preference for methodologies that combine the contextual richness of qualitative data with the empirical rigor of quantitative analysis.

This trend emphasizes the necessity of methodological flexibility in educational evaluation research particularly when addressing complex, multidimensional programs commonly assessed through the CIPP framework. Furthermore, the increasing use of mixed and qualitative approaches signals a paradigm shift toward more holistic understandings of educational processes, moving beyond strictly numerical outcomes to embrace the nuanced dynamics of learning and instruction.

The majority of studies reviewed in this analysis consistently adopt the CIPP model (Context, Input, Process, Product) as the primary framework for evaluating educational program effectiveness. The model is regarded as comprehensive, encompassing key dimensions such as planning, implementation, and outcomes. Its adaptability is evidenced by its widespread application across various sectors and educational levels, including healthcare education

(Rooholamini et al., 2017) ; (Agus et al., 2023) ; (AbdiShahshahani et al., 2015) ., 2023, higher education (Sokhanvar et al., 2021), and early childhood education (Al-Shanawani et al., 2029). This demonstrates the model's capacity to effectively address the complexities of diverse educational contexts.

Several studies further enhance the CIPP framework by integrating it with other evaluation theories, such as the Kirkpatrick model and authentic assessment approaches. This integrative strategy enriches the analytical depth of the evaluations, particularly in addressing affective and social dimensions. For instance, (Suryana et al., 2023) and Keinänen et al. (2019) combined the CIPP model with 21st-century skills frameworks, while Stocks and Trevitt (2016) linked it with reflective learning theories to support professional development. Moreover, many researchers employ constructivist and contextual learning approaches to strengthen the model's relevance, as seen in studies on innovation education (Wang, 2024), pharmaceutical training (Lee et al., 2024), and community-based learning (Molope & Oduaran, 2019).

Authentic assessment also plays a prominent role in studies emphasizing real-world learning and the development of soft skills such as collaboration and critical thinking ((Huang et al., 2025); (Basaran et al., 2021); Farrell et al., 2020). The (Darma, 2019) se approaches reinforce the process and product components of the CIPP model. Additionally, CIPP is increasingly applied as a quality assurance instrument in vocational and professional education, focusing on evaluating employability skills and workforce readiness (Keinänen et al., 2018); (Sokhanvar et al., 2021). In the social domain, humanistic and socio-cultural perspectives are incorporated to emphasize the significance of local context and participatory evaluation practices (Dwi et al., 2018); Stocks & Trevitt, 2016). Lastly, recent studies highlight the model's adaptability to technological advancements through digital tools such as LMS platforms, virtual simulations, and generative AI, which further extend CIPP's relevance to 21st-century digital learning environments (Salinas-Navarro et al., 2024); (Way et al., 2021)Way, ; (Jopp, 2020).

One of the major debates surrounding the implementation of the CIPP (Context, Input, Process, Product) evaluation model revolves around its comprehensiveness **versus** specificity. While the model is widely recognized for its broad scope encompassing planning (Context and Input), execution, and outcomes (Process and Product) critics argue that it lacks the precision needed to assess complex behavioral changes and long-term program impacts effectively (Zhao et al., 2025) , (El Khammari & Hasnaoui, 2025) ; (Tuna & Başdal, 2021). As a result, several scholars recommend combining CIPP with more outcome-oriented or domain-specific models, such as the Kirkpatrick framework, to address these gaps (Gerayllo et al., 2025); (Omid & Avizhgan, 2024).

A second point of contention relates to the adaptability of the CIPP framework within the evolving landscape of modern education, especially in terms of technological integration and interdisciplinary approaches. In the context of digital learning, blended instruction, and artificial intelligence-driven teaching, the traditional structure of the CIPP model is sometimes perceived as rigid and insufficiently responsive to rapid pedagogical innovations (Salinas-Navarro et al., 2024);(Jopp, 2020) ; Way, 2021). Consequently, some researchers have proposed hybrid or adaptive versions of CIPP that are more suited to contemporary digital learning environments.

The third debate centers on methodological orientation: should CIPP based evaluations prioritize objective indicators such as exam scores, resource sufficiency, and process efficiency, or should they incorporate more subjective and participatory dimensions such as student satisfaction, teacher perceptions, and learner motivation? Studies by Stocks and Trevitt (2016) and Molope and Oduaran (2019) advocate for a more humanistic and culturally sensitive approach, particularly in community education and programs involving marginalized populations.

