

Instructional Decoupling in Character Education Policy: An Evaluation of Elementary Schools in Yogyakarta

¹Izzul Kiram Suardi, ²Muhammad Nursa' ban, ³Uldiyanti Sya'diah Syahdan

^{1,2}Universitas Negeri Yogyakarta, Indonesia

³Universitas Negeri Makassar, Indonesia

CORRESPONDING AUTHOR'S EMAIL

izzul0010pasca.2023@student.uny.ac.id

MANUSCRIPT HISTORY

Submitted: 24/09/2025

final revision: 15/05/2026

accepted: 08/07/2026

ABSTRACT

This study examines the implementation of the *akhlak kepada manusia* (morality toward others) character element within Indonesia's Pancasila Student Profile (P5) program across eight Sekolah Penggerak elementary schools in Yogyakarta City. A concurrent embedded mixed-methods design guided by Stake's Countenance model combined questionnaires across all eight schools with document analysis and interviews with principals, coordinators, and teachers at five purposively selected schools. The findings revealed at antecedents, strong collective cohesion (78.1%) paradoxically weakens teachers' instructional agency, evident in inadequate assessment-design documentation (40%); at transactions, teachers offset these deficits through improvisational pedagogy, producing an instructional decoupling between written plans and classroom practice; and at outcomes, a perfect compliance score (100%) reflects administrative performance rather than genuine instructional quality, even as students show measurable empathetic growth through local-wisdom-based experiential projects. This study reveals that institutional cohesion can undermine individual pedagogical agency. This paradox has not been documented in previous P5 evaluation research. Internalization of character depends on the teacher's ability to create authentic social learning opportunities through learning spaces that go beyond the application of a textual curriculum. Supervision policies need to be directed at direct observation of authentic learning practices, not just at auditing document completeness.

KEYWORDS

pancasila student profile; morality toward others; countenance stake evaluation; instructional decoupling; administrative compliance

Introduction

Indonesia's national education mandate prioritizes the development of intelligent and character-driven individuals (Rohana, 2019). To realize this objective, the government implemented the *Projek Penguatan Profil Pelajar Pancasila* (Project for Strengthening the Pancasila Student Profile) or P5 (Rachman et al., 2024). This program is an elaboration of national educational goals aimed at shaping students' character and competence (BSKAP, 2022). However, two years post-launch, significant implementation challenges persist at the school level. The primary constraint lies in conceptual and procedural misunderstandings.

For instance, Dyahningtyas and Muthmainah (2023) found that most early childhood educators failed to grasp the core concepts of P5. This issue extends to elementary schools, where Mulyani et al. (2023) indicate instances in which schools implemented all dimensions simultaneously. This practice directly contradicts official guidelines that recommend focusing on two or three dimensions to ensure a deeper impact (Satria et al., 2024).

Beyond procedural misunderstandings, limited collaboration among educators further hampers the program. Raihani et al. (2024) identified in their ethnographic study that a lack of principal initiative frequently hinders collaboration. Preliminary observations conducted by the researcher at an elementary school in Yogyakarta City corroborated this systemic issue. Project modules remained underdeveloped, and educators felt overwhelmed and ill-equipped to manage inclusive classrooms. Consequently, teachers experienced burnout, which led to an uneven distribution of attention among students.

The need to evaluate and strengthen character education programs has become increasingly urgent amid growing concerns about student behavior and school climate. In the Special Region of Yogyakarta, data released by the Ministry of Women's Empowerment and Child Protection (KemenPPPA) in early 2025 reveals a notable paradox. Although the number of general cases of violence against children in the City of Yogyakarta declined from 130 in 2023 to 71 in 2024, the city continues to register the highest number of cases relative to its compact geographical area (32.5 km²) and its total population. Of greater concern is the fact that incidents of violence occurring specifically within school environments across the province have risen consistently, increasing from 26 cases in 2021 to 71 cases in 2024. Reports indicate that these incidents frequently involve peers and educators as perpetrators, which demonstrates that educational institutions have not yet fully transformed into safe and conducive environments. This situation necessitates targeted character education interventions.

