JURNAL CINTA NUSANTARA
Volume 04 Nomor 04 2026
E-ISSN: 3025-4469
https://jurnalbundaratu.org/journal/index.php/cintanusantarajournal
Lisensi Creative Commons Atribusi 4.0 Internasional.

EVALUATING THE IMPLEMENTATION OF A CHARACTER EDUCATION PROGRAM USING THE CIPP MODEL: A CASE STUDY IN AN INDONESIAN ELEMENTARY SCHOOL

Nuraini[1], Bunyamin[2], Fetrimen[3]*
[1,2,3] *Master's Program in Educational Administration, Postgraduate School, Universitas Muhammadiyah Prof. DR. HAMKA, Jakarta, Indonesia*
**Corresponding author: fetrimen@uhamka.ac.id*

Abstract

This study evaluates the implementation of the character education program at an Indonesian elementary school that has run for more than five decades without ever being formally evaluated. Using the CIPP (Context, Input, Process, Product) evaluation model, the study aims to determine the alignment of the program with the school vision, the readiness of supporting resources, the consistency of implementation, and the resulting changes in student behavior. A qualitative single-case study design grounded in the constructivist paradigm was employed. Twenty-nine informants across eight purposively selected clusters, comprising the principal, vice-principals, teachers, a counselor, non-teaching staff, parents, and students, were interviewed, observed, and their related documents analyzed. Data were examined using the Miles and Huberman interactive model and verified through triangulation, negative case analysis, and member checking. The findings show that the Context dimension was partially aligned, Input was reasonably prepared, Process was reasonably effective, and Product was partially aligned. The program's strength lies in consistent daily habituation and teachers' exemplary conduct, while its weakness concentrates on formal documentation, namely unstandardized attitude-assessment rubrics and a school-family partnership that remains underdeveloped. The school's character culture has grown faster than its documentation and measurement systems, indicating that the program should be continued with targeted improvements rather than discontinued.

Keywords: *program evaluation; character education; CIPP model; elementary school; case study*

1. INTRODUCTION

Character formation is central to Indonesia's national education goals, which aim to develop learners who are faithful, of noble character, independent, and responsible (Undang-Undang Nomor 20 Tahun 2003, 2003). However, national crime data suggest that this aspiration has yet to be fully realized. The Central Statistics Agency recorded 561,993 criminal incidents nationwide in 2024, equivalent to a crime rate of 204 per 100,000 residents (Badan Pusat Statistik, 2025). In addition, the National Police Criminal Information Center reported that, in less than two months of early 2025, hundreds of children were named as suspects in theft, assault, and narcotics cases (Pusat Informasi Kriminal Nasional Bareskrim Polri, 2025). These figures raise a long-standing question in Indonesian education policy: whether school-based character programs, many of which have operated informally for decades, actually produce the outcomes they claim.

Character itself, as a construct, has been defined across the literature as an integrated system of moral values, ranging from cognition and affect to observable moral action, that forms gradually through habituation, social interaction, and psychological development (Heidari et al., 2016). Lickona (1991) similarly frames character education as the deliberate cultivation of virtues that are good not only for the individual but for society, a formulation that anchors much of the *CIPP-based* evaluation literature reviewed below, since it treats character as something that must be intentionally designed for, not merely hoped for.

A substantial body of empirical work now supports the general premise that character

education can shape students' moral and social development. A comprehensive *meta-analysis* of character education programs found small but consistent positive effects on students' moral and ethical development across diverse program designs (Brown et al., 2023), and an earlier synthesis of over two hundred studies identified recurring features of effective practice, including adult modeling, *relationship-based* pedagogy, and sustained rather than episodic implementation (Berkowitz & Bier, 2007). In elementary and *pre-primary* settings specifically, a study across eighteen schools in Ethiopia found that parents tend to emphasize honesty and responsibility while teachers emphasize fairness and respect, with *advice-giving* and modeling as the dominant pedagogies used by both groups (Birhan et al., 2021). Religious education teachers in England likewise reported that character formation is most durable when it is embedded in daily school culture rather than delivered as a discrete curriculum unit (Metcalfe & Moulin-Stozek, 2020).

The suitability of the CIPP model for evaluating moral and *character-related* programs is not confined to the Indonesian context. A recent evaluation index system built on the CIPP framework for moral cultivation in *blended learning* courses at Chinese vocational colleges similarly demonstrated the model's capacity to structure context, input, process, and product indicators around *values-based* educational aims, lending further international support to its suitability for the present evaluation (Zhao et al., 2024).

Within Indonesia, the CIPP (Context, Input, Process, Product) evaluation model has become the dominant framework for assessing school character programs, appearing in a growing number of *institution-level* case studies. Several of these evaluations, conducted within the same university network as the present study, illustrate both the model's versatility and its recurring findings. An evaluation of the character education program at a Jakarta public elementary school found that the program had met most CIPP criteria but required a more formal implementation structure and closer *school-family* collaboration (Astuti, 2024). A study of an Islamic junior high school found that facility limitations and time constraints were the main obstacles to consistent implementation, even as the program continued to function adequately overall (Lestari, 2022). At an elementary school in South Sumatra, a CIPP evaluation similarly concluded that all four dimensions were fulfilled to a reasonable degree, with teacher readiness and school habituation programs identified as key enabling factors (Qibtiyah et al., 2026). Comparable patterns have been reported in a *project-based* character evaluation at an elementary school, where the product and process dimensions scored highest and the input dimension stood out as the clear weak point because of funding constraints (Christanti & Kusuma, 2025), and in an evaluation of a *five-value* cultural habituation program, where context was well aligned with the school's mission and stakeholder involvement was strong, even though limited funding and infrastructure constrained implementation and classroom modeling by teachers proved inconsistent (Putri & Sanoto, 2025). At the junior secondary level, a CIPP evaluation of a character program in Central Java concluded that the basis for the program was sound but that supporting resources and achievement indicators required further development (Dwi W. et al., 2018).