Furthermore, some scholars criticize the CIPP model for placing too much emphasis on formal and administrative structures, often at the expense of emotional and social dynamics in educational settings. In reality, learning is a complex process that cannot be fully captured by technical indicators alone. Therefore, integrating values such as trust, collaboration, and reflective practice is essential for a more holistic CIPP-based evaluation particularly in the Process and Product components (Stocks & Trevitt, 2016; Quinton et al., 2022).

Another area of debate concerns the contextual relevance of the CIPP model in local educational settings. Although it is designed as a universal evaluation framework, its successful application often depends on the degree of alignment with local values, social structures, and learner needs. This is particularly evident in studies conducted among indigenous communities, minority groups, or community-based training programs, where success is not only measured through formal outputs but also through social

empowerment and cultural responsiveness (Molope & Oduaran, 2019; Quinton et al., 2022). Finally, questions have been raised about the sustainability and long-term impact assessment capabilities of the CIPP model. Many researchers express concern over whether CIPP is sufficiently equipped to capture enduring effects of educational programs, particularly in the context of systemic change and policy reform. As a result, there is growing advocacy for supplementing CIPP with more longitudinal and transformation-oriented models, such as outcome-based or participatory evaluation frameworks, to ensure future relevance and effectiveness in educational research (Desilets, 2018; Dizon, 2023).

The application of the CIPP evaluation model (Context, Input, Process, Product) across various educational programs, particularly at the junior high school and higher education levels, has demonstrated its practical value in assessing program effectiveness while simultaneously offering empirical foundations for policy development and instructional improvement. For instance, (Gerayllo et al., 2025) illustrated how the CIPP model can identify divergent perceptions between instructors and students in online learning, thereby informing the creation of more inclusive and adaptive educational strategies.

Moreover, (Kurucz et al., 2025) highlighted the model's utility in uncovering structural challenges within STEM education reforms, such as misaligned teacher workloads, curriculum constraints, and resource disparities. These findings provide critical insights for policymakers aiming to design data-driven interventions that are better aligned with field-level realities.

In the realm of educational technology innovation, the CIPP model has proven to be highly relevant. Salinas-Navarro et al. (2024) demonstrated that generative artificial intelligence can be effectively evaluated using the CIPP framework, particularly by adjusting input and process components to optimize technology-enhanced learning outcomes. Such evaluations are crucial in the context of rapid digital transformation within the education sector.

Additional studies by Jopp (2019) and Way (2021) further affirm that the CIPP model can effectively pinpoint strengths and weaknesses in digital learning implementation. Components such as student engagement, platform usability, and content integration can be systematically assessed using CIPP, allowing institutions to make informed improvements.

In the domain of workforce development and professional training, CIPP has contributed to enhancing program relevance and efficacy. For example, (Jang et al., 2025). (2024) utilized the model to evaluate safety training programs for migrant workers, factoring in linguistic and cultural barriers. This has supported the creation of more equitable and participant-centered training approaches. Within community-based education, CIPP also serves as a tool to assess the social impact and sustainability of training initiatives. Molope and Oduaran (2019) found that program success depends not only on formal outputs but also on alignment with local contexts and community involvement, underlining the model's adaptability to informal and marginalized education settings.

Collectively, these studies underscore the broad practical implications of the CIPP model for educational policy-making, program design, and systemic reform. Its flexibility allows for meaningful application across educational levels and sectors, enabling the evaluation of learning effectiveness, technological integration, and social impact. As a result, embedding the CIPP model within future education policy frameworks may foster more responsive, participatory, and data-driven education systems (Gerayllo et al., 2025; Salinas-Navarro et al., 2024; Jang et al., 2024; Molope & Oduaran, 2019).

This Systematic Literature Review (SLR) offers critical insights into prevailing research trends concerning the application of the CIPP (Context, Input, Process, Product) evaluation model across diverse domains such as education, training, and healthcare. However, several limitations must be acknowledged. Firstly, a significant portion of the reviewed studies employed quantitative methods (47.7%), which, while valuable for producing objective data, often fail to capture more nuanced qualitative aspects like participants' perceptions and experiences (Mohebi et al., 2018) ; (Perry et al., 2020) This methodological bias may hinder a comprehensive understanding of program effectiveness, particularly in complex environments such as interprofessional education or technology-based learning ((AbdiShahshahani et al., 2015) ; (Molope & Oduaran, 2020).

