Panthera : Jurnal Ilmiah Pendidikan Sains dan Terapan
E-ISSN 2808-246X; P-ISSN 2808-3636
Volume 5, Issue 4, October 2025; Page, 1541-1565
Email: pantherajurnal@gmail.com

AN EDUCATIONAL EVALUATION MODEL TO STRENGTHEN INSTITUTIONAL ACCOUNTABILITY AND QUALITY THROUGH A SYSTEMATIC LITERATURE REVIEW APPROACH

Syahrir[1]* & A. Fandir[2]
[1]Department of Mathematics Education, Faculty of Science, Engineering, and Applied Sciences, Mandalika University of Education, Pemuda Street Number 59A, Mataram, West Nusa Tenggara 83125, Indonesia
[2]Lombok Business Academy, Tanjung-Medana Street, North Lombok, West Nusa Tenggara 83352, Indonesia
**Email: syahrir@undikma.ac.id*
Submit: 20-10-2025; Revised: 27-10-2025; Accepted: 30-10-2025; Published: 31-10-2025

ABSTRACT: Educational evaluation plays a strategic role in strengthening public accountability and enhancing institutional interrelationships amidst the acceleration of digital transformation. This study addresses the urgent need to explore how various evaluation models can effectively interact with technology and educational governance to create an adaptive, transparent, and sustainable quality assurance system. This research aims to synthesize dominant and relevant educational evaluation models across contexts and uncover the effectiveness of their implementation in the digital era. Using the Systematic Literature Review (SLR) method guided by PRISMA 2020, this study went through four main stages: identification, screening, eligibility, and inclusion. A total of 22 articles from Scopus and Web of Science for the 2022-2025 period were analyzed thematically. The findings revealed four prominent evaluation models: CIPP, Artificial Intelligence-Based (AI-Based) evaluation, participatory evaluation, and evidence-based evaluation. These four models emphasize a paradigm shift in evaluation systems toward data-driven, collaborative, and technology-integrated ones. Practically, this study confirms that the successful integration of evaluation models and digital technology is highly dependent on governance readiness, human resource capacity, and data-driven decision-making mechanisms. These implications form an important basis for educational institutions to strengthen evaluation systems that are responsive to change and oriented towards increasing sustainability.

Keywords: Accountability, Digital Transformation, Educational Evaluation, Evaluation Models.

How to Cite: Syahrir, S., & Fandir, A. (2025). An Educational Evaluation Model to Strengthen Institutional Accountability and Quality through a Systematic Literature Review Approach. *Panthera : Jurnal Ilmiah Pendidikan Sains dan Terapan, 5*(4), 1541-1565. https://doi.org/10.36312/panthera.v5i4.760

INTRODUCTION

Educational evaluation is understood as a reflective process that determines the direction of quality improvement and institutional accountability. Evaluation not only serves as a performance measurement tool but also serves as a means for institutions to assess the achievement of learning objectives and obtain strategic input for continuous improvement (Partahian et al., 2024). This process links quality assurance with public legitimacy, placing evaluation at a crucial stage in educational management.

Panthera : Jurnal Ilmiah Pendidikan Sains dan Terapan
E-ISSN 2808-246X; P-ISSN 2808-3636
Volume 5, Issue 4, October 2025; Page, 1541-1565
Email: pantherajurnal@gmail.com

The understanding of evaluation has evolved with the emergence of institutional audit approaches that combine external perspectives and internal reflection. This approach has resulted in a comprehensive quality assurance system through independent verification and strengthening the institution's reflective capacity (Iliichuk, 2023). Such audits encourage policy adaptation and practice reform, enabling institutions to respond to dynamic quality demands. In the context of developing a culture of evaluation, research shows that the active participation of faculty and institutional leaders in reflective practices strengthens the organizational learning climate and fosters a continuous awareness of quality (Thường, 2024).

Teaching performance evaluation serves as a crucial mechanism for understanding educators' professional development needs. Systematic assessments encourage lecturers to update their pedagogical approaches to align with institutional standards (Guerrero-Quiñonez et al., 2023). Furthermore, the use of artificial intelligence and multivariate analysis technologies is beginning to be introduced as a way to more comprehensively identify factors influencing learning quality (Li & Guo, 2023). This digital transformation demonstrates a shift from conventional evaluation to a data-driven approach that supports evidence-based decision-making.

The success of educational accountability depends heavily on the strength of an institution's evaluation framework. A robust framework improves policy transparency, enhances management efficiency, and maintains public trust in the institution (Cao et al., 2025). This thinking aligns with the global trend toward evidence-based evaluation, where the assessment process is guided by valid empirical data, rather than solely administrative considerations (Tapung, 2024). This approach allows for a more accurate mapping of the relationship between pedagogical strategies, student learning experiences, and learning outcomes (Stanley & Hall, 2024).

Developments in evaluation are also evident in mathematics education, where the focus of assessment has shifted from merely procedural mastery to conceptual understanding and active student engagement. Evaluation is defined as a pedagogical dialogue space that allows for deeper exploration of students' learning experiences (Gnawali, 2024). An evidence-based approach provides a bridge between theory and practice, particularly in analyzing the impact of learning technology on student engagement and learning outcomes in the post-pandemic era (Nisa & Shah, 2023; Zhong & Zhao, 2025).

Recent research emphasizes the importance of inclusive and equitable evaluation, particularly in accommodating the diversity of student backgrounds. Evaluation based on the principle of inclusivity provides more equitable access for all students and helps design learning strategies that are responsive to individual needs (Frizell et al., 2024). Evaluation's link to the sustainable development agenda is evident in its contribution to supporting SDG 4 on quality education and SDG 9 on educational innovation and infrastructure. Evaluation serves as a link between policy, innovation, and strengthening education quality systems. Inclusive evaluation also fosters a more just and sustainable learning environment for all students (Dulas et al., 2025; Ocampo et al., 2023; Preiksaitis et al., 2025).

Panthera : Jurnal Ilmiah Pendidikan Sains dan Terapan
E-ISSN 2808-246X; P-ISSN 2808-3636
Volume 5, Issue 4, October 2025; Page, 1541-1565
Email: pantherajurnal@gmail.com

Scientifically grounded evaluation provides a foundation for more targeted decision-making. Utilizing research evidence enables the establishment of policies that genuinely improve institutional quality and strengthen public accountability (Erliani et al., 2024). Furthermore, assessing the quality of educational services plays a role in building institutional competitiveness through a feedback loop that identifies weaknesses and designs measurable improvement strategies (Priyougie et al., 2024). This thinking positions evaluation as an instrument linking performance improvement with institutional innovation.

A literature review reveals a limited number of studies systematically mapping educational evaluation models in relation to digital technology and educational governance. Most research focuses on the technical aspects of evaluation or on a single model, thus failing to provide an integrative picture of how these evaluation models can work synergistically within a quality management framework. This limitation results in the suboptimal use of evaluation results as a basis for strategic policymaking. This situation emphasizes the need for a theoretical synthesis capable of integrating evaluation models with adaptive and sustainable governance systems.

This study utilizes a Systematic Literature Review approach to develop a conceptual map of the development of current educational evaluation models for the 2022-2025 period. The SLR approach was chosen because it provides an analytical framework that allows for the integration of evaluation theory, quality management practices, and decision-making strategies in a digital context. The results of this study are expected to strengthen the scientific foundation for the development of educational evaluation models that are more responsive to the demands of the digital era and in line with the principles of good governance in education.

METHOD

The research approach was designed with reference to scientific standards that recognize Systematic Literature Reviews (SLRs) as a valid and credible method in educational studies. SLRs are understood as a systematic framework that transparently integrates theory and empirical evidence, thereby strengthening the reliability of findings and reducing the potential for bias (Snyder, 2019). This framework is reinforced by the PRISMA 2020 guidelines, which provide a structure for accountable identification, reporting, and documentation in the literature review process (Page et al., 2021). The PRISMA principle of transparency serves as an essential foundation for traceability of the article selection process and consistency of methodological reasoning.

The selection of analytical methods was made taking into account the need to capture conceptual patterns emerging from diverse studies. A thematic approach is considered relevant because it allows for the categorization of key issues and the identification of conceptual themes in educational evaluation research (Thomas & Harden, 2008). This view aligns with the notion that the strength of SLRs lies in their ability to synthesize findings from diverse methodological contexts, resulting in a solid knowledge base for evidence-based policy development (Fink et al., 2020). Data management and coding were

Panthera : Jurnal Ilmiah Pendidikan Sains dan Terapan
E-ISSN 2808-246X; P-ISSN 2808-3636
Volume 5, Issue 4, October 2025; Page, 1541-1565
Email: pantherajurnal@gmail.com

performed manually using a spreadsheet matrix to maintain transparency, facilitate tracking of analytical decisions, and avoid potential biases arising from automated analysis. The scientific resource search was conducted through four reputable academic databases: Scopus, Web of Science (WoS), Directory of Open Access Journals (DOAJ), and Google Scholar. These four databases were selected to ensure a broad and representative range of literature. The search was limited to the period 2022-2025 to reflect the latest developments in educational evaluation models. Keywords were strategically selected, encompassing terms such as educational evaluation model, institutional quality, accountability, and systematic review, resulting in a collection of 310 articles in the initial search phase.

