International Journal of Educational Qualitative Quantitative Research (IJE-QQR)
ISSN: 2962-9918 (Online)
Vol. 5, No. 1, pp. 170-183

Bridging the Training-to-Classroom Gap: A Kirkpatrick Evaluation of Teacher Professional Development in Rural Indonesia

Zikri Rakhman[1], Muh Yusup[2]
[1,2]Institut Elkatarie, Lombok Timur, Indonesia

Article Info
Article history:
Received: Feb 21, 2026
Revised: Jun 15, 2026
Accepted: Jul 3, 2026

DOI: 10.58418/ijeqqr.v5i1.206

How to cite this article:
Rakhman, Z., & Yusup, M. (2026). Bridging the Training-to-Classroom Gap: A Kirkpatrick Evaluation of Teacher Professional Development in Rural Indonesia. *International Journal of Educational Qualitative Quantitative Research*, *5*(1), 170–183. https://doi.org/10.58418/ijeqqr.v5i1.206

Read online:

Scan this QR code with your smart phone or mobile device to read online.

ABSTRACT

While global investments in teacher professional development (TPD) have surged, the "training-to-classroom" gap remains a persistent challenge in under-resourced, rural settings. This study evaluates the systemic impact of three primary TPD modalities, comprising formal certification (PPG), decentralized teacher working groups (MGMP), and centralized in-service workshops (Diklat), on instructional quality by utilizing Kirkpatrick's Four-Level Evaluation Model. Adopting a qualitative bounded case study design centered in East Lombok as a region emblematic of the infrastructural and pedagogical challenges in rural Indonesia, data were generated from 38 key educational stakeholders through semi-structured interviews, 48 rigorous classroom observations, and systematic document analysis. The findings unpack a non-linear decoupling across Kirkpatrick's evaluative levels. While TPD initiatives successfully generated positive teacher reactions (Level 1) and conceptual knowledge gains (Level 2), behavioral transfer into actual classroom practice (Level 3) was severely bottlenecked by rigid institutional accountability, resource scarcity, and generic program designs. Crucially, measurable improvements in teaching quality (Level 4) did not occur linearly but were ecologically contingent upon active principal instructional leadership and localized peer-learning networks, with the decentralized MGMP model showing the highest adaptive potential. Ultimately, this study challenges the linear efficacy assumptions of standardized TPD models in the Global South, demonstrating that professional growth in rural contexts is structurally mediated. It offers a socio-contextual framework for policy-makers to transition from standardized in-service training toward school-embedded, leadership-supported professional development ecosystems.

Keywords: Training-to-classroom Gap, Teacher Professional Development, Kirkpatrick, Rural Education, Instructional Quality

Corresponding Author:
Zikri Rakhman
Institut Elkatarie, Lombok Timur, Indonesia
Email: zickrysengav@gmail.com

1. INTRODUCTION

Teacher quality is consistently recognized as one of the most influential school-level determinants of student learning outcomes. Evidence from recent systematic reviews and meta-analyses indicates that high-quality teacher professional development (TPD) significantly improves instructional quality and student achievement, particularly when it is sustained, collaborative, and practice-based (Ventista & Brown, 2023; Visscher et al., 2025). Nevertheless, substantial disparities in instructional capacity and access to effective professional development remain, especially in low- and middle-income countries, posing continuing challenges to achieving Sustainable Development Goal 4 (Ambon et al., 2024; Zickafoose et al., 2024). Despite massive financial mobilization, the translation of TPD investments into actual classroom practice

remains deeply inconsistent worldwide. Developing countries grapple with a persistent training-to-classroom gap where heavily funded capacity-building programs are rarely evaluated with empirical rigor, resulting in highly variable instructional outcomes (Popova et al., 2022). This structural disconnect poses a critical challenge in peripheral ecologies where geographic remoteness and institutional fragmentation heavily compromise implementation fidelity.

In Indonesia, Teacher Professional Development (TPD) was formally institutionalized through the Law on Teachers and Lecturers Number 14 of 2005, which mandates the continuous development of teachers' pedagogical, professional, social, and personal competencies. To operationalize this mandate, the government has implemented several professional development pathways, including the Teacher Professional Education Program (PPG), Subject Teacher Working Groups (MGMP), and centralized in-service training (Diklat). However, empirical evidence indicates that many national TPD initiatives remain centrally designed, generic, and insufficiently responsive to teachers' local classroom contexts, particularly in rural and under-resourced schools (Revina et al., 2023). Recent systematic reviews further demonstrate that effective professional development requires sustained duration, active learning, collaborative engagement, coaching, and strong alignment with teachers' instructional contexts, whereas one-off, generic training has limited impact on classroom practice and student learning (Z. Li et al., 2023; Liao et al., 2023; Surahman & Wang, 2023; Ventista & Brown, 2023).

This policy disconnect is particularly pronounced in socioeconomically vulnerable regions, making East Lombok Regency in West Nusa Tenggara Province an analytically significant setting for investigation. Teacher shortages have become a growing global concern, with recruitment difficulties being especially severe in rural and disadvantaged areas where attracting and retaining qualified teachers remains challenging (OECD, 2024; Pholphirul et al., 2023). East Lombok exemplifies these structural constraints as a densely populated yet resource-constrained district characterized by geographic isolation, limited access to professional learning opportunities, and persistent shortages of qualified teachers. These challenges are reflected in West Nusa Tenggara's relatively low human development outcomes and continuing educational disparities reported in official regional statistics (BPS NTB, 2023). Collectively, these contextual conditions raise an important question regarding whether nationally standardized teacher professional development programmes can be effectively adapted to highly constrained rural environments to generate meaningful improvements in classroom practice.

Although scholarship on TPD evaluation has expanded, a major theoretical lacuna persists. International consensus establishes that professional development improves instruction only when it is continuous, content-focused, and embedded within active peer-learning communities (Kart, 2026; Ventista & Brown, 2023). Yet, empirical evidence from the Global South demonstrates that positive training outcomes are highly unstable, frequently evaporating shortly after the program ends (Brunetti et al., 2024). While Indonesian literature has documented national systemic constraints and specific design gaps within single programs like PPG (Attabik & Zainiza, 2025; Chang et al., 2025), these studies operate almost exclusively at aggregate national or provincial levels. Consequently, the localized institutional dynamics governing district-level execution remain invisible. Existing research fails to evaluate the systemic cycle of rural TPD through multi-level frameworks, leaving a profound gap in our understanding of how regional socioeconomic, geographic, and administrative architectures collectively mediate the path between teacher learning and actual behavioral transfer.

To bridge these interconnected gaps, this study evaluates the systemic impact of Indonesia's three primary TPD modalities, which are PPG, MGMP, and Diklat, on instructional quality within rural classrooms. Utilizing Kirkpatrick's Four-Level Evaluation Model (Alsalamah & Callinan, 2022; Paul et al., 2024), we interrogate the path of professional growth across four distinct dimensions: reaction, learning, behavior, and results. The scientific novelty of this research lies in its empirical interrogation of the linear assumptions of evaluation theory when applied to under-resourced, peripheral settings.

This study advances three original contributions to the global literature on educational reform. First, it delivers the first systematic, multi-level evaluation of multi-tiered training programs in a peripheral Indonesian regency, producing actionable data for district policymakers. Second, it extends the theoretical boundary of the Kirkpatrick model by demonstrating how structural bottlenecks disrupt its linear transfer assumptions, thereby enriching evaluation science in resource-constrained ecologies. Third, by mapping the specific institutional enablers and structural barriers that moderate classroom application, this study provides a socio-contextual framework for TPD design in the Global South.