Secondly, the predominance of the CIPP model in 51 studies indicates a continued reliance on established evaluation frameworks. While the CIPP model remains effective, it has been critiqued for its limitations in assessing long-term outcomes and capturing systemic educational changes (El Khammari & Hasnaoui, 2025); (Li & Hu, 2022). Moreover, its application is often seen as insufficiently adaptable to localized and context-specific needs, especially in developing countries (Kurucz et al., 2025); (Perry et al., 2020). Another

concern lies in the uneven geographical representation of studies, which are heavily concentrated in developed nations such as the

United States, the United Kingdom, and Australia. Meanwhile, research from developing countries like Iran and India remains underrepresented (Rooholamini et al., 2017) ; (Tavani et al., 2024). This imbalance presents risks in generalizing findings, as educational and training outcomes are deeply influenced by social, cultural, and economic contexts (Suharno et al., 2022); (Li & Hu, 2022).

Furthermore, the proportion of studies utilizing mixed methods (24.2%) or qualitative approaches (8.1%) remains relatively limited, despite the potential of these methodologies to yield deeper insights into program implementation and outcomes (Aleandri, 2015) ; (AbdiShahshahani et al., 2015). This highlights a pressing need for future research to adopt more integrative approaches combining both quantitative and qualitative techniques to offer a more holistic perspective. Another limitation concerns the underexplored integration of technology into evaluation practices. Emerging tools such as online learning systems, simulations, and big data analytics are still not widely represented in the studies reviewed (Al-Shanawani, 2019); (Warju, 2016). Although some research, such as (Way et al., 2021), has begun to explore virtual training in nursing education, the potential remains vast for developing innovative and adaptive evaluation frameworks. Additionally, the low citation counts in some studies for example, (Aleandri, 2015) and (Farrell, 2020) suggest that newer findings often require time to gain academic recognition. This underscores the importance of maintaining research quality and contextual relevance to enhance scholarly impact.

In conclusion, while this SLR provides a comprehensive overview of the use of the CIPP model in program evaluation research, it also identifies several critical areas for improvement. Addressing methodological limitations, ensuring more equitable geographic representation, and fostering greater integration of digital tools will be essential in future studies. By confronting these challenges, research in this domain can evolve more effectively and make meaningful contributions to advancing the quality of education, training, and public service delivery across various sectors capital letter for the initial word.

## Conclusion

This systematic review of 51 empirical studies (2015–2024) highlights the CIPP (Context, Input, Process, Product) evaluation model as a comprehensive and adaptable framework for assessing educational programs, particularly in junior high school settings. The model effectively supports both formative and summative evaluations, and its integration with other theories enhances responsiveness to contextual, cultural, and affective factors. Its application in technology-enhanced environments underscores its relevance in addressing modern educational challenges. However, limitations include an overreliance on quantitative methods, limited assessment of long-term impacts, geographical concentration in developed countries, and minimal integration of emerging technologies like AI and big data. Future research should adopt mixed methods, promote culturally responsive adaptations, incorporate digital tools and analytics, and conduct longitudinal studies to strengthen the model's utility and global applicability.

## References


**AbdiShahshahani, M., Ehsanpour, S., Yamani, N., Kohan, S., & Hamidfar, B.** (2015). The Evaluation of Reproductive Health PhD Program in Iran: A CIPP Model Approach. Procedia - Social and Behavioral Sciences, 197(February), 88–97. https://doi.org/10.1016/j.sbspro.2015.07.059

**Agus, A., Juliadharma, M., & Djamaluddin, M.** (2023). Application of the CIPP Model in Evaluation of The Inclusive Education Curriculum in Madrasah Aliyah. Nidhomul Haq : Jurnal Manajemen Pendidikan Islam, 8(1), 31–50. https://doi.org/10.31538/ndh.v8i1.2705

**Al-Shanawani, H. M.** (2019). Evaluation of Self-Learning Curriculum for Kindergarten Using Stufflebeam's CIPP Model. SAGE Open, 9(1). https://doi.org/10.1177/2158244018822380

**Aleandri, G.** (2015). Autobiography as Educational Scientific Tool in Penitentiary. Procedia - Social and Behavioral Sciences, 197(February), 4–9. https://doi.org/10.1016/j.sbspro.2015.07.038

**Basaran, M., Dursun, B., Gur Dortok, H. D., & Yilmaz, G.** (2021). Evaluation of Preschool Education Program According to CIPP Model. Pedagogical Research, 6(2), em0091. https://doi.org/10.29333/pr/9701

**Chong, D. Y. K., Tam, B., Yau, S. Y., & Wong, A. Y. L.** (2020). Learning to prescribe and instruct exercise in physiotherapy education through authentic continuous assessment and rubrics. BMC Medical Education, 20(1), 1–11. https://doi.org/10.1186/s12909-020-02163-9

