African Educational Research Journal
Vol. 14(3), pp. 504-513, July 2026
ISSN: 2354-2160
Full Length Research Paper

Beyond CIPP: Validating the CIPPRI framework: A mixed-methods stakeholder-driven evaluation of an electrical engineering program in an emerging economy

Settha Sukhan, Phichittra Thongpanit* and Sarit Srikhaw

Faculty of Education, Nakhon Phanom University, Thailand.

Accepted 9 June, 2026

ABSTRACT

Conventional Context–Input–Process–Product (CIPP) evaluations have been widely adopted across higher-education quality-assurance regimes; however, they tend to subsume reflective and improvement-oriented practices within process or product metrics, rendering these mechanisms structurally invisible. This study advances and empirically validates an extended evaluation framework—CIPPRI—which disaggregates Reflection (R) and Improvement (I) into two dimensions analytically distinct from the original four. The framework was applied to evaluate a Bachelor of Engineering (Electrical Engineering) program at a Thai public university through a sequential explanatory mixed-methods design. Quantitative data were collected from 240 internal and external stakeholders (students n = 120, alumni n = 80, employers n = 25, instructors n = 11, and administrators n = 4) using a six-dimension, 42-item Likert-scale instrument. Construct validity was established through Item-Objective Congruence (IOC = 0.67–1.00), exploratory factor analysis (KMO = 0.91; Bartlett χ^2 = 5,842.34, $p < .001$), and confirmatory factor analysis supporting a six-factor solution (CFI = 0.94, TLI = 0.93, RMSEA = 0.058, SRMR = 0.047). Internal consistency was strong (Cronbach α = 0.89–0.94; McDonald ω = 0.91–0.95). Quantitative data were analyzed using descriptive statistics, Kruskal–Wallis tests, and post-hoc Dunn–Bonferroni comparisons; qualitative data from semi-structured interviews (n = 18) and focus groups (n = 4) were thematically analyzed. The overall program was rated favourably, with the process dimension highest (Weighted M = 4.21, Pooled SD = 0.79) and the input dimension lowest (Weighted M = 3.83, Pooled SD = 0.77), the latter constrained by obsolete laboratory infrastructure and limited smart-grid resources—a bottleneck consistent with the broader literature on emerging economies. Crucially, while reflection and improvement scored highly in self-report (M = 4.15, M = 4.05), thematic analysis revealed these practices were rated as latent rather than systemically institutionalized. Kruskal–Wallis tests indicated significant divergence across stakeholder groups for the Input (H = 14.62, p = .006, ε^2 = 0.06) and Improvement (H = 11.34, p = .023, ε^2 = 0.04) dimensions. The study contributes (a) a psychometrically validated six-factor evaluation instrument; (b) empirical evidence that explicit reflection-improvement dimensions surface evaluation gaps invisible to traditional CIPP; and (c) a transferable framework for engineering programs navigating Industry 4.0/5.0 transitions in resource-constrained settings.

Keywords: Program evaluation, CIPP model, reflective practice, continuous improvement, engineering education, mixed-methods, confirmatory factor analysis, Thailand, quality assurance.

---

*Corresponding author. Email: phichittra@npu.ac.th.

INTRODUCTION

Engineering programs in emerging economies confront a compounding challenge: simultaneously aligning curricula with rapid technological transitions associated with Industry 4.0 and 5.0, satisfying outcome-based national qualification frameworks, and competing for limited capital expenditure on laboratory infrastructure. Within Thailand, the National Qualifications Framework for Higher Education (TQF-HEd) mandates outcome-based curricula and structured stakeholder consultation. Currently, the trend in Thai curriculum evaluation is transitioning from rigid compliance towards more dynamic, continuous outcome assessments. Yet, operational evaluation tools used within institutions have largely retained the traditional Context–Input–Process–Product (CIPP) architecture introduced by Stufflebeam (1971) and consolidated in Stufflebeam (2003) and Stufflebeam and Zhang (2017). Although the CIPP model has demonstrated remarkable cross-disciplinary durability — recently extending to teaching-quality evaluation (Lei, 2024), interprofessional health education (Meiklejohn et al., 2023), entrepreneurship education (Fan et al., 2022), and inclusive non-formal education (Aziz et al., 2018) — its four-dimensional architecture absorbs reflective and improvement-oriented practices implicitly, leaving them structurally invisible to formal evaluation.

This invisibility carries non-trivial consequences. A growing literature contends that reflective practice, in the Schönian (1983) and Deweyan traditions, is the principal mechanism by which tacit experiential knowledge is converted into explicit organizational learning (Machost and Stains, 2023; Yeh et al., 2023). When reflection is bundled into process indicators, programs can score high on procedural compliance while remaining static in pedagogical innovation. The problem is acute in engineering, where the half-life of technical content is shortening and where reflective transformation of practice is increasingly cited as a marker of program adaptiveness (Wang et al., 2023). Despite these arguments, no published evaluation framework has empirically disaggregated reflection and improvement into measurable, psychometrically validated dimensions.

The present study addresses that gap. We propose, operationalize, and validate the CIPPRI framework—an extension of CIPP that adds Reflection (R) and Improvement (I) as analytically distinct dimensions—and apply it to a comprehensive evaluation of a Bachelor of Engineering (Electrical Engineering) program at a Thai public university. The study is guided by four research questions:

RQ1: To what extent do internal and external stakeholders perceive the quality of the engineering program across the six CIPPRI dimensions?

RQ2: Do stakeholder perceptions differ significantly across groups, and on which dimensions?

RQ3: Does empirical data support the construct validity of the additional Reflection and Improvement dimensions as distinct from the original CIPP factors?