Although empirical evidence consistently demonstrates that well-structured character education programs are effective in reducing school violence and improving the overall school climate (Kang et al., 2024; Palade & Pascal, 2023; Wang et al., 2023; Zhang et al., 2025), the success of these international intervention studies depends substantially on their capacity to design and evaluate specific behavioral parameters. This stands in sharp contrast to the current implementation of character education in Indonesia, which is often evaluated in overly general terms. While international literature emphasizes the need to measure specific traits such as empathy or grit, research examining a single P5 dimension remains limited. Consequently, precise diagnostic information regarding the successes or obstacles of implementing specific character elements remains obscured. This gap is highly problematic because the lack of monitoring from relevant authorities renders internal school evaluations insufficient for external quality assurance (Indah & Muhson, 2025). Therefore, a comprehensive external evaluation is necessary to provide a clear diagnostic evidence for policy designers.

To fulfill this diagnostic need and address the aforementioned research gap, this study employs Stake's Countenance Evaluation Model. Stake's Countenance Evaluation Model systematically evaluates educational programs by comparing intended conditions, implementation processes, and outcomes (Stake, 1967). To better explain implementation gaps, this study also draws on two institutional theory concepts: *instructional decoupling*, the divergence between formal instructional plans and classroom practice (Hasse & Krücken, 2014), and *administrative compliance*, the fulfillment of reporting requirements regardless of instructional quality (Hudson et al., 2019; Mockler & Stacey, 2021). Together, these concepts provide an analytical lens for examining whether P5 implementation reflects authentic instructional practice or primarily administrative compliance.

This model was selected because it systematically evaluates educational programs across three stages—antecedents, transactions, and outcomes—while comparing intended program design with actual implementation and outcomes (Stake, 1967). Instead of conducting a broad evaluation, this research focuses exclusively on one crucial element: *akhlak kepada manusia* (morality toward others). Based on this rationalization, this study poses three primary questions regarding the implementation of the morality toward others element in Yogyakarta elementary schools: (1) What is the achievement of the antecedents? (2) What is the achievement of the transactions? (3) What is the achievement of the outcomes?

By addressing these research questions, this study makes three distinct contributions. At the practical level, the findings provide schools with empirical diagnostic data to optimize program execution, enabling clearer task delegation and targeted capacity building. At the policy level, the evidence-based recommendations offer guidance to education authorities on resource allocation and the design of standardized teacher training. At the theoretical level, the study enriches the

program evaluation literature by applying Stake's Countenance framework to a single character dimension, establishing a replicable methodological template for future researchers evaluating other P5 elements.

Literature Review

Character Education Interventions and the P5 Framework

Character education programs have been shown to be effective for addressing school violence, as demonstrated by various intervention studies. These studies show success through both comprehensive interventions and those focused on specific traits. For instance, Zhang et al. (2025) tested the PurposeFull People program, which broadly integrates social-emotional learning, and found significant improvements in behavioral expectations across elementary schools. Additionally, targeted interventions such as the six-week grit training program investigated by Wang et al. (2023) have shown positive results in reducing bullying behavior.

Focusing on specific traits such as empathy is also highly effective in educational settings. Kang et al. (2024) reported that an empathy-based violence prevention program successfully increased students' empathy levels and reduced permissive attitudes toward violence. Similarly, Palade and Pascal (2023) found that intensive empathy training significantly reduced verbal bullying. However, their study also noted that positive impacts were only observed when the teacher was present during the intervention, highlighting the crucial role of educators. This research affirms that structured character education interventions possess measurable potential to improve school climates.