2. METHOD

2.1. Research Design

This study adopts a qualitative bounded case study design to investigate how teacher professional development (TPD) modalities shape instructional practice within rural ecologies. Grounded in an interpretivist epistemological framework, this research views professional growth and behavioral transfer not as linear mechanics, but as phenomena constructed through the interactions between educators and their

institutional environments. East Lombok Regency serves as the bounded case, functioning as an analytical representative of the structural, infrastructural, and pedagogical challenges characterising rural Indonesia.

While Kirkpatrick's Four-Level Evaluation Model (Alsalamah & Callinan, 2022; Paul et al., 2024) historically features positivistic applications, this study adapts the framework as a qualitative heuristic scaffold. Rather than quantifying outcomes, the four levels (reaction, learning, behavior, and results) guide the qualitative interrogation of mechanisms, barriers, and enablers that mediate the trajectory from training attendance to classroom application.

2.2. Research Site and Participants

Fieldwork was executed across a purposively selected network of educational institutions in East Lombok Regency, West Nusa Tenggara Province, Indonesia. To capture the full spectrum of rural educational constraints, a maximum variation sampling strategy altered three core dimensions: school level (primary, junior secondary, and senior secondary), geographic location (urban administrative centers and isolated rural sub-districts), and institutional governance (public and private schools). Participant selection utilized criterion-based purposive sampling across three institutional tiers to establish a systemic vantage point. Table 1 establishes the operational dimensions of teaching quality used during field tracking, while Table 2 outlines the finalized participant matrix.

Table 1. Operational Framework for Classroom Observations.

Teaching Quality Dimension	Observed Field Indicators	Aligned Kirkpatrick Tier
Pedagogical Content Knowledge	Mastery of subject matter, contextualization of concepts to rural environments, integration of student-centered activities.	Level 2 (Learning) & Level 3 (Behavior)
Instructional Management	Classroom time allocation, formative assessment execution, responsiveness to diverse student learning paces.	Level 3 (Behavior)
Student Engagement	Frequency of active student questioning, peer-to-peer collaborative task completion, sustained focus during instruction.	Level 4 (Results)

Table 2. Participant Categories, Selection Criteria, and Sample Size.

Participant Category	Selection Criteria	Sample Size (N)
Classroom Teachers	Minimum of 2 years of experience; completion of at least one formal TPD modality (PPG, MGMP, or Diklat) within the past 3 years.	24
School Principals	Currently active in a participating school; holds direct supervisory and instructional leadership responsibility.	8
Policymakers & Facilitators	Direct involvement in the administrative design, fiscal delivery, or facilitation of district TPD programs.	6
Total Sample	Thematic saturation achieved at 34; 4 additional interviews conducted to confirm stability.	38

2.3. Data Collection Instruments

Data generation relied on the strategic triangulation of semi-structured interviews, classroom observations, and systematic document analysis to maximize interpretive depth and eliminate self-reporting bias. Semi-structured interviews (N=38) lasted between 45 and 90 minutes, were conducted in Indonesian, audio-recorded with consent, and transcribed verbatim. Interview protocols were mapped directly onto the adapted Kirkpatrick tiers. Teacher protocols investigated affective satisfaction (Level 1), cognitive gains (Level 2), classroom implementation barriers (Level 3), and perceived instructional outcomes (Level 4). Principal and policymaker protocols focused on structural enablers, including fiscal support, resource provision, accountability metrics, and post-training monitoring systems.

Classroom observations (N=48) provided direct empirical verification of actual instructional practices, mitigating the discrepancy between teachers' narrative accounts and real classroom execution. Individual observation sessions spanned a complete instructional block of 35 to 45 minutes. Driven by the protocols established in Table 1, the field researchers recorded extensive descriptive and reflective field notes, logging real-time teacher pedagogical behaviors, student engagement levels, and material adaptations.

Document analysis contextualized the primary data. The analyzed archival corpus comprised centralized TPD modules, training attendance registers, regional education strategic development plans, school-level professional development logs, and localized student performance metrics. This layer facilitated the cross-referencing of official policy designs against lived field realities.

2.4. Data Analysis Procedure

Thematic analysis followed a structured hybrid inductive-deductive workflow managed via NVivo 14 software. The four primary levels of the Kirkpatrick model functioned as the deductive organizational framework, while inductive coding was simultaneously utilized to capture unanticipated contextual barriers and institutional enablers specific to rural Indonesia. The systemic data analysis workflow, which progresses from raw data triangulation to the final thematic matrix, is conceptualized in Figure 2.

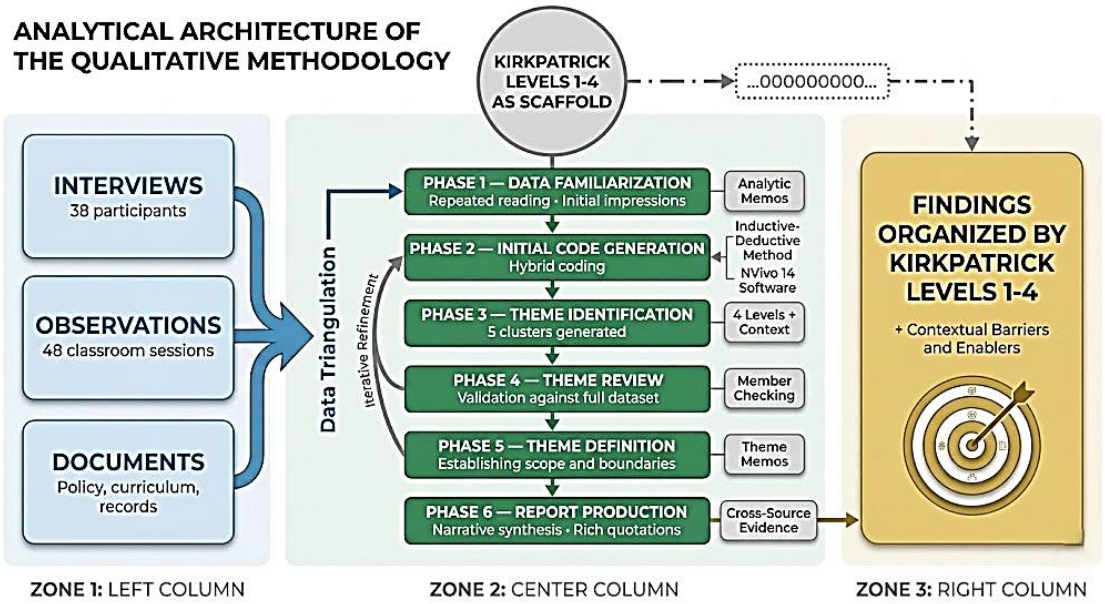

Figure 2. Data Analysis Workflow: Hybrid Thematic Processing, Mapping, Data Triangulation Through Multi-Cycle Coding Nodes to the Final Kirkpatrick Evaluative Matrix

Analysis proceeded through a rigorous six-phase iterative workflow. Initial open coding established semantic tags across all 38 transcripts and 48 observation records during the familiarization phase. Secondary axial coding clustered these tags into broader conceptual categories, focusing on patterns of structural decoupling between teacher knowledge and classroom behavior.