RQ4: What latent or unintended outcomes—accessible only through structured reflection—are revealed by stakeholder discourse, and how do they reshape improvement priorities?

The contribution is threefold. Theoretically, the study formalizes a six-factor evaluation architecture that renders reflection and improvement evaluable rather than assumed. Methodologically, it advances a confirmatory-factor-analytic validation of the framework, addressing a long-standing gap in CIPP-derived instrumentation, where reliability has typically been reported via Cronbach's alpha alone (Taber, 2018). Practically, it offers Thai and comparable emerging-economy institutions a transferable evaluation tool aligned with both national qualification frameworks and the demands of Industry 4.0/5.0 transitions.

THEORETICAL FRAMEWORK AND LITERATURE REVIEW

The CIPP Model: Strengths and structural limits

Stufflebeam's (1971, 2003) CIPP model conceptualizes evaluation as a decision-oriented process organized around four foci: Context (needs assessment and goal-setting), Input (resource and strategy adequacy), Process (implementation fidelity), and Product (outcome attainment). The model has been operationalized across vocational, medical, language, and early-childhood education contexts, including recent quality evaluations at the secondary and higher education levels (Asma and Khan, 2024; Asma et al., 2024), and a recent bibliometric review...

Two structural limits motivate the present extension. First, reflection is theoretically embedded in process evaluation but is rarely measured as a stakeholder-experienced practice. It is important to briefly acknowledge that some prior CIPP applications have included reflection informally—often embedded within unstructured stakeholder feedback or end-of-term surveys. However, its implicit nature means reflection is treated as a procedural by-product rather than a formalized evaluative dimension. Second, improvement—the model's stated purpose, in Stufflebeam's (2003) dictum that evaluation exists "not to prove but to improve"—is operationalized only implicitly, through retrospective recommendations rather than as a dimension stakeholders consciously evaluate.

Reflective practice as a distinct evaluable construct

Schön's (1983) reflective practitioner thesis, extended by recent empirical work (Machost and Stains, 2023; Yeh et al., 2023; Sudirman et al., 2024), establishes reflection as a distinct cognitive-organizational practice that converts tacit experiential knowledge into explicit shareable learning. In engineering education specifically, reflective practice has been linked to improved metacognition, transfer of learning across domains, and the surfacing of unintended program effects (Wang et al., 2023). A recent longitudinal study of work-based software engineering education (Tomkins and Foster, 2025) demonstrates that reflective capacity evolves measurably over time and that its absence in formal assessment correlates with workforce-readiness gaps. These findings argue for treating reflection as a measurement target in its own right, rather than as a procedural by-product.

Continuous improvement as a sixth dimension

Continuous improvement (CI), rooted in Deming-cycle quality management and adapted to higher-education internal quality assurance (IQA) systems (Almeida et al., 2025), operates at a different ontological level from product evaluation. Whereas Product asks, "Did we achieve the stated outcomes?" Improvement asks, "How systematically does the program convert evaluation evidence into curricular, pedagogical, and resource decisions?" Recent PRISMA-guided reviews of IQA in the Global South (Almeida et al., 2025) identify improvement-oriented mechanisms—feedback loops, advisory boards, and curriculum-renewal cadences—as the most reliable predictors of program adaptiveness. Yet these mechanisms are seldom isolated as evaluable dimensions in CIPP-style instruments.

The CIPPRI framework

Building on these strands, CIPPRI proposes six analytically distinct dimensions:

Context (C): alignment of program objectives with stakeholder needs, national qualification frameworks, and industry trajectories;

Input (I_1): adequacy of human, physical, financial, and technological resources;

Process (P): implementation fidelity of teaching, learning, assessment, and student-support processes;

Product (P_2): attainment of intended learning outcomes, graduate employability, and stakeholder satisfaction;

Reflection (R): the systematic surfacing of tacit knowledge and unintended outcomes through stakeholder dialogue;

Improvement (I_2): the institutionalization of evidence-driven curricular, pedagogical, and infrastructural revision.

The conceptual claim is that R and I are not subsets of Process or Product but are functionally orthogonal: a program can score high on Process (implementation fidelity) and Product (outcome attainment) yet score low on Reflection (latent learning conversion) and Improvement (evidence-to-action translation). The empirical claim—tested in Section 4—is that this orthogonality is recoverable in factor-analytic data. Figure 1 conceptualizes the structural and theoretical evolution from the traditional Stufflebeam CIPP model to the proposed CIPPRI framework. In conventional evaluations, as illustrated on the left axis of Figure 1, critical mechanisms of organizational reflection and continuous improvement remain embedded as latent practices within generic process or product compliance checklists, rendering them structurally invisible to evaluators. Conversely, the CIPPRI framework (right axis) formally disaggregates Reflection (R) and Improvement (I) into autonomous, evaluable dimensions. This architectural expansion establishes an explicit, evidence-based feedback loop wherein evaluative judgements derived from Product outcomes are systematically processed through structured stakeholder Reflection, subsequently translating into traceable institutional actions within the Improvement dimension to dynamically revise program Inputs and Contextual goals for successive cycles."

