

KEMI Journal of Educational Leadership and Management

KJELM

[ISSN 3079-4048]

Volume: 2 Issue: 3 | Nov-2026

From induction to impact: The theory-anchored CIPP-Plus (TA-CIPP+) model for evaluating school leadership training programmes

Raphael Ondeko Nyonje[1]

Open University of Kenya[1]

Corresponding Author Email: rnyonje@ouk.ac.ke

Cite this article in APA

https://doi.org/10.68147/8f632c5d

Copyright: ©2026 by the author(s). Open access under CC BY.

Abstract

School leadership training has become a policy priority across Africa, yet the programmes that prepare principals are seldom evaluated with frameworks that are theory-driven, context-sensitive and oriented towards improvement. This paper documents, theorises and refines an evaluation model first developed and tested by Nyonje (2008) in the evaluation of the Kenya Education Staff Institute (KESI) Induction Course in Educational Management for Secondary School Principals (ICEMSSP). The original model adapted Stufflebeam's Context–Input–Process–Product (CIPP) framework by anchoring its four stages in human capital theory and in systematic and blended training theory. Using a qualitative conceptual design, the paper combines documentary analysis of the original thesis with an integrative review of contemporary literature on programme evaluation and school leadership development. Reflexive thematic analysis was used to map the thesis findings onto the CIPP stages and to derive principles for refinement. The analysis shows that the theory-anchored model exposed the programme's strengths, including improved principal performance across six management domains, as well as its weaknesses in resource-person selection, venue quality, training duration, content standardization, delivery methods and equitable access. Building on these lessons, the paper proposes the Theory-Anchored CIPP-Plus (TA-CIPP+) Model. The model retains a theory core, adds a fifth Impact and Sustainability stage and introduces two cross-cutting lenses, policy–governance and equity–inclusion, bound together by a continuous monitoring, evaluation and learning loop. It offers ministries, leadership institutes such as the Kenya Education Management Institute, universities and development partners a

coherent tool for designing, monitoring and judging leadership programmes. Recommendations for adoption, capacity building and validation are provided.

Key terms: Theory-Anchored CIPP-Plus (TA-CIPP+) Model, school leadership training, programme evaluation, educational leadership, human capital theory, blended training, educational management, Kenya

1. Introduction and Background

The quality of school leadership is now widely recognized as second only to classroom teaching among the school-related factors that influence what students learn (Leithwood et al., 2020). A systematic synthesis of two decades of research concluded that an effective principal has an effect on student achievement comparable in size to that of an effective teacher, but spread across an entire school (Grissom et al., 2021). International bodies have translated this evidence into policy. The OECD's Improving School Leadership review urged systems to treat leadership development as a continuum spanning preparation, induction and in-service growth (Pont et al., 2008). More recently, UNESCO (2024) devoted its Global Education Monitoring Report to leadership in education, arguing that leaders must be deliberately prepared to "lead for learning".

In sub-Saharan Africa, principals have historically been appointed from among successful classroom teachers on the strength of teaching record and seniority rather than preparation for management (Bush & Oduro, 2006). Once appointed, they manage budgets, human resources, curriculum delivery, learner welfare, physical infrastructure and community relations, frequently with limited formal training (Bush, 2018; Wanzare, 2012). Recent reviews confirm that these challenges persist and have grown more complex because of accountability pressures, resource constraints and disruption such as the COVID-19 pandemic (Bush, 2022; Harris & Jones, 2020; Tintoré et al., 2022).

Kenya responded to this challenge in the early 2000s through the Kenya Education Staff Institute (KESI), which offered an Induction Course in Educational Management for Secondary School Principals (ICEMSSP). KESI was later reconstituted as the Kenya Education Management Institute (KEMI), which today offers a Diploma in Educational Leadership and Management to serving and aspiring school leaders (Kenya Education Management Institute [KEMI], n.d.). Leadership training has since been linked to the Teachers Service Commission's career progression framework (Teachers Service Commission [TSC], 2018) and to the national policy agenda for competency-based education reform (Republic of Kenya, 2019). The scale of public and private investment in such programmes makes a persuasive case for rigorous evaluation. Governments, sponsors and participants need to know whether the programmes are well designed, properly resourced, faithfully implemented and capable of changing how schools are led.

The question of how such programmes should be evaluated remains open. Much practice in training evaluation still relies on post-course reaction forms or on outcome-only designs. Such designs say little about why a programme succeeds or fails (Guskey, 2000; Kirkpatrick & Kirkpatrick, 2016). In 2008, Nyonje evaluated the ICEMSSP using an adapted version of Stufflebeam's Context–Input–Process–Product (CIPP) model. The adaptation placed human capital theory and systematic and blended training theory at the centre of the evaluation, so that every stage was judged against explicit theoretical expectations (Nyonje, 2008). That study generated rich empirical evidence on the programme, but the evaluation model it produced has not been systematically documented, refined or disseminated as a stand-alone tool. This paper addresses that gap. It sets out the model, examines its theoretical and conceptual foundations, draws lessons from its original application, and proposes a refined version for adoption by institutions that train and evaluate school leaders.

2. Literature Review

2.1 The Evolution of Programme Evaluation Approaches for Training

Programme evaluation has developed along several traditions, which Fitzpatrick et al. (2011) classify as objectives-oriented, management-oriented, consumer-oriented, expertise-oriented and participant-oriented approaches. Objectives-oriented designs judge a programme by the extent to which pre-specified objectives are met. They are parsimonious but tend to treat the programme as a "black box". Participant-oriented and developmental approaches, by contrast, foreground stakeholder meaning-making and adaptation in complex settings (Patton, 2011). Management-oriented approaches, of which the CIPP model is the most influential, take a middle path: they aim to serve decision makers by supplying timely information about each stage of a programme's life cycle (Stufflebeam & Coryn, 2014; Worthen et al., 1997).

Within the narrower field of training evaluation, Kirkpatrick's four levels of reaction, learning, behaviour and results remain the most widely cited framework (Kirkpatrick & Kirkpatrick, 2016). Phillips and Phillips (2016) extended the hierarchy to include return on investment. Guskey (2000) adapted it for educator professional development by adding a level on organisational support and change. These frameworks are strong on outcomes but comparatively silent on the context in which training needs arise, on the adequacy of inputs such as trainers and venues, and on the fidelity of delivery. For programmes that are publicly funded and nationally scaled, this silence is costly. When results disappoint, decision makers cannot tell whether the problem lay in the diagnosis of needs, the resources committed or the way the programme was delivered (Mertens & Wilson, 2019).