Candidate themes were systematically verified against the entire dataset to ensure empirical grounding and distinct boundaries, which was further strengthened through member-checking procedures. Finalized themes were defined through analytic memos, which synthesized narrative interview data, direct observation field logs, and documentary evidence into an integrated explanatory account organized around the adapted Kirkpatrick tiers.

2.5. Trustworthiness and Rigor

Methodological rigor was maintained through the operationalization of Lincoln and Guba's four validation criteria (Enworo, 2023). Credibility was secured through data source triangulation, cross-referencing interviews against the 48 observation logs and policy documents. Prolonged field engagement over the data collection cycle minimized researcher reactivity. Member checking was executed with a strategic subsample of eight participants (four teachers, two principals, two policymakers) who reviewed and validated the preliminary thematic structures. Peer debriefing with two external qualitative experts provided critical evaluations of the coding architecture.

Transferability is facilitated through the provision of thick description regarding the socioeconomic traits of East Lombok, participant demographics, and localized school operational conditions, allowing analytical generalization to analogous rural contexts in the Global South. Dependability was maintained through a detailed audit trail tracking all analytical adjustments, raw code definitions, and theme evolution logs in NVivo 14. Confirmability was established via systematic reflexivity practices, where a researcher journal was used to isolate personal assumptions and institutional biases from the analytical process.

3. RESULTS

The empirical findings generated from the triangulation of 38 semi-structured interviews, 48 systematic classroom observations, and archival document analysis are presented in this chapter. In strict accordance with the qualitative application of the adapted evaluation model, this section focuses exclusively on primary field data and localized document metrics, omitting external literature comparisons to maintain empirical transparency. The findings map the operational realities of teacher professional development

(TPD) within rural Indonesia, tracking the journey from initial program administration to field-level instructional execution.

3.1. Baseline Typology of TPD Modalities in Rural Indonesia

Archival data and interviews with administrative stakeholders indicate that professional capacity building in the rural periphery is structured around three distinct regulatory modalities. These modalities comprise formal postgraduate teacher certification (PPG), decentralized subject-matter peer networks (MGMP), and centralized short-term in-service workshops (Diklat). Each modality operates under a unique funding structure, services a different phase of the professional career cycle, and possesses distinct administrative expectations.

Systematic document analysis of district-level training logs spanning the 2021 to 2023 period reveals a profound geographic stratification regarding program access and workforce participation. At the aggregate regency level, official documentation indicates that 58 percent of the teacher workforce participated in at least one formal TPD activity within the three-year window. However, this baseline metric collapses when disaggregated by sub-district geography. Archival records confirm that teacher participation rates dropped below 40 percent within the three most geographically isolated sub-districts, exposing a severe structural disparity between urbanized administrative centers and peripheral rural schools.

This spatial inequity is heavily driven by the centralized, urban-centric layout of formal certification programs like PPG. As a high-stakes qualification mechanism tied to state salary supplements, PPG mandates extensive engagement with accredited higher education provider institutions. For educators stationed in under-resourced rural zones, this centralized design introduces significant logistical and financial friction points. A district education officer described these structural constraints with considerable candor:

"The nearest PPG provider is in Mataram. For teachers in remote sub-districts, that means leaving their family, renting a room, and spending money they don't have. Many simply choose not to go, or they go once and never return for the continuation modules." (P-PO2, District Education Officer)

In contrast to the residential demands of PPG, the MGMP modality functions as a localized alternative, convening subject-specific teachers for bi-weekly or monthly sessions hosted on a rotating basis across cluster schools. Document analysis of school-level logs confirms that MGMP achieves a higher frequency of individual teacher touchpoints in comparison to other modalities. Nevertheless, the archival data also show that the operational continuity of these peer networks is highly volatile, depending heavily on localized parent-teacher association subsidies and fluctuating district budget allocations. In remote sub-districts, these financial constraints frequently lead to prolonged periods of operational dormancy, compounding the isolation of rural educators.

The third modality, Diklat, is deployed top-down by provincial or district authorities as an episodic response to national curriculum overhauls, most recently exemplified by workshops dedicated to the *Merdeka* (Independent) Curriculum framework. Document analysis of Diklat rosters demonstrates that these workshops are disproportionately staged in centralized urban hotels, which inherently excludes teachers from infrastructure-poor areas due to transit limitations. A training facilitator characterized this severe drop in remote participation, where workforce engagement consistently plummets below the 40 percent threshold, as a structural failure that remains unaddressed by regional administrators:

"a quiet crisis that nobody in the provincial office wants to talk about." (P-TF1, Training Facilitator)

This baseline profile proves that the TPD infrastructure in rural Indonesia does not offer a uniform intervention. Instead, it functions as an ecologically unequal system where standardized institutional designs intersect with geographic barriers, restricting professional access long before any instructional transfer can occur in the classroom.

3.2. Levels 1 and 2: Affective Reactions and Cognitive Acquisition Divergences

Empirical data across the three teacher professional development (TPD) modalities reveal stark contrasts in teacher reactions, establishing a clear divergence between localized peer learning and centralized, top-down instruction. Teachers who participated in decentralized working groups (MGMP) expressed consistently positive reactions. They emphasized the collegial atmosphere and immediate contextual relevance of the sessions, which allowed for immediate collaborative problem solving. A senior secondary mathematics teacher contrasted this community-driven approach against urban-centered workshops:

"MGMP is the only training where I feel the facilitator understands what my classroom actually looks like. We discuss real problems, my students cannot do fractions, your students cannot understand word problems. We help each other find solutions. That is different from being sent to a hotel in town and told to sit and listen for three days." (P-T07, Senior Secondary Mathematics Teacher)

Conversely, teacher reactions to centralized workshops (Diklat) were highly ambivalent. While participants acknowledged the conceptual importance of the new *Merdeka* (Independent) Curriculum, they expressed widespread dissatisfaction with the passive transmission format. A junior secondary science teacher described the frustration of receiving standardized content detached from raw classroom limitations:

"Every Diklat is the same. They show us PowerPoint slides about the new curriculum. We copy notes. We take a photo of the materials. Then we go home and nothing changes because we don't know how to actually do what they showed us in our own classroom, with forty students and no laboratory equipment." (P-T14, Junior Secondary Science Teacher)

The formal certification program (PPG) generated the most polarized reactions among the cohort. While teachers recognized its immense career value, statutory credentials, and robust theoretical foundation, its demanding structural framework imposed severe personal, financial, and domestic burdens on rural educators. A primary school teacher recounted the compounding pressures that ultimately forced her withdrawal from the program:

"I believe in PPG. I know it will make me a better teacher. But I could not continue. My husband works in the rice fields, we have three children, and the provider institution is four hours away. Who will teach my class while I am gone for weeks? The school cannot always find a substitute." (P-T03, Primary School Teacher)

At the learning level, the data uncover a critical divergence between the acquisition of conceptual knowledge and actual procedural competence. Across all TPD modalities, teachers achieved significant cognitive gains regarding educational philosophies and curriculum nomenclature. However, this conceptual understanding rarely translated into procedural mastery, leaving a substantial operational void regarding classroom execution under resource-constrained conditions. A junior secondary social studies teacher captured this sudden shift from post-training enthusiasm to immediate operational paralysis:

"I learned a lot about the philosophy of the new curriculum, why we should give students more freedom to explore, why we should not just focus on examinations. Honestly, after that training I felt excited. I came back to school wanting to try new things... But when I came back and stood in front of my class, I realized I did not know the steps. I knew the concept. I knew it was good. But nobody showed me how to do it with forty-two students, no projector, and three different learning levels in the same room." (P-T19, Junior Secondary Social Studies Teacher)

Document analysis of official Diklat performance logs corroborated this procedural vacancy. While standardized post-test scores showed consistent, uniform increases in theoretical content recall, corresponding pedagogical design tasks showed negligible growth. This learning imbalance remains hidden due to a complete absence of institutional mechanisms to verify actual skill acquisition. A junior secondary school principal highlighted this total lack of accountability and monitoring:

"We send teachers to training and they come back with certificates and folders full of materials. But I never see any assessment of what they actually learned. There is no test, no demonstration, no follow-up visit from the trainer. So how does anyone know if the training worked?" (P-PR4, School Principal, Junior Secondary)

This empirical evidence shows that while the lower tiers of the evaluation model indicate successful execution on paper, they mask a profound pedagogical gap. Teachers exit training with high affective satisfaction or theoretical familiarity, yet they remain ill-equipped with the practical, concrete strategies required to navigate the structural realities of rural classrooms.

3.3. Level 3: The Behavioral Transfer Bottleneck in Actual Classroom Practices

The third level of the evaluation model examines the behavioral translation of professional learning into actual classroom practice, exposing the definitive core of the training-to-classroom gap. Systemic synthesis of the 48 classroom observation sessions reveals a persistent and troubling decoupling between teachers' stated pedagogical intentions and their actual instructional execution. Despite 100 percent of the observed teacher cohort having attended at least one recent TPD program explicitly championing student-centered, active learning paradigms, teacher-centered transmission methods remained overwhelmingly dominant across 41 out of the 48 observed sessions. Instructional blocks were characterized by extended teacher monologues, rigid rows of desks facing the blackboard, minimal student questioning, and almost no differentiation for varied student learning paces.

Follow-up interviews and member-checking reflections reveal that this behavioral stagnation is heavily driven by a structural conflict between reformist pedagogy and localized accountability mandates. Teachers face acute pressure from high-stakes assessment metrics, forcing a strategic retreat to traditional rote-learning modalities that prioritize exam preparation over conceptual exploration. A senior secondary language teacher detailed this systemic contradiction:

"In the training, they told us: ask questions, let students discuss, don't just lecture. I agreed completely. But then I go back and think: if I stop lecturing, will they pass the national exam? The curriculum says one thing, the exam says another. I go with the exam because that is what parents and the principal measure me on." (P-T22, Senior Secondary Language Teacher)

The observation logs demonstrate a stark contrast in behavioral transfer based on TPD modality design, revealing a complete failure of transfer among exclusive Diklat graduates compared to a targeted breakthrough among active MGMP participants. In classrooms helmed by teachers whose TPD was confined to centralized Diklat workshops, no observable traces of training content survived past the six-

month mark. Conversely, sustained behavioral transformation was explicitly verified in rural schools where sub-district MGMP clusters had institutionalized localized peer observation and lesson study protocols. This ongoing collaborative reflection provided the psychological safety and practical modeling necessary to sustain instructional change. A junior secondary mathematics teacher contrasted the powerful impact of this peer learning mechanism against conventional workshops:

"In our MGMP, we started observing each other's classes and discussing what we saw not to judge, but to learn. That changed how I teach. I saw my colleague use a group problem-solving approach that I had never tried. I copied it, adapted it for my class, and my students became more active. That observation changed me more than any Diklat I have attended." (P-T11, Junior Secondary Mathematics Teacher)

Beyond structural accountability and modality design, the behavioral transfer bottleneck is exacerbated by severe material friction. Centralized training architectures routinely operate under the flawed assumption of resource abundance, demonstrating advanced tools and pedagogical manipulatives that are completely unavailable in peripheral rural environments. When faced with this infrastructure deficit, teachers abandon the innovative strategies learned due to an inability to bridge the gap between training ideals and material scarcity. A primary school teacher described this immediate operational barrier:

"They showed us how to use manipulatives for mathematics. Beautiful lesson. But when I go to my school, we have no manipulatives. I cannot create them from nothing." (P-T06, Primary School Teacher)

3.4. Level 4: Ecological Variation in Instructional Quality and Student Engagement

The fourth level of the evaluation framework assesses how teacher behavioral changes alter systemic instructional quality and student engagement. The field data reveal that macro-level results are not uniform across the regency, displaying instead profound ecological variations dictated by school-level micro-climates. The 48 classroom observation sessions document two highly contrasting operational archetypes, demonstrating that identical professional development inputs produce vastly divergent classroom realities depending on institutional leadership and peer structures.

The first archetype, observed in schools characterized by collaborative micro-cultures and active principal support, exemplifies an optimized instructional environment. In these settings, observers documented a high integration of pedagogical content knowledge and active student engagement. For example, during a junior secondary science block, a teacher who regularly participated in decentralized peer networks led students in constructing simple circuit diagrams using locally salvaged materials. The instructor systematically connected empirical observations to abstract physics principles through targeted, open-ended questioning. Student participation remained continuous and collaborative throughout the 40-minute session, with zero recorded instances of passive or off-task behavior. Interviews confirmed that this instructional risk-taking is directly enabled by supportive, non-punitive supervision from school management. A senior secondary school principal explained this proactive instructional leadership mechanism:

"I make it my responsibility to visit classrooms every week. Not to inspect, but to learn together. When teachers know I will come and we will discuss what I saw, they prepare differently. They try new things because they know there is a conversation waiting, not a punishment." (P-PR2, School Principal, Senior Secondary)

The second archetype presents a stark contrast, occurring predominantly in geographically isolated, under-resourced schools managed by passive administrative leadership. In these environments, classroom observations revealed a complete stagnation of teaching quality, where lessons routinely reverted to rigid, textbook-bound monologues. During a parallel science session at a comparably resourced but isolated school, an instructor who had attended identical national curriculum workshops delivered 35 minutes of uninterrupted rote exposition straight from a manual. Student interaction was limited to three closed-ended questions at the end of the period, while the majority of the class exhibited disengaged, passive behavior. Field tracking confirms that without localized institutional scaffolding, external training inputs fail to penetrate traditional teaching routines. A training facilitator summarized this phenomenon of isolated training delivery:

"I go back to schools six months after training and I observe. Sometimes I cannot find a single trace of what we taught. The teachers remember the training, they remember it was good, but the classroom looks exactly the same as before. The training happened in a bubble. It did not touch the school." (P-TF3, Training Facilitator)

These contrasting observation vignettes prove that Level 4 instructional outcomes are ecologically contingent. Standardized training content delivered uniformly across a region fails to yield uniform improvements in classroom quality. Instead, the conversion of teacher learning into measurable student engagement is fundamentally moderated by the instructional leadership of the principal and the presence of active peer learning networks within the school infrastructure.

3.5. Operationalizing the Matrix of Structural Barriers and Institutional Enablers

To systematically synthesize the macro-level constraints and micro-level support systems operating across rural schools, the cross-cutting thematic analysis yields an integrated matrix of structural barriers

and institutional enablers. These factors represent the multidimensional ecosystem that mediates the trajectory from teacher training to real classroom application. Table 3 details these parameters across geographic, institutional, material, personal, and systemic dimensions.