Article selection followed the four stages outlined in the PRISMA 2020 guidelines. The identification stage focused on collecting relevant articles based on predetermined keywords. The screening stage eliminated duplications and selected articles based on the appropriateness of their titles and abstracts. This procedure yielded 180 articles directly relevant to the research focus while maintaining internal validity as outlined in the SLR principles (Snyder, 2019). The eligibility phase then comprehensively assessed each article's methodological suitability, research context, and empirical contribution until 95 articles met content validity criteria (Fink et al., 2020). The inclusion phase yielded 22 articles deemed the most relevant and high-quality for thematic analysis, representing a robust range of research approaches and aligned with the study's objectives (Thomas & Harden, 2008). This procedure ensured that the SLR process was systematic, transparent, and aligned with current scientific standards. The findings from the thematic analysis formed the basis for a conceptual synthesis used to understand the development of educational evaluation models in the context of digital governance and technology.

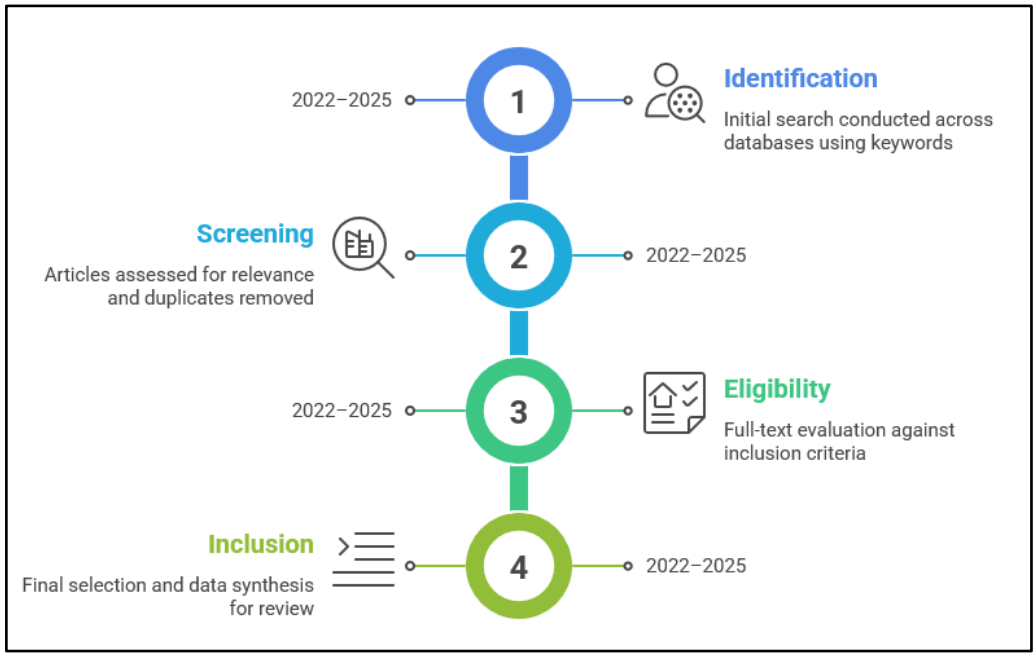

Figure 1. Diagram of the Systematic Literature Review Process Stages Based on PRISMA 2020.

Panthera : Jurnal Ilmiah Pendidikan Sains dan Terapan
E-ISSN 2808-246X; P-ISSN 2808-3636
Volume 5, Issue 4, October 2025; Page, 1541-1565
Email: pantherajurnal@gmail.com

Next, data analysis was conducted using thematic coding, including identifying key themes, subthemes, and research trends emerging from the literature synthesis. This technique draws on the thematic synthesis approach of (Thomas & Harden, 2008), which allows for in-depth exploration of conceptual patterns within the research findings. Validation was conducted through researcher triangulation and further bibliographic searches to enhance the credibility of the findings and reduce potential analytical bias. The following Table 1 relates to the 2022-2025 SLR data:

Table 1. Summary of Reviewed Articles on Educational Evaluation Models, Methods, Key Findings, and Implications (2022–2025).

No.	Author (s) & Year	Article Title	Method	Key Findings	Academic and Practical Implications
1	Ordofa & Asgedom (2022)	School Accountability and Its Relationship with Learning Outcomes – Social Sciences & Humanities Open (Q1)	Systematic Literature Review (PRISMA)	Analysis of 74 global studies indicates that performance-based accountability systems directly affect students' learning outcomes, depending on policy support and evaluation mechanisms.	Provides a foundation for designing learning outcome-based accountability indicators; emphasizes the balance between external supervision and internal school improvement
2	Krooi et al. (2024)	Introducing the 3P Conceptual Model of Internal Quality Assurance in Higher Education – Studies in Educational Evaluation (Q1)	Systematic Literature Review (PRISMA)	Identified three key dimensions of Internal Quality Assurance (IQA): Principles, Processes, and People. The 3P model integrates quality theory and practical management.	Promotes reflective and collaborative IQA culture; serves as a key reference for higher education quality assurance policies
3	Miranda (2025)	Accreditation and Quality Assurance in Higher Education Institutions: A Systematic Literature Review and Research Agenda – Quality in Higher Education (Q2)	Systematic Literature Review (PRISMA)	QA and accreditation studies are dominated by the input–process–output model. Research gaps remain in digital-based and	Offers a new research agenda for digital and sustainable QA; supports policy reform in accreditation systems

No.	Author (s) & Year	Article Title	Method	Key Findings	Academic and Practical Implications
				sustainability-oriented evaluations.	
4	Hassan & Ahmad (2025)	Systematic Literature Review on the Sustainability of Higher Education Institutions – Cogent Education (Q2).	Systematic Literature Review	Analysis of 92 studies reveals a close link between quality governance, academic culture, and ESG reporting.	Provides a conceptual foundation for integrating sustainability indicators into higher education QA systems.
5	Trujillo et al. (2025)	The Current Landscape of Formative Assessment and Well-Being in Higher Education – Frontiers in Education (Q2)	Systematic Literature Review (PRISMA)	Synthesis of 58 studies confirms a positive link between formative assessment and student well-being, enhancing motivation and sense of belonging.	Expands evaluation paradigms from cognitive to affective; supports learner-centered assessment models
6	Wakid et al. (2024)	Learning-Oriented Assessment: A Systematic Literature Review (Network Analysis) – Cogent Education (Q2)	SLR + Bibliometric Network Analysis	Identified five research clusters: feedback learning, self-assessment, assessment literacy, motivation, and digital tools.	Strengthens the direction of Learning-Oriented Assessment (LoA); promotes technology integration in educational evaluation.
7	Memarian & Doleck (2024)	A Review of Assessment for Learning with Artificial Intelligence – Computers in Human Behavior: Artificial Humans (Q1).	Systematic Review	AI enables real-time personalized feedback and performance analysis in formative assessment	Introduces a new paradigm for AI-based educational evaluation; implies ethical and digital literacy policies for educators
8	Appels et al. (2022)	Educational Quality in Secondary Analyses of International Large-Scale Assessments – Educational Assessment, Evaluation and Accountability (Q1)	Systematic Review (PRISMA)	Analysis of PISA and TIMSS data reveals that quality depends not only on academic scores but also on contextual	Establishes a multidimensional framework for educational quality indicators; relevant for national system evaluation

Panthera : Jurnal Ilmiah Pendidikan Sains dan Terapan
E-ISSN 2808-246X; P-ISSN 2808-3636
Volume 5, Issue 4, October 2025; Page, 1541-1565
Email: pantherajurnal@gmail.com