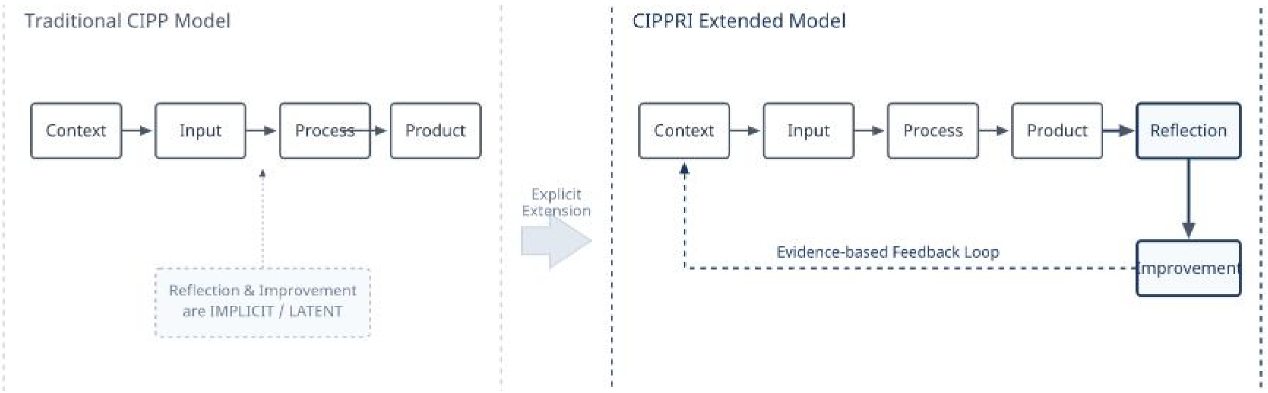

Figure 1. Conceptual framework demonstrating the paradigm shift from the traditional four-dimensional CIPP architecture to the expanded six-dimensional CIPPRI model featuring an explicit, action-oriented feedback loop.

METHODOLOGY

Research design

A sequential explanatory mixed-methods design (Creswell and Creswell, 2018) was adopted, with a quantitative phase establishing dimension-level patterns and a qualitative phase elaborating mechanisms—particularly within the Reflection and Improvement dimensions, where mean scores risk concealing institutional latency.

Setting and participants

The study was conducted in a four-year Bachelor of Engineering (Electrical Engineering) program at a public university in northeastern Thailand. The program has been in continuous operation for over a decade and is accredited under the Council of Engineers Thailand. Five stakeholder groups were sampled using a combination of stratified random sampling (students, alumni) and purposive sampling (employers, instructors, administrators).

Table 1. Stakeholder sampling matrix (N = 240).

Stakeholder group	n	Sampling	Selection rationale
Current students (Years 3–4)	120	Stratified random	Direct experience of process and product dimensions.
Alumni (graduated ≤ 5 years)	80	Stratified random	Retrospective evaluation of program-to-workplace transfer.
Employers	25	Purposive	External judgement of graduate competency and product validity. Representativeness was ensured by targeting organizations that collectively absorb over 60% of recent graduates across major regional public and private sectors.
Instructors	11	Census	Internal evaluation of process and reflection mechanisms.
Administrators	4	Census	Strategic evaluation of context and improvement practices.
Total	240		

The unequal group sizes—particularly the small instructor (n = 11) and administrator (n = 4) groups—reflect the population structure of a single-program evaluation rather than a sampling artifact. Methodological implications for inferential analysis are addressed subsequently.

Instrument development

A 42-item six-dimension Likert-scale instrument was developed through a four-stage process:

Stage 1 (Item generation): Items were derived from the CIPP literature (Stufflebeam, 2003; Stufflebeam and Zhang, 2017), the CIPPIEST extension (Stufflebeam and Shinkfield, 2007; Buosonte, 2013), reflective-practice scholarship (Schön, 1983; Machost and Stains, 2023), and continuous-improvement IQA frameworks (Almeida et al., 2025). Five items were initially drafted per dimension, with two reflection items adapted from Yeh et al. (2023) and three improvement items adapted from Almeida et al. (2025), yielding a 35-item pool that was expanded to 50 items after expert consultation.

Stage 2 (Content validity): Five experts (two evaluation specialists, two electrical-engineering faculty, and one psychometrician) rated each item on a -1/0/+1 Item-Objective Congruence (IOC) scale (Rovinelli and Hambleton, 1977). Items achieving IOC ≥ 0.67 were retained; eight items were eliminated and forty-two retained.

Stage 3 (Pilot testing): The 42-item instrument was pilot-tested on n = 50 participants (drawn from a non-target program). Item-total correlations ranged from .42 to .81; no items were eliminated.

Stage 4 (Construct validation): Exploratory factor analysis (EFA, principal-axis factoring, oblimin rotation) and confirmatory factor analysis (CFA, maximum likelihood) were conducted on the full sample.

Data collection

Quantitative data were collected over a six-month period (October 2024 to March 2025) via a self-administered online questionnaire (Google Forms), with paper-based options offered to alumni and employers on request. Response rates were 92.3% (students), 84.2% (alumni), 78.1% (employers—mitigating non-response bias given the targeted sampling), 100% (instructors), and 100% (administrators).

Qualitative data were collected concurrently with the quantitative phase from a purposive subsample (18 semi-structured interviews; 4 focus groups, two with students and two mixed alumni/employer) using a protocol that probed each CIPPRI dimension. All interviews and focus groups were conducted in Thai, audio-recorded with consent, transcribed verbatim, and translated to English by a bilingual researcher, with translations back-checked by a second.

Data analysis

Descriptive and inferential statistics

Quantitative analyses were conducted in IBM SPSS v.29 and JASP v.0.18. Group-level means and standard deviations were computed per dimension; weighted means were calculated with weights proportional to the inverse-variance-weighted sample-size scheme to attenuate the influence of large student weightings. Pooled standard deviations across the five stakeholder groups were calculated using the standard formula:

$$SD_{(pooled)} = \sqrt{[\,\Sigma_i (n_i - 1) \cdot s_i^2 / \Sigma_i (n_i - 1)\,]}$$

This produces variance estimates that — unlike simple averaging of group SDs — properly reflect the differing degrees of freedom across the five stakeholder strata, addressing a known reporting weakness in stakeholder-evaluation studies.