**Chong, E. J. M., Lim, J. S. W., Liu, Y., Lau, Y. Y. L., & Wu, V. X.** (2016). Improvement of learning domains of nursing students with the use of authentic assessment pedagogy in clinical practice. Nurse Education in Practice, 20, 125–130. https://doi.org/10.1016/j.nepr.2016.08.002

**Darma, I. K.** (2019). The effectiveness of teaching program of CIPP evaluation model. International Research Journal of Engineering, IT & Scientific Research, 5(3), 1–13. https://doi.org/10.21744/irjeis.v5n3.619

**Dwi, Y. R., Haryono, & Florentinus, T. S. ;** (2018). The Evaluation of the CIPP Model in The Implementation of Character Education at Junior High School. Innovative Journal of Curriculum and Educational Technology, 7(2), 65–77.

**El Khammari, N., & Hasnaoui, R.** (2025). Evaluation of the Impact of Educational Programs at Secondary School: Systematic and Bibliometric Review of Literature. Educational Process: International Journal, 15. https://doi.org/10.22521/edupij.2025.15.141

**Farrell, C.** (2020). Do international marketing simulations provide an authentic assessment of learning? A student perspective. International Journal of Management Education, 18(1), 100362. https://doi.org/10.1016/j.ijme.2020.100362

**Gandomkar, R.** (2018). Comparing Kirkpatrick's original and new model with CIPP evaluation model. Journal of Advances in Medical Education & Professionalism, 6(2), 94–95.

**Geraylloo, S., Vakili, M., Jouybari, L., Moghadam, Z., Jafari, A., & Heidari, A.** (2025). Using the CIPP Model to elicit perceptions of health professions faculty and students about virtual learning. BMC Medical Education, 25(1), 1–12. https://doi.org/10.1186/s12909-025-06747-1

**Hakan, K., & Seval, F.** (2011). CIPP evaluation model scale: Development, reliability and validity. Procedia - Social and Behavioral Sciences, 15, 592–599. https://doi.org/10.1016/j.sbspro.2011.03.146

**Heard-Lauréote, K., & Buckley, C.** (2025). Building trustful relationships through intrapreneurship in Academic Development. Journal of Learning Development in Higher Education, 33. https://doi.org/10.47408/jldhe.vi33.1177

**Huang, M., He, Z., Zhang, J., Deng, Z., & Tang, D.** (2025). Course Evaluation of Advanced Structural Dynamics Based on Improved SAPSO and FAHP. Buildings, 15(1), 1–22. https://doi.org/10.3390/buildings15010072

**Ismail, S. M., Nikpoo, I., & Prasad, K. D. V.** (2023). Promoting self-regulated learning, autonomy, and self-efficacy of EFL learners through authentic assessment in EFL classrooms. Language Testing in Asia, 13(1). https://doi.org/10.1186/s40468-023-00239-z

**Jang, J. H., Noh, Y., & Kim, J. S.** (2025). Validation of a Digital Human-Based Safety Education Framework for Migrant Construction Workers in Korea Using the CIPP Model and a Modified Delphi Study. Sustainability (Switzerland), 17(1), 1–18. https://doi.org/10.3390/su17010169

**Jopp, R.** (2020). A case study of a technology enhanced learning initiative that supports authentic assessment. Teaching in Higher Education, 25(8), 942–958. https://doi.org/10.1080/13562517.2019.1613637

**Keenan, K. M., & Stewart-Wells, A. G.** (2021). Moving Beyond Comprehensive Exams: Implementing Authentic Assessments to Enhance Doctoral Student Learning. Christian Higher Education, 20(1–2), 69–86. https://doi.org/10.1080/15363759.2020.1851815

**Keinänen, M., Ursin, J., & Nissinen, K.** (2018). How to measure students' innovation competences in higher education: Evaluation of an assessment tool in authentic learning environments. Studies in Educational Evaluation, 58(May), 30–36. https://doi.org/10.1016/j.stueduc.2018.05.007

**Kurucz, E., Spencer-Mueller, E. K., Hands, C., Gudz, N., & Archer, K.** (2025). Co-constructing educational innovations for an uncertain future: Design thinking and developmental evaluation in the school reform process. Evaluation and Program Planning, 111(February), 102559. https://doi.org/10.1016/j.evalprogplan.2025.102559

**Li, Y., & Hu, C.** (2022). The Evaluation Index System of Teaching Quality in Colleges and Universities: Based on the CIPP Model. Mathematical Problems in Engineering, 2022. https://doi.org/10.1155/2022/3694744