2.2 The CIPP Model: Origins, Logic and Contemporary Applications

Stufflebeam introduced the CIPP model in the late 1960s and early 1970s as an alternative to narrowly objectives-based testing. He defined evaluation as "the process of delineating, obtaining and providing useful information for judging decision alternatives" (Stufflebeam, 1971). The model distinguishes four complementary evaluations. Context evaluation assesses needs, problems, assets and opportunities to define goals. Input evaluation appraises competing strategies, resources and work plans. Process evaluation monitors, documents and gives feedback on implementation. Product evaluation identifies and judges intended and unintended outcomes (Stufflebeam, 2003; Stufflebeam & Shinkfield, 1985). In later formulations, product evaluation was subdivided into impact, effectiveness, sustainability and transportability, reflecting a growing concern with whether gains last and can be transferred to other settings (Stufflebeam & Zhang, 2017).

The model rests on a systems view of organisations. Organisations take inputs from the environment, transform them through throughput processes and export outputs, and each part must function well if the system is to survive (French et al., 2005; von Bertalanffy, 1968). By making the decision maker the primary audience, CIPP clarifies who will use evaluation findings, how, and about which aspects of the system (Worthen et al., 1997). Its formative orientation, "not to prove but to improve", distinguishes it from purely summative accountability designs (Stufflebeam & Zhang, 2017).

Contemporary scholarship confirms the model's adaptability. A systematic review of CIPP use in medical education found it valued for its comprehensiveness and its capacity to guide programme improvement (Toosi et al., 2021). CIPP has been applied to judge the quality and merit of nursing programmes (Lippe & Carter, 2018), to evaluate school-level quality (Aziz et al., 2018) and to assess teacher training programmes (Khaksar et al., 2023). In the last of these studies, applying the model to 296 stakeholders showed that a teacher-training programme met its core objectives. It also exposed weaknesses in course

duration, classroom-management content and practical teaching opportunities that an outcome-only design would have missed (Khaksar et al., 2023). Critics note, however, that CIPP can become descriptive and resource-intensive when it is applied without an explicit theory of how the programme is expected to work (Mertens & Wilson, 2019; Stufflebeam & Coryn, 2014). This critique is the entry point for the theory-anchored adaptation examined in this paper.

2.3 School Leadership Development and the Evaluation Gap

Research on principal preparation identifies features that distinguish effective programmes. They are grounded in clear standards and needs assessment. They combine coursework with authentic, problem-based learning and mentoring. They are delivered by qualified facilitators, and they are followed by sustained in-service support (Darling-Hammond et al., 2022; Pont et al., 2008). Context matters: leadership practices that succeed in one system may not transfer unchanged to another, and evaluators must take account of the institutional, cultural and resource conditions in which principals work (Hallinger, 2018). Bush and Glover (2014) further caution that leadership models are often imported without adequate attention to local realities.

In African settings, reviews consistently report that preparation is uneven, induction is short and follow-up is weak (Bush, 2018; Bush & Oduro, 2006). Studies from Kenya document gaps in principals' instructional leadership and financial management competencies (Wanzare, 2012). Much of the available evidence on leadership programmes consists of participant satisfaction surveys or single-point outcome studies, which rarely explain the mechanisms through which training does or does not change practice. This paper refers to this as the evaluation gap: leadership training is expanding faster than the frameworks available to evaluate it in a comprehensive, theory-informed and improvement-oriented way.

2.4 The ICEMSSP Evaluation as a Point of Departure

Nyonje's (2008) evaluation of the ICEMSSP was one of the earliest systematic attempts in Kenya to evaluate a national school leadership programme across its whole life cycle. The study was guided by four questions. They asked how effective the programme inputs were in meeting objectives, how effective the implementation process was, to what extent the programme improved principals' performance in educational management, and what strategies could improve it (Nyonje, 2008). Programme effectiveness was operationalised through six indicators of school management: financial management, curriculum management, staff management, student management, school plant management and school–community relations. The study's distinctive contribution was methodological. It placed an explicit theory core inside the CIPP framework and treated evaluation as a two-way, iterative flow between stages rather than a linear sequence. The sections that follow document and refine that contribution.

3. Objective and Research Questions

The objective of this paper is to document, theorise, refine and popularise the Theory-Anchored CIPP-Plus (TA-CIPP+) Model developed and tested by Nyonje (2008) as a practical and theory-informed framework for evaluating school leadership training programmes. The paper seeks to demonstrate how the model can be used to strengthen programme planning, implementation, monitoring, evaluation and continuous improvement in educational leadership development initiatives.

How was the CIPP model conceptualised, adapted and applied in the evaluation of the KESI Induction Course in Educational Management for Secondary School Principals?

What theoretical and conceptual foundations underpin the adapted model, and how do they shape the meaning and conduct of each evaluation stage?

What lessons emerge from the original application of the model that point to its strengths and to areas requiring refinement?

How can a refined version of the model be articulated, and what conditions would support its adoption and popularisation among institutions that train and evaluate school leaders?

4. Methodology

4.1 Research Approach and Design

The paper adopts a qualitative, interpretive approach and a conceptual model-building design. Conceptual papers do not report new primary data. They develop theory by integrating existing theories, frameworks and empirical evidence into new or refined constructs (Jaakkola, 2020). Of the four conceptual designs identified by Jaakkola (2020), namely theory synthesis, theory adaptation, typology and model, this paper combines theory adaptation, which extends the scope of an established framework (CIPP) by importing insights from other theories, with a model design that specifies relationships among constructs. The qualitative orientation is appropriate because the aim is to understand and communicate how an evaluation model works, why it works and how it may be improved, rather than to test statistical hypotheses (Creswell & Poth, 2018; Patton, 2015).