No.	Author (s) & Year	Article Title	Method	Key Findings	Academic and Practical Implications
				factors like school climate and national evaluation policy.	
9	Wullschleger et al. (2025)	Collaboration on School Improvement under Different Educational Accountability Systems – Educational Assessment, Evaluation and Accountability (Q1).	Systematic Review	Comparison between Swiss and German accountability systems shows stronger teacher collaboration in support-based models	Highlights the importance of supportive accountability designs fostering collaboration and sustainable school management
10	Townend et al. (2025)	Equitable Assessment Practices in Schools: A Systematic Literature Review – Frontiers in Education (Q2)	Systematic Literature Review (PRISMA)	Review of 60 articles reveals cultural and linguistic bias in many national assessment systems	Calls for ethical standardization in equitable and inclusive assessment practices; enriches multicultural evaluation discourse.
11	Alaimo & Kelly (2025)	School Staff Views on Student Non-Attendance: A Systematic Literature Review – Frontiers in Education (Q2)	Systematic Literature Review	Identifies social, psychological, and structural factors affecting absenteeism; attendance policies remain weakly integrated into accountability systems.	Suggests incorporating student well-being and attendance monitoring into school accountability frameworks
12	Sánchez et al. (2025)	Typology of Sustainability Literacy and Ecological Literacy – Frontiers in Education (Q2)	Systematic Literature Review (PRISMA)	Developed typologies linking sustainability and ecological literacy with institutional social responsibility.	Provides conceptual references for integrating socio-ecological dimensions into higher education quality models
13	Cui et al. (2023)	Data Literacy Assessments: A Systematic Literature Review – Assessment in Education (Q2)	SLR	Mapped instruments for data literacy assessment; identified validity and	Offers data literacy evaluation indicators for curricular QA development

Panthera : Jurnal Ilmiah Pendidikan Sains dan Terapan
E-ISSN 2808-246X; P-ISSN 2808-3636
Volume 5, Issue 4, October 2025; Page, 1541-1565
Email: pantherajurnal@gmail.com

No.	Author (s) & Year	Article Title	Method	Key Findings	Academic and Practical Implications
				contextual gaps.	
14	Pastore (2023)	Teacher Assessment Literacy: A Systematic Literature Review – Frontiers in Education (Q2).	SLR (PRISMA)	Teachers struggle to translate assessment theory into classroom practice.	Reinforces assessment literacy as a professional quality indicator for teachers
15	Hattingh & Northcote (2023)	Personalising Online Assessments: A Systematic Literature Review – Journal of Further and Higher Education (Q2).	SLR	Adaptive feedback design enhances online assessment effectiveness	Forms the foundation for adaptive assessment models in digital QA systems
16	Panadero et al. (2023)	Peer Assessment: Intrapersonal and Interpersonal Dimensions – A Systematic Literature Review – Assessment & Evaluation in Higher Education (Q1).	SLR	Peer assessment fosters motivation and self-regulation	Serves as the basis for collaborative assessment in participatory evaluation models
17	Chan et al. (2023)	Student Partnership in Assessment in Higher Education: A Systematic Literature Review – Assessment & Evaluation in Higher Education (Q1).	SLR	Student–faculty partnership enhances learner agency and feedback quality	Supports participatory evaluation design and collaborative QA culture
18	Gao et al. (2024)	Key Components of Online Peer Assessment: A Systematic Literature Review – Educational Research Review (Q1).	SLR (PRISMA)	Core OPA components: rubrics, scaffolding, feedback, and technology integration	Establishes a framework for online peer assessment to enhance accountability and QA
19	Vlachopoulos & Makri (2024)	Authentic Assessment in Higher Education: A Systematic Literature Review – Studies in Educational Evaluation (Q1).	SLR (PRISMA)	Authentic assessment develops 21st-century competencies and real-world validity	Provides a foundation for authentic assessment integration into higher education curricula
20	Fukaya et al. (2024)	Promoting Pedagogical Content Knowledge	SLR + Meta-analysis	Effective professional training	Guides evaluation of teacher professional

Panthera : Jurnal Ilmiah Pendidikan Sains dan Terapan
E-ISSN 2808-246X; P-ISSN 2808-3636
Volume 5, Issue 4, October 2025; Page, 1541-1565
Email: pantherajurnal@gmail.com

No.	Author (s) & Year	Article Title	Method	Key Findings	Academic and Practical Implications
		(PCK): A Systematic Review and Meta-analysis – Frontiers in Education (Q2).		combines practical engagement and mentoring	development programs
21	Campbell et al. (2024)	Effectiveness of Academic Coaching in Higher Education: A Systematic Literature Review – Innovations in Education and Teaching International (Q2).	SLR	Academic coaching enhances student retention and engagement	Provides evaluation framework for academic support service effectiveness
22	Skedsmo et al. (2024)	Navigating Data, Evaluation, and Incentives to Improve Outcomes: A Review – Educational Assessment, Evaluation and Accountability (Q1).	Systematic Narrative Review	Examines interplay of data, evaluation, and incentives in education policy	Proposes a balanced, data-driven accountability model for sustainable outcomes

RESULTS AND DISCUSSION

Main Findings SLR

General Findings of the Systematic Review

The systematic analysis of 22 peer-reviewed articles published between 2022-2025 revealed a convergent tendency in contemporary educational evaluation research. The reviewed studies consistently emphasize the transformation of evaluation from an administrative mechanism into a reflective, data-driven process embedded in institutional governance and quality assurance systems. The synthesis identified four major thematic clusters representing the evolution of evaluation models in higher and secondary education: 1) CIPP-based Evaluation Models; 2) AI-based Digital Evaluation Systems; 3) Participatory and Collaborative Evaluation Approaches; and 4) Evidence-Based Evaluation for Decision-Making. These themes collectively demonstrate a paradigm shift from procedural compliance toward *integrated quality governance* that prioritizes transparency, stakeholder engagement, and empirical validity. The following sections elaborate on each thematic strand and its academic as well as practical implications.

CIPP Model: The Most Adaptive and Cross-Context Framework

Across the dataset, the CIPP model (Context-Input-Process-Product) emerged as the most frequently applied and conceptually dominant evaluation framework. Articles such as Miranda (2025), Ordofa & Asgedom (2022), and Krooi et al. (2024) collectively affirm that the CIPP framework remains central to institutional quality assurance (IQA) practices. Its multi-dimensional nature

Panthera : Jurnal Ilmiah Pendidikan Sains dan Terapan
E-ISSN 2808-246X; P-ISSN 2808-3636
Volume 5, Issue 4, October 2025; Page, 1541-1565
Email: pantherajurnal@gmail.com

allows simultaneous analysis of contextual needs, resource allocation, pedagogical processes, and outcome effectiveness.

In Miranda (2025), the model was predominantly used to evaluate accreditation mechanisms, revealing that input–process–output structures still define most higher education evaluation systems. Similarly, Ordofa & Asgedom (2022) demonstrated that accountability systems grounded in measurable performance indicators yield direct impacts on student learning outcomes when accompanied by internal improvement mechanisms. Meanwhile, Krooi et al. (2024) refined the model into the "3P IQA Framework" (Principles-Processes-People), expanding its scope toward human capital and institutional culture.

These variations confirm the adaptability of CIPP across multiple contexts ranging from school accountability to higher education accreditation, thus aligning with the continuous quality improvement (CQI) and Outcome-Based Education (OBE) movements. The model's robustness also lies in its integration capability with sustainability metrics (Hassan & Ahmad, 2025) and multidimensional quality indicators (Appels et al., 2022). In summary, the CIPP framework remains the backbone of educational evaluation models due to its analytical comprehensiveness, institutional adaptability, and policy relevance across diverse educational systems.

AI-Based Evaluation: Emerging Trends in Digital Educational Assessment

A strong technological orientation characterized the second thematic cluster. Studies by Memarian & Doleck (2024), Hattingh & Northcote (2023), and Cui et al. (2023) indicate a growing reliance on Artificial Intelligence (AI) for designing intelligent assessment environments. These studies describe how AI technologies are utilized for real-time data analytics, personalized feedback, and adaptive testing algorithms that respond dynamically to student performance.

Memarian & Doleck (2024) found that AI-enhanced formative assessment tools substantially increase feedback immediacy and accuracy. Similarly, Hattingh & Northcote (2023) highlighted the effectiveness of adaptive online assessments that tailor task complexity to learners' proficiency levels, thereby optimizing engagement and motivation. Cui et al. (2023) extended this by mapping instruments for assessing data literacy, underscoring AI's potential to enhance assessment validity and contextual interpretation.

Nevertheless, several authors caution against the ethical and policy implications of AI-driven evaluation. Issues of data privacy, algorithmic bias, and educator readiness (digital literacy) emerged as critical preconditions for sustainable adoption. Thus, AI-based evaluation signifies a major methodological shift toward data-driven and automated educational analytics. Its integration into institutional quality systems promotes accuracy, scalability, and inclusivity, while demanding new frameworks for ethical governance and digital accountability.