Table 3. Structural Barriers and Institutional Enablers of TPD Effectiveness

Dimension	Barriers	Enablers
Geographic & Infrastructural	Physical distance from training venues, limited transportation access, and poor internet connectivity in remote sub-districts.	Decentralized MGMP meetings rotating across cluster schools and mobile training delivery models.
Institutional & Administrative	Heavy non-teaching workloads, complete absence of post-training follow-up systems, and weak accountability mechanisms for skill application.	Active principal instructional leadership, structured peer observation cycles, lesson study protocols, and school-level planning.
Resource & Material	Absence of teaching materials matching training content, lack of digital infrastructure, and inadequate classroom equipment.	Creative resource adaptation by motivated teachers, community resource mobilization, and NGO-supported material provision.
Motivational & Personal	Perceived disconnect between training content and classroom reality, certification-oriented rather than learning-oriented participation, and fatigue from multiple role demands.	Intrinsic professional motivation, a collaborative peer learning culture within the MGMP network, and tangible recognition of professional growth.
Policy & Systemic	Generic, non-contextualized program design, insufficient training budgets for remote areas, and lack of coordination between PPG, MGMP, and Diklat systems.	District-level policy commitment, alignment with the *Merdeka* (Independent) Curriculum reform priorities, and evidence-based program monitoring.

The empirical data indicate that among these multidimensional factors, institutional leadership operates as the primary mediating variable that determines the success or failure of behavioral transfer. While material and geographic barriers heavily restrict initial access, the internal administrative climate of the school dictates whether newly acquired pedagogical skills are enacted or abandoned. In schools where resource constraints are severe, active instructional leadership routinely mitigates these deficits by fostering localized collaborative structures, authorizing creative curriculum adaptations, and shielding teachers from administrative overburdens. Conversely, in schools where principals remain detached from instructional supervision, even highly resourced teachers fail to implement innovative practices due to a lack of accountability and psychological safety. A senior district education officer with extensive longitudinal oversight of regional teacher development programs confirmed this decisive institutional dynamic:

"We have been sending teachers to training for twenty years. The schools that improve are not the ones that sent the most teachers. They are the ones with principals who actually lead learning. The principal is the key. Without that, training money flows into a drain." (P-PO1, Senior District Education Officer)

This empirical synthesis proves that the training-to-classroom gap is not an isolated instructional failure, but rather a structurally mediated phenomenon. Standardized training inputs are constantly filtered through the institutional constraints summarized in Table 3. Without targeted policy interventions that reinforce principal instructional leadership and institutionalize post-training follow-up ecosystems at the school level, centralized TPD expenditures remain decoupled from the actual quality of classroom instruction.

4. DISCUSSION

4.1. Unpacking the Non-Linear Decoupling in Kirkpatrick's Evaluation Framework

The classical application of Kirkpatrick's Four-Level Evaluation Model operates under a foundational assumption of hierarchical linearity, presupposing that positive affective reactions (Level 1) function as a causal gateway to cognitive learning (Level 2), which sequentially drives behavioral transfer (Level 3) and systemic results (Level 4). However, critical scholarship increasingly questions this deterministic path, arguing that satisfaction and knowledge acquisition do not automatically guarantee the structural

modification of workplace practices (Alsalamah & Callinan, 2021). Evidence from this study challenges the assumption of a linear progression across Kirkpatrick's evaluation levels, instead revealing a profound non-linear decoupling within the educational ecosystem of rural Indonesia. While the studied teacher professional development (TPD) initiatives successfully generated positive participant reactions and strengthened teachers' conceptual understanding, the progression toward classroom implementation was significantly constrained. This disconnect reinforces broader international evidence showing that substantial investments in teacher professional development often fail to produce corresponding improvements in classroom practice when implementation conditions and organizational support are insufficient (da Silva et al., 2024; Popova et al., 2022).

This non-linear decoupling reveals that individual teacher cognition is insufficient to bridge the training-to-classroom gap when confronted with hostile institutional environments. As documented in the findings, rural teachers displayed intense conceptual enthusiasm for student-centered instructional models during localized workshops, yet returned to rigid, teacher-centered transmission methods under actual field observations. This severe breakdown between Level 2 learning and Level 3 behavior indicates that professional growth is not a simple internal psychological mechanic but an ecologically mediated process (Daly et al., 2020; Liu et al., 2024). Even when teachers possess strong general pedagogical knowledge, their capacity to alter instruction is actively suppressed by external variables, a reality that renders standard linear evaluation frameworks analytically blind to the operational realities of under-resourced schools (Leijen et al., 2024). Consequently, evaluating TPD in low-resource settings requires an expansion of evaluation theory, transitioning from evaluating individual teacher compliance to examining the structural barriers that govern the workspace.

The vulnerability of training outcomes in peripheral contexts highlights the necessity of embedding professional development within continuous, collaborative networks rather than relying on episodic interventions (Yoon & Goddard, 2025). In rural territories, the immediate cognitive gains captured at Level 2 function in an institutional vacuum, meaning that without localized scaffolding, any positive behavioral shift degrades rapidly (Brunetti et al., 2024; Chen & Cao, 2022; Jian Li et al., 2020). Evidence from Indonesia suggests that centralized teacher professional development remains ineffective when it assumes schools are passive recipients of reform, whereas implementation is fundamentally shaped by school-level organizational conditions, leadership, and contextual constraints (Kong & Wang, 2024; Revina et al., 2023). Rural instructional practices are deeply fossilized within institutional cultures that actively resist alternative approaches, especially when state-mandated curriculum reforms conflict with rigid, examination-oriented local accountability metrics (Burhan et al., 2024; Fadilah et al., 2023; Wahidmurni et al., 2024). Therefore, as confirmed by the behavioral variations between exclusive workshop graduates and active peer-network participants in this study, the presence of localized school-level collaborative reflection and supportive micro-cultures represents the primary determinant of whether theoretical learning transforms into durable classroom practice (McChesney & Cross, 2023).

4.2. The Paradox of Performativity: Pedagogical Reform vs Accountability Systems

The acute instructional dilemma articulated by teacher P-T22 exposes a severe structural friction known in global education policy as the paradox of performativity. This institutional phenomenon occurs when state-mandated pedagogical ideals directly collide with localized, high-stakes accountability mechanisms. On a macro-policy level, the Indonesian state champions the *Merdeka* (Independent) Curriculum, which is an ambitious initiative designed to foster progressive, student-centered, 21st-century learning environments. However, the micro-political climate at the sub-district and school levels remains aggressively anchored to traditional, metrics-driven performance indicators. This misalignment forces rural educators into an untenable position where they must navigate two conflicting definitions of educational quality. While TPD modules evaluate them on their capacity to facilitate student dialogue and critical thinking, their immediate community, school principals, and regional offices evaluate them almost exclusively on their students' examination performance.

Faced with this systemic contradiction, the retreat of rural teachers back to traditional, teacher-centered lecture methods represents a logical defense mechanism rather than a failure of individual motivation. This behavioral pattern constitutes strategic compliance, which is a pragmatic survival strategy where educators nominally accept reform rhetoric during training workshops but prioritize conventional, exam-oriented routines within their actual classrooms. This institutional decoupling is deeply embedded within the historical architecture of the country's educational governance, where four decades of TPD reforms have consistently incorporated progressive programmatic features on paper while falling short in practice (Rasmitadila et al., 2025; Revina et al., 2023). When education reforms are introduced without corresponding changes in quality assurance, assessment, and implementation systems, teacher professional development is unlikely to produce sustained improvements in instructional practice because teachers remain constrained by incoherent institutional conditions (Richter & Richter, 2024; Tatik et al., 2024).