4.2 Data Sources

Three bodies of text constituted the data corpus. The primary source was the doctoral thesis in which the model was first developed and tested (Nyonje, 2008), specifically the chapters on the evaluation model and conceptual framework, the evaluation design and methodology, and the summary, conclusions and recommendations. The thesis itself drew on programme documents, including ICEMSSP field reports from 2004, 2006 and 2007. It also used questionnaires administered to secondary school principals, interviews with District Education Officers (DEOs), Quality Assurance and Standards Officers (QASOs) and KESI staff, and a multi-stage random sample of 120 principals drawn from 15 districts in four provinces. The second source was the foundational evaluation literature on the CIPP model and related approaches. The third was recent international, African and Kenyan scholarship on school leadership development and training evaluation, including current policy documents.

4.3 Document Selection and Literature Search

The thesis chapters were selected purposively because they contain the model's conceptual statement and the empirical evidence of its application. The wider literature was identified through an integrative review strategy (Snyder, 2019). This combined targeted searches of scholarly databases and publisher platforms, using terms such as "CIPP model", "training evaluation", "principal preparation" and "school leadership development", with backward citation tracking from seminal sources. Priority was given to peer-reviewed studies and authoritative reports published since 2014, while seminal works were retained where they define the constructs in use.

4.4 Analytical Procedure

Analysis followed Bowen's (2009) procedure for qualitative document analysis, which involves skimming, thorough reading and interpretation. Reflexive thematic analysis was layered on top of this procedure (Braun & Clarke, 2022). The analysis proceeded in four iterative phases. In the first phase, familiarization, the thesis chapters were read repeatedly, and analytic memos were written on how the model was defined and applied. In the second phase, deductive coding, the findings were coded against the four CIPP stages and the two anchoring theories, which produced a stage-by-theory evidence matrix. In the

third phase, inductive theme development, patterns that did not fit neatly within existing stages were identified as candidate refinements; examples were inequitable access, the influence of government policy and the absence of post-training follow-up. In the fourth phase, abductive model refinement, these themes were interrogated against contemporary literature to judge whether they warranted structural changes to the model, such as a new stage or lens, or operational changes, such as new indicators. The refined model and its evaluation matrix are the outputs of this final phase.

4.5 Trustworthiness

Trustworthiness was pursued through the criteria of credibility, transferability, dependability and confirmability (Lincoln & Guba, 1985; Nowell et al., 2017). Credibility was enhanced by triangulating the thesis evidence, which itself combined principals' self-reports, supervisors' ratings and documents, with independent literature. Dependability and confirmability were supported by an audit trail of memos and coding matrices, and by quoting the thesis statistics exactly. Transferability is addressed through thick description of the Kenyan context, so that readers can judge the model's fit to their own settings. Because the author also conducted the original study, reflexivity was essential. The analysis deliberately sought disconfirming evidence and gave the programme's weaknesses the same attention as its strengths.

4.6 Strategy for Popularizing the Model

Popularization is treated here as a methodological concern, not an afterthought. Drawing on diffusion of innovations theory, the adoption of a new practice depends on its perceived relative advantage, compatibility with existing values and systems, simplicity, trialability and observability (Rogers, 2003). The paper therefore presents the model in three forms that serve these attributes: a visual diagram (Figure 2), which supports simplicity and observability; a stage-by-stage evaluation matrix (Table 2), which supports trialability; and a comparative positioning against familiar frameworks (Table 3), which demonstrates relative advantage and compatibility. Section 9 sets out a knowledge-translation pathway for institutional uptake.

5. Theoretical and Conceptual Foundations of the Model

The distinctive feature of the model is that evaluation at every stage is anchored in explicit theory. Three bodies of theory, together with the decision-oriented philosophy of evaluation, supply the lenses through which each stage is judged.

5.1 Systems Theory and the Decision-Oriented Management Approach

The CIPP model and Alkin's related work form what Worthen et al. (1997) describe as the management-oriented approach to evaluation. The approach assumes that evaluators are most useful when they serve administrators, policy makers, boards and practitioners who need good evaluative information. It rests on a systems conception of education in which decisions are made about inputs, processes and outputs (French et al., 2005; von Bertalanffy, 1968). For a leadership programme, this means that the programme itself is a system. Its inputs are trainers, modules, venues, equipment and finance. Its throughputs are delivery methods, schedules and procedures. Its outputs are principals whose management practice has changed. A weakness in any sub-system threatens the whole (Nyonje, 2008).

5.2 Human Capital Theory

Human capital theory holds that investment in the knowledge, skills and health of people yields returns in productivity, much as investment in physical capital does (Becker, 1993; Schultz, 1961). In the original

evaluation, human capital was regarded as “the engine” of the ICEMSSP programme. Both resource persons and programme implementers were seen as key to the programme’s success, and their ability to organise, plan and implement it was a major concern of the study (Nyonje, 2008). The theory operates at two levels in the model. First, the programme is itself an investment in the human capital of principals, whose enhanced competence is expected to yield returns in better-managed schools. Second, the quality of the programme depends on the human capital of those who design and deliver it. The theory therefore generates evaluative criteria at every stage. At the context stage, highly qualified personnel should conduct the needs analysis. At the input stage, resource persons should be selected for demonstrable expertise and experience. At the product stage, graduates should add measurable value to school management.

5.3 Systematic and Blended Training Theory

Systematic training theory conceives training as a planned cycle of needs assessment, design, delivery and evaluation, in which each phase is logically linked to the next (Goldstein & Ford, 2002). Blended training theory adds that effective delivery combines several methods, such as face-to-face instruction, discussion, case study, project work, field study and, increasingly, technology-mediated learning, so as to exploit the strengths of each (Garrison & Kanuka, 2004; Graham, 2006). In the original model, the theory had two meanings. At the design level it implied borrowing and blending methodologies, strategies and procedures from well-researched programmes that had proved effective in training educational managers. At the delivery level it implied resource persons skilled in employing a blend of methods for maximum impact (Nyonje, 2008). This dual meaning gave the evaluator a normative standard: a programme dominated by a single method, such as lecturing, would fall short of theoretical expectations regardless of participant satisfaction.