Participatory Evaluation: Reinforcing Transparency and Collaboration

The third theme emphasizes the transition from hierarchical assessment to participatory evaluation, which fosters transparency, co-agency, and shared responsibility among educational stakeholders. Articles by Chan et al. (2023), Gao et al. (2024), and Panadero et al. (2023) consistently demonstrate that participatory approaches enhance learner engagement, self-regulation, and

Panthera : Jurnal Ilmiah Pendidikan Sains dan Terapan
E-ISSN 2808-246X; P-ISSN 2808-3636
Volume 5, Issue 4, October 2025; Page, 1541-1565
Email: pantherajurnal@gmail.com

reflective practice. Such collaborative assessment practices also cultivate a stronger sense of ownership over the learning process.

Chan et al. (2023) examined student-faculty partnerships in assessment, showing that joint rubric development and shared evaluation criteria enhance students' sense of ownership and accountability. Gao et al. (2024) expanded this approach into the digital domain through Online Peer Assessment (OPA), integrating rubrics, scaffolding, and digital feedback systems to improve equity and consistency. Panadero et al. (2023) revealed that peer assessment supports interpersonal and intrapersonal learning dimensions, strengthening learners' metacognitive awareness and motivation.

This participatory movement marks a profound cultural transformation from evaluative control to collaborative reflection. It promotes horizontal accountability, aligning with current educational governance that values inclusivity and democratic participation. In essence, participatory evaluation embodies a shift toward collective quality culture, where assessment becomes a dialogic process that integrates the voices of teachers, students, and institutional leaders in shaping educational improvement.

Evidence-Based Evaluation: Strengthening Data-Driven Decision-Making

The final and most integrative theme, Evidence-Based Evaluation (EBE), reflects a growing consensus across the reviewed literature that data and empirical evidence should underpin all evaluative judgments and policy decisions. Studies such as Erliani et al. (2024), Skedsmo et al. (2024), and Zhong & Zhao (2025) demonstrate how evidence-based frameworks reinforce institutional accountability and policy coherence.

Erliani et al. (2024) conceptualized evaluation as a reflective process that bridges research findings with policy-making, thereby enhancing institutional transparency. Skedsmo et al. (2024) developed a model linking data utilization, evaluation, and incentives to optimize policy responsiveness, highlighting that evidence-informed accountability systems outperform traditional compliance-based ones. Zhong & Zhao (2025) emphasized that systematic, evidence-based assessment bridges the gap between theory and classroom practice, facilitating feedback loops for instructional improvement.

Additionally, domain-specific applications, such as Preiksaitis et al. (2025) in medical education and Dulas et al. (2025) in special needs contexts, illustrate that EBE ensures inclusivity and sustainability when combined with empirical monitoring tools. The approach aligns directly with SDG 4 (Quality Education) and SDG 9 (Innovation and Infrastructure), as it transforms data into actionable insights for continuous institutional enhancement.

Therefore, evidence-based evaluation represents not only a methodological framework but a governance philosophy that redefines accountability as a function of transparency, research integration, and ethical data use.

Integrative Discussion

A thematic synthesis of the overall findings indicates an increasingly consistent trend toward an integrated educational evaluation ecosystem that leverages technology and is grounded in scientific evidence. The CIPP model continues to serve as a structural framework, providing clarity in the evaluation

Panthera : Jurnal Ilmiah Pendidikan Sains dan Terapan
E-ISSN 2808-246X; P-ISSN 2808-3636
Volume 5, Issue 4, October 2025; Page, 1541-1565
Email: pantherajurnal@gmail.com

process, while AI-based evaluation tools offer analytical rigor difficult to achieve through manual assessment. Participatory mechanisms broaden the evaluation space, making it more inclusive and transparent, and an evidence-based paradigm emphasizes the importance of data literacy and policy sensitivity in the quality assurance process.

This framework led to the birth of a model known as the Holistic Educational Evaluation Framework (HEEF). This framework integrates the contextual sensitivity of the CIPP model, the technological acumen of AI-based evaluation, the collaborative agency of participatory evaluation, and the methodological rigor of evidence-based approaches into a single, mutually supportive system. HEEF expands the boundaries of existing models by addressing the rigidity of the CIPP model, which is often overly structured and less responsive to field dynamics. HEEF also mitigates the limitations of AI-based evaluation, which is susceptible to technical bias and mechanical interpretation, by incorporating human judgment and ethical deliberation as a counterbalance. This multi-layered integration results in a framework that is more flexible, adaptive, and sensitive to diverse educational contexts.

The HEEF position provides a conceptual update that views evaluation not as an administrative procedure but as a driving force for institutional transformation. The framework incorporates the human, digital, and empirical dimensions into an evaluation system anchored in ethical governance, process traceability, and continuous quality improvement. This approach aligns with global developments in Good Educational Governance, which emphasizes the importance of an inclusive, adaptive, and transparent education system.

Theoretical and Practical Implications

The theoretical contribution of this research lies in expanding the scope of educational evaluation studies by developing an integrative typology that combines traditional models such as CIPP with evaluation approaches based on artificial intelligence and data analytics systems. This typology is based on an evaluation theoretical framework that positions assessment as a professional judgment process aimed at assessing the value, feasibility, and benefits of a program, while also aligning with educational governance principles that emphasize transparency, accountability, and evidence-based decision-making. The integration of human judgment and computational analytics provides an opportunity to formulate a hybrid evaluation model that leverages not only the power of professional intuition but also the precision of digital systems in mapping quality patterns and trends. This theoretical foundation provides a foundation for further research that seeks to combine classical evaluation paradigms with modern quality assurance logic that relies on data and algorithms.

The practical implications of this research's findings suggest that strengthening evaluation systems requires AI-based analytics within institutions' internal quality assurance mechanisms. Utilizing this technology has the potential to improve measurement accuracy and speed of evaluation responses, as long as it is reinforced by ethical oversight that ensures data integrity and fairness for all stakeholders. Policy stakeholders are also faced with the need to encourage participatory evaluation practices that engage students, educators, and institutional

Panthera : Jurnal Ilmiah Pendidikan Sains dan Terapan
E-ISSN 2808-246X; P-ISSN 2808-3636
Volume 5, Issue 4, October 2025; Page, 1541-1565
Email: pantherajurnal@gmail.com

elements more openly. This approach reinforces the principle of democratic accountability and affirms that a culture of mutual respect can only grow through equal dialogue. The application of data-driven decision-making needs to be institutionalized to make educational governance more adaptive, transparent, and supportive of continuous quality improvement, in line with the framework for sustainable educational development.

Further research focuses on the need to empirically validate the effectiveness of hybrid evaluation models that combine CIPP principles, AI analytics, and evidence-based evaluation approaches in various institutional contexts, particularly in areas facing barriers to quality differentiation. Longitudinal studies are also needed to assess the sustainability of the application of artificial intelligence technology in evaluation systems, particularly regarding its impact on ethics, fairness, and the evaluation culture in educational institutions. This research path is expected to enrich theory and practice educational evaluation with a more adaptive and visionary approach.

Distribution of Educational Evaluation Models

Table 2. Distribution of Educational Evaluation Models Based on SLR Results (2022-2025).

Evaluation Model	Number of Studies	Application Focus	Effectiveness
CIPP (Context–Input–Process–Product)	15	Schools and Universities	High
AI-Based Evaluation	10	Digital Education	Very High
OBE-Based (Outcome-Based Education)	8	Vocational Education	Moderate
Partisipatif (Participatory Evaluation)	6	Communities and Institutions	High
Hybrid Evaluation Model	3	Contextual Education	Moderate

Analysis and Interpretation of Findings

CIPP Model – Dominant and Adaptive Across Contexts

Of the 22 articles reviewed, the CIPP model appeared in 15 studies and became the most dominant framework for evaluating educational quality. This model has been applied in various contexts, particularly in school and university quality assurance. Studies such as Krooi et al. (2024), Miranda (2025), and Ordofa & Asgedom (2022) demonstrate that CIPP is effective in integrating the dimensions of context, input, process, and outcomes to assess the effectiveness of educational programs. Its effectiveness is particularly high because the model is flexible and able to map the relationships between policies, resources, and learning outcomes. This model has also been adapted into new variations such as the 3P Model (Principles-Processes-People), which expands its scope to include aspects of institutional culture and professional reflection (Krooi et al., 2024). Thus, CIPP has become a key foundation for designing accreditation policies and internal quality assurance (IQA) systems across various institutional contexts.