Despite this growing body of work, three gaps remain. *Institution-level* CIPP evaluations typically rely on teachers and students as informants, while *non-teaching* personnel such as administrative staff and security officers are rarely included, even though they interact with students throughout the school day. The alignment between a school's vision, mission, and objectives is also often assumed rather than examined, and few studies trace how each value in a vision statement is carried into policy documents. In addition, the evaluation of character education still suffers from limited measurement and *cross-validation*. A thematic review of 222 empirical studies found that validated instruments make up only 4.5% of the literature and that most intervention studies were never replicated (Oldham & McLoughlin, 2026). Similarly, the persistent underperformance of character education in Indonesian schools

has been linked to the lack of an operational and effective evaluation model (Salirawati, 2021). Consequently, few case studies combine qualitative evidence with external quantitative indicators, such as national education report data, to verify their conclusions. This study responds to these gaps by involving *non-teaching* staff as informants, mapping the school's vision against its mission and objectives through a *vertical-logic* matrix, and *cross-validating* its qualitative findings with three years of national education report data.

National-level studies reinforce the relevance of these gaps to the Indonesian context specifically. An analysis of *Pancasila-based* character education in a *pesantren-affiliated* secondary school found that the program was most effectively implemented when elaborated through both the school's vision and mission and delivered through intracurricular and extracurricular activity alike, rather than through either channel alone (Murdiono et al., 2017), a finding that anticipates the *vertical-logic* approach adopted in the present study. Separately, a conceptual analysis of the ongoing crisis in national role models argued that character education cannot succeed as an isolated school subject and instead requires *whole-institution* commitment, including from actors outside the classroom (Sidik, 2024), while a mixed *cross-sectional* study of Pancasila *student-profile* development identified inconsistent teacher understanding of character indicators as a persistent constraint even where school policy was well formulated (Zakso et al., 2022). A study of character curriculum management in a *faith-based* boarding school likewise found that the durability of character outcomes depended heavily on whether supporting staff, not only classroom teachers, understood and modeled the institution's stated values (Woenardi et al., 2022). None of these studies, however, evaluate a single program across all four CIPP dimensions while also triangulating findings against independent, *government-issued* quantitative indicators, which is the approach adopted here.

The present study addresses these gaps by evaluating the character education program of a private elementary school in West Jakarta that has operated since 1972 without ever undergoing a comprehensive evaluation. The study applies the CIPP model in full, across all four dimensions, and extends the model with a *vertical-logic* alignment analysis of the school's vision, mission, and objectives (Gasper, 2000), drawing on the view that *public-serving* organizations, including schools, must maintain a consistent logical chain from abstract vision to daily operational strategy (Bryson, 2018). The evaluation deliberately incorporates eight informant clusters, including *non-teaching* staff, parents, and students from both lower and upper grades, and triangulates *classroom-level* qualitative findings with three years of the school's national education report data. This combination of design elements distinguishes the present evaluation from the *single-cluster* or *teacher-and-student-only* designs that characterize most of the CIPP studies reviewed above, and allows findings at one CIPP stage to be *cross-checked* against findings at another, an approach that becomes central to the *cross-dimensional* synthesis presented later in this article.

This study offers empirical, methodological, and practical contributions. Empirically, it provides the first systematic evaluation of a character education program that has operated for more than five decades, offering a rare *long-duration* case within the character education literature. Methodologically, it combines the CIPP model with an explicit *vision-mission-objective* alignment matrix and a wide informant base rarely achieved in *single-institution* evaluations. Practically and in terms of policy, the findings are translated into a concrete set of recommendations for the school's management, its supervising foundation, teachers, and parents. The purpose of this study is therefore to evaluate the implementation of the school's character education program through four questions: (1) to what extent the program's design and strategy align with the school's vision and mission (Context); (2) to what extent supporting resources are ready to sustain the program (Input); (3) how the program is implemented and integrated into classroom learning and school culture, and what obstacles are encountered (Process); and (4) what results the program has produced in terms of student behavioral change

(Product).

2. METHOD

Research Design

This study employed a qualitative evaluative design within a *single-case study* framework (Yin, 2018), grounded in the constructivist paradigm, which treats the success of a character education program not as a fixed, quantifiable outcome but as a socially constructed reality shaped by the interpretations of those who enact it daily (Guba & Lincoln, 1994). The evaluation model applied was the CIPP model (Stufflebeam & Zhang, 2017), and the evaluation process was guided by the Joint Committee's program evaluation standards covering utility, feasibility, propriety, accuracy, and accountability (Yarbrough et al., 2010). A *single-case* design was chosen because the study sought an *in-depth*, holistic understanding of program implementation within one institutional context rather than statistical generalization across schools.