5.4 The Original Conceptual Framework

Figure 1 reproduces the conceptual framework through which these theories were integrated with the CIPP stages. The framework proposes that an effective programme design passes through four stages, context, input, process and products, all guided by a central theory box (Nyonje, 2008). Arrows from the theory box to each stage signify that theoretical principles govern every evaluative judgement. The arrows linking the four stages point in both directions, which denotes the kind of evaluation that takes place during the programme. Training needs assessment takes place at the context stage and leads to the identification of objectives. Inputs must be congruent with those objectives, and their adequacy can in turn reveal whether the objectives are sufficiently responsive to assessed needs. Monitoring and formative evaluation of the process lead to improvement in inputs and training strategies. Summative evaluation of products leads to an assessment of the effectiveness of both the graduates and the programme design. This logic aligns with Ogula’s (2002) five-level view of programme evaluation: needs assessment, monitoring, formative evaluation, summative evaluation and impact evaluation.

Figure 1

Conceptual Framework of the Original Theory-Anchored CIPP Evaluation Model

Note. Adapted from Nyonje (2008). ICEMSSP = Induction Course in Educational Management for Secondary School Principals; M&E = monitoring and evaluation.

6. A Detailed Account of the School Leadership Training Evaluation Model

This section describes how each stage of the model was operationalised in the ICEMSSP evaluation and what the stage revealed. The account shows both the practical workings of the model and the evidence

base from which the refinements in Section 7 are derived.

6.1 Context Evaluation: Diagnosing Needs and Setting Direction

Context evaluation is the most basic type of evaluation. Its purpose is to provide a rationale for objectives by defining the relevant environment, identifying unmet needs and unused opportunities, and diagnosing the problems that prevent needs from being met (Stufflebeam & Shinkfield, 1985). It is macro-analytic: it sets the boundaries of the system, describes its values and goals, and asks whether practice is consistent with the validated principles of relevant theory. Two methodological modes were distinguished in the original study. The contingency mode searches outside the system for opportunities and pressures, for example through study visits, literature, consultants and assessment of community values, and projects future societal, technological, political, economic and demographic trends. The congruence mode compares actual with intended system performance (Nyonje, 2008).

In the ICEMSSP evaluation, context evaluation relied on document review. It established the target population as principals of secondary schools, the programme objectives, and the implementation strategies and procedures. Principals were located within a context of societal needs and values, economic and political trends and technological change, all analysed for their effect on school management (Nyonje, 2008). Applying the human capital lens at this stage meant asking whether needs had been identified by suitably qualified people. Applying the systematic training lens meant asking whether objectives flowed logically from those needs.

6.2 Input Evaluation: Judging Resources and Strategies

Input evaluation provides information for deciding how to use resources to meet programme goals. It identifies and assesses the capabilities of the responsible agency, the strategies for achieving goals and the designs for implementing a selected strategy. Alternative designs are appraised in terms of staffing, time and budget requirements, procedural barriers, relevance to objectives and overall potential (Stufflebeam, 2003). Input evaluation in the ICEMSSP study examined resource persons, training modules, venues, equipment, financial and administrative support, and the adequacy of monitoring and evaluation.

The input evaluation produced mixed findings. Most participants found resource persons effective, and 47% described them as knowledgeable in their areas of facilitation. Yet implementers were unclear about the criteria for selecting resource persons, and some facilitators did not follow facilitation guidelines, communicated poorly, had limited content mastery or could not use presentation technology (Nyonje, 2008). Sixty percent of implementers held a master’s degree in education and were experienced, a view confirmed by 58.1% of participants. Just over half of participants (52.3%) considered the venues appropriate, while 34.1% did not. Complaints concerned seating, sanitation, catering, hygiene and customer care. These problems were attributed partly to a government directive requiring training to be held in public institutions, some of which were poorly equipped. Although 65.9% of participants judged materials and equipment adequate, implementers reported shortages of essential equipment and frequent computer breakdowns. Participant fees kept some principals out of the programme, even though 59.1% of participants considered it well funded (Nyonje, 2008).

6.3 Process Evaluation: Monitoring Fidelity and Quality of Delivery

Process evaluation provides periodic feedback to those responsible for implementation. It has three objectives: to detect or predict defects in procedural design or implementation, to provide information for programmed decisions, and to maintain a record of the procedure as it occurs (Stufflebeam & Shinkfield, 1985). The original study applied three strategies. The first was continuous monitoring of potential

sources of failure, including interpersonal relationships, communication, logistics, shared understanding of programme intent and the adequacy of resources and time. The second was serving pre-programmed decisions. The third was documenting what actually took place, so that later successes or failures could be explained (Nyonje, 2008).

Process evaluation revealed several implementation weaknesses. Participants and resource persons were not invited in good time. Some resource persons started late or overran the two-hour session allocation, and topics such as financial management received insufficient time. A substantial minority of principals (40.9%) judged the two-week duration inadequate, which led presenters to rush through content (Nyonje, 2008). Most participants found the content relevant and well sequenced. However, field reports from 2004, 2006 and 2007 showed that different cohorts had been exposed to different amounts of content, creating imbalances in the knowledge, skills and attitudes acquired. Lecturing was the dominant method, followed by subject-expert presentations (56.8%) and group discussion (47.7%). Case studies, projects, field studies and role play were rarely used, largely because of time pressure (Nyonje, 2008). Judged against blended training theory, this narrow repertoire was a significant shortfall.

6.4 Product Evaluation: Judging Outcomes Against Theory-Derived Indicators

Product evaluation measures and interprets attainments, not only at the end of a cycle but as often as necessary. It involves operationally defining objectives, measuring criteria associated with them, comparing measurements with absolute or relative standards, and interpreting outcomes in the light of the recorded context, input and process information (Stufflebeam, 2003). The ICEMSSP study used an ex post facto, causal-comparative design (Cohen & Manion, 1994; Kerlinger & Lee, 2000). It compared principals who had been exposed to the programme with those who had not, across the six management indicators. Ratings were supplied both by principals and by DEOs and QASOs, who, as immediate supervisors, were regarded as credible key informants.

The product evidence was favourable. On supervisors' ratings, the mean performance of exposed principals across all six domains was 3.11, compared with 2.13 for non-exposed principals. The largest differences were in financial management (1.26), staff management (1.24) and curriculum management (1.03). Principals' self-ratings rose from a mean of 2.373 before training to 3.257 after training and paired-sample t-tests showed that these differences were statistically significant. Qualitatively, exposed principals were described as undertaking their roles with greater confidence, more systematically and with a clearer understanding of expectations (Nyonje, 2008). Two cautions emerged. Financial management, although most improved, remained the lowest-rated domain for both groups. Access was also uneven: among exposed principals, 18.5% were female and 34.1% were male, and 42.3% of principals of district schools had not yet been exposed to the programme (Nyonje, 2008).