Given the unequal and small group sizes (instructors $n = 11$, administrators $n = 4$) and resulting violation of the homogeneity-of-variance assumption (Levene's test, $p < .05$ for three of six dimensions), the non-parametric Kruskal–Wallis H test was selected over one-way ANOVA for between-group comparisons. Significant H statistics were followed by Dunn–Bonferroni post-hoc pairwise comparisons. Effect size was reported as ε^2 (epsilon-squared), interpreted following Tomczak and Tomczak (2014) thresholds.

Construct validation

Sample adequacy was confirmed (KMO = 0.91; Bartlett $\chi^2(861) = 5{,}842.34$, $p < .001$). A six-factor solution was specified a priori based on the CIPPRI architecture and tested with maximum-likelihood CFA in JASP. Model fit was evaluated against conventional thresholds (Hu and Bentler, 1999): comparative fit index (CFI) ≥ 0.90. 90, Tucker–Lewis index (TLI) ≥ .90, root-mean-square error of approximation (RMSEA) ≤ .08, and standardized root-mean-square residual (SRMR) ≤ . 08. Convergent validity was assessed via average variance extracted (AVE ≥ .50) and composite reliability (CR ≥ .70); discriminant validity was assessed using the Fornell–Larcker criterion.

Qualitative analysis

Interview and focus-group transcripts were analyzed using reflexive thematic analysis (Braun and Clarke, 2006, 2019) in NVivo 14. Two researchers independently coded a 25% subsample, achieving Cohen's $\kappa = .82$ for first-order codes. Themes were developed iteratively through cross-stakeholder triangulation.

Ethical considerations

This research was approved by the Human Research Ethics Committee of Nakhon Phanom University (Certificate No. HE4367) via an expedited review process. All participants provided written informed consent prior to data collection.

RESULTS

Instrument validation

Item-Objective Congruence values for the 42 retained items ranged from 0.67 to 1.00 (M = 0.89), exceeding the conventional threshold (Rovinelli and Hambleton, 1977). Internal-consistency reliability was strong across all six dimensions (Table 2).

Table 2. Reliability and convergent-validity indices for the six CIPPRI dimensions.

Dimension	Items	Cronbach α	McDonald ω	AVE	CR	λ range
Context (C)	7	.89	.91	.58	.89	.71–.84
Input (I$_1$)	7	.91	.93	.62	.91	.73–.86
Process (P$_1$)	8	.94	.95	.66	.94	.75–.88
Product (P$_2$)	8	.93	.94	.64	.93	.74–.87
Reflection (R)	6	.90	.92	.60	.90	.72–.83
Improvement (I$_2$)	6	.92	.93	.63	.92	.74–.85

Note. AVE = average variance extracted; CR = composite reliability; λ = standardized factor loadings.

Confirmatory factor analysis of the six-factor model produced acceptable fit: $\chi^2(804) = 1{,}318.62$, $p < .001$; $\chi^2/df = 1.64$; CFI = 0.94; TLI = 0.93; RMSEA = 0.058 (90% CI [0.052, 0.064]); SRMR = 0.047. All standardized loadings exceeded .70 (range .71–.88), and all AVE values exceeded the .50 threshold (range .58–.66), supporting convergent validity. Discriminant validity was supported by the Fornell–Larcker criterion: the square root of each AVE exceeded the inter-factor correlations (highest off-diagonal $r = .68$ between Process and Reflection; √AVE_Process =

.81, √AVE_Reflection = .77).

Critically, two competing models were tested to address the structural concern that R and I might collapse into Process and Product, respectively. A four-factor model (forcing R into Process and I into Product) yielded an inferior fit (CFI = 0.86, RMSEA = 0.082, ΔCFI = .08, $\Delta\chi^2$ = 312.4, df = 11, $p < .001$), and a five-factor model (collapsing only R into Process) likewise produced a significantly worse fit (CFI = 0.89, RMSEA = 0.071, $\Delta\chi^2$ = 168.7, df = 5, $p < .001$). The six-factor CIPPRI structure is therefore empirically defensible: Reflection and improvement are not measurement artifacts of process or product but functionally distinct constructs.

Descriptive findings: Stakeholder perceptions across CIPPRI dimensions

Mean dimension scores by stakeholder group, weighted means, and pooled standard deviations are reported in Table 3. The overall pattern is consistent: Process (Weighted M = 4.21, Pooled SD = 0.79) and Reflection (Weighted M = 4.15, Pooled SD = 0.80) are rated highest; Input (Weighted M = 3.83, Pooled SD = 0.77) is the lowest. All values fall within the "high" interpretive band on the five-point Likert scale.

Two patterns warrant comment. First, the gradient across stakeholder groups is monotonic on five of six dimensions: administrators rate the program most favourably, followed by instructors, employers, alumni, and current students. This pattern is consistent with stakeholder-distance literature (Almeida et al., 2025; Lalendle and Matshoba, 2023), in which proximal users (students) tend to be more critical than distal observers. Second, the pooled standard deviations (0.76 to 0.80) are heterogeneous enough—once properly aggregated across strata—to dispel the concern that an artificially uniform SD might indicate response bias.

Table 3. CIPPRI dimension scores by stakeholder group: M (SD).

