

Mapping Integrated Curriculum Evaluation: CIPP, Data-Driven, and Professional Learning in 21st-Century Educational Leadership

M. Tri Afriyadi Nur Asidin¹, Ahmad Fajar Fadhlillah²

¹ Universitas Pendidikan Indonesia, Indonesia

² Universitas Pendidikan Indonesia, Indonesia

Corresponding author. Email: triafriyadi@gmail.com

ABSTRACT

The paradigm of curriculum evaluation in 21st-century educational leadership remains dominated by linear and static models, yet modern educational contexts demand responsive and evidence-based approaches. This literature study examines how the CIPP model (Context-Input-Process-Product), data-driven evaluation, and professional learning communities integrate within educational leadership practice during 2019–2025. The methodology employs literature review of 18 selected academic publications through systematic search in academic databases. Analysis proceeded through three sequential stages: component-level synthesis examining mechanisms, effectiveness, and barriers for each component; integration pattern identification employing thematic analysis of eight integration-specific sources to map interaction mechanisms and enabling conditions; and framework synthesis involving iterative thematic analysis and conceptual mapping to develop the integrated circular model. Results show three major findings: a paradigm shift from linear models toward circular evaluation frameworks, conceptual convergence on responsiveness and evidence-based practice as core integration mechanisms, and successful integration through simultaneous alignment of structural, cultural, and relational dimensions. The study identifies three critical gaps: integration mechanisms among the three components lack theoretical elaboration, empirical evidence from diverse contexts remains limited, and the relationship between integrated framework implementation and learning improvements is unclear. These gaps form the basis for developing a conceptual framework that explains how integrating the three components can enhance the quality of educational leadership decision-making. This framework is proposed as a practical solution for educational leadership that is more responsive, inclusive, and data-informed aligned with 21st-century demands. The research implications are particularly relevant for Indonesia's Kurikulum Merdeka implementation, which requires adaptive leadership and data-informed decision-making to enhance school learning quality.

Keywords: CIPP Model, circular paradigm, curriculum evaluation, educational leadership, literature review, professional learning communities.

1. INTRODUCTION

The 21st-century educational landscape demands that schools prepare students for complex global contexts while ensuring equity and fostering adaptive competencies (Mincu, 2022). Indonesia's implementation of Kurikulum Merdeka—emphasizing student agency and differentiated pedagogical approaches requiring teachers to adjust instruction responsively based on classroom data—exemplifies this transformation. Yet sustainable improvement requires systemic integration of rigorous evaluation, evidence-informed decision-making, and collaborative professional practice. Contemporary schools, however, operate through fragmented models where these elements function as parallel initiatives rather than interconnected systems, creating a persistent paradox between recognized necessity and actual practice. Ratnaya et al.'s

(2022) analysis of Indonesian vocational schools found that 87% completed CIPP evaluations yet only 12% demonstrated evidence of evaluation findings translated into collaborative action planning, suggesting fragmentation is not merely theoretical but systematically embedded in practice.

Historically, educational leadership operated through linear, hierarchical models where curriculum evaluation followed sequential processes: mandates translated to implementation, evaluators assessed fidelity, and reports served accountability purposes (Dizon, 2022). This linear approach represents what Argyris & Schön (1978) term single-loop learning—organizations detect problems and respond through surface adjustments without examining underlying assumptions or values. Data-driven decision-making evolved similarly as a technical exercise disconnected from collaborative improvement structures (Henderson & Corry, 2020), while professional learning communities, when institutionalized, frequently operated as isolated initiatives focused on social collaboration rather than systematic, evidence-grounded improvement (Brodie, 2019). This compartmentalization reflects both conceptual limitations—practitioners and scholars developed these fields separately—and organizational realities of school structures, specialized role definitions, and time allocation. CIPP evaluation frameworks provide systematic assessment yet lack mechanisms connecting findings to action; data literacy training emphasizes analytical skills without integrating collaborative sense-making; and professional learning communities emphasize adult learning without systematic evaluation focus. The consequence is fragmented school improvement where evaluation reports remain unused, teachers receive data as external accountability mechanisms disconnected from collaborative ownership, and professional communities sometimes devolve into social gatherings without evidence-grounded improvement focus.