6.5 What the Application Revealed

Table 1 synthesizes the evidence by stage and states the implication of each finding for the model. Three insights stand out. First, the theory core did real analytic work. It allowed the evaluator to judge the dominance of lecturing as a design flaw, and not merely a matter of preference, and to treat the selection of resource persons as a human capital question. Second, the bi-directional arrows proved apt. Process findings such as insufficient time explained product findings such as persistent weakness in financial management, and input constraints such as the government venue directive explained process difficulties. Third, several important findings did not belong comfortably to any single stage. Inequitable access by gender and school type, the effect of government procurement and remuneration policies, and the absence of structured follow-up after training all cut across stages. These cross-cutting findings provide the empirical warrant for the refinements proposed in the next section.

Table 1

Synthesis of ICEMSSP Evaluation Findings by CIPP Stage and Implications for Model Refinement

CIPP stage	Key findings from the original evaluation (Nyonje, 2008)	Implication for the refined model
Context	Target group, objectives and strategies clarified through document review; 42.3% of district-school principals not yet reached; female principals under-represented among participants (18.5% vs 34.1% male).	Add access-equity analysis and policy/career-path alignment to context evaluation; introduce a cross-cutting equity–inclusion lens.
Input	Resource persons largely effective (47% knowledgeable) but no uniform selection criteria; 60% of implementers held a master's degree; 34.1% judged venues inappropriate; government venue and remuneration directives constrained quality; equipment breakdowns; fees deterred some principals.	Specify transparent trainer selection and induction criteria, ICT readiness and financing/cost-sharing as input criteria; introduce a policy–governance lens.
Process	Late invitations; overrunning sessions; 40.9% judged two-week duration inadequate; unequal content across cohorts (2004–2007); lecture method dominant, participatory methods rarely used.	Monitor time use, content fidelity and delivery blend; recommend phased, thematic and blended designs.
Product	Exposed principals outperformed non-exposed (M = 3.11 vs 2.13); largest gains in financial (1.26), staff (1.24) and curriculum (1.03) management; self-rated performance rose from 2.373 to 3.257 (significant); financial management remained weakest.	Retain six-domain indicators and comparison/pre–post logic; add formal assessment of participants' knowledge, skills and attitudes.
Beyond product	No structured follow-up; further research recommended on learner outcomes and on cost–benefit of mixed-mode delivery.	Add a fifth Impact and Sustainability stage and a continuous M&E and learning loop.

7. The Refined Model: The Theory-Anchored CIPP-Plus (TA-CIPP+) Model

Drawing on the lessons summarised in Table 1 and on contemporary evaluation literature, the original model is refined into the Theory-Anchored CIPP-Plus (TA-CIPP+) Model for evaluating school leadership training programmes. The "Theory-Anchored" label preserves the model's defining contribution. The "Plus" signals three extensions: a fifth stage, two cross-cutting lenses and an explicit learning loop. Figure 2 presents the model.

Figure 2

The Theory-Anchored CIPP-Plus (TA-CIPP+) Model for Evaluating School Leadership Training Programmes

Note. Developed by the author from Nyonje (2008), Stufflebeam and Zhang (2017), Guskey (2000) and Ogula (2002). KSA = knowledge, skills and attitudes; M&E = monitoring and evaluation.

7.1 Design Principles

Five principles guided the refinement. (a) Theory first: every evaluative criterion should be traceable to an explicit theory of how leadership training produces change. (b) Improvement over proof: the model should primarily serve programme decisions, consistent with Stufflebeam and Zhang (2017). (c) Whole life cycle: evaluation should extend beyond course completion to transfer, school outcomes and sustainability (Guskey, 2000; Kirkpatrick & Kirkpatrick, 2016). (d) Context sensitivity: policy, financing and equity conditions should be examined explicitly rather than treated as background (Hallinger, 2018). (e) Utility and parsimony: the model should be simple enough to be adopted by institutions with modest evaluation capacity (Rogers, 2003).

7.2 Components of the Refined Model

The theory core. At the centre of the model sit four mutually reinforcing theoretical anchors. Human capital theory frames training as investment and makes the competence of trainers and graduates the focal criterion. Systematic and blended training theory requires logical linkage across the training cycle and a varied, participatory delivery repertoire. Systems theory treats the programme as interdependent sub-systems. The decision-oriented philosophy directs every stage towards identifiable decisions and decision makers. Arrows from the core to each stage indicate that stage-specific criteria are derived from these theories.

Stage 1: Context. Context evaluation retains its original functions of needs assessment, target-group definition and objective setting. It adds explicit attention to alignment with policy and career pathways, for example national leadership standards and teacher career progression requirements (TSC, 2018), and to the equity of access. Objectives should be specific, measurable and time-bound so that later stages have clear standards.

Stage 2: Input. Input evaluation continues to appraise trainers, modules, venues, equipment and finance. It adds three explicit criteria derived from the original findings: documented and transparent criteria for selecting and inducting resource persons; standardised, thematically organised modules so that all cohorts receive equivalent content; and financing models, including cost-sharing and subsidies, that protect access. Readiness of information and communication technology (ICT) becomes a core input because of the shift to blended and online delivery.

Stage 3: Process. Process evaluation monitors fidelity, time use, scheduling and the blend of delivery methods, and it feeds formative information back to implementers during delivery. The refined model sets a normative expectation that no single method should dominate. Participatory methods such as case analysis, school-based projects, peer learning and mentoring are expected to feature prominently (Darling-Hammond et al., 2022; Garrison & Kanuka, 2004).

Stage 4: Product. Product evaluation judges immediate outcomes: assessed gains in knowledge, skills and attitudes, and changes in performance across the six management domains of finance, curriculum, staff, students, physical plant and school–community relations. Following the original design, it combines pre–post and comparison-group evidence with supervisor ratings. It adds formal participant assessment during and at the end of training, which the original study recommended (Nyonje, 2008).