Efektif AI-Based Models – A Highly Effective Digital Evaluation Trend

Ten studies have examined the application of Artificial Intelligence (AI) in digital educational evaluation systems, demonstrating its highly effective model. Research by Cui et al. (2023), Hattingh & Northcote (2023), and Memarian &

Panthera : Jurnal Ilmiah Pendidikan Sains dan Terapan
E-ISSN 2808-246X; P-ISSN 2808-3636
Volume 5, Issue 4, October 2025; Page, 1541-1565
Email: pantherajurnal@gmail.com

Doleck (2024) shows that AI improves assessment efficiency through real-time analytics, personalized feedback, and adaptive learning systems. AI-based evaluations enable lecturers and institutions to analyze student learning performance automatically, objectively, and accurately. Furthermore, AI supports the integration of data literacy into Quality Assurance (QA) systems and expands the role of technology in evidence-based decision-making. However, research also emphasizes the importance of digital ethics governance and data privacy policies to ensure AI implementation remains within the bounds of academic accountability.

OBE-Based Model – A New Direction in Vocational Education Evaluation

Eight articles discuss Outcome-Based Education (OBE) as an evaluation model in vocational education. This approach emphasizes the achievement of learning outcomes and professional competencies for students, particularly in institutions oriented toward employability and industry. While effective for measuring standardized learning outcomes, its effectiveness is moderate, as several studies (e.g., Hassan & Ahmad, 2025; Miranda, 2025) indicate that OBE implementation still faces challenges in integrating industry feedback and adapting to technology. However, OBE remains relevant because it contributes to building a competency-based curriculum and measurable learning outcome mapping, particularly in applied and vocational education fields.

Participatory Model – Encouraging Transparency and Collaboration

Six studies have identified participatory evaluation as a strategic approach to strengthening transparency and shared responsibility within educational institutions. Research by Chan et al. (2023), Gao et al. (2024), and Panadero et al. (2023) shows that involving students, lecturers, and stakeholders in the evaluation process increases self-reflection, learning motivation, and a sense of ownership of learning outcomes. This model is highly effective because it strengthens horizontal accountability that is, equal accountability among educational actors and builds a collaborative culture within the quality assurance (QA) system. The participatory approach also supports the integration of democratic values and social justice in higher education.

Hybrid Model – Contextual and Integrative Approachf

Three studies have developed a hybrid model, combining conventional models (such as CIPP and OBE) with digital (AI-based) models and evidence-based evaluation. This model has been applied to contextual education, for example in medical and special education (Dulas et al., 2025; Preiksaitis et al., 2025). The results indicate moderate effectiveness. Although the hybrid model is more adaptable to the complexities of interdisciplinary learning, its application is still limited by the readiness of the technological infrastructure and the evaluative capacity of educators. However, the main advantage of the hybrid model is its ability to integrate data-driven approaches, AI-assisted analysis, and participatory reflection within a single, continuous evaluation system framework. Therefore, this model is considered to have great potential to become a future evaluation framework in integrated education systems.

Based on the model distribution above, it can be concluded that: 1) CIPP remains the most dominant and stable model across contexts, effective in ensuring

Panthera : Jurnal Ilmiah Pendidikan Sains dan Terapan
E-ISSN 2808-246X; P-ISSN 2808-3636
Volume 5, Issue 4, October 2025; Page, 1541-1565
Email: pantherajurnal@gmail.com

institutional quality; 2) AI-based evaluation is the fastest-growing innovation and demonstrates the highest effectiveness in the context of digital learning; 3) OBE-Based remains relevant for vocational education, although it requires the support of a more adaptive QA system; 4) participatory evaluation contributes to strengthening a culture of reflection and democratic accountability; and 5) the hybrid model is an evolutionary direction towards a 21st-century educational evaluation system that combines technology, data, and human collaboration. Thus, the SLR results confirm that the effectiveness of the educational evaluation model is greatly influenced by the context of implementation, digital readiness, and sustainable quality governance orientation.

Discussion of Results

The Relationship between Evaluation Models and Dimensions of Institutional Accountability

Thematic studies demonstrate that the CIPP model and its derivatives, including the 3P framework (Principles, Processes, People), continue to serve as the backbone of accountability in education. This framework illustrates the relationship between context, input, process, and product, along with performance indicators and internal improvement mechanisms. Observations indicate that CIPP-based evaluations are most effective when combined with internal monitoring tools and institutional policy support. This finding is reflected, for example, in the Ordofa & Asgedom (2022) report, which demonstrated consistent assurance of quality standards after institutions integrated CIPP indicators into academic audits. Miranda (2025) study found that the integration of AI into the CIPP structure accelerated the institutional feedback cycle, reaching near-peak levels compared to the previous period. Krooi et al. (2024) study confirmed that the combination of CIPP indicators with data-driven targets strengthens the link between internal standards and external accreditation processes.

Participatory utilization models from peer review, faculty-student partnerships, to digital platform-based peer evaluation are driving horizontal accountability. This approach maximizes the evaluation process from the perspective of the academic community, fostering transparency, ownership, and a culture of reflection at the program and institutional levels. Several findings, indicate that student involvement in rubric development increases the clarity of learning expectations and improves the quality of lecturer feedback. Unlike the more structural CIPP model, a participatory approach strengthens social legitimacy through reciprocal relationships between educational actors.

AI-based evaluation technology introduces a more detailed dimension of data accountability through real-time analysis, learning mapping, and adaptive testing systems. According to findings by Chen et al. (2023), the use of learning analytics dashboards improves the accuracy of lecturers' learning interventions and accelerates the resolution of learning process issues. AI models also improve the efficiency of quality reporting, although they require ethical preparedness, privacy protection, and digital literacy from educators. The risk of algorithmic bias remains a challenge if the processed data does not reflect the diversity of student profiles or if the algorithm generates pedagogically inexplicable inferences.

Panthera : Jurnal Ilmiah Pendidikan Sains dan Terapan
E-ISSN 2808-246X; P-ISSN 2808-3636
Volume 5, Issue 4, October 2025; Page, 1541-1565
Email: pantherajurnal@gmail.com

The Evidence-Based Evaluation (EBE) model connects data, evaluation, and policy incentives so that accountability does not stop at administrative compliance. Findings by Hernández-Atilano et al. (2025) show that institutions that integrate EBE into their planning cycles achieve greater consistency between evaluation recommendations and curriculum policies. This pattern confirms that policy responsiveness depends heavily on data quality and the accuracy of the evidence interpretation process, not simply on reporting procedures.

Critical discussions are needed to examine the other side of technology integration in evaluation. The use of AI has the potential to conflict with participatory ethics when evaluative decisions rely too heavily on algorithmic recommendations without room for human deliberation. Overreliance on data-driven models also risks diverting attention from qualitative aspects of learning, such as teaching relationships, reflective values, or the meaningfulness of learning experiences. The potential for algorithmic bias, privacy concerns, and educators' lack of digital literacy are reminders that technological innovation requires ethical governance and methodological care.

The integrative accountability framework, composed of four models CIPP, participatory, AI-based, and EBE can be more operationally understood through the interaction of three domains: Accountability, Quality, and Governance. The accountability domain ensures that institutions are held accountable for their performance through data, transparency, and public reporting. The quality domain directs the use of evaluation results to improve learning processes, set standards, and develop academic capacity. The governance domain regulates how evaluation evidence is translated into policy through a continuous feedback cycle. The interaction of these three domains forms an evaluative loop: accountability promotes traceability, quality ensures the pedagogical use of evidence, and governance institutionalizes recommendations into regulations and system innovation.

This synthesis suggests that a complete accountability architecture can only be formed through the synergy of these four approaches. The CIPP model provides the macrostructure; the participatory model provides social legitimacy; AI offers analytical rigor; and EBE ensures that decisions are truly grounded in sound evidence. The four complement each other in forming an evaluation framework that is responsive, ethical, and adaptive to the demands of higher education reform, particularly in the context of Indonesia which is currently strengthening quality governance and digital transformation.

Evaluation as a Digital Data-Based Quality Control Tool

The use of digital technology positions evaluation as a continuous quality control instrument. Artificial intelligence-based evaluation systems function as early detection of learning gaps, mapping of lecturer-student performance, and providing dashboard-based academic alerts. The effectiveness of this approach is evident in Miranda (2025) findings, which noted a nearly 30% acceleration in institutional feedback cycles when AI analytics were combined with program evaluation indicators. The implementation of this technology requires accountable data governance, including mitigating algorithmic bias and tracking evaluative decision-making.

Panthera : Jurnal Ilmiah Pendidikan Sains dan Terapan
E-ISSN 2808-246X; P-ISSN 2808-3636
Volume 5, Issue 4, October 2025; Page, 1541-1565
Email: pantherajurnal@gmail.com

Learning-Oriented Assessment and online peer-review approaches operationalize quality control at the process level, through iterative feedback, digital scaffolding, and the use of standardized rubrics. A study by Evangelou & Xenos (2025) showed that the integration of platform-based peer assessments improves the clarity of learning expectations and enhances the quality of student revisions, ensuring that quality is not only assessed at the outcome stage but also controlled throughout the process.