Research Setting and Participants

The study was conducted at a private, nationally accredited elementary school in West Jakarta, Indonesia, established in 1972 under a foundation originally formed by *tax-office* employees' spouses. In the 2025/2026 academic year, the school enrolled 523 students, employed 47 teachers, and was supported by 16 *non-teaching* staff, delivering Indonesia's national curriculum enriched with an international curriculum strand. *Twenty-nine* informants across eight *purposively* selected clusters participated in the study (Table 1), chosen because each cluster held a distinct vantage point on the program: *policy-setting* authority, technical coordination, daily classroom interaction, individualized behavioral guidance, *non-classroom* interaction, *out-of-school* reinforcement, and lived experience as program recipients.

Sampling followed a *purposive* logic rather than random selection, on the premise that depth and relevance of insight, not statistical representativeness, were required to answer the evaluation questions. For clusters occupied by a single office holder, such as the principal and the counselor, no further selection procedure was needed. For teachers, selection prioritized variation across grade levels to capture how character habituation is enacted at different developmental stages. For students, class teachers recommended *lower-grade* participants based on adequate verbal communication ability, while *upper-grade* participants were selected for their capacity to offer reflective *self-report*, since the evaluation sought *first-hand* data on character experience rather than relying solely on adult perception.

Table 1. Characteristics of Research Participants

Informant cluster	n	Rationale for inclusion
Principal	1	Holds ultimate policy authority; key informant on program direction
Vice principals	2	Student affairs (1) and public relations (1); technical coordination
Teachers (homeroom)	8	Direct daily interaction with students across lower to upper grades
Guidance and counseling teacher	1	Handles individualized behavioral and attitudinal cases
Non-teaching staff	6	Administrative head/staff, security, cleaning staff, and librarian
Parents	6	Represent grades 1, 5, and 6; perceive character change outside school
Lower-grade students	3	Grades 1-3; earliest stage of value internalization
Upper-grade students	2	Grades 4-6; capable of reflective self-report

Informant cluster	n	Rationale for inclusion
Total	29	

Source: research data (2026).

Data Collection

Primary data were collected through *in-depth* interviews and both *in-class* and *out-of-class* observations, with the researcher serving as the principal instrument, supported by *semistructured* interview and observation guides organized around the four CIPP dimensions. Each informant cluster received a distinct set of questions; teachers, for instance, were asked about *lesson-plan* integration and classroom assessment, while *non-teaching* staff were asked about *character-related* briefing and their daily interaction with students. Observations covered scheduled habituation activities, such as morning assembly and communal prayer, as well as unscheduled interactions in corridors, the canteen, and at the school gate. Secondary data were drawn from the school handbook, curriculum documents, lesson plans, *positive-culture* standard operating procedures, the code of conduct, the *conflict-mitigation* protocol, samples of the *home-school* communication diary, student report cards, and three years (2023–2025) of the national education report (*Rapor Pendidikan*) data. Fieldwork began in late July 2026 and continued as an extended observation period through August 2026, during which emerging findings were verified against further observation.

Because the researcher held a governance role within the foundation supervising the school, positional bias was addressed through source, method, and time triangulation, negative case analysis, and *member checking*. Differing accounts across informant clusters were *crosschecked* against documents, observations, and national report data rather than resolved in favor of any single source.

Data Analysis and Trustworthiness

Data were analyzed using the Miles and Huberman interactive model of data reduction, data display, and conclusion drawing and verification (Miles et al., 2014). Interview transcripts and field notes were first reduced to identify segments relevant to each CIPP dimension, then displayed in matrices organized by informant cluster and dimension to allow patterns and contradictions across clusters to be compared directly, before conclusions for each dimension were drawn and revisited as additional data were collected, consistent with the iterative character of the model. Trustworthiness was established through prolonged engagement, persistent observation, source, method, and time triangulation, negative case analysis, and *member checking*, following established criteria for naturalistic inquiry (Lincoln & Guba, 1985); the originally planned observation window was extended by several months specifically to allow the researcher to verify emerging findings against further fieldwork rather than relying on a single *data-collection* pass. *Vision-mission-objective* alignment was mapped using a *vertical-logic* matrix, and qualitative findings on program outcomes were *cross-validated* against the school's quantitative national education report scores, allowing each qualitative conclusion at the process stage to be checked against an independent *government-issued* indicator at the product stage.

Ethical Considerations

The study was conducted with institutional permission and informants' consent, and informants' identities were anonymized using *cluster-based* codes. Because the principal researcher held a governance role within the foundation supervising the school, triangulation, negative case analysis, and *member checking* were deliberately emphasized to reduce the risk of positional bias in the interpretation of findings.

3. RESULTS AND DISCUSSION

Findings across the four CIPP dimensions are summarized in Table 2 and discussed in relation to the reviewed literature below.