Stage 5: Impact and sustainability (new). The added stage asks whether learning transfers into sustained leadership practice and, ultimately, into improved school functioning and learner outcomes. It also asks whether benefits justify costs and whether the programme can be scaled or transported. The stage operationalises Stufflebeam and Zhang's (2017) subdivision of product evaluation, Guskey's (2000) levels of organisational support and student learning outcomes, and Ogula's (2002) impact evaluation. It also

responds directly to the original recommendations for a structured follow-up mechanism, cost–benefit analysis of mixed-mode delivery, and studies of effects on academic performance (Nyonje, 2008).

Cross-cutting lens 1: Policy, governance and financing environment. The ICEMSSP evaluation showed that government directives on venue procurement and trainer remuneration shaped input quality and, through it, process quality (Nyonje, 2008). The refined model therefore requires evaluators to examine, at every stage, how laws, regulations, financing arrangements and institutional mandates enable or constrain the programme.

Cross-cutting lens 2: Equity and inclusion. Access disparities by gender, school type and ability to pay were among the most consequential findings of the original study. The equity lens requires disaggregated analysis at every stage: who is targeted, who is resourced, who participates, who benefits, and whose schools improve.

The continuous M&E and learning loop. The bi-directional arrows of the original framework are formalised as a cyclical loop running from needs assessment through monitoring, formative, summative and impact evaluation to structured lesson-sharing, consistent with Ogula (2002) and Nyonje et al. (2012). Findings at any stage can trigger redesign at any other.

7.3 Operationalising the Model: An Evaluation Matrix

To support trialability, Table 2 translates the model into an evaluation matrix that specifies, for each stage, guiding evaluation questions, illustrative indicators, data sources and methods, and the decisions the stage serves. The matrix is illustrative rather than prescriptive and should be adapted to the scale and purpose of each programme.

Table 2

Evaluation Matrix for Applying the TA-CIPP+ Model

Stage	Guiding evaluation questions	Illustrative indicators	Data sources and methods	Decisions served
1. Context	What leadership needs exist? Are objectives aligned to needs, standards and policy? Who is excluded?	Documented needs; clarity of objectives; policy alignment; access by gender, region and school type	Needs surveys; policy and document review; focus groups; supervisor interviews	Programme goals, target groups, priorities
2. Input	Are trainers, modules, venues, ICT and finance adequate and appropriate? Is the design the best available option?	Trainer qualifications and selection criteria; module standardisation; venue and ICT audits; unit cost; fee burden	Records and CV review; venue checklists; budget analysis; expert review	Resource allocation, trainer selection, design choice

3. Process	Is the programme delivered as designed? Is the delivery blend adequate? Is time sufficient?	Attendance; punctuality; content coverage; method mix; participant engagement	Observation; session logs; participant feedback; implementer interviews	Mid-course corrections; scheduling; trainer support
4. Product	What knowledge, skills and attitudes were gained? Has performance improved in the six management domains?	Assessment scores; pre–post and comparison differences; supervisor ratings in six domains	Tests and assignments; questionnaires; supervisor rating scales; t-tests/ANOVA; thematic analysis	Certification; programme continuation or revision
5. Impact & sustainability	Is learning transferred and sustained? Are schools and learners better off? Is the programme cost-effective and scalable?	Observed leadership practice; school management audits; learner outcome trends; cost–benefit ratios	Follow-up visits; tracer studies; case studies; administrative data; cost analysis	Scale-up, policy adoption, reinvestment
Cross-cutting lenses	How do policy, governance and financing shape each stage? Are benefits equitably distributed?	Policy constraints identified; disaggregated participation and outcome data	Policy analysis; disaggregated statistics; stakeholder dialogue	Policy reform; targeting and subsidy decisions

7.4 Positioning the Model Among Established Frameworks

Table 3 compares the TA-CIPP+ model with the classical CIPP model and with the two most widely used training evaluation frameworks. The comparison shows that the refined model does not replace these frameworks. It integrates their strengths: CIPP's decision orientation and whole-programme scope, and the Kirkpatrick and Guskey attention to transfer and results. It adds features that are especially salient for leadership training in resource-constrained, policy-driven systems.

Table 3

Comparison of the TA-CIPP+ Model With Established Evaluation Frameworks

Feature	Kirkpatrick (2016)	Guskey (2000)	Classical CIPP (Stufflebeam & Zhang, 2017)	TA-CIPP+ (this paper)

Primary focus	Training outcomes	Professional development outcomes	Programme decisions across life cycle	Leadership programme decisions across life cycle
Needs and context analysis	Limited	Limited	Yes	Yes, with policy and equity alignment
Evaluation of inputs and design	No	Partial (organisational support)	Yes	Yes, with explicit trainer, ICT and financing criteria
Implementation (process) monitoring	No	No	Yes	Yes, with delivery-blend and fidelity criteria
Transfer and results / impact	Yes (Levels 3–4)	Yes (Levels 4–5)	Yes (within product)	Yes (separate Stage 5)
Explicit theory anchoring	No	Implicit	Not specified	Yes (theory core)
Equity lens	No	No	Not explicit	Yes (cross-cutting)
Policy–governance lens	No	Partial	Not explicit	Yes (cross-cutting)

8. Conclusion

This paper set out to document, theorise, refine and popularise the school leadership training evaluation model first developed and tested by Nyonje (2008). Four conclusions follow from the analysis.

First, in relation to the first research question, the original model was a principled adaptation of Stufflebeam's CIPP framework. It retained CIPP's four stages and decision orientation but placed an explicit theory core at the centre and treated the links between stages as iterative and bi-directional. Applied to the ICEMSSP, it generated a comprehensive account of the programme that combined document analysis, principals' self-reports and supervisors' ratings across six management domains.

Second, in relation to the second research question, human capital theory and systematic and blended training theory did substantive evaluative work. They converted descriptive observations, such as the dominance of lecturing, unclear criteria for selecting trainers and uneven content across cohorts, into judgements about the programme's fidelity to its own theory of change. This theory anchoring directly answers the common criticism that CIPP evaluations can become descriptive and atheoretical.

Third, in relation to the third research question, the original application demonstrated the model's power to explain as well as to judge. The programme improved principals' performance in all six domains. The model also revealed why gains in financial management remained limited and why some principals, notably women and heads of district schools, benefited less. Those findings that cut across stages showed where the model itself needed to grow.