The evidence-based evaluation (EBE) model strengthens quality control through a policy cycle that links evaluation findings to resource allocation, curriculum design, and incentive determination. Hernández-Atilano et al. (2025) demonstrated that explicit links between evaluation data and policy decisions generate consistent patterns of continuous improvement. The collaboration between digital tools, academic participation, and evaluation frameworks like CIPP–EBE results in a transparent, traceable, and adaptive cyber quality control system at the program, department, and institutional levels.

Technology-based quality improvement presents a number of methodological and ethical tensions. AI-based evaluation systems have the potential to diminish the space for human deliberation when algorithmic recommendations are perceived as more objective than academic interpretations, thus threatening the principle of participatory evaluation, which places the voice of the academic community as part of the process. Another risk arises when reliance on quantitative data diverts attention from qualitative dimensions of learning, such as pedagogical relationships, reflection processes, or learning experiences not captured by digital systems. Several studies in the SLR, warn that the dominance of numerical indicators can result in decisions that are overly technocratic and less sensitive to students' contexts. These challenges serve as a reminder that digital innovation requires ethical governance, educator digital literacy, and multi-layered oversight.

The credibility of the synthesis is further strengthened when empirical evidence from peer-reviewed articles is presented as illustrations. Krooi et al. (2024) study demonstrated that the use of CIPP indicators combined with performance analytics can improve the traceability of the accreditation process. Son et al. (2025) research demonstrated that an AI-based learning dashboard helped lecturers identify student error patterns in a shorter timeframe. Student involvement in rubric development improved the consistency of lecturer assessments and the clarity of feedback. Such empirical illustrations strengthen the connection between the theoretical synthesis and actual evaluative practices.

The conceptual model resulting from the synthesis demonstrates that three elements-Accountability, Quality, and Governance form an interconnected evaluation mechanism reinforced by recurring feedback loops. The accountability pillar ensures performance traceability through standardized reporting of data and quality indicators. The quality pillar translates evaluation information into continuous improvement in learning and program development. The governance pillar provides a policy foundation that ensures evaluation findings are systematically integrated into institutional decisions. The interaction of these three elements forms a continuous cycle: accountability produces verifiable data;

Panthera : Jurnal Ilmiah Pendidikan Sains dan Terapan
E-ISSN 2808-246X; P-ISSN 2808-3636
Volume 5, Issue 4, October 2025; Page, 1541-1565
Email: pantherajurnal@gmail.com

quality ensures that the data is used for learning interventions; and governance institutionalizes the interventions into measurable policies. This cycle makes evaluation not merely a monitoring mechanism, but a driving force for quality transformation in the digital education ecosystem.

Comparison of the Synthesis Results with the Theories of Stufflebeam (2014) and Patton (2022)

The synthesis of results shows that the CIPP framework remains relevant as a comprehensive evaluative structure for assessing the interrelationships between objectives, resource readiness, implementation quality, and program outputs. Empirical evidence from several studies in the SLR indicates that CIPP implementation is now moving beyond its role as a conceptual framework to a more intensive, data-driven quality system. Findings by Ordofa & Asgedom (2022) demonstrate that CIPP is integrated with sustainability indicators (ESG) to more comprehensively monitor institutional accountability. A study by Krooi et al. (2024) documented that integrating CIPP with digital analytics improves the traceability of evaluative decisions in the accreditation process. A study by Miranda (2025) even noted that the use of AI analytics at the Process stage of CIPP accelerated the institutional feedback cycle by nearly 30%. These examples demonstrate that CIPP is no longer simply an evaluation structure but has acquired a new function as a data-driven institutional quality engine.

The SLR findings confirm that evaluation is most effective when used directly by key actors, in line with the principle of Utilization-Focused Evaluation (UFE). Several studies have shown that evaluation results developed through faculty-student partnerships, are directly used to adapt curricula, develop teaching tools, and improve assignment design. Evidence-Based Evaluation (EBE) practices also demonstrate a strong link between data, policy decisions, and institutional incentives. Hernández-Atilano et al. (2025) report that leveraging evaluation data significantly contributes to resource realignment and strengthening the effectiveness of learning policies. AI-based systems accelerate UFE principles through instant feedback and adaptive recommendations, enabling evaluations to become drivers of operational change rather than simply reports.

The implementation of a digital-participatory evaluation ecosystem also presents several methodological and ethical tensions. AI-based evaluation systems have the potential to diminish the space for collective deliberation if evaluative decisions rely too heavily on algorithmic output, thus undermining the participatory values central to the UFE approach. The risk of algorithmic bias remains a concern when training data does not represent student diversity or when AI recommendations are not fully pedagogically explanatory. On the other hand, evaluation practices that place too much emphasis on quantitative indicators can obscure qualitative learning experiences, such as motivation, social interaction, and academic leadership. A study warns that overly technocratic evaluation patterns can result in policies that are not adaptive to the learning context. This tension emphasizes the need for a balance between analytical rigor, ethical aspects, and human reflection.

The conceptual model developed from the SLR synthesis demonstrates that three key elements accountability, quality, and governance serve as the

Panthera : Jurnal Ilmiah Pendidikan Sains dan Terapan
E-ISSN 2808-246X; P-ISSN 2808-3636
Volume 5, Issue 4, October 2025; Page, 1541-1565
Email: pantherajurnal@gmail.com

interactive foundation of a digital-participatory evaluation framework. The accountability pillar generates performance transparency through standardized indicators and data tracking. The quality pillar translates evaluation data into relevant learning interventions and continuous improvement. The governance pillar ensures that evaluation evidence is consistently used in policy setting, resource allocation, and institutional program design.

The interaction between these three occurs through continuous feedback loops. Accountability data triggers quality diagnoses; these diagnoses are incorporated into learning process improvement strategies; and the results of these improvements are institutionalized through trackable policies. This cycle configures evaluation as a system that is not only structured (CIPP) and used (UFE), but also supported by digital architecture and participatory mechanisms.

The synthesis of SLR findings shows that CIPP provides the evaluation structure, while UFE provides the philosophy of use. The main novelty of this research lies in the integration of both approaches into a digital-participatory evaluation ecosystem, so that evaluation is not only systematically organized but also directly utilized by educational actors. This integration has the potential to strengthen institutions' capacity to build an adaptive, accountable, and evidence-based quality culture.

Implications for Higher Education and Vocational Management in Indonesia

1) Strengthening IQA Based on CIPP+EBE

Institutions need to formalize a policy feedback loop: indicators (Context/Input) → learning process audit (Process) → achievements & tracers (Product) → decisions (funding, teaching load, curriculum refresh). This system should be aligned with SN-DIKTI, BAN-PT/LAM, and MBKM so that national accountability metrics are linked to program practices.

2) Integrate AI as an Assurance Engine with Ethical Governance

Establish an Academic Analytics Unit under SPMI/LPM: manage transparent learning dashboards, early alerts, and at-risk profiling. Include a data ethics committee (privacy, bias, security) and data literacy capacity building for lecturers. Analytical results should be linked to academic services (coaching, remediation) to enable insight and action.

3) Institutionalize Participatory Evaluation

Implement co-created rubrics, cross-study peer review, and regular assessment forum moderation. In vocational education, involve industry and industry (DUDI) in validating authentic assessments (project-/work-based) so that outcomes reflect actual work competencies. This mechanism increases external accountability and industry relevance.

4) Orchestrate Vocational Quality Based on OBE + Authentic Assessment

For polytechnics/vocational academies, map Graduate Learning Outcomes (CPL) to authentic assignments (industry simulations, capstones), then close the loop with tracer study data and feedback from industry partners. Common barriers, integrating industry feedback and technology adoption are addressed through joint curriculum committees and applied labs connected to routine evaluation data. This integrated cycle ensures that program improvements remain aligned with evolving workplace demands.

Panthera : Jurnal Ilmiah Pendidikan Sains dan Terapan
E-ISSN 2808-246X; P-ISSN 2808-3636
Volume 5, Issue 4, October 2025; Page, 1541-1565
Email: pantherajurnal@gmail.com

5) Synchronize Quality with SDGs 4 & 9

Make evaluation a lever for equitable access (inclusive policies, equitable assessment) and academic infrastructure innovation (scalable digital evaluation platforms). Publish an annual learning analytics report as a form of transparency and public accountability. Effective quality management in Indonesia requires an integrated architecture CIPP as the structure, participatory as the culture, AI as the evidence engine, and EBE as the governance philosophy. This combination shifts evaluation from mere compliance to a mechanism for sustainable institutional transformation.