Table 2. Summary of CIPP Evaluation Results

Dimension	Category	Key strength	Key weakness
Context	Partially aligned	Legal basis and program objectives fully aligned; the "noble character" pillar is clearly operationalized in the mission and objectives; practice has continued since 1972	The "faith and piety" pillar of the vision is not explicitly translated into the mission or objectives
Input	Reasonably prepared	Strong positive-culture SOP and supporting regulations; representative facilities; high affective role-modeling by teachers	Lesson-plan format and attitude-assessment instruments are not standardized; non-teaching staff are not briefed or covered by a code of conduct; no dedicated budget
Process	Reasonably effective	Consistent daily-to-annual habituation; restorative discipline; all 18 recorded conflict cases fully resolved	The home-school communication diary is reduced to a signature ritual; teachers face limited time for individualized affective assessment
Product	Partially aligned	Character Index, School Safety Climate, and Diversity Climate ranked in the top national percentile (2023-2024); behavioral change corroborated across informant clusters	The 2025 Character dimension ranking declined in relative percentile; measurement instruments remain subjective; parental participation is only moderate

Source: research data (2026).

Context Evaluation

A *vertical-logic* analysis mapping the school's vision against its mission and objectives showed that the value of "noble character" is explicitly and consistently operationalized: the mission commits to instilling good conduct and noble morals and to cultivating tolerance, responsibility, independence, and emotional competence, while the school's objectives add integrity as a further indicator. By contrast, the vision's pillar of "faith and piety" is not translated into any of the school's seven mission statements or nine objectives, even though the school has operated since 1972 and religious habituation activities, such as communal prayer and short religious talks, occur daily. Two additional values, love of country and environmental care, appear in the mission and objectives but were never anticipated in the vision statement itself. Consistent with the logic advanced by Gasper (2000), a pillar of the vision that is not carried through to the mission and objectives constitutes a broken *vertical logic* that risks producing a downstream deviation at the input and process stages, because teachers lack an explicit institutional target for that value. The school's legal basis, however, was found to be fully aligned: national legislation on education, the presidential regulation on character strengthening, and the ministerial regulation on formal character education were consistently reflected in the program's objectives, and the school's founding predates these regulations by more than four decades, indicating that character practice at the school grew organically from institutional philosophy rather than administrative compliance alone.

The school's legal and historical foundation reinforces this *partial-alignment* picture. *Character-related* provisions in national, presidential, and ministerial regulations are reflected

in the program's objectives, indicating full normative alignment at the regulatory level even where internal *vision-mission* alignment remained incomplete. Because the school predates these regulations by more than four decades, its character practice cannot be read solely as administrative compliance. Rather, the habituation, teacher exemplary conduct, and school culture observed at the process stage appear to have grown organically from the institution's founding philosophy, later reinforced rather than initiated by national policy. This dual legitimacy provides a more durable foundation than compliance alone, though it does not resolve the unsupported vision pillar.

Bates (2019) argues that character education succeeds only when a young person's recognition as a moral agent is embedded in the institution's stated purpose. The absence of an explicit *faith-and-piety* pathway in the mission and objectives is consistent with this risk, as it leaves one dimension of students' moral identity institutionally *under-recognized* despite being practiced daily. However, the finding that this pillar achieved the highest measurable outcome at the product stage requires cautious interpretation. Strong daily habituation may explain this pattern, but the explanation remains conjectural and untested. Alternative explanations—such as how the *Rapor Pendidikan* instrument measures character or *cohort-specific* student characteristics—also need consideration. The finding is therefore an initial indication requiring further research, not causal evidence.

This study is one of the first systematic evaluations of a character program running for more than five decades, though this claim does not suggest that no other research exists. Several CIPP evaluations of Indonesian character programs have been conducted, but they generally focus on younger schools or do not combine *vision-mission* alignment analysis with external quantitative validation as this study does.

This *partial-alignment* finding is broadly consistent with prior CIPP evaluations of Indonesian schools, several of which report that the context dimension is generally the strongest of the four (Christanti & Kusuma, 2025; Putri & Sanoto, 2025), yet the present analysis goes further by identifying precisely which value pillar is unsupported rather than treating context alignment as a single composite judgment. This distinction matters because, as later findings on the product dimension will show, the unsupported pillar is also the one on which the school subsequently achieved its strongest measurable outcome, a pattern not previously documented in the reviewed CIPP literature on Indonesian character programs.

Input Evaluation

Policy-level readiness was found to be strong. The school publishes an annual handbook covering its profile, curriculum management, and a code of conduct for good behavior, and its curriculum explicitly lists noble character, faith and piety, and environmental care as indicators of vision attainment, with attitude scores formally required for grade promotion and graduation. The school's code of conduct functions not merely as a set of prohibitions but as a structured *behavioral-conditioning* system that assigns shared responsibility to students, parents, and the school itself. *Operational-level* readiness, however, was weaker. *Lesson-plan* formats and *attitude-assessment* rubrics were not standardized across teachers, leaving their depth and consistency dependent on individual initiative. Teachers displayed strong affective *role-modeling*, consistently demonstrating courtesy and accountability in daily interaction with students, yet their methodological capacity to design structured *value-internalization* strategies varied considerably, a pattern one may describe as a paradox between natural exemplary and limited pedagogical standardization. *Non-teaching* staff, including administrative personnel, security officers, and cleaning staff, had never been included in *vision-mission* briefings and were not covered by any written code of conduct; security personnel in particular reported never having received any *character-related* orientation. No dedicated budget line existed for character education activities, although general program funding and physical facilities, such as a place of worship and segregated waste bins, were adequate.