This friction is heavily intensified in peripheral, under-resourced schools where institutional vulnerability magnifies the pressure to perform on traditional metrics. Overly teacher-centered pedagogy is reinforced by examination-oriented accountability systems that encourage teachers to prioritize

examination performance over interactive classroom practices (Wang & McLaughlin, 2025), while evidence also indicates that student-centered pedagogy provides greater opportunities for pedagogical growth and active learning than teacher-centered instruction (Kong & Wang, 2024; Woods & Copur-Gencturk, 2024). Global educational scholarship confirms that when high-stakes performative pressures are intense, teachers routinely compromise their progressive pedagogical beliefs to avoid institutional sanctions or community disapproval (OECD, 2024). Disparities in implementation quality between urban and rural settings are driven by these structural inequities in local support and conflicting administrative expectations rather than variations in teacher motivation (Baharuddin & Burhan, 2025). Therefore, the training-to-classroom gap cannot be resolved by upgrading teacher content knowledge alone (Fukaya et al., 2024; Kong & Wang, 2024; Jingxian Li & Copur-Gencturk, 2024). Unless systemic accountability structures are synchronized with progressive pedagogical mandates, standardized TPD investments will continue to produce strategic compliance instead of authentic instructional reform (Popova et al., 2022).

4.3. Ecological Contingency and the Micro-Politics of Principal Leadership

The profound divergence in classroom outcomes documented between supported and unsupported schools underscores that Level 4 results are fundamentally governed by ecological contingency. Teacher knowledge is never transferred in an institutional vacuum, but is instead heavily filtered through the micro-politics and structural conditions of the school environment. Within this micro-ecosystem, the school principal operates as the absolute determinant of whether professional development inputs yield genuine improvements in instructional quality and student engagement. The proactive stance of principal P-PR2 demonstrates that when school leaders practice active instructional leadership, they significantly reduce the systemic risk of training failure. By shifting the administrative paradigm from punitive inspection to shared professional inquiry, such leaders cultivate an institutional micro-climate characterized by psychological safety. This supportive oversight creates a crucial space where teachers feel authorized to experiment with innovative methods and navigate initial classroom failures without fearing professional retaliation or administrative sanctions.

This critical role of organizational dynamics aligns directly with the multi-dimensional architecture of Guskey's professional development evaluation framework (Garone et al., 2022; Govender & Juggernath, 2025), which positions organizational support as a non-negotiable intermediate bridge between teacher learning and systemic change. When a school lacks active administrative scaffolding, individual cognitive gains achieved during external workshops remain trapped in an institutional bottleneck, failing to disrupt entrenched pedagogical routines. Conversely, a supportive school culture actively facilitates behavioral transfer by aligning internal incentives with reform goals (McChesney & Cross, 2023). School-level collaborative structures, including peer observations and lesson study cycles, provide the continuous feedback and reinforcement necessary to help teachers overcome the practical friction of implementing new strategies.

Furthermore, sustained teacher collaboration contributes to positive professional outcomes by strengthening collaborative school cultures that enhance teachers' job satisfaction and occupational well-being (Kutyłowska et al., 2023), while international evidence also indicates that supportive professional environments are associated with teachers' successful fulfilment of lesson aims and overall professional well-being (Nwoko et al., 2023). In rural and under-resourced territories, where external infrastructure and material resources are scarce, this internal social capital becomes the primary engine for instructional resilience. Principals who actively participate in the learning community transform the school from a collection of isolated classrooms into an integrated professional learning ecosystem. Ultimately, these findings indicate that investing in teacher training without a simultaneous, co-equal commitment to developing principal instructional leadership is an inefficient policy strategy that systematically diminishes the return on professional development investments.

4.4. Structural Inequities and the Failure of Standardized TPD Modalities in the Global South

The systematic failure of centralized, one-size-fits-all training architectures highlights a profound macroeconomic inefficiency within the educational governance of the Global South. National standardization frameworks routinely ignore the severe infrastructural and geographic realities of peripheral schools, operating under a flawed assumption of resource abundance that does not exist in rural communities. As vividly illustrated by teacher P-T14, who recounted the complete futility of copying digital PowerPoint slides when faced with an absolute lack of laboratory equipment and high student density, top-down models offer abstract pedagogical theories completely detached from local classroom constraints. This severe decoupling represents one of the largest misallocations of public education budgets across developing countries, as the temporary cognitive gains captured during luxury off-site hotel workshops rapidly decay once teachers return to their unsupportive local environments (Brunetti et al., 2024). This rapid obsolescence proves that standardized, episodic interventions fail to establish sustainable roots when embedded in low-resource environments where post-training reinforcement is non-existent.

Furthermore, these implementation deficits are driven by deeply entrenched urban-rural geographic disparities rather than deficits in individual teacher motivation or professional intent (Baharuddin & Burhan, 2025). Centralized TPD architectures systematically privilege urban administrative centers while placing severe personal, emotional, and financial burdens on rural educators who must navigate extreme physical distance and broken transportation infrastructure to access state-mandated training. To reduce systemic inequalities in education, regional policies should prioritise contextualised, school-embedded professional learning supported by local governance and school-based implementation rather than relying on standardised delivery models (Greany et al., 2025; Kong & Wang, 2024).

Consequently, the grassroots alternative embodied by decentralized teacher working groups (MGMP) demonstrates a far higher level of ecological resilience within rural territories. By anchoring the professional development cycle directly inside the actual socio-material constraints of the rural district, the peer reflection and lesson study components of MGMP bypass the limitations of centralized bureaucracy. Rural teachers working within these localized networks collaborate to co-create creative pedagogical adaptations, utilizing salvaged local materials to bridge the infrastructure gap. Transitioning funding away from standardized in-service workshops toward these sustainable, community-driven professional development ecosystems is therefore an urgent policy requirement to effectively close the training-to-classroom gap in the Global South.

5. CONCLUSION

This study evaluated the systemic impacts of three dominant teacher professional development (TPD) modalities in rural Indonesia by utilizing Kirkpatrick's Four-Level Evaluation Model. The empirical evidence generated from 38 stakeholders and 48 classroom observations reveals a profound non-linear decoupling across the evaluative spectrum. While TPD initiatives successfully achieve positive reactions and conceptual learning gains at the lower levels, behavioral transfer into actual instructional practice remains heavily bottlenecked by rigid administrative accountability, resource scarcity, and generic program designs. Meaningful improvements in classroom quality are ecologically contingent upon active principal instructional leadership and localized peer-learning networks rather than standardized training inputs. This research directly challenges the linear efficacy assumptions of classical evaluation theory in peripheral settings, demonstrating that the training-to-classroom gap in the Global South is a structurally mediated phenomenon rather than a consequence of individual teacher deficits.