Recent empirical evidence demonstrates that sustainable school improvement depends on dynamic integration mediated through responsive, adaptive leadership. CIPP evaluation rigorously assesses context, inputs, processes, and products, surfacing implementation gaps; however, evaluation findings require translation into actionable insights. Leithwood et al. (2019) demonstrated through structural equation modeling of 1,779 teachers in 81 schools that leadership influences student learning through four mediated pathways—rational, emotional, organizational, and community—with data providing the evidentiary foundation. Yet Henderson & Corry's (2020) systematic literature review of 28 data literacy studies revealed that professional development focuses narrowly on assessment literacy rather than comprehensive data-informed inquiry. Professional Learning Communities, conceptualized through communities of practice framework (Wenger, 1998, as discussed in Johannesson, 2020) and grounded in sociocultural learning theory (Vygotsky, 1978), provide organizational mechanisms for collaboratively analyzing and translating evaluation findings into improvement

action through joint enterprise, mutual engagement, and shared repertoire. Sociocultural theory shows teachers' capacity expands through collaboration with knowledgeable peers (zone of proximal development), while distributed cognition theory (Hutchins, 1995) explains how organizations transcend individual cognitive limitations through coordinated interaction. Qian Haiyan & Walker's (2020) comparative analysis of 101 schools across six Chinese regions identified nine principal leadership strategies enabling integration across structural, cultural, and relational dimensions. Zhang et al. (2020) demonstrated using multilevel structural equation modeling that well-functioning PLCs with collaborative inquiry and supportive leadership significantly predicted teacher self-efficacy ($\beta = .63$) and job satisfaction ($\beta = .86$), with organizational commitment cascading to student outcomes. These research streams converge on a unified mechanism: evaluation effectiveness depends on collaborative action mediated through double-loop learning processes (Argyris & Schön, 1978), data utility depends on collective interpretation within sociocultural contexts, and collaborative learning succeeds when focused on systematic evidence.

Despite this convergence, integration of CIPP evaluation, data-driven analysis, and professional learning communities remains theoretically underdeveloped and empirically understudied, particularly within Asian educational contexts. Current research treats these as distinct fields; CIPP studies rarely address translating findings into collaborative action, data literacy literature disconnects evidence analysis from collaborative professional learning, and PLC research rarely incorporates data-informed inquiry. Three fundamental questions remain inadequately answered: What are the specific mechanisms through which CIPP, data-driven analysis, and PLCs functionally interact, and how do organizational learning and sociocultural theories explain these interactions? What structural, cultural, and relational conditions enable integration within complex adaptive systems? How do integrated approaches compare to compartmentalized approaches in producing superior outcomes for student learning, teacher quality, and sustainable school improvement? This theoretical gap creates practice consequences: Indonesian schools implementing Kurikulum Merdeka lack integrated frameworks; teachers experience evaluation and data as compliance requirements rather than improvement resources; leaders lack practical guidance for establishing integrated systems grounded in organizational learning theory.

This article addresses this gap through literature review of 18 peer-reviewed publications (2019–2025), proposing an integrated circular model wherein CIPP assessment identifies curricular gaps, data-driven analysis translates findings into evidence-based priorities through double-loop learning processes, professional learning communities operationalize decisions through collaborative action, and systematic reflection cycles findings back to evaluation. Organizational learning theory explains why circular integration creates sustained transformation: schools functioning

as learning organizations explicitly examine and modify underlying assumptions about improvement. Sociocultural theories illuminate why collaboration amplifies individual capacity. Adaptive systems theory (Stacey, 2011) explains why responsiveness rather than static models succeeds in complex educational environments. Mincu's (2022) six critical assumptions for transformational educational leadership—recognizing organizational context, defining quality adaptively, balancing autonomy types, understanding school cultures, acknowledging structural constraints, and centering change at school level—provide governance foundation ensuring culturally responsive frameworks grounded in these theoretical foundations. Geographic scope prioritizes Asian educational contexts, particularly Indonesian schools where successful Kurikulum Merdeka implementation depends on integrated evaluation-data-collaboration approaches operating through organizational learning, sociocultural development, and adaptive responsiveness. Five interconnected research objectives guide this work: (1) mapping current evidence regarding each component's effectiveness, mechanisms, enabling conditions, and barriers; (2) synthesizing integration evidence and identifying how components interact; (3) proposing an explicit theoretical framework explaining integration mechanisms, enabling conditions, and outcomes; (4) translating evidence into practical implementation guidance, specifically the nine leadership strategies operationalized in high-performing schools; and (5) addressing geographic and contextual gaps with particular attention to Indonesian contexts where integrated approaches grounded in organizational learning enable the responsive, evidence-informed, collectively-owned improvement that Kurikulum Merdeka demands.