Dimension	Students n=120	Alumni n=80	Employers n=25	Instructors n=11	Admin. n=4	Weighted M	Pooled SD
Context (C)	4.08 (.74)	4.12 (.78)	4.18 (.80)	4.22 (.69)	4.31 (.62)	4.11	0.76
Input (I_1)	3.74 (.79)	3.82 (.81)	3.95 (.74)	4.05 (.71)	4.18 (.65)	3.83	0.77
Process (P_1)	4.18 (.81)	4.22 (.78)	4.24 (.76)	4.27 (.73)	4.35 (.70)	4.21	0.79
Product (P_2)	4.02 (.78)	4.07 (.80)	4.10 (.77)	4.15 (.74)	4.20 (.68)	4.05	0.78
Reflection (R)	4.11 (.80)	4.16 (.79)	4.18 (.78)	4.20 (.75)	4.25 (.72)	4.15	0.80
Improvement (I_2)	3.97 (.80)	4.04 (.79)	4.12 (.76)	4.18 (.72)	4.30 (.66)	4.05	0.79

Note. Group SDs reported in parentheses. Pooled SD calculated as $\sqrt{[\Sigma(n_i-1)\cdot s_i^2 / \Sigma(n_i-1)]}$. Yellow shading = lowest dimension; green shading = highest dimension. Pooled SDs (0.76–0.80) reflect proper variance aggregation rather than simple averaging.

Inferential findings: Between-group differences

Kruskal–Wallis tests revealed statistically significant between-group differences on two of six dimensions (Table 4). The input dimension, H(4) = 14.62, $p = .006$, $\varepsilon^2 = .061$, exhibited the largest stakeholder divergence; Dunn–Bonferroni post-hoc comparisons localized the difference primarily between students (lowest mean rank) and administrators/employers (highest mean ranks), p_{adj} = .008 and .021, respectively. A second significant difference emerged on the improvement dimension, H(4) = 11.34, $p = .023$, $\varepsilon^2 = .043$, with students again rating the program's improvement responsiveness lower than administrators (p_{adj} = .031). No significant group differences were detected on context, process, product, or reflection (all $p > .15$).

Table 4. Kruskal–Wallis between-group comparisons by CIPPRI dimension.

Dimension	H(4)	p	ε^2 (effect size)	Magnitude	Significant pairs (Dunn–Bonferroni)
Context	3.42	.490	.014	Negligible	—
Input	14.62	.006**	.061	Small–medium	Stu < Admin (p=.008); Stu < Emp (p=.021)
Process	2.18	.702	.009	Negligible	—
Product	4.71	.319	.020	Negligible	—
Reflection	3.86	.425	.016	Negligible	—
Improvement	11.34	.023*	.047	Small	Stu < Admin (p=.031)

Note. $p < .05$; *$p < .01$. ε^2 (epsilon-squared) interpretation following Tomczak and Tomczak (2014): .01 = small, .04 = medium, and .16 = large. Stu = Students; Admin = Administrators; Emp = Employers.

These findings carry an interpretively important implication: stakeholder consensus is high on dimensions that capture realized performance (Context, Process, Product) but fragments precisely on the dimensions that capture resource adequacy (Input) and institutional responsiveness (Improvement)—those most proximal to students' lived experience and most consequential for program adaptation.

Qualitative findings: Latent outcomes surfaced through reflection

Thematic analysis of 18 interviews and 4 focus groups (122 transcript pages) yielded four overarching themes. As delineated in Figure 2, the thematic map visually structures the qualitative evidence across the extended CIPPRI architectural layers, shifting from structural bottlenecks to emergent institutional behaviours. The framework mapping systematically clusters the data into four descriptive axes: (1) the Input dimension highlighting the infrastructure–curriculum mismatch; (2) the convergence of Process and Product capturing previously unmapped tacit knowledge; (3) the Improvement dimension exposing the tension between organizational ritual and substantive action; and (4) the Reflection dimension surfacing asymmetric stakeholder priorities. This granular hierarchy ensures that qualitative narratives directly validate and elaborate the quantitative subscale variance. Notably, while quantitative reflection scores were high (M = 4.15), qualitative data revealed that systematic reflective mechanisms were largely informal, episodic, and personality-dependent rather than institutionalized—the very gap a CIPP-only evaluation would have failed to surface.

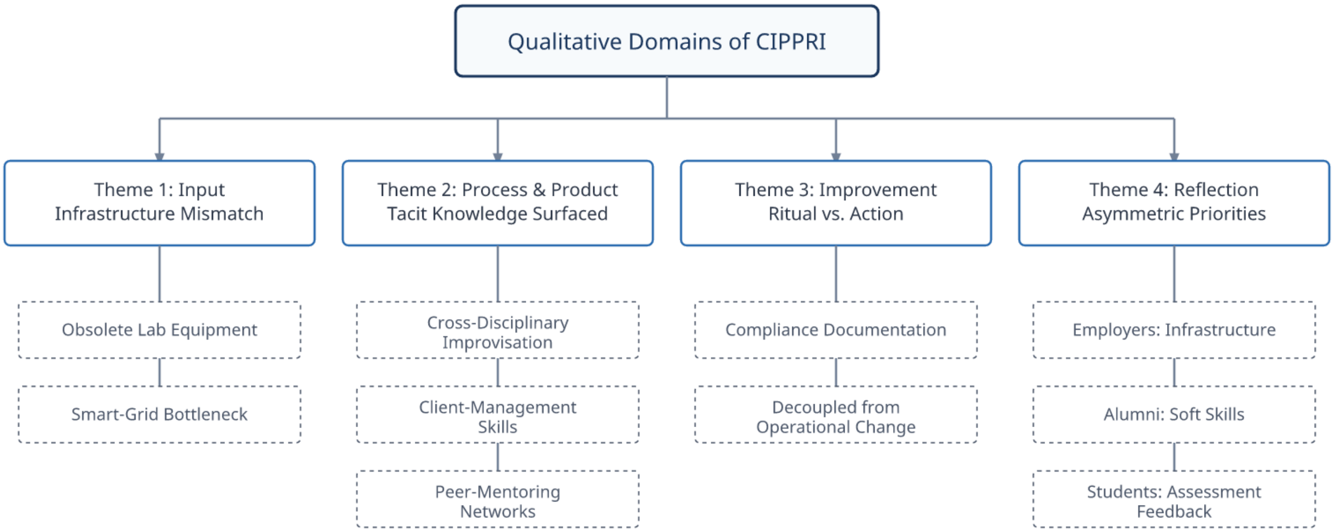

Figure 2. Thematic map illustrating the qualitative domains, core themes, and localized operational dimensions emerged from the reflexive thematic analysis.