Fourth, in relation to the fourth research question, the refined TA-CIPP+ model extends the original through a fifth Impact and Sustainability stage, policy–governance and equity–inclusion lenses, and a formalised learning loop. It is operationalised through an evaluation matrix and positioned against established frameworks. The model offers a coherent, context-sensitive and practical tool for evaluating school leadership training at a time when such training is expanding rapidly in Kenya and across Africa.

Its main limitation is that the refinements are conceptually derived and have not yet been tested empirically; the recommendations below therefore include validation as a priority.

The study demonstrates that effective leadership training requires evaluation approaches that extend beyond participant satisfaction and immediate learning outcomes. By integrating theory, context, programme inputs, implementation processes, outcomes, and long-term sustainability considerations, the TA-CIPP+ Model provides a comprehensive framework for evaluating leadership development initiatives. As investments in educational leadership continue to expand across Kenya and Africa, the model offers policymakers, training institutions and development partners a practical tool for strengthening accountability, promoting continuous improvement and enhancing the overall effectiveness of school leadership programmes.

9. Recommendations for the Adoption of the Model

The following recommendations are addressed to the institutions best placed to adopt and popularise the TA-CIPP+ model.

Institutional Adoption by KEMI and Similar Institutions

The Kenya Education Management Institute (KEMI) and other leadership development institutions should adopt the TA-CIPP+ Model as the organising framework for monitoring and evaluating school leadership programmes. This will ensure that context, input, process, product, and impact dimensions are integrated into programme design and implementation.

Policy Integration by the Ministry of Education and the Teachers Service Commission

The Ministry of Education and the Teachers Service Commission should institutionalise periodic evaluations of educational leadership programmes using the TA-CIPP+ Model and utilise the equity lens to monitor participation and outcomes across gender, region, school category and socio-economic status.

Development of Quality Standards for Leadership Training

Training providers should establish and implement quality assurance standards covering trainer selection, curriculum design, learning resources, ICT integration, monitoring and evaluation procedures, and participant support mechanisms.

Promotion of Blended and Phased Learning Approaches

Leadership development programmes should move beyond intensive lecture-based delivery and adopt phased and blended learning approaches that incorporate mentoring, school-based projects, peer learning, reflective practice and online learning environments.

Institutionalisation of Follow-up and Impact Tracking

Training institutions should establish formal follow-up systems to monitor leadership practice after programme completion through coaching, mentoring, school visits, professional learning communities and tracer studies.

Capacity Building in Educational Evaluation

Universities and training institutions should integrate the TA-CIPP+ Model into postgraduate programmes, professional development courses and monitoring and evaluation training initiatives to strengthen evaluation capacity within the education sector.

Dissemination and Knowledge Translation

The model should be disseminated through policy briefs, practitioner guides, scholarly publications,

professional conferences and leadership forums to support wider adoption and utilisation across educational institutions.

Empirical Validation and Further Research

Future research should test the TA-CIPP+ Model in different educational settings and evaluate its effectiveness in improving leadership development programmes, educational outcomes and organisational performance.

This paper contributes to the field of educational leadership and programme evaluation by documenting, refining and operationalising the Theory-Anchored CIPP-Plus (TA-CIPP+) Model for evaluating school leadership training programmes. The study extends the classical CIPP framework by embedding Human Capital Theory, Systems Theory, Systematic and Blended Training Theory, and the Decision-Oriented Evaluation perspective into a unified evaluation framework. The paper further advances existing evaluation approaches by introducing an explicit Impact and Sustainability stage, integrating policy-governance and equity-inclusion lenses, and establishing a continuous monitoring, evaluation and learning loop. The resulting model provides educational institutions, policymakers and leadership development providers with a practical framework for designing, monitoring and evaluating leadership development initiatives while contributing a contextually relevant evaluation model for use in Kenya, Africa and other similar educational contexts.

References

1. Aziz, S., Mahmood, M., & Rehman, Z. (2018). Implementation of CIPP model for quality evaluation at school level: A case study. Journal of Education and Educational Development, 5(1), 189–206.
2. Becker, G. S. (1993). Human capital: A theoretical and empirical analysis with special reference to education (3rd ed.). University of Chicago Press.
3. Bowen, G. A. (2009). Document analysis as a qualitative research method. Qualitative Research Journal, 9(2), 27–40. https://doi.org/10.3316/QRJ0902027
4. Braun, V., & Clarke, V. (2022). Thematic analysis: A practical guide. SAGE.
5. Bush, T. (2018). Preparation and induction for school principals: Global perspectives. Management in Education, 32(2), 66–71.
6. Bush, T. (2022). Challenges facing school principals: Problems and solutions. Educational Management Administration & Leadership, 50(4), 533–535. https://doi.org/10.1177/17411432221096238
7. Bush, T., & Glover, D. (2014). School leadership models: What do we know? School Leadership & Management, 34(5), 553–571. https://doi.org/10.1080/13632434.2014.928680
8. Bush, T., & Oduro, G. K. T. (2006). New principals in Africa: Preparation, induction and practice. Journal of Educational Administration, 44(4), 359–375. https://doi.org/10.1108/09578230610676587
9. Cohen, L., & Manion, L. (1994). Research methods in education (4th ed.). Routledge.
10. Creswell, J. W., & Poth, C. N. (2018). Qualitative inquiry and research design: Choosing among five approaches (4th ed.). SAGE.
11. Darling-Hammond, L., Wechsler, M. E., Levin, S., Leung-Gagné, M., & Tozer, S. (2022). Developing effective principals: What kind of learning matters? Learning Policy Institute.
12. Fitzpatrick, J. L., Sanders, J. R., & Worthen, B. R. (2011). Program evaluation: Alternative approaches and practical guidelines (4th ed.). Pearson.
13. French, W. L., Bell, C. H., & Zawacki, R. A. (2005). Organization development and transformation:

Managing effective change (6th ed.). McGraw-Hill/Irwin.