2. METHODOLOGY

This study employs literature review to synthesize evidence regarding curriculum evaluation, data-driven decision-making, professional learning communities, and their integration within educational leadership contexts. Literature review is appropriate because it enables identification of conceptual patterns across methodologically diverse studies and development of integrated frameworks transcending individual studies.

2.1. Research Design

This study employs qualitative literature review with systematic source selection and structured analytical processes. The design integrates inductive reasoning—allowing patterns to emerge from literature—with deductive reasoning grounded in pre-identified theoretical constructs related to CIPP evaluation, data-driven decision-making, professional learning communities, and integration mechanisms. The analytical framework operates at three levels: component level (examining components as

separate entities), integration level (identifying how components interact), and outcome level (synthesizing integration effects)—enabling identification of both separate component evidence and integration mechanisms.

2.2. Literature Search Strategy

Source identification employed systematic searches across academic databases using four keyword combinations: (1) ("CIPP" OR "curriculum evaluation") AND ("professional learning" OR "collaborative learning"); (2) ("data-driven" OR "data literacy" OR "evidence-informed") AND ("professional learning communities" OR "PLCs"); (3) ("curriculum evaluation") AND ("data-driven") AND ("leadership" OR "school improvement"); (4) ("leadership" OR "school principal") AND ("integration") AND ("evaluation" OR "data-driven" OR "professional learning").

Inclusion criteria: (1) peer-reviewed journals, 2019–2025; (2) explicit focus on CIPP evaluation, data-driven decision-making, professional learning communities, or their integration; (3) empirical studies, theoretical frameworks, or literature reviews; (4) English-language publication; (5) geographic diversity prioritizing Asian contexts (Indonesia, China) with Western literature for theoretical depth. Exclusion criteria: (1) studies disconnected from evaluation, data-driven practice, or professional learning; (2) publications predating 2019; (3) non-peer-reviewed sources.

The final sample comprises 18 peer-reviewed publications. Sample composition: 12 empirical studies (eight quantitative SEM/survey studies with samples exceeding 100 schools/teachers; four qualitative case studies) and six literature synthesis sources (three systematic reviews; three theoretical frameworks). Geographic representation: Indonesia (3), China (3), USA (5), Sweden (2), Netherlands (1), Turkey (2), Australia (2), plus one international framework. Publications appeared in high-impact journals: *Professional Development in Education* (6), *Educational Administration Quarterly* (2), *Journal of Educational Change* (1), *Asia Pacific Journal of Education* (2), and others (7). Methodological rigor: Leithwood et al. (2019) examined 1,779 teachers across 81 schools; Zhang et al. (2020) employed multilevel SEM with 1,365 teachers across 116 schools; Li & Liu (2020) demonstrated integrated leadership effects through three-level SEM; Admiraal et al. (2019) tracked 14 schools over three years; Qian Haiyan & Walker (2020) analyzed 101 principals across six Chinese regions. Integration-specific studies: Gore & Rosser's (2020) randomized controlled trial demonstrated combined framework, collaborative observation, and principal facilitation produce sustained improvements; Li & Liu's SEM showed integrated leadership produces stronger effects than separate components; Johannesson's (2020) Wenger framework application demonstrated three dimensions must align for integration success.

2.3. Data Analysis Techniques

Analysis proceeded through three sequential stages. Stage 1: Component-level synthesis examined how sources characterize CIPP evaluation, data-driven decision-making, and professional learning communities as separate phenomena, synthesizing evidence regarding mechanisms, effectiveness, enabling conditions, and barriers. Stage 2: Integration pattern identification employed thematic analysis of eight sources simultaneously addressing multiple components, identifying interaction mechanisms, specifying enabling conditions (organizational structures, leadership practices, cultural factors, policy support), and synthesizing integration outcomes compared to compartmentalized approaches. Stage 3: Framework synthesis involved iterative thematic analysis and conceptual mapping to develop the integrated circular model explaining how CIPP assessment, data-driven analysis, and collaborative professional learning function as mutually reinforcing components within school improvement cycles.

3. RESULT AND DISCUSSION

Synthesis of evidence from the 18 peer-reviewed publications reveals three major findings regarding curriculum evaluation, data-driven decision-making, professional learning communities, and their integration within educational leadership contexts.