A closer look at the *human-resource* dimension reveals a more specific paradox than the general finding above suggests. Teachers consistently positioned themselves as natural role models through small daily acts, courteous language, admitting mistakes, punctuality, yet when asked about their pedagogical strategy for integrating values into instruction, their answers were highly individual and unstandardized, ranging from simple classroom discussion to *sticker-based* reward systems to structured *project- or problem-based* learning. Several teachers reported having received some prior training related to character curriculum implementation but regarded it as insufficient for building a consistent, *evidence-based* approach to values integration. This pattern, strong personal exemplary paired with fragmented pedagogical technique, mirrors the finding that inconsistent teacher understanding of character indicators persists even in schools with *well-formulated* policy (Zakso et al., 2022), and suggests that policy clarity alone does not guarantee pedagogical consistency. A comparable asymmetry appeared in the physical and financial dimension of input readiness: the school possessed representative facilities aligned with its character pillars, including a place of worship and *organic-inorganic* waste segregation at every corridor, yet lacked dedicated visual instructional media, such as *character-themed* posters, and had no *ring-fenced* budget line for activities like *character-themed* reading materials or periodic behavioral incentives, even though general programmatic funding was otherwise adequate. Taken together, these findings indicate that input readiness at the school is uneven not because any single resource is absent, but because *policy-level* and *facility-level* readiness has outpaced the operational instruments, training, and dedicated funding needed to make that readiness fully actionable

These findings echo several elements of the broader literature. The pattern of strong policy documentation paired with weaker operational standardization mirrors the general finding that character education initiatives with over two hundred studies reviewed tend to succeed when adult modeling is consistent and pedagogy is sustained, but tend to *under-deliver* when instruments for measuring outcomes are absent or inconsistent (Berkowitz & Bier, 2007). The exclusion of *non-teaching* staff from *character-related* briefing is a gap rarely documented explicitly in the reviewed Indonesian CIPP literature, most of which focuses evaluation on teachers and students alone; this omission is notable given that *religious-education* teachers in a separate international context reported that character formation is most durable when the entire school environment, not only the classroom, reinforces the same values (Metcalfe & Moulin-Stozek, 2020). Budget limitation as an input constraint is also consistent with prior evaluations of Indonesian character programs, where input was repeatedly the *weakest-scoring* CIPP dimension because of funding and infrastructure shortfalls even where stakeholder commitment was otherwise strong (Christanti & Kusuma, 2025; Putri & Sanoto, 2025), suggesting this is a structural rather than *institution-specific* weakness.

Process Evaluation

Character habituation was implemented in tightly layered cycles. At the daily level, students experienced a welcoming ritual at the school gate, communal prayer before and after lessons, orderly queuing, and punctuality expectations. At the weekly level, habituation was reinforced through the flag ceremony, a short literacy period, congregational Friday prayer, and a savings program. At the monthly and annual level, reinforcement continued through a dedicated character month, religious commemorations, and *community-service* camps. Teachers applied restorative rather than purely punitive discipline through negotiated classroom agreements, and all 18 recorded conflict cases were fully resolved without recurrence involving the same students, with no open conflict observed during the fieldwork period. Teachers also embedded character indicators into their lesson plans and used daily observation notes rather than *end-of-semester* recall to assess student attitudes, although they reported that abstract assessment indicators, limited observation time, and large *student-to-teacher* ratios constrained the depth of this assessment. The clearest weakness emerged in the

home-school communication diary: across six sampled diaries from three grade levels, the subject and task columns were consistently completed, but the *teacher-comment*, *parent-comment*, reflection, and *character-attitude* columns were uniformly left blank, indicating that parental and teacher engagement had been reduced to a signature ritual rather than substantive dialogue. This gap was compensated informally through class WhatsApp groups and direct phone communication, which teachers and parents both reported as more responsive than the physical diary.

The habituation cycle observed at the school was notably dense and *multi-layered* rather than a single flagship activity. Beyond the daily and weekly rituals already described, monthly and annual reinforcement included a dedicated *character-themed* month, religious commemorations, a Ramadan boarding program, a *livestock-sacrifice* activity during the Islamic festival of sacrifice, joint scouting camps, parent meetings, and *end-of-semester* class gatherings, indicating that character habituation was distributed across the full academic calendar rather than concentrated in a single term. *Non-teaching* staff, including administrative personnel and security officers, were also observed intervening spontaneously and educationally when students misbehaved in corridors or near the school gate, extending character reinforcement beyond the classroom even without formal guidance to do so, a pattern consistent with the general finding that durable character outcomes depend on *whole-institution* modeling rather than *teacher-only* reinforcement (Woenardi et al., 2022). When formal instruments proved cumbersome, actors across the school adapted pragmatically: teachers simplified their observation notes into brief daily logs, class WhatsApp groups substituted for the underused physical communication diary, and a case involving a forged parental signature was resolved through reflective moral dialogue rather than punitive sanction, illustrating a broader shift from rigid procedural administration toward flexible, restorative practice. These adaptive strategies were, however, informally sustained by individual initiative rather than codified into institutional policy, leaving their continuity dependent on the specific actors currently in place rather than guaranteed by written procedure.