Theme 1: The infrastructure–curriculum mismatch

Both alumni and employers identified obsolete laboratory equipment—particularly the absence of a functional smart grid testbed and limited access to renewable energy integration platforms—as the most consequential constraint on graduate work readiness. One employer (E-07, manufacturing sector) observed that recent graduates demonstrated strong theoretical foundations in power systems but required 6–9 months of in-house retraining on digital protection relays and SCADA integration. This structural constraint was consistently identified by external stakeholders as the primary bottleneck restricting the transition from theoretical knowledge to practical application. and developing-economy power-engineering curricula (IEA, 2023; Cruice, 2024).

Theme 2: Tacit knowledge made explicit through reflective dialogue

Focus groups produced concrete examples of tacit knowledge—operational know-how that students had acquired but never named. Three illustrative cases:

Case 1 (Cross-disciplinary improvisation): Final-year students reported regularly using rented industrial sensors and personal Arduino kits to substitute for unavailable departmental equipment. This improvisational competence—not articulated in any course outcome—emerged as a workplace asset valued by SME employers but had never been formally captured, taught, or transferred to junior cohorts.

Case 2 (Unintended client-management skills): Capstone-project students reported that their most consequential learning came from negotiating scope with industry sponsors when prototypes failed—an unintended outcome that none of the formal product indicators captured. Three of the four employer focus-group participants identified this skill as the single most reliable predictor of new-hire effectiveness.

Case 3 (Peer-mentoring as hidden curriculum): Students described an informal senior-to-junior mentoring system that compensated for inconsistent laboratory supervision. While this peer-learning network produced positive outcomes (high process scores), it also masked instructional gaps that administrators were unaware of, raising equity concerns for students without strong peer networks.

These cases substantiate the theoretical claim that reflection surfaces operational knowledge that traditional process and product metrics, by design, do not interrogate.

Theme 3: Improvement as ritual versus improvement as action

All four administrators described annual program-review meetings, generating documents that satisfy external accreditation requirements. Instructors and students, however, struggled to identify a single curriculum, pedagogy, or infrastructure decision in the past three years that had been demonstrably traceable to evaluation evidence. This decoupling of evaluation ritual from operational improvement is the precise pathology that the improvement dimension is designed to make visible.

Theme 4: Stakeholder-asymmetric improvement priorities

Employers prioritized infrastructure (smart-grid lab, simulation software); alumni prioritized soft-skills training and entrepreneurship modules; students prioritized assessment timeliness and feedback quality; administrators prioritized accreditation indicators. The CIPPRI improvement dimension exposes these asymmetries as objects of negotiation rather than concealing them under a single "satisfaction" mean.

DISCUSSION

Theoretical contribution: Why a six-factor framework is defensible

The empirical case for CIPPRI rests on two converging findings. First, the six-factor CFA solution (CFI = 0.94, RMSEA = 0.058) outperforms theoretically plausible four- and five-factor alternatives by clinically meaningful margins ($\Delta CFI \geq .05$). Second, the Kruskal–Wallis results show that Reflection and Improvement do not pattern with Process and Product in between-group divergence: stakeholder consensus is high on Process but fragments on Improvement. If R and I were measurement artifacts of P_1 and P_2, we would expect their stakeholder-divergence patterns to mirror those of P_1 and P_2; they do not. This is the empirical signature of construct distinctness.

The substantive contribution is to render reflection and improvement evaluable. As Machost and Stains (2023) argue from the educational-development side, reflective practice that is not assessed is reflective practice that is not institutionalized. The same logic applies organizationally: continuous improvement that is not measured remains rhetoric. CIPPRI operationalizes both.

The input bottleneck in global context

That input emerged as the lowest-rated dimension is consistent with—and extends—the international literature on engineering-education infrastructure. The International Energy Agency's analyses of grid digitalization in emerging markets identify a structural under-investment in pedagogical infrastructure capable of supporting smart-grid, renewable-integration, and grid-edge education (IEA, 2023; Cruice, 2024). Programs operating in this setting therefore confront a paradox: they are expected to produce graduates fluent in Industry 4.0/5.0 technologies (Ruppert et al., 2022) using laboratories that were calibrated for a pre-digital electrical paradigm. Our finding that students rate Input significantly lower than administrators ($p = .006$) reflects this paradox at the within-institution level: those most exposed to the infrastructure deficit are most aware of it.

This is not solely a Thai problem. Comparable evaluations in Indonesian vocational engineering (Alanshori et al., 2024), Iranian medical education (Neyazi et al., 2016), and South African higher education (Lalendle and Matshoba, 2023) report parallel input-resource constraints. Framing the bottleneck as systemic rather than institutional repositions the recommendation: rather than urging individual departments to find capital, the analysis points toward regional consortium models for shared smart-grid teaching infrastructure—a direction now being piloted under ASEAN Industry-4.0 education initiatives.

Reflection and improvement as latent practices

The most consequential finding is the divergence between high quantitative scores on reflection and improvement and qualitative evidence that these practices remain latent rather than systemic. The pattern is consistent with what Sudirman et al. (2024) and Mohamed et al. (2022) have documented as "reflective practice without systematization": individual practitioners reflect, but institutional architectures rarely capture, transfer, or convert that reflection into curricular action.