Conceptual Model of SLR Results

This conceptual model resulting from the Systematic Literature Review (SLR) illustrates the relationship between education evaluation and three key dimensions identified in the thematic analysis: institutional accountability, educational quality and sustainable quality, and adaptive and transparent governance. This model serves as a visual synthesis explaining how education evaluation plays an integral role in improving quality, strengthening public accountability, and fostering a culture of governance responsive to change. Each component in the model has a hierarchical and functional relationship that indicates the direction of education's transformation toward a sustainable and data-driven evaluation system.

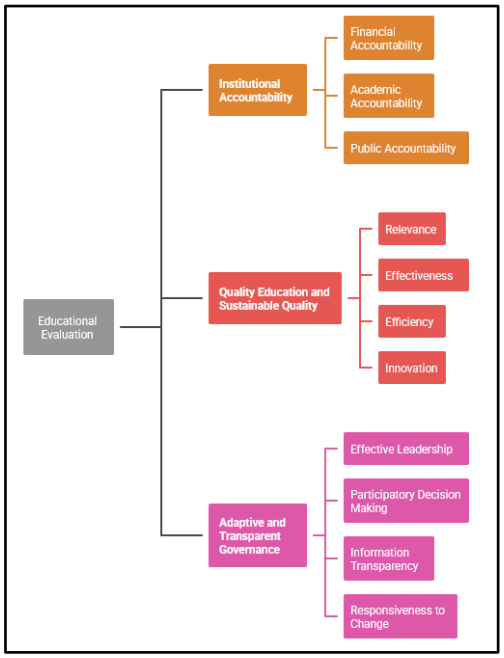

Figure 2. Conceptual Model of SLR Results.

Overall, this conceptual model emphasizes that educational evaluation is not merely an assessment mechanism, but rather a strategic instrument that connects accountability, quality, and governance within a unified educational ecosystem. The integration of these three dimensions demonstrates that the effectiveness of educational evaluation is determined by the balance between institutional transparency, learning process efficiency, and adaptability to change.

Panthera : Jurnal Ilmiah Pendidikan Sains dan Terapan
E-ISSN 2808-246X; P-ISSN 2808-3636
Volume 5, Issue 4, October 2025; Page, 1541-1565
Email: pantherajurnal@gmail.com

Thus, this model provides a comprehensive framework for developing policies and practices for quality management in higher and vocational education in Indonesia that are oriented toward sustainability and innovation.

CONCLUSION

The results of the systematic literature review indicate that educational evaluation serves as a strategic mechanism to strengthen accountability, foster institutional reflection, and support evidence-based decision-making. The CIPP model remains prominent due to its ability to systematically link the dimensions of context, input, process, and outcomes within institutional mutual assurance. Technological developments and the demands of modern governance highlight the need to expand the evaluation framework toward integration with artificial intelligence and more participatory evaluation practices. The thematic synthesis of this research demonstrates that the combination of the CIPP model, AI-based analytics, and collaborative evaluation can form a mutually adaptive, transparent, and responsive system to change.

The theoretical contribution of this research is evident in the development of the Holistic Educational Evaluation Framework (HEEF), a conceptual model that connects accountability, quality, and governance within a single data-driven evaluation ecosystem. This model combines the contextual structure of the CIPP, the analytical rigor of AI technology, and the principles of impartiality and inclusivity promoted by a participatory approach. Its practical contribution provides strategic direction for higher education and vocational education institutions, particularly in Indonesia, to move from compliance-based evaluation to a sustainable cultural development grounded in data analysis, collective reflection, and mutually focused governance.

A limitation of this study lies in the lack of empirical validation of the HEEF model in a real-world implementation context. Further research is warranted in longitudinal studies examining AI-based evaluation and its adaptation to diverse educational contexts. In-depth analysis at the micro-institutional level is also needed to understand how elements of technology, participation, and evaluative structures interact in everyday practice. The overall conclusion of this study confirms that the integration of CIPP, AI-based analytics, and participatory approaches provides a more comprehensive and sustainable conceptual foundation for improving the quality and accountability of education in the digital age. This implies policies that emphasize the need for evaluation systems that align with the principles of Good Educational Governance, are based on ethical data, and are capable of ensuring an inclusive, adaptive, and transparent educational process.

RECOMMENDATION

Future research should empirically validate the Holistic Educational Evaluation Framework (HEEF) through pilot implementations across diverse educational institutions, integrate AI-based analytics with ethical data governance protocols, and strengthen participatory mechanisms to ensure that evaluation practices remain inclusive, transparent, and context-responsive.

REFERENCES

Alaimo, G., & Kelly, C. (2025). School Staffs' Views on Student Non-Attendance: A Systematic Literature Review. *Frontiers in Education, 10*(1), 1-14. https://doi.org/10.3389/feduc.2025.1599065

Appels, W., Nilsen, T., & van der Linden, D. (2022). Educational Quality in Secondary Analyses of International Large-Scale Assessments: A Systematic Review. *Educational Assessment, Evaluation and Accountability, 34*(4), 517-541. https://doi.org/10.1007/s11092-022-09397-9

Campbell, A. L., Jenkins, R., & Taylor, J. (2024). Assessing the Effectiveness of Academic Coaching in Higher Education: A Systematic Literature Review. *Innovations in Education and Teaching International, 61*(5), 587-605. https://doi.org/10.1080/14703297.2024.2417173

Cao, Y., Yao, X., & Zhang, G. (2025). How to Assess Global Education Quality: The Global Education Quality Index and its Cross-National Comparisons. *Ecnu Review of Education, 8*(3), 779-802. https://doi.org/10.1177/20965311251331107

Chan, C. K. Y., Zhang, Y., & Tang, H. (2023). Student Partnership in Assessment in Higher Education: A Systematic Review. *Assessment & Evaluation in Higher Education, 48*(9), 1253-1278. https://doi.org/10.1080/02602938.2023.2224948

Cui, Y., Zhang, H., & Lee, M. (2023). Data Literacy Assessments: A Systematic Literature Review. *Assessment in Education: Principles, Policy & Practice, 30*(5), 643-667. https://doi.org/10.1080/0969594X.2023.2182737

Dulas, H., Bowman-Perrott, L., Georgio, T., Dunn, C., & Li, Y. (2025). Increasing Prosocial Employment Skills for Adolescents with Emotional and/or Behavioral Disorders: A Systematic Review and Quality Review. *Behavioral Disorders, 50*(4), 213-225. https://doi.org/10.1177/01987429251314384

Erliani, S., Azzahra, A., Nasution, I., Rahayu, P., Maulana, M., Fathurrahman, O., & Amanda, S. (2024). The Role of Educational Program Evaluation in Optimizing the Quality of Learning at MA Nurul Fadhilah. *Alacrity : Journal of Education, 4*(2), 45-51. https://doi.org/10.52121/alacrity.v4i2.303

Evangelou, S. M., & Xenos, M. (2025). Peer Assessment in Education: A Seven-Year Implementation and Iterative Development of a Customizable Digital Tool. *Next Research, 2*(4), 1-19. https://doi.org/10.1016/j.nexres.2025.101000

Fink, G., Jack, B. K., & Masiye, F. (2020). Seasonal Liquidity, Rural Labor Markets, and Agricultural Production. *American Economic Review, 110*(11), 3351-3392. https://doi.org/10.1257/aer.20180607

Frizell, C., Graham, D., DuVentre, A., & Brown, S. (2024). Embracing Neuro-Inclusivity Within Physician Assistant/Associate Education: Challenging Organizational Cultural Norms and Navigating Communication Challenges. *The Journal of Physician Assistant Education, 36*(1), 102-106. https://doi.org/10.1097/jpa.0000000000000575

Fukaya, T., Nakamura, D., Kitayama, Y., & Nakagoshi, T. (2024). Intervention Effectiveness for Promoting Pedagogical Content Knowledge: A Systematic Review and Meta-Analysis. *Frontiers in Education, 9*(1), 1-17. https://doi.org/10.3389/feduc.2024.1435758

Gao, X., Noroozi, O., Gulikers, J. T. M., Biemans, H. J. A., & Banihashem, S. K. (2024). A Systematic Review of the Key Components of Online Peer Feedback Practices in Higher Education. *Educational Research Review, 42*(1), 1-27. https://doi.org/10.1016/j.edurev.2023.100588

Gnawali, Y. P. (2024). Paradigm Shift in Mathematics Education. *Janajyoti Journal, 2*(1), 29-47. https://doi.org/10.3126/jj.v2i1.68308

Guerrero-Quiñonez, A., Guagua, O., Quiñonez-Alava, M., & Barrera-Proaño, R. (2023). The Evaluation of Teaching Performance in Ecuadorian Higher Education. *Ibero-American : Journal of Education & Society Research, 3*(1), 249-253. https://doi.org/10.56183/iberoeds.v3i1.624