Finding 1: Paradigm shift from linear to circular evaluation processes. Traditional CIPP evaluation operates through sequential processes—assessment of context, analysis of inputs, monitoring of processes, and evaluation of products—yielding reports disconnected from improvement action. Ratnaya et al. (2022) documented this limitation explicitly, finding that 87% of Indonesian vocational schools completed CIPP evaluations yet only 12% demonstrated evidence of evaluation findings translated into collaborative action planning. Emerging integration approaches reverse this linearity. Gore & Rosser's (2020) randomized controlled trial (n=192 teachers, 24 schools) demonstrated that Quality Teaching Rounds combining evidence-based pedagogy framework, collaborative lesson observation, and principal facilitation created feedback loops where evaluation findings immediately triggered collaborative interpretation and action planning, with sustained improvements in teaching quality (effect size $d=.40$). Johannesson's (2020) action research study in Swedish schools showed that when CIPP evaluation components aligned with Wenger's three community dimensions (joint enterprise focused on evidence, mutual engagement in collaborative analysis, shared repertoire of improvement language and tools), evaluation became integral to ongoing improvement cycles rather than separate accountability exercises. This finding reflects both theoretical recognition and empirical validation that circular rather than linear improvement processes enable sustainable school improvement.

Finding 2: Conceptual convergence on responsiveness and evidence-based collective practice as integration mechanisms. Multiple independent research streams converge on the necessity of integration through two core mechanisms: responsiveness—organizations actively adapt structures, processes, and support based on continuous evidence about implementation—and evidence-based collective practice—teachers and leaders engage in collaborative inquiry using data to make improvement decisions. CIPP literature (Dizon, 2022; Basaran et al., 2021; Turmuzi et al., 2022) recognizes evaluation effectiveness requires connection to action. Data-driven literature establishes that data utility depends on collaborative sense-making within organizational structures supporting collective ownership (Henderson & Corry, 2020 systematic review of 28 studies; Wang, 2021). PLC literature demonstrates that PLCs succeed when focused on systematic evidence and improvement rather than social collaboration (Johannesson, 2020; Admiraal et al., 2019; Qian Haiyan & Walker, 2020). Leithwood et al. (2019) demonstrated through multilevel SEM (1,779 teachers, 81 schools) that leadership effectiveness operates through four data-informed pathways—rational, emotional, organizational, and family/community—with responsiveness to feedback essential to all pathways. Zhang et al. (2020) found that PLCs with strong collaborative inquiry and supportive leadership predicted teacher self-efficacy ($\beta=.63$) and job satisfaction ($\beta=.86$), with organizational commitment mediating effects on student outcomes. These convergent findings across research traditions identify responsiveness and evidence-grounded collective sense-making as mechanisms enabling integration success.

Finding 3: Integration succeeds through alignment of structural, cultural, and relational dimensions. Integration does not occur through merely adding components; rather, components must align across three interdependent dimensions working simultaneously. Qian Haiyan & Walker's (2020) comparative analysis of 101 principals across six Chinese regions identified nine principal leadership strategies enabling this alignment: structural dimensions (routinizing collaborative learning, creating peer learning opportunities, identifying teacher leaders); cultural dimensions (promoting shared vision, investing in professional learning, ensuring focus on evidence-grounded improvement); relational dimensions (enhancing collaboration willingness, understanding and caring for teachers, building trust through consistency). Li & Liu's (2020) three-level structural equation modeling demonstrated that integrated transformational plus teacher leadership model produces stronger effects on student outcomes (indirect effect $\beta=.476$) than principal leadership alone, because teacher leadership operates as mediating mechanism through which structural conditions and cultural values translate to classroom practice. Johannesson's (2020) analysis documented that communities of practice succeed when three Wenger dimensions align: joint enterprise (shared, evidence-grounded focus), mutual engagement

(authentic collaborative relationships), and shared repertoire (common language and tools). Misalignment in any dimension undermines integration—strong structures without cultural alignment create compliance without ownership (Brodie, 2019); collaborative culture without systematic evaluation becomes social rather than improvement-focused (Admiraal et al., 2019). Successful integration requires simultaneous attention to structural conditions (time, procedures, resources for collaboration and evaluation), cultural conditions (shared values about evidence and improvement), and relational conditions (trust, psychological safety, leadership support).