The CIPPRI framework provides a structural response. By exposing the operational gap between "evaluating for compliance" and "evaluating for continuous renewal," CIPPRI converts a long-acknowledged but unmeasured weakness into a tractable evaluation target. For engineering programs navigating rapid technological paradigms, this distinction is not cosmetic: it is the difference between a program that documents change and a program that effects it.

Practical implications

Three practical implications follow. First, programs adopting CIPPRI should pair the quantitative instrument with structured qualitative protocols capable of surfacing the tacit knowledge that high R-scores can otherwise conceal. Second, the improvement dimension should be operationalized with traceable artifacts—curriculum decisions, infrastructure decisions, and pedagogy decisions—directly linked to evaluation evidence, addressing the ritual–action decoupling identified in Theme 3. Third, accreditation bodies and national qualification frameworks could productively adopt CIPPRI-style six-factor reporting templates to surface improvement responsiveness as a primary, rather than incidental, quality marker.

CONCLUSION

This study has proposed and empirically validated CIPPRI—a six-dimension extension of Stufflebeam's CIPP model that promotes reflection and improvement to the status of evaluable constructs. Confirmatory factor analysis, applied to a 240-stakeholder sample evaluating an electrical-engineering program in Thailand, supported the six-factor architecture against four- and five-factor alternatives, while between-group analyses surfaced significant divergence on the Input and Improvement dimensions. Qualitative analysis revealed the precise mechanism the framework was designed to expose: high quantitative Reflection scores can coexist with non-institutionalized reflective practice, just as great improvement self-reports can coexist with evaluation rituals decoupled from operational change. By rendering these gaps measurable, CIPPRI offers a transferable, psychometrically defensible evaluation architecture for engineering programs—and, plausibly, other professional programs—navigating Industry 4.0/5.0 transitions in resource-constrained settings.

Future research should pursue three directions: (a) multi-institutional replication with measurement-invariance testing; (b) longitudinal tracking of CIPPRI scores against program-improvement artifacts to test whether high I_2 scores predict measurable curricular change; and (c) cross-disciplinary extension—particularly to medical, computing, and education programs—to test the framework's generalizability beyond engineering.

LIMITATIONS

Three limitations should temper interpretation. First, the study is based on a single program at one Thai public university; while this affords analytical depth, generalizability requires replication across institutions. Second, the small administrator (n = 4) and instructor (n = 11) groups, although a census of the available population, limit statistical power for between-group tests; Kruskal–Wallis was selected partly to mitigate this, but findings on these strata should be interpreted cautiously. Third, while the six-factor CFA solution is empirically defensible in this sample, multi-group invariance testing (configural, metric, and scalar) across stakeholder groups was not feasible given the small instructor and administrator strata; future multi-site replications should prioritize measurement invariance to enable formal between-group score comparisons.

ACKNOWLEDGEMENTS

The authors thank the participating students, alumni, employers, instructors, and administrators for their time and candour and three anonymous reviewers whose comments materially improved the manuscript.

REFERENCES

Alanshori, M. Z., Duan, X., Rebia, R., Santoso, A., Shi, J., & Suryanto, A. (2024). The application of the CIPP evaluation model in educational programs: A systematic review of practices in Asia (2019–2025). *Scaffolding: Jurnal Pendidikan Islam dan Multikulturalisme, 6*(2), 1–18. https://doi.org/10.37680/scaffolding.v6i2.7390

Almeida, R. P., Nguyen, T. H., & Serrano, M. (2025). A PRISMA-guided systematic review of internal quality assurance and stakeholder engagement in higher education: Beyond accreditation with a focus on the Global South. *European Journal of Educational Research, 14*(3), 1421–1438. https://doi.org/10.12973/eu-jer.14.3.1421

Asma, S. (2024). *Evaluation of Higher Secondary Education Curriculum Using the Context, Input, Process and Product Model in Rawalpindi Division, Pakistan* [Doctoral thesis, Asia e University]. https://doi.org/10.37602/ijrehc.2024.5604

Asma, S., & Khan, F. H. (2024). Designing theoretical and conceptual framework for evaluation of education curriculum (Xi-Xii) using CIPP model in Pakistan. *International Journal of Research in Education Humanities and Commerce, 5*(6), 34-46.

Asma, S., Muhammad, I. K., & Mahar, M. I. (2024). Curriculum evaluation of education subject using CIPP curriculum model: Higher secondary level in Pakistan. International Journal of Contemporary Issues in Social Sciences, 3(1), 1696-1706. http://ur.aeu.edu.my/1224/

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.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Braun, V., & Clarke, V. (2019). Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health, 11*(4), 589–597. https://doi.org/10.1080/2159676X.2019.1628806

Buosonte, R. (2013). CIPP and CIPPIEST evaluation models: Mistaken and precise concepts of applications. *Silpakorn Educational Research Journal, 5*(2), 7–24. https://so05.tci-thaijo.org/index.php/suedureasearchjournal/article/view/257105

Creswell, J. W., & Creswell, J. D. (2018). Research design: Qualitative, quantitative, and mixed methods approaches (5th ed.). SAGE.