By mapping this multidimensional matrix of constraints, this study provides an original socio-contextual framework illustrating that grassroots, decentralized peer networks (MGMP) possess significantly higher ecological resilience in resource-constrained environments than centralized workshops (Diklat). These insights yield immediate, actionable implications for educational authorities seeking a higher return on public TPD investments. Resolving the training-to-classroom gap requires a systemic shift toward school-embedded, contextually responsive capacity building, which can be achieved by structurally reinforcing decentralized peer networks with continuous lesson study protocols. Most crucially, state investments in principal instructional leadership must be treated as a co-equal priority alongside teacher training, as the administrative climate of the school ultimately dictates whether newly acquired pedagogical skills are enacted or abandoned.

While this qualitative case study offers rich analytical transferability to analogous under-resourced contexts in developing nations, specific limitations remain. Isolating the long-term impacts of training on student learning outcomes at Level 4 remains a complex methodological challenge due to confounding socioeconomic variables, and the findings lack macro-level statistical generalizability. Future inquiry should deploy longitudinal mixed-methods designs to track the behavioral durability of teacher cohorts over multiple years across varying regional administrative structures. Additionally, future investigations must explore the efficacy of digital and hybrid delivery models in overcoming the geographic isolation that restricts equitable professional development access across rural communities.

ACKNOWLEDGMENTS

The authors express sincere gratitude to all participants who contributed their time, insights, and candid reflections to this study, including the classroom teachers, school principals, district education officers, and training facilitators across East Lombok Regency whose experiences and perspectives constitute the empirical foundation of this research. The authors also acknowledge the cooperation of the East Lombok Regency Department of Education and Culture *(Dinas Pendidikan dan Kebudayaan Kabupaten Lombok Timur)* for facilitating access to research sites and documentary records. Special appreciation is extended to the MGMP coordinators and school principals who generously accommodated observation sessions within their schools and professional development activities. The authors further thank the two external qualitative researchers who provided peer debriefing support during the analytical phase of the study, and the eight participants who contributed to member-checking sessions whose critical feedback strengthened the accuracy and credibility of the final findings. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

REFERENCES

Alsalamah, A., & Callinan, C. (2021). Adaptation of Kirkpatrick's Four-Level Model of Training Criteria to Evaluate Training Programmes for Head Teachers. *Education Sciences*, *11*(3), 116. https://doi.org/10.3390/educsci11030116

Alsalamah, A., & Callinan, C. (2022). The Kirkpatrick model for training evaluation: bibliometric analysis after 60 years (1959–2020). *Industrial and Commercial Training*, *54*(1), 36–63. https://doi.org/10.1108/ICT-12-2020-0115

Ambon, J., Alias, B. S., Komariah, A., & Mansor, A. N. (2024). The impact of continuous professional development on teaching quality: a systematic review. *International Journal of Evaluation and Research in Education (IJERE)*, *13*(6), 3838. https://doi.org/10.11591/ijere.v13i6.30427

Attabik, M. H., & Zainiza, M. (2025). Pre-Service English Teachers' Goals and Beliefs in the Indonesian PPG Program. *Language Circle: Journal of Language and Literature*, *20*(1), 121–137. https://doi.org/10.15294/lc.v20i1.34552

Baharuddin, & Burhan. (2025). Urban and rural teacher perspectives on Indonesian educational reform: challenges and policy implications. *Cogent Education*, *12*(1). https://doi.org/10.1080/2331186X.2025.2497142

BPS NTB. (2023). *Indeks Pembangunan Manusia (IPM) Provinsi NTB Menurut Kabupaten/Kota*. Central Statistics Agency of West Nusa Tenggara Province. https://ntb.bps.go.id/en/pressrelease/2023/12/01/946/indeks-pembangunan-manusia--ipm--provinsi-ntb-menurut-kabupaten-kota.html

Brunetti, A., Büchel, K., Jakob, M., Jann, B., & Steffen, D. (2024). Inadequate teacher content knowledge and what could be done about it: evidence from El Salvador. *Journal of Development Effectiveness*, *16*(2), 206–229. https://doi.org/10.1080/19439342.2023.2213681

Burhan, Mirayanti, Marzuki, Saugadi, & Rusdin, D. (2024). Redefining rural education: exploring EFL teachers' perceptions and recommendations for enhanced human resource management practices. *Cogent Education*, *11*(1). https://doi.org/10.1080/2331186X.2024.2317503

Chang, C.-F., Annisa, N., & Chen, K.-Z. (2025). Pre-service teacher professional education program (PPG) and Indonesian science teachers' TPACK development: A career-path comparative study. *Education and Information Technologies*, *30*(7), 8689–8711. https://doi.org/10.1007/s10639-024-13160-6

Chen, Y., & Cao, L. (2022). Promoting maker-centred instruction through virtual professional development activities for K-12 teachers in low-income rural areas. *British Journal of Educational Technology*, *53*(4), 1025–1048. https://doi.org/10.1111/bjet.13183

da Silva, C., Pereira, F., & Amorim, J. P. (2024). The integration of indigenous knowledge in school: a systematic review. *Compare: A Journal of Comparative and International Education*, *54*(7), 1210–1228. https://doi.org/10.1080/03057925.2023.2184200

Daly, C., Milton, E., & Langdon, F. (2020). How do ecological perspectives help understand schools as sites for teacher learning? *Professional Development in Education*, *46*(4), 652–663. https://doi.org/10.1080/19415257.2020.1787208

Enworo, O. C. (2023). Application of Guba and Lincoln's parallel criteria to assess trustworthiness of qualitative research on indigenous social protection systems. *Qualitative Research Journal*, *23*(4), 372–384. https://doi.org/10.1108/QRJ-08-2022-0116

Fadilah, R., Ayudhia, H. Y., Pratama, V. M., & Harmawan, V. (2023). Teachers' Teaching Reflection: English Teachers' Challenges in Rural Areas of Indonesia. *LET: Linguistics, Literature and English Teaching Journal*, *13*(1), 85. https://doi.org/10.18592/let.v13i1.8923

Fukaya, T., Nakamura, D., Kitayama, Y., & Nakagoshi, T. (2024). A systematic review and meta-analysis of intervention studies on mathematics and science pedagogical content knowledge. *Frontiers in Education*, *9*, 1435758. https://doi.org/10.3389/feduc.2024.1435758

Garone, A., Bruggeman, B., Philipsen, B., Pynoo, B., Tondeur, J., & Struyven, K. (2022). Evaluating professional development for blended learning in higher education: a synthesis of qualitative evidence. *Education and Information Technologies*, *27*(6), 7599–7628. https://doi.org/10.1007/s10639-022-10928-6

Govender, N., & Juggernath, A. (2025). An analysis of Guskey's framework of teachers' perceptions and experiences in integrating ICTs for lesson study. *International Journal for Lesson & Learning Studies*, *14*(1), 55–67. https://doi.org/10.1108/IJLLS-05-2024-0092

Greany, T., Cowhitt, T., Noyes, A., Gripton, C., & Hudson, G. (2025). Local learning landscapes: conceptualising place-based professional learning by teachers and schools in decentralised education systems. *Journal of Educational Change*, *26*(1), 1–28. https://doi.org/10.1007/s10833-024-09508-x

Kart, M. (2026). Implementing Inclusive Education in the Classroom: Teacher Competencies, Beliefs, and Continuous Professional Development. In P. Jandrić (Ed.), *Chapters in Education: Volume 1* (pp. 239–266). Springer Nature Switzerland. https://doi.org/10.1007/60490_2026_9