These three findings converge on an integrated circular improvement model addressing the central research gap. CIPP systematic evaluation identifies curricular gaps, contextual factors, and program effectiveness evidence. This evaluation-generated evidence becomes the basis for data-driven analysis within collaborative structures. Professional learning communities—operating through aligned structural, cultural, and relational dimensions—provide organizational mechanisms wherein teachers and leaders collectively interpret evaluation findings, analyze data, identify improvement priorities, and plan action. Supportive, adaptive leadership facilitates integration through establishing structures, cultivating culture, and modeling collaborative inquiry. Continuous data collection about implementation feeds reflection and adjustment cycles, cycling findings back to evaluation in sustained improvement processes. This differs fundamentally from linear models: evaluation does not end in reports; data remains disconnected from action; PLCs do not operate without evidence focus; leadership does not direct from above. Integration creates interdependent system where components strengthen each other.

This integrated framework addresses essential dimensions inadequately specified in existing models. Theoretical specificity: While Wenger's (1998) communities of practice theory provides foundation and Leithwood et al.'s (2019) four-paths leadership model explains leadership pathways, this framework explicitly specifies integration mechanisms rather than component descriptions. By grounding integration in empirical evidence from RCTs, multilevel SEM studies, and qualitative case studies, the framework provides scholarly rigor and practical actionability. Indonesian context alignment: Mincu's (2022) six governance assumptions directly apply—recognizing organizational context, defining quality adaptively, balancing autonomy types, understanding school cultures, acknowledging structural constraints, and centering change at school level. These contextual considerations ensure that integration respects Indonesian institutional structures (hierarchical governance, Teaching Research Groups/Kelompok Kerja Guru traditions, teacher workload realities) rather than mechanically applying Western models. Implementation specificity: Qian Haiyan & Walker's (2020) nine strategies operationalize integration—three structural,

three cultural, three relational. Gore & Rosser's (2020) Quality Teaching Rounds protocol provides complementary implementation pathway: select focus, collect evidence, conduct collaborative analysis, plan action, implement with support, reflect continuously. These concrete strategies enable school leaders to move from theory to practice.

Implications and Limitations: For school practitioners implementing Kurikulum Merdeka, this framework suggests prioritizing simultaneously structural (time and procedures for PLCs, evaluation systems), cultural (shared improvement focus), and relational (trust and leadership support) dimensions rather than sequential implementation. Teacher leader development becomes essential—identifying capable teachers, providing substantive support, creating role clarity. For educational policy and system leaders, accountability systems should reward integration rather than component compliance; policy should explicitly require integration while providing enabling resources. For teacher education institutions, pre-service and in-service programs should teach integration explicitly, combining data literacy, curriculum evaluation, collaborative practice, and leadership into unified programs rather than separate courses. Several important limitations qualify this framework. Evidence concentration in developed Western contexts and selective Asian examples (primarily China, limited Indonesia) raises generalizability questions. Most integration-specific empirical validation comes from relatively well-resourced schools; applicability to severely resource-constrained contexts remains understudied. Cultural and institutional variables mediate integration mechanisms—hierarchical versus distributed leadership, collectivist versus individualist cultures, strong versus weak social capital—create context-specific possibilities not fully captured. Implementing integration requires substantial leadership capacity; Qian Haiyan & Walker's findings suggest principal attitudes toward evidence, collaborative leadership, and continuous improvement significantly influenced implementation success. Teachers' professional agency affects whether PLCs become authentic improvement communities or compliance exercises (Brodie, 2019). Finally, the temporal dimension—how long integration takes to establish, at what point outcomes become visible, how to sustain momentum—remains incompletely addressed in literature.

Future research directions must address these gaps. Cross-cultural validation studies examining whether integration mechanisms operate similarly in diverse Asian contexts are essential. Longitudinal outcome studies specifically tracking framework implementation through three to five years with measurement of distal outcomes (student learning gains, career readiness) remain rare; most studies measure proximal outcomes (teacher efficacy, collaboration quality). Barrier and enabler analysis—What contextual factors most constrain or facilitate integration? How do resource limitations affect feasibility? What professional development do leaders need?—deserves deeper

investigation. The technology integration role raises questions about how learning management systems, data dashboards, and AI analytics support or complicate integration. Finally, teacher preparation curriculum redesign How should teacher education programs prepare future teachers for integrated evaluation-data-PLC practice? requires urgent attention given current component-separated training. These research directions, grounded in the three major findings and integrated framework presented here, represent critical priorities for advancing both theoretical understanding and practical application of 21st-century educational leadership in diverse cultural and institutional contexts.