14. Garrison, D. R., & Kanuka, H. (2004). Blended learning: Uncovering its transformative potential in higher education. The Internet and Higher Education, 7(2), 95–105. https://doi.org/10.1016/j.iheduc.2004.02.001
15. Goldstein, I. L., & Ford, J. K. (2002). Training in organizations: Needs assessment, development, and evaluation (4th ed.). Wadsworth.
16. Graham, C. R. (2006). Blended learning systems: Definition, current trends, and future directions. In C. J. Bonk & C. R. Graham (Eds.), The handbook of blended learning: Global perspectives, local designs (pp. 3–21). Pfeiffer.
17. Grissom, J. A., Egalite, A. J., & Lindsay, C. A. (2021). How principals affect students and schools: A systematic synthesis of two decades of research. The Wallace Foundation.
18. Guskey, T. R. (2000). Evaluating professional development. Corwin Press.
19. Hallinger, P. (2018). Bringing context out of the shadows of leadership. Educational Management Administration & Leadership, 46(1), 5–24. https://doi.org/10.1177/1741143216670652
20. Harris, A., & Jones, M. (2020). COVID 19 – school leadership in disruptive times. School Leadership & Management, 40(4), 243–247. https://doi.org/10.1080/13632434.2020.1811479
21. Jaakkola, E. (2020). Designing conceptual articles: Four approaches. AMS Review, 10(1–2), 18–26. https://doi.org/10.1007/s13162-020-00161-0 Kenya Education Management Institute. (n.d.). Diploma in educational leadership and management. Retrieved September 24, 2026, from https://www.kemi.ac.ke/diploma-courses/diploma-in-educational-leadership-and-management/
22. Kerlinger, F. N., & Lee, H. B. (2000). Foundations of behavioral research (4th ed.). Harcourt College Publishers.
23. Khaksar, M., Kiany, G. R., & ShayesteFar, P. (2023). Using a CIPP-based model for evaluation of teacher training programs in a private-sector EFL institute. Language Teaching Research Quarterly, 38, 65–91. https://doi.org/10.32038/ltrq.2023.38.04
24. Kirkpatrick, J. D., & Kirkpatrick, W. K. (2016). Kirkpatrick's four levels of training evaluation. ATD Press.
25. Leithwood, K., Harris, A., & Hopkins, D. (2020). Seven strong claims about successful school leadership revisited. School Leadership & Management, 40(1), 5–22. https://doi.org/10.1080/13632434.2019.1596077
26. Lincoln, Y. S., & Guba, E. G. (1985). Naturalistic inquiry. SAGE.
27. Lippe, M., & Carter, P. (2018). Using the CIPP model to assess nursing education program quality and merit. Teaching and Learning in Nursing, 13(1), 9–13.
28. Mertens, D. M., & Wilson, A. T. (2019). Program evaluation theory and practice: A comprehensive guide (2nd ed.). Guilford Press.
29. Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic analysis: Striving to meet the trustworthiness criteria. International Journal of Qualitative Methods, 16(1), 1–13. https://doi.org/10.1177/1609406917733847
30. Nyonje, R. O. (2008). Evaluation of the Kenya Education Staff Institute induction course in educational management for secondary school principals [Unpublished doctoral thesis]. University of Nairobi.
31. Nyonje, R. O., Ndunge, K. D., & Mulwa, A. S. (2012). Monitoring and evaluation of projects and programs: A handbook for students and practitioners. Aura Publishers.

32. Ogula, P. A. (2002). Monitoring and evaluation of educational projects and programmes. New Kemit Publishers.
33. Patton, M. Q. (2011). Developmental evaluation: Applying complexity concepts to enhance innovation and use. Guilford Press.
34. Patton, M. Q. (2015). Qualitative research and evaluation methods (4th ed.). SAGE.
35. Phillips, J. J., & Phillips, P. P. (2016). Handbook of training evaluation and measurement methods (4th ed.). Routledge.
36. Pont, B., Nusche, D., & Moorman, H. (2008). Improving school leadership: Vol. 1. Policy and practice. OECD Publishing. https://doi.org/10.1787/9789264044715-en
37. Republic of Kenya. (2019). Sessional Paper No. 1 of 2019 on a policy framework for reforming education and training for sustainable development in Kenya. Ministry of Education.
38. Rogers, E. M. (2003). Diffusion of innovations (5th ed.). Free Press.
39. Schultz, T. W. (1961). Investment in human capital. The American Economic Review, 51(1), 1–17.
40. Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. Journal of Business Research, 104, 333–339. https://doi.org/10.1016/j.jbusres.2019.07.039
41. Stufflebeam, D. L. (1971). The relevance of the CIPP evaluation model for educational accountability. Journal of Research and Development in Education, 5(1), 19–25.
42. Stufflebeam, D. L. (2003). The CIPP model for evaluation. In T. Kellaghan & D. L. Stufflebeam (Eds.), International handbook of educational evaluation (pp. 31–62). Kluwer Academic.
43. Stufflebeam, D. L., & Coryn, C. L. S. (2014). Evaluation theory, models, and applications (2nd ed.). Jossey-Bass.
44. Stufflebeam, D. L., & Shinkfield, A. J. (1985). Systematic evaluation: A self-instructional guide to theory and practice. Kluwer-Nijhoff.
45. Stufflebeam, D. L., & Zhang, G. (2017). The CIPP evaluation model: How to evaluate for improvement and accountability. Guilford Press.
46. Teachers Service Commission. (2018). Career progression guidelines for teachers. Teachers Service Commission.
47. Tintoré, M., Cunha, R. S., Cabral, I., & Alves, J. J. M. (2022). A scoping review of problems and challenges faced by school leaders (2003–2019). Educational Management Administration & Leadership, 50(4), 536–573. https://doi.org/10.1177/1741143220942527
48. Toosi, M., Modarres, M., Amini, M., & Geranmayeh, M. (2021). Context, input, process, and product evaluation model in medical education: A systematic review. Journal of Education and Health Promotion, 10, Article 199.
49. UNESCO. (2024). Global education monitoring report 2024/5: Leadership in education – Lead for learning. UNESCO. von Bertalanffy, L. (1968). General system theory: Foundations, development, applications. George Braziller.
50. Wanzare, Z. (2012). Instructional leadership in public secondary schools in Kenya. Educational Management Administration & Leadership, 40(2), 188–216.
51. Worthen, B. R., Sanders, J. R., & Fitzpatrick, J. L. (1997). Program evaluation: Alternative approaches and practical guidelines (2nd ed.). Longman.