Kong, S.-C., & Wang, Y.-Q. (2024). The impact of school support for professional development on teachers' adoption of student-centered pedagogy, students' cognitive learning and abilities: A three-level analysis. *Computers & Education*, *215*, 105016. https://doi.org/10.1016/j.compedu.2024.105016

Kutyłowska, K., Sitek, M., Wągrowska, U., & Paczuska, K. (2023). School climate as a predictor of teacher job satisfaction and occupational well-being: TALIS 2018 evidence from Central and Eastern Europe. *EDUKACJA*, *2023*(4(167)), 94–115. https://doi.org/10.24131/3724.230407

Leijen, Ä., Pedaste, M., Baucal, A., Poom-Valickis, K., & Lepp, L. (2024). What predicts instructional quality and commitments to teaching: self-efficacy, pedagogical knowledge or integration of the two? *Frontiers in Psychology*, *15*, 1287313. https://doi.org/10.3389/fpsyg.2024.1287313

Li, Jian, Shi, Z., & Xue, E. (2020). The problems, needs and strategies of rural teacher development at deep poverty areas in China: Rural schooling stakeholder perspectives. *International Journal of Educational Research*, *99*, 101496. https://doi.org/10.1016/j.ijer.2019.101496

Li, Jingxian, & Copur-Gencturk, Y. (2024). Learning through teaching: the development of pedagogical content knowledge among novice mathematics teachers. *Journal of Education for Teaching*, *50*(4), 582–597. https://doi.org/10.1080/02607476.2024.2358041

Li, Z., Hassan, N. C., & Jalil, H. A. (2023). The Effectiveness of Face-to-Face versus Online Delivery of Continuing Professional Development for Science Teachers: A Systematic Review. *Education Sciences*, *13*(12), 1251. https://doi.org/10.3390/educsci13121251

Liao, W., Li, X., Dong, Q., & Wang, Z. (2023). Non-university-based teacher educators' professional learning: A systematic review. *Teaching and Teacher Education*, *136*, 104374. https://doi.org/10.1016/j.tate.2023.104374

Liu, M., Wang, H., Shi, H., Zhang, B., Sun, X., & Wang, J. (2024). Examining the personal growth of college teacher educators through the lens of human development ecology: An approach utilizing artificial neural networks (ANNs) modeling. *Heliyon*, *10*(21), e39372. https://doi.org/10.1016/j.heliyon.2024.e39372

McChesney, K., & Cross, J. (2023). How school culture affects teachers' classroom implementation of learning from professional development. *Learning Environments Research*, *26*(3), 785–801. https://doi.org/10.1007/s10984-023-09454-0

Nwoko, J. C., Emeto, T. I., Malau-Aduli, A. E. O., & Malau-Aduli, B. S. (2023). A Systematic Review of the Factors That Influence Teachers' Occupational Wellbeing. *International Journal of Environmental Research and Public Health*, *20*(12), 6070. https://doi.org/10.3390/ijerph20126070

OECD. (2024). *Education Policy Outlook 2024*. OECD Publishing. https://doi.org/10.1787/dd5140e4-en

Paul, S., Burman, R. R., & Singh, R. (2024). Training effectiveness evaluation: Advancing a Kirkpatrick model based composite framework. *Evaluation and Program Planning*, *107*, 102494. https://doi.org/10.1016/j.evalprogplan.2024.102494

Pholphirul, P., Rukumnuaykit, P., & Teimtad, S. (2023). Teacher shortages and educational outcomes in developing countries: Empirical evidence from PISA-Thailand. *Cogent Education*, *10*(2). https://doi.org/10.1080/2331186X.2023.2243126

Popova, A., Evans, D. K., Breeding, M. E., & Arancibia, V. (2022). Teacher Professional Development around the World: The Gap between Evidence and Practice. *The World Bank Research Observer*, *37*(1), 107–136. https://doi.org/10.1093/wbro/lkab006

Rasmitadila, R., Rachmadtullah, R., Prasetyo, T., Humaira, M. A., Sari, D. A., Samsudin, A., Nurtanto, M., . F., & ZamZam, R. (2025). Professional development for Indonesian elementary school teachers: Increased competency and sustainable teacher development programs. *F1000Research*, *13*, 1375. https://doi.org/10.12688/f1000research.156946.3

Revina, S., Pramana, R. P., Bjork, C., & Suryadarma, D. (2023). Replacing the old with the new: long-term issues of teacher professional development reforms in Indonesia. *Asian Education and Development Studies*, *12*(4/5), 262–274. https://doi.org/10.1108/AEDS-12-2022-0148

Richter, E., & Richter, D. (2024). Measuring the quality of teacher professional development: A large-scale validation study of an 18-item instrument for daily use. *Studies in Educational Evaluation*, *81*, 101357. https://doi.org/10.1016/j.stueduc.2024.101357

Surahman, E., & Wang, T.-H. (2023). In-service STEM teachers professional development programmes: A systematic literature review 2018–2022. *Teaching and Teacher Education*, *135*, 104326. https://doi.org/10.1016/j.tate.2023.104326

Tatik, T., Nguyen, H., & Loughland, T. (2024). The impact of standards on novice teachers during mandated teaching induction: lessons from the Indonesian context. *Asia-Pacific Journal of Teacher Education*, *52*(1), 28–46. https://doi.org/10.1080/1359866X.2023.2298303

Ventista, O. M., & Brown, C. (2023). Teachers' professional learning and its impact on students' learning outcomes: Findings from a systematic review. *Social Sciences & Humanities Open*, *8*(1), 100565. https://doi.org/10.1016/j.ssaho.2023.100565

Visscher, A. J., Dmoshinskaia, N., Pellegrini, M., & Rey-Naizaque, A. (2025). (When) do teacher professional development interventions improve student Achievement? A meta-analysis of 128 high-

quality studies. *Educational Research Review*, *49*, 100742. https://doi.org/10.1016/j.edurev.2025.100742

Wahidmurni, W., Susilawati, S., & Abidin, A. A. (2024). Resistance to Curriculum Changes among Teachers of Madrasah Ibtidaiyah in Indonesia. *Nidhomul Haq : Jurnal Manajemen Pendidikan Islam*, *9*(1), 16–27. https://doi.org/10.31538/ndh.v9i1.4300

Wang, T., & McLaughlin, C. (2025). Promoting learner-centred education amid the culture of test-based accountability: insights from a cross-cultural teacher education programme. *Comparative Education*, *61*(4), 586–605. https://doi.org/10.1080/03050068.2024.2423445

Woods, P. J., & Copur-Gencturk, Y. (2024). Examining the role of student-centered versus teacher-centered pedagogical approaches to self-directed learning through teaching. *Teaching and Teacher Education*, *138*, 104415. https://doi.org/10.1016/j.tate.2023.104415

Yoon, I., & Goddard, R. D. (2025). Professional development quality and instructional effectiveness: Testing the mediating role of teacher self-efficacy beliefs. *Professional Development in Education*, *51*(7), 1471–1485. https://doi.org/10.1080/19415257.2023.2264309

Zickafoose, A., Ilesanmi, O., Diaz-Manrique, M., Adeyemi, A. E., Walumbe, B., Strong, R., Wingenbach, G., Rodriguez, M. T., & Dooley, K. (2024). Barriers and Challenges Affecting Quality Education (Sustainable Development Goal #4) in Sub-Saharan Africa by 2030. *Sustainability*, *16*(7), 2657. https://doi.org/10.3390/su16072657