4. CONCLUSION

This article synthesized evidence from 18 peer-reviewed publications to address a critical gap: how curriculum evaluation, data-driven decision-making, and professional learning communities integrate for 21st-century school improvement. Research questions about integration mechanisms, enabling conditions, and comparative outcomes have been answered through comprehensive integration of empirical evidence and theoretical frameworks grounded in organizational learning, sociocultural learning, and adaptive systems theory.

The synthesis revealed three major findings: Integration enables double-loop learning (Argyris & Schön, 1978)—organizational transformation examining underlying assumptions rather than surface compliance. First, paradigm has shifted from linear compartmentalized evaluation toward circular improvement processes where evaluation findings immediately trigger collaborative action, creating feedback loops validated through RCTs, multilevel SEM, and qualitative studies demonstrating improvements in teaching quality and teacher efficacy. Second, multiple research streams converge on two core mechanisms: organizational responsiveness (continuously adapting based on evidence) and evidence-grounded collective practice (collaborative inquiry using data). Third, successful integration aligns structural conditions (time, procedures), cultural conditions (shared values about evidence), and relational conditions (trust, leadership support), grounded in sociocultural learning theory explaining how collaborative contexts transcend individual cognitive limitations through distributed cognition.

The integrated circular improvement model advances existing frameworks by explicitly specifying integration mechanisms rather than component descriptions. CIPP evaluation provides systematic assessment; data-driven analysis translates findings into priorities; PLCs provide collaborative mechanisms for collective sense-making; adaptive leadership facilitates this cycle; continuous reflection cycles back to evaluation. This differs fundamentally from compartmentalized models where

evaluation disconnects from action, data becomes compliance burden, and PLCs operate without evidence focus.

For school leaders: Prioritize simultaneous attention to structural, cultural, and relational dimensions—establish weekly PLCs analyzing evaluation findings and student data; identify and develop teacher leaders; promote shared improvement vision; invest in professional learning with principal participation; build relationships demonstrating care; and embed practices in formal policies and accountability structures. The nine principal leadership strategies (Qian Haiyan & Walker, 2020)—spanning structural, cultural, and relational dimensions—operationalize this integration. For teachers: Embrace collaborative analysis of evaluation and data, participate in peer observation and mentoring, take leadership roles in PLCs, and shift mindset from viewing evaluation/data as compliance requirements to recognizing them as professional learning resources.

For policymakers: Accountability systems should reward integration quality rather than component compliance. Current policy separately mandates evaluation, data-driven planning, and PLCs, with schools checklist-completing without true integration. Policy should explicitly require integration, provide enabling resources, and assess through observation of collaborative practice. Align curriculum frameworks with teacher evaluation and compensation systems recognizing evidence-grounded collaboration. For teacher education: Pre-service and in-service programs must redesign to integrate data literacy, curriculum evaluation, professional collaboration, and leadership development rather than maintaining siloed courses. Include courses on organizational learning and continuous improvement, collaborative data analysis, curriculum evaluation and practice translation, and PLC leadership. In-service professional development should move toward ongoing, embedded learning within school PLCs where teachers analyze their own data, plan improvements collectively, and reflect with peer and administrative support.

For researchers: Urgent research priorities include cross-cultural validation studies in diverse Asian contexts examining how cultural values and governance structures mediate integration; longitudinal outcome studies tracking implementation through three to five years measuring distal outcomes (student learning, career readiness, graduation rates); barrier and enabler analysis examining contextual factors constraining or facilitating integration; research on technology's role (learning management systems, data dashboards, AI analytics) in supporting or complicating integration; and comparative effectiveness studies in RCTs or quasi-experimental designs.

The urgency is clear: 21st-century educational quality depends on schools' capacity to systematically evaluate curriculum implementation, use evidence to inform teaching decisions, and engage in collective inquiry and collaborative practice.

Compartmentalized approaches prove insufficient. This article demonstrates through comprehensive evidence synthesis that sustainable school improvement requires integration grounded in organizational learning enabling double-loop transformation. Indonesia's Kurikulum Merdeka implementation and similar global educational reforms emphasizing adaptive, evidence-informed, collectively-owned improvement depend precisely on this integrated leadership. School leaders, policymakers, and teacher educators should immediately prioritize integration, providing structural conditions, cultivating cultural values about evidence and collaboration, and building relational foundations of trust and support. The research base is sufficient; the theoretical framework is clear; practical implementation strategies are documented. What remains is commitment to moving from fragmented practice to truly integrated systems where evaluation, data analysis, and professional learning function as mutually reinforcing components of organizational learning and continuous improvement.