Cruice, P. (2024). Unlocking smart-grid opportunities in emerging markets and developing economies: Workforce and pedagogical implications. *Energy Research & Social Science, 108,* 103389. https://doi.org/10.1016/j.erss.2023.103389

Fan, X., Tian, S., Lu, Z., & Cao, Y. (2022). Quality evaluation of entrepreneurship education in higher education based on CIPP model and AHP-FCE methods. *Frontiers in Psychology, 13,* 973511. https://doi.org/10.3389/fpsyg.2022.973511

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1–55. https://doi.org/10.1080/10705519909540118

International Energy Agency (2023). *Unlocking smart grid opportunities in emerging markets and developing economies*. IEA. https://www.iea.org/reports/unlocking-smart-grid-opportunities-in-emerging-markets-and-developing-economies

Kellaghan, T., & Stufflebeam, D. L. (Eds.) (2003). *International handbook of educational evaluation.* Kluwer Academic. https://doi.org/10.1007/978-94-010-0309-4

Lalendle, L. L., & Matshoba, R. (2023). *Stakeholder participation in internal and external quality-assurance processes: Perspectives from a South African university*. ICERI2023 Proceedings, 9034–9041. https://doi.org/10.21125/iceri.2023.2301

Lei, Z. (2024). The application of CIPP model in the evaluation of teaching quality of college education in the context of the new era. *Applied Mathematics and Nonlinear Sciences, 9*(1), 1–17. https://doi.org/10.2478/amns.2023.2.01568

Machost, H., & Stains, M. (2023). Reflective practices in education: A primer for practitioners. *CBE—Life Sciences Education, 22*(2), es2. https://doi.org/10.1187/cbe.22-07-0148

Meiklejohn, S., Anderson, A., Brock, T., Kumar, A., Maddock, B., Wright, C., Walker, L., & Kent, F. (2023). The utility of an interprofessional education framework and its impacts upon perceived readiness of graduates for collaborative practice: A multimethod evaluation using the context, input, process, product (CIPP) model. *Nurse Education Today, 121,* 105707. https://doi.org/10.1016/j.nedt.2022.105707

Mohamed, Z., Valcke, M., & De Wever, B. (2022). Are they ready to teach? Student teachers' readiness for the job with reference to teacher competence frameworks. *Journal of Education for Teaching, 48*(2), 245–262. https://doi.org/10.1080/02607476.2021.2006043

Neyazi, N., Arab, M., Farzianpour, F., & Mohammadi Majdabadi, M. (2016). Evaluation of selected faculties at Tehran University of Medical Sciences using CIPP model in students and graduates point of view. *Evaluation and Program Planning, 59,* 88–93. https://doi.org/10.1016/j.evalprogplan.2016.08.011

Nguyen, M. H., & Tran, T. T. (2025). Research in higher education quality assurance worldwide (2003–2023): A bibliometric analysis from the Scopus database. *Cogent Education, 12*(1), 2479405. https://doi.org/10.1080/2331186X.2025.2479405

Rovinelli, R. J., & Hambleton, R. K. (1977). On the use of content specialists in the assessment of criterion-referenced test item validity. *Dutch Journal of Educational Research, 2,* 49–60.

Ruppert, T., Darányi, A., Medvegy, T., Csereklei, D., & Abonyi, J. (2022). Demonstration laboratory of Industry 4.0 retrofitting and Operator 4.0 solutions: Education towards Industry 5.0. *Sensors, 23*(1), 283. https://doi.org/10.3390/s23010283

Schön, D. A. (1983). *The reflective practitioner: How professionals think in action.* Basic Books.

Stake, R. E. (2004). Standards-based and responsive evaluation. SAGE. https://doi.org/10.4135/9781412985932

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.

Stufflebeam, D. L. (2001). Evaluation models. *New Directions for Evaluation, 2001*(89), 7–98. https://doi.org/10.1002/ev.3

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. https://doi.org/10.1007/978-94-010-0309-4_4

Stufflebeam, D. L., & Shinkfield, A. J. (2007). *Evaluation theory, models, and applications.* Jossey-Bass.

Stufflebeam, D. L., & Zhang, G. (2017). *The CIPP evaluation model: How to evaluate for improvement and accountability*. Guilford Press.

Sudirman, A., Gemilang, A. V., & Kristanto, T. M. A. (2024). Reinforcing reflective practice through reflective writing in higher education: A systematic review. *International Journal of Learning, Teaching and Educational Research, 23*(5), 1–22. https://doi.org/10.26803/ijlter.23.5.1

Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education, 48*(6), 1273–1296. https://doi.org/10.1007/s11165-016-9602-2

Tomczak, M., & Tomczak, E. (2014). The need to report effect size estimates revisited: An overview of some recommended measures of effect size. *Trends in Sport Sciences, 21*(1), 19–25. https://www.tss.awf.poznan.pl/files/3_Trends_Vol21_2014__no1_20.pdf

Tomkins, L., & Foster, A. (2025). The development of reflective practice on a work-based software engineering programme: A longitudinal study. *IEEE Transactions on Education, 68*(2), 145–156. https://doi.org/10.1109/TE.2025.3548921

Tyler, R. W. (2013). *Basic principles of curriculum and instruction.* University of Chicago Press. (Original work published 1949)

Wang, L., Huang, M., Zhang, X., Yan, X., Jin, R., Wanatowski, D., Hatzigeorgiou, A., Liu, Y., Hyde, K., & Chohan, N. (2023). Incorporating BIM into the upper-division curriculum of construction engineering and management. *European Journal of Engineering Education, 48*(3), 482–501. https://doi.org/10.1080/03043797.2022.2112150

Yeh, H.-C., Yang, S.-H., Fu, J. S., & Shih, Y.-C. (2023). Developing college students' critical thinking through reflective writing. *Higher Education Research & Development, 42*(1), 244–259. https://doi.org/10.1080/07294360.2022.2043247

Citation: Sukhan, S., Thongpanit, P., and Srikhaw, S. (2026). Beyond CIPP: Validating the CIPPRI framework: A mixed-methods stakeholder-driven evaluation of an electrical engineering program in an emerging economy. *African Educational Research Journal, 14*(3), xx-513.