The combination of strong daily habituation with a weak formal communication channel closely parallels a finding from a CIPP evaluation of a character program in an East Jakarta elementary school, where facility and time constraints, rather than a lack of commitment, were identified as the principal barrier to fully consistent implementation (Lestari, 2022). It also aligns with the broader observation that character education programs demonstrate the largest effects when implementation is sustained and embedded in routine rather than treated as an occasional activity (Brown et al., 2023). The finding on *non-teaching* staff's spontaneous, *unprompted* character modeling outside the classroom, despite lacking any formal guidance, is consistent with evidence that adult modeling, more than formal curriculum content, is among the most consistently effective mechanisms identified across the character education literature (Berkowitz & Bier, 2007).

Product Evaluation

National education report (*Rapor Pendidikan*) data for 2023–2025, obtained from the school's documents, served as an independent indicator to complement the qualitative findings. The report presents both indicator scores and the school's relative position among schools nationally (percentile band). The Character Index rose from 77.45 in 2023 to 78.84 in 2024, placing the school in the top national percentile, while the School Safety Climate and Diversity Climate indicators increased from 79.08 to 86.84 and from 80.32 to 86.23, respectively. In 2025, the Character dimension scored 61.21 ("good"), but the school's relative position fell to the *upper-middle* percentile band. Because the government revised the report's instrument and weighting in 2025, results for that year are not directly comparable with earlier years; therefore, the decline is interpreted through percentile position rather than score differences.

The 2025 *sub-indicators* reveal a notable pattern. Faith and piety, the pillar left unsupported in the mission and objectives at the context stage, scored highest and ranked in the top national percentile, whereas independence and global diversity, both stated in the mission and objectives, ranked only in the *upper-middle* percentile. This suggests that the school's *religious-character* outcome may be driven more by daily habituation, such as the congregational prayer and religious talks documented at the input and process stages, than by the clarity of its planning documents. Context and product thus diverge in the same direction, a pattern this study treats as a productive paradox within the CIPP framework.

These figures should nevertheless be interpreted with caution. As a qualitative *single-case study* without a comparison group, this study uses the national report data only as a supplementary indicator. Changes in scores or rankings cannot be attributed directly to the program, as they may also reflect other factors, such as differences between student cohorts and the revised instrument.

Qualitative evidence corroborated this pattern. The principal, *vice-principals,* the counselor, *non-teaching* staff, and four of six interviewed parents consistently reported visible behavioral change, including improved discipline, empathy, and *public-speaking* confidence; the remaining two parents attributed limited change to their children's temperament rather than the program, indicating that outcomes are not uniform across individuals and that attribution to the school program alone should be treated cautiously. *Upper-grade* students *self-reported* a shift from shyness and social hesitancy toward greater confidence and independence, while a *lower-grade* student consistently invoked internalized values, honesty, tolerance, and environmental care, when asked about everyday situations, evidence that internalization is occurring from an early stage rather than only among older, more reflective students. A *semester-long* case study documented by the school counselor traced one student's progression from repeated verbal conflict in August 2025 to sustained positive peer relationships by January 2026, offering longitudinal evidence that the program's restorative approach produces measurable *individual-level* change over a defined period. However, the school's formal measurement instruments, a monthly character theme with *peer-nominated* recognition and a *semester-based* behavior medal, remain observational and were acknowledged by the principal as not yet objective or quantifiable, directly echoing the *input-stage* finding on unstandardized *attitude-assessment* rubrics. Parental participation on the national education report was also rated only "moderate," the *lowest-performing* indicator in its cluster, corroborating the *process-stage* finding on the underused communication diary. Character attainment is also formally institutionalized through the school's monthly *character-theme* recognition and *semester-based* behavior medal, and through narrative descriptions of spiritual and social attitudes recorded on every student's report card each semester; a sampled *report-card* description, for instance, noted that a student was "good in religious observance, prayer before and after activities, and tolerance in worship" for the spiritual domain, and "good in honesty, discipline, responsibility, courtesy, care, and confidence" for the social domain. While such descriptions demonstrate that character attainment is consistently reported to every family each semester, their repetitive and generic phrasing, "good in..." without *behavior-specific* evidence, limits their diagnostic value compared with the more detailed, *incident-based* documentation produced by the counselor's individual case notes, reinforcing the broader *input-stage* finding that assessment instruments remain underdeveloped relative to the richness of the program's daily practice.

This dual pattern, strong outcomes in areas driven by habituation and weak measurement in areas requiring formal documentation, closely mirrors the broader empirical literature's observation that validated, standardized instruments for measuring character outcomes remain one of the most persistent gaps in the global character education research base (Oldham & McLoughlin, 2026). It is also consistent with survey evidence from Ethiopian *pre-primary* and primary schools, where *advice-giving* and modeling, rather than *curriculum-*

based instruction, were found to be the dominant and most trusted pedagogical mechanisms among both parents and teachers (Birhan et al., 2021).

Cross-Dimensional Synthesis

Read across all four dimensions, the school's character education program exhibits a coherent pattern rather than four isolated results. The single vision pillar left unsupported at the context stage, faith and piety, is the same pillar that produced the strongest measurable result at the product stage, while the *input-stage* gap in standardized assessment instruments reappears at the product stage as a limitation on measurement objectivity, and the *process-stage* weakness in the communication diary reappears in the national report's *parental-participation* indicator. Taken together, these findings indicate that the school's character culture, sustained through decades of daily habituation and consistent adult modeling, has developed faster than its formal systems for documenting, standardizing, and measuring that culture. This interpretation is consistent with the general conclusion that the character education field as a whole continues to struggle with instrument validation and *cross-context* replication even as implementation evidence accumulates (Oldham & McLoughlin, 2026), and it extends the CIPP evaluation literature on Indonesian schools by demonstrating, through explicit *cross-referencing* of findings across all four dimensions, exactly where culture and system diverge.