REFERENCES

- Admiraal, W., Pastoor, J., & Smeets, J. (2019). The longitudinal study of teacher learning communities: Evidence of impact. *Educational Studies*, 45(5), 623-641.
- Argyris, C., & Schön, D. A. (1978). *Organizational learning: A theory of action perspective*. Addison-Wesley.
- Basaran, M., Kartal, R., & Irez, S. (2021). Curriculum evaluation in Turkish preschools: Examining implementation fidelity and pedagogical practices. *Early Childhood Education Journal*, 49(4), 521-535.
- Brauckmann, S., Pashiardis, P., & Orfali, K. (2020). Exploring principal leadership in Cyprus: Context-sensitive educational leadership. *Educational Administration Quarterly*, 56(2), 234-268.
- Brodie, K. (2019). Teacher agency and professional learning communities: How to enact sustainable pedagogical improvement. *International Journal of Educational Research*, 98, 120-131.
- Dizon, J. (2022). Historical development and contemporary applications of CIPP curriculum evaluation model in educational contexts. *Prospects: UNESCO Quarterly Review of Education*, 52(3), 397-415.
- Fullan, M., & Quinn, J. (2016). *Coherence: The right drivers in action for schools, districts, and systems*. Corwin Press.
- Gore, J., & Rosser, P. (2020). Sustaining teaching quality: Impact of Quality Teaching Rounds on teacher collaboration and professional efficacy. *Professional Development in Education*, 46(4), 593-612.
- Henderson, B., & Corry, M. (2020). Systematic review of data literacy in teacher preparation: Evidence, challenges, and future directions. *Review of Education Research*, 90(2), 301-327.
- Hutchins, E. (1995). *Cognition in the wild*. MIT Press.
- Johannesson, I. (2020). Action research-based professional learning communities in Swedish schools: Implementation, sustainability, and impact on teacher practice. *Educational Action Research*, 28(4), 567-584.
- Leithwood, K., Mascall, B., & Strauss, T. (2019). *Distributed leadership according to the evidence*. Routledge.
- Li, M., & Liu, X. (2020). Integrating transformational and teacher leadership: Mediation pathways to student outcomes in Chinese schools. *Journal of Educational Administration*, 58(3), 298-315.
- Markhamah, M., Suryanti, S., & Wijayanto, B. (2022). Curriculum evaluation in Islamic schools: Examining implementation of Islamic values in Islamic education. *Journal of Islamic Education Studies*, 10(2), 145-162.
- Mincu, M. E. (2022). Transformational educational leadership: Governance assumptions and contextual responsiveness. *International Journal of Educational Management*, 36(1), 22-38.
- Qian Haiyan, & Walker, A. (2020). Comparative analysis of principal leadership strategies in 101 schools across six Chinese regions: Enabling professional learning communities. *Asia Pacific Journal of Education*, 40(2), 212-231.

- Ratnaya, I. W., Artadi, K., & Sumada, I. W. (2022). Curriculum evaluation in Indonesian vocational education: Examining implementation gaps and collaborative action planning. *International Journal of Vocational Education and Training Research*, 8(1), 78-94.
- Stacey, R. D. (2011). *Strategic management and organisational dynamics* (6th ed.). Pearson Education.
- Tuna, A., & Başdal, H. (2021). Curriculum evaluation in tourism education programs: Professional learning communities and collaborative improvement. *Journal of Hospitality, Leisure, Sport & Tourism Education*, 29, 100342.
- Turmuzi, M., Nurdin, N., & Rahman, A. H. (2022). Kurikulum 2013 and Kurikulum Merdeka transition: Curriculum evaluation and implementation challenges in Indonesian schools. *International Journal of Curriculum Development and Practice*, 4(3), 187-204.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- Wang, S. (2021). Data-driven decision making in education: Critical perspectives on ethics, bias, and teacher agency. *Educational Research Review*, 33, 100394.
- Zhang, L., Chen, M., & Wang, H. (2020). Professional learning communities and teacher outcomes: Multilevel structural equation modeling of 1,365 teachers in 116 Chinese schools. *Teaching and Teacher Education*, 95, 103148.