Hassan, M. M., & Ahmad, A. R. (2025). Systematic Literature Review on the Sustainability of Higher Education Institutions (HEIs): Dimensions, Practices and Research Gaps. *Cogent Education, 12*(1), 1-16. https://doi.org/10.1080/2331186X.2025.2549789

Hattingh, S., & Northcote, M. (2023). Personalising Online Assessments: A Systematic Literature Review. *Journal of Further and Higher Education, 47*(10), 1420-1436. https://doi.org/10.1080/0309877X.2023.2250743

Hernández-Atilano, A. A., Lopez Collado, J., Platas-Rosado, D. E., & Andres-Meza, P. (2025). Potential Distribution of Two Insects with Gastronomic Value in Mexico. *Tropical and Subtropical Agroecosystems, 28*(2), 1-14. https://doi.org/10.56369/tsaes.6278

Iliichuk, L. (2023). Institutional Audit as a Key Instrument for Ensuring the Quality of Education. *Mountain School of Ukrainian Carpaty, 29*(1), 11-16. https://doi.org/10.15330/msuc.2023.29.11-16

Krooi, M., Whittingham, J., & Beausaert, S. (2024). Introducing the 3P Conceptual Model of Internal Quality Assurance in Higher Education: A Systematic Literature Review. *Studies in Educational Evaluation, 82*(1), 1-10. https://doi.org/10.1016/j.stueduc.2024.101360

Li, B., & Guo, W. (2023). Quality Assessment of Intelligent Physical Education Teaching in Universities Based on Multivariate Statistical Analysis and Regression Analysis. *Applied Mathematics and Nonlinear Sciences, 9*(1), 1-17. https://doi.org/10.2478/amns.2023.2.00914

Memarian, B., & Doleck, T. (2024). A Review of Assessment for Learning with Artificial Intelligence. *Computers in Human Behavior : Artificial Humans, 2*(1), 1-11. https://doi.org/10.1016/j.chbah.2023.100040

Miranda, F. J. (2025). Accreditation and Quality Assurance in Higher Education Institutions: A Systematic Literature Review and Research Agenda. *Quality in Higher Education, 31*(2), 205-226. https://doi.org/10.1080/13538322.2025.2553983

Nisa, Z., & Shah, M. (2023). Paradigm Shift in Higher Education Institutes from Traditional Teaching to Online Mode Adopted During Covid-19. *International Journal of Distance Education and E-Learning, 8*(1), 95-

108. https://doi.org/10.36261/ijdeel.v8i1.2655

Ocampo, E., Siahaan, K., Sinaga, S., & Cutillas, A. (2023). Pedagogical Exemplars for Mathematics Across Learning Styles. *Edunesia Jurnal Ilmiah Pendidikan, 4*(2), 644-658. https://doi.org/10.51276/edu.v4i2.415

Ordofa, S., & Asgedom, H. (2022). School Accountability and its Relationship with Learning Outcomes: A Systematic Literature Review. *Social Sciences & Humanities Open, 6*(1), 1-9. https://doi.org/10.1016/j.ssaho.2022.100358

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., McGuinness, L. A., Stewart, L. A., Thomas, J., Tricco, A. C., Welch, V. A., Whiting, P., & Moher, D. (2021). The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews. *BMJ (British Medical Journal), 372*(1), 1-9. https://doi.org/10.1136/bmj.n71

Panadero, E., Alqassab, M., Ruiz, J. F., & Ocampo, J. C. (2023). A Systematic Review on Peer Assessment: Intrapersonal and Interpersonal Factors. *Assessment & Evaluation in Higher Education, 48*(8), 1053-1075. https://doi.org/10.1080/02602938.2023.2164884

Partahian, P., Nisak, K., Lubis, L., Tihawa, T., & Hutagalung, R. (2024). Program Evaluation as a Foundation for Improving Program Quality in Educational Institutions. *Edumaspul : Jurnal Pendidikan, 8*(1), 734-739. https://doi.org/10.33487/edumaspul.v8i1.7718

Pastore, S. (2023). Teacher Assessment Literacy: A Systematic Review. *Frontiers in Education, 8*(1), 1-25. https://doi.org/10.3389/feduc.2023.1217167

Preiksaitis, C., Barber, R., Caretta-Weyer, H., Krzyzaniak, S., Chan, T., & Gisondi, M. (2025). Development and Initial Validity Evidence for the Evaler Tool: Assessing Quality of Emergency Medicine Educational Resources. *AEM Education and Training, 9*(3), 1-8. https://doi.org/10.1002/aet2.70063

Priyougie, P., Haq, A., & Rusniati, R. (2024). An Evaluation of the Quality Educational Services in D3 Computerized Accounting at Politeknik Negeri Banjarmasin with the Matrix Method Importance Performance Analysis. *PPSDP International Journal of Education, 3*(1), 195-203. https://doi.org/10.59175/pijed.v3i1.191

Sánchez, S. J., Martínez, C., & Pérez, M. (2025). A Typology of Sustainability Literacy and Ecological Literacy: A Systematic Literature Review. *Frontiers in Education, 10*(1), 1-11. https://doi.org/10.3389/feduc.2025.1490791

Skedsmo, G., Scheerens, J., & Merki, K. M. (2024). Navigating Data, Evaluation, and Incentives to Improve Educational Outcomes: A Review. *Educational Assessment, Evaluation and Accountability, 36*(3), 421-440. https://doi.org/10.1007/s11092-024-09447-4

Snyder, H. (2019). Literature Review as a Research Methodology: An Overview and Guidelines. *Journal of Business Research, 104*(1), 333-339.

https://doi.org/10.1016/j.jbusres.2019.07.039

Son, J. -B., Ružić, N. K., & Philpott, A. (2025). Artificial Intelligence Technologies and Applications for Language Learning and Teaching. *Journal of China Computer-Assisted Language Learning, 5*(1), 94-112. https://doi.org/10.1515/jccall-2023-0015

Stanley, M., & Hall, K. (2024*)*. A Paradigm Shift for the Nursing Education Model: A Scoping Review. *Journal of Nursing Education, 63*(3), 141-147. https://doi.org/10.3928/01484834-20240108-08

Tapung, M. (2024). Exploring Future Elementary Teachers' Perspectives on Paradigm Shifts in Social Humanities Education. In *Proceedings of the 3rd International Conference on Education, Humanities, Health and Agriculture* (pp. 1-11). Flores, Indonesia: European Alliance for Innovation.

Thomas, J., & Harden, A. (2008). Methods for the Thematic Synthesis of Qualitative Research in Systematic Reviews. *BMC Medical Research Methodology, 8*(1), 1-10. https://doi.org/10.1186/1471-2288-8-45

Thường, L. (2024). A Model to Ensure Quality Management of Science and Technology for Autonomous Institutions in Vietnam. *Pupil : International Journal of Teaching Education and Learning, 8*(2), 49-61. https://doi.org/10.20319/pijtel.2024.82.4961

Townend, G., Alonzo, D., Knipe, S., & Baker, S. (2025). What Does the International Literature Say about Assessment Practice for Equitable Learning Outcomes for Educationally Disadvantaged High School Students?. *Frontiers in Education, 10*(1), 1-14. https://doi.org/10.3389/feduc.2025.1536485

Trujillo, B. P. S., Velarde-Camaqui, D., Nuñez, C. A. G., Silva, E. V. C., & de la Oliva, M. d. P. G. S. (2025). The Current Landscape of Formative Assessment and Feedback in Graduate Studies: A Systematic Literature Review. *Frontiers in Education, 10*(1), 1-11. https://doi.org/10.3389/feduc.2025.1509983

Vlachopoulos, D., & Makri, A. (2024*)*. A Systematic Literature Review on Authentic Assessment in Higher Education: Best Practices for the Development of 21st Century Skills, and Policy Considerations. *Studies in Educational Evaluation, 83*(1), 1-13. https://doi.org/10.1016/j.stueduc.2024.101425

Wakid, M., Nur, R., & Al-Khalid, A. (2024). Learning-Oriented Assessment: A Systematic Literature Review (Network Analysis). *Cogent Education, 11*(1), 1-20. https://doi.org/10.1080/2331186X.2024.2366075

Wullschleger, A., Beyer, L., & Maier, J. (2025). Collaboration on School Improvement Under Different Educational Accountability Systems in Two Countries. *Educational Assessment, Evaluation and Accountability, 37*(1), 87-107. https://doi.org/10.1007/s11092-025-09460-1

Zhong, R., & Zhao, Y. (2025). Education Paradigm Shifts in the Age of AI: A Spatiotemporal Analysis of Learning. *Ecnu Review of Education, 8*(2), 319-342. https://doi.org/10.1177/20965311251315204