4. CONCLUSION

This study evaluated the implementation of a character education program at a private Indonesian elementary school that has operated for more than five decades without a prior comprehensive evaluation, using the CIPP model across all four dimensions. The Context dimension was found to be partially aligned, primarily because the vision's *faith-and-piety* pillar was not carried through into the school's mission and objectives, even though its legal basis was fully aligned. The Input dimension was reasonably prepared, supported by strong policy documentation and high teacher *role-modeling* but constrained by unstandardized assessment instruments, the exclusion of *non-teaching* staff from *character-related* briefing, and the absence of a dedicated budget. The Process dimension was reasonably effective, sustained by consistent *daily-to-annual* habituation and restorative discipline, though undermined by a *home-school* communication channel reduced to a signature ritual. The Product dimension was partially aligned, with nationally *top-ranked* outcomes in several indicators alongside a relative decline in the school's 2025 character ranking and persistently subjective measurement tools. Overall, the program's strengths rest on elements that are not formally documented, namely daily habituation and teacher exemplarity, while its weaknesses concentrate on documentation and instrumentation, indicating that the school's character culture has outpaced its own measurement systems. The program is accordingly recommended to continue, with targeted improvement rather than discontinuation or wholesale replacement.

The findings extend the CIPP evaluation literature on Indonesian character education by showing that the four dimensions are empirically interdependent rather than separable: a single unsupported vision pillar identified at the context stage reappeared as an unexpectedly strong *product-stage* outcome, and an instrumentation gap identified at the input stage reappeared as a measurement limitation at the product stage. Practically, the school's management is advised to formally incorporate *faith-and-piety* strategy into its mission and objectives, develop a standardized *attitude-assessment* rubric, issue a written code of conduct and briefing schedule for *non-teaching* staff, codify informally successful adaptive practices such as class messaging groups into formal policy, and redesign the *home-school* communication diary to strengthen parental engagement. The supervising foundation is

advised to allocate a dedicated budget line for character education and to provide methodological training in affective assessment for teachers.

Theoretically, this evaluation contributes an explicit demonstration of how the CIPP model's four dimensions interact empirically rather than functioning as separate checkpoints, and it extends the model with a *vertical-logic* alignment matrix that other *institution-level* evaluations could adopt to test *vision-mission-objective* consistency more systematically than has typically been done in the reviewed literature. In Stufflebeam's original formulation, product evaluation exists to inform a recycling decision, whether a program should be continued, modified, terminated, or installed elsewhere; the pattern identified here, strong practice outpacing weak documentation, points toward a *continue-with-modification* decision rather than either extreme.

This study is limited by its *single-case* design and *cross-sectional* data collection at one point in time, which constrains generalizability beyond the studied institution. Because the principal researcher held a governance role within the school's supervising foundation, the risk of positional bias cannot be fully eliminated despite the triangulation, negative case analysis, and *member-checking* procedures applied. Future research should develop and validate quantitative instruments for measuring character outcomes, conduct longitudinal studies following the same student cohort across grade levels to test the durability of behavioral change, examine the underlying causes of limited parental participation identified but not fully explained in this study, and replicate this evaluation design in other schools to test whether the *culture-versus-system* pattern identified here is *institution-specific* or more broadly generalizable.

5. ACKNOWLEDGEMENT

The authors thank the principal, teachers, staff, parents, and students of the school studied for their openness and participation throughout this research.

6. REFERENCES

Astuti, U. W. (2024). Evaluation of character education program at SDN Tanjung Duren Utara 02 Jakarta. *Jurnal Kepemimpinan Pendidikan, 7*(2), 999–1004. https://doi.org/10.22236/jkpuhamka.v7i2.18229

Badan Pusat Statistik. (2025). *Statistik kriminal 2024/2025*. Badan Pusat Statistik.

Bates, A. (2019). Character education and the 'priority of recognition'. *Cambridge Journal of Education, 49*(6), 695–710. https://doi.org/10.1080/0305764X.2019.1590537

Berkowitz, M. W., & Bier, M. C. (2007). What works in character education. *Journal of Research in Character Education, 5*(1), 29–48.

Birhan, W., Shiferaw, G., Amsalu, A., Tamiru, M., & Tiruye, H. (2021). Exploring the context of teaching character education to children in preprimary and primary schools. *Social Sciences & Humanities Open, 4*(1), 100171. https://doi.org/10.1016/j.ssaho.2021.100171

Brown, M., McGrath, R. E., Bier, M. C., Johnson, K., & Berkowitz, M. W. (2023). A comprehensive meta-analysis of character education programs. *Journal of Moral Education, 52*(2), 119–138. https://doi.org/10.1080/03057240.2022.2060196

Bryson, J. M. (2018). *Strategic planning for public and nonprofit organizations: A guide to strengthening and sustaining organizational achievement* (5th ed.). John Wiley & Sons.