**Mohebi, S., Parham, M., Sharifirad, G., & Gharlipour, Z.** (2018). Social Support and Self - Care Behavior Study. January, 1–6. https://doi.org/10.4103/jehp.jehp

**Molope, M., & Oduaran, A.** (2020). Evaluation of the community development practitioners' professional development programme: CIPP model application. Development in Practice, 30(2), 194–206. https://doi.org/10.1080/09614524.2019.1650894

**Omid, A., & Avizhgan, M.** (2024). Integrating a New Model of Community-Based Education in the General Medicine: An Evaluation Study. Journal of Health System Research, 20(3), 246–256. https://doi.org/10.48305/jhsr.v20i3.1552

**Perry, N. E., Lisaingo, S., Yee, N., Parent, N., Wan, X., & Muis, K.** (2020). Collaborating with teachers to design and implement assessments for self-regulated learning in the context of authentic classroom writing tasks. Assessment in Education: Principles, Policy and Practice, 27(4), 416–443. https://doi.org/10.1080/0969594X.2020.1801576

**Rooholamini, A., Amini, M., Bazrafkan, L., Dehghani, M. R., Esmaeilzadeh, Z., Nabeiei, P., Rezaee, R., & Kojuri, J.** (2017). Program evaluation of an Integrated Basic Science Medical Curriculum in Shiraz Medical School, Using CIPP Evaluation Model. Journal of Advances in Medical Education & Professionalism, 5(3), 148–154.

**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). https://doi.org/10.3390/educsci14010083

**Sokhanvar, Z., Salehi, K., & Sokhanvar, F.** (2021). Advantages of authentic assessment for improving the learning experience and employability skills of higher education students:

A systematic literature review. Studies in Educational Evaluation, 70(April), 101030. https://doi.org/10.1016/j.stueduc.2021.101030

**Suharno, S., Pardiman, P., Harijanto, D., & Ashar, A.** (2022). Monitoring Strategy and Evaluation of Education and Training Program with The CIPP Method. Nazhruna: Jurnal Pendidikan Islam, 5(3), 1228–1241. https://doi.org/10.31538/nzh.v5i3.2566

**Suhonen, J., & Sutinen, E.** (2022). Critical events and pivotal factors as tools for analyzing the sustainability of online study programs. Educational Technology Research and Development, 70(5), 1551–1573. https://doi.org/10.1007/s11423-022-10133-9

**Suryana, D., Husna, A., & Mahyuddin, N.** (2023). CIPP Evaluation Model: Analysis of Education Implementation in PAUD Based on Government Policy on Implementation of Learning During the Covid-19 Pandemic. Jurnal Obsesi : Jurnal Pendidikan Anak Usia Dini, 7(4), 4386–4396. https://doi.org/10.31004/obsesi.v7i4.3722

**Tavani, M. E., Poursaki, T., & Gharibi, F.** (2024). The Necessity of Promoting Technology Commercialization in Iran's Universities of Medical Sciences: A Policy Brief. Koomesh, 26(1), 1–7. https://doi.org/10.69107/koomesh-153982

**Tuna, H., & Başdal, M.** (2021). Curriculum evaluation of tourism undergraduate programs in Turkey: A CIPP model-based framework. Journal of Hospitality, Leisure, Sport and Tourism Education, 29(October 2020), 1–9. https://doi.org/10.1016/j.jhlste.2021.100324

**Warju, W.** (2016). Educational Program Evaluation using CIPP Model. Innovation of Vocational Technology Education, 12(1), 36–42. https://doi.org/10.17509/invotec.v12i1.4502

**Way, K. A., Burrell, L., D'Allura, L., & Ashford-Rowe, K.** (2021). Empirical investigation of authentic assessment theory: An application in online courses using mimetic simulation created in university learning management ecosystems. Assessment and Evaluation in Higher Education, 46(1), 17–35. https://doi.org/10.1080/02602938.2020.1740647

**Yazdani, N., & Moradi, M.** (2017). Evaluation of the Quality of Undergraduate Nursing Education Program in Ahvaz Based on CIPP Evaluation Model. Sadra Medical Journal, 5(3), 159–172.

**Zhao, T., Yu, L., Jiang, X., Yang, S., Zhang, X., Zhang, M., Zhang, Y., Yan, T., Wu, H., & Xu, F.** (2025). Construction of a public health practice teaching quality evaluation system based on the CIPP model using the Analytic Hierarchy Process (AHP) and entropy weight method. BMC Medical Education, 25(1). https://doi.org/10.1186/s12909-025-06943-z