Christanti, M. S. O., & Kusuma, D. (2025). Evaluation of Pancasila Student Profile Strengthening Project implementation in elementary school grade 1 using the CIPP model. *Journal of Innovation and Research in Primary Education, 4*(4), 2036–2048. https://doi.org/10.56916/jirpe.v4i4.2026

Creswell, J. W., & Poth, C. N. (2018). *Qualitative inquiry and research design: Choosing among five approaches* (4th ed.). SAGE Publications.

Dwi W., Y. R., Haryono, & Florentinus, T. S. (2018). The evaluation of the CIPP model in the implementation of character education at junior high school. *Innovative Journal of Curriculum and Educational Technology, 7*(2), 65–77. https://doi.org/10.15294/ijcet.v7i2.28391

Gasper, D. (2000). Evaluating the "logical framework approach" towards learning-oriented development evaluation. *Public Administration and Development, 20*(1), 17–28.

Guba, E. G., & Lincoln, Y. S. (1994). Competing paradigms in qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 105–117). SAGE Publications.

Heidari, M. H., Nowrozi, R. A., & Ahmadpoor, P. (2016). Recognition and applying character education approaches in schools. *Review of European Studies, 8*(3), 125–133. https://doi.org/10.5539/res.v8n3p125

Lestari, E. (2022). Evaluation of character education programs. *Jurnal Kepemimpinan Pendidikan, 5*(2), 747–751. https://doi.org/10.22236/jpkuhamka.v5i2.11218

Lickona, T. (1991). *Educating for character: How our schools can teach respect and responsibility*. Bantam Books.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. SAGE Publications.

Metcalfe, J., & Moulin-Stozek, D. (2020). Religious education teachers' perspectives on character education. *British Journal of Religious Education, 43*(3), 349–360. https://doi.org/10.1080/01416200.2020.1732868

Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook* (3rd ed.). SAGE Publications.

Murdiono, M., Miftahuddin, M., & Kuncorowati, P. W. (2017). The education of the national character of Pancasila in secondary school based on pesantren. *Jurnal Cakrawala Pendidikan, 36*(3), 423–434. https://doi.org/10.21831/cp.v36i3.15399

Oldham, P., & McLoughlin, S. (2026). Character education empirical research: A thematic review and comparative content analysis. *Journal of Moral Education, 55*(2), 313–341. https://doi.org/10.1080/03057240.2025.2480185

Peraturan Menteri Pendidikan dan Kebudayaan Nomor 20 Tahun 2018 tentang Penguatan Pendidikan Karakter pada Satuan Pendidikan Formal.

Peraturan Presiden Nomor 87 Tahun 2017 tentang Penguatan Pendidikan Karakter.

Pusat Informasi Kriminal Nasional Bareskrim Polri. (2025). *Data anak sebagai terlapor tindak kejahatan* [Data accessed 20 February 2025].

Putri, D. A. A., & Sanoto, H. (2025). Evaluating 5S cultural program for character education. *Journal of Innovation and Research in Primary Education, 4*(4), 2753–2763. https://doi.org/10.56916/jirpe.v4i4.1875

Qibtiyah, S., Ulfa, R. T., Masruhi, S., & Irawan, G. B. (2026). Evaluation of the implementation of character education based on national values using the CIPP model in elementary schools. *Journal of Educational Review and Research, 9*(1), 65–76.

Salirawati, D. (2021). Identifikasi problematika evaluasi pendidikan karakter di sekolah [Identifying problems in character education evaluation at school]. *Jurnal Sains dan Edukasi Sains, 4*(1), 17–27. https://doi.org/10.24246/juses.v4i1p17-27

Sidik, P. (2024). Character education and the crisis of national role models. *Jurnal Pendidikan Karakter, 15*(2). https://doi.org/10.21831/jpka.v15i2.73676

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

Sugiyono. (2023). *Metode penelitian kombinasi (mixed methods) dengan 9 desain* (2nd ed.). Alfabeta.

Undang-Undang Nomor 20 Tahun 2003 tentang Sistem Pendidikan Nasional.

Wholey, J. S., Hatry, H. P., & Newcomer, K. E. (Eds.). (2015). *Handbook of practical program evaluation* (3rd ed.). Jossey-Bass.

Woenardi, T. N., Supratno, H., & Mudjito. (2022). Management of Christian character education curriculum based on dormitory Abdi-Allah Theological College Trawas-Mojokerto. *Studies in Learning and Teaching, 3*(1), 74–96.

Yarbrough, D. B., Shulha, L. M., Hopson, R. K., & Caruthers, F. A. (2010). *The program evaluation standards: A guide for evaluators and evaluation users* (3rd ed.). Corwin Press.

Yin, R. K. (2018). *Case study research and applications: Design and methods* (6th ed.). SAGE Publications.

Zakso, A., Agung, I., Sofyatinngrum, E., & Capnary, M. C. (2022). Factors affecting character education in the development of the profile of Pancasila students: The case of Indonesia. *Journal of Positive School Psychology, 6*(2), 2254–2273.

Zhao, H., Gootjes, D. C., & Gong, F. (2024). A CIPP-based evaluation index system for moral cultivation in blended learning courses for vocational colleges in China. *International Journal of New Developments in Education, 6*(8), 71–78. https://doi.org/10.25236/IJNDE.2024.060810