In Indonesia, this structured intervention is actualized through the *Projek Penguatan Profil Pelajar Pancasila* (Project for Strengthening the Pancasila Student Profile) or P5. P5 serves as a cross-disciplinary, project-based framework designed to instill six core dimensions of character (BSKAP, 2022) while promoting contextual learning that connects students to real-world societal issues (Molyneux et al., 2023). Among these dimensions, morality toward others is highly relevant to addressing social friction and violence. This element requires students to prioritize equality, respect diversity, and cultivate empathy (Satria et al., 2024; Ulandari & Rapita, 2023). As Jannah (2018) noted, social responsibility (*kesalehan sosial* or social piety) and morality toward others are fundamental to healthy interpersonal relationships. Furthermore, Mahanani et al. (2022) emphasized that character remains the most crucial element for individuals living in a society, making this specific character element an important protective factor against school-based violence.

Research Gap and Novelty

As prior literature has established, the efficacy of global intervention studies derives from their targeted approach to specific behavioral characteristics (e.g., Kang et al., 2024; Wang et al., 2023). However, existing evaluations of the P5 framework in Indonesia deviate from this standard by employing overly broad diagnostic metrics that assess multiple dimensions simultaneously. Research focusing on the implementation barriers of any single P5 element remains exceedingly scarce. Consequently, actionable data concerning the specific pedagogical or institutional obstacles to cultivating targeted behaviors remains largely unclear. The present study directly addresses this methodological gap by evaluating the morality toward others dimension as a distinct character construct. In doing so, it shifts the analytical focus away from generic program review toward a novel, focused diagnostic analysis of the specific character attributes relevant to preventing school-based violence.

Theoretical Framework: Stake's Countenance Evaluation Model

To support this targeted evaluation and analyze the collected data, this study employs Stake's Countenance Evaluation Model (Stake, 1967) as the primary theoretical framework. Originally developed to provide a comprehensive overview of educational programs, this model directs evaluators to systematically analyze programs across three chronological stages: antecedents, transactions, and outcomes (Stake, 1967). The antecedents stage assesses the preliminary conditions and institutional readiness before the program begins. The transactions stage

focuses on the actual execution, peer interactions, and dynamic pedagogical processes during the program. Finally, the outcomes stage assesses program results and impacts and behavioral achievements (Indah & Muhson, 2025).

This model is highly suitable for evaluating complex character education programs because it accommodates the nuances of social dynamics and allows for a deep analysis of implementation quality (Purwati & Sukirman, 2024). Harjanti et al. (2019) explained that Stake's framework enables evaluators to answer research questions systematically across every phase of a character education program. Similarly, Masruraini et al. (2023) stated that this approach provides a more comprehensive integration for measuring the outcomes and impacts of character strengthening. By mapping the morality toward others element onto Stake's framework, this research can systematically identify whether implementation failures occur at the preparation, execution, or final achievement stage.

Research Methods

This study employed a program evaluation design using a concurrent embedded mixed-methods approach, in which qualitative data served as the primary source supported by quantitative data. This evaluation applied Stake's (1967) Countenance model, which examines the program through three stages: antecedents, transactions, and outcomes. This model was selected because of its comprehensive framework for systematically comparing field implementation outcomes with the official guidelines from the ministry.

The research was conducted from March to June 2025 across eight public elementary schools designated as *Sekolah Penggerak* in Yogyakarta City. Quantitative data were collected from all eight schools to capture information from all participating schools. Meanwhile, in-depth qualitative data collection was focused on five target schools, purposively selected as *information-rich* cases to achieve data saturation (Table 1). To ensure strict confidentiality, the precise identities of these institutions are masked using alphabetical pseudonyms.

Table 1

List of Research Locations and Type of Data Collection

Research Location	Data Collection Methods
School A	Questionnaires, interviews, and document studies
School B	Questionnaires, interviews, and document studies
School C	Questionnaires, interviews, and document studies
School D	Questionnaires, interviews, and document studies
School E	Questionnaires, interviews, and document studies
School F	Questionnaires
School G	Questionnaires
School H	Questionnaires

The primary research participants encompassed principals, program coordinators, teachers, and staff. Schools were selected using purposive sampling, whereas individual participants were recruited through convenience sampling. Additionally, student developmental outcomes were assessed indirectly using secondary data (student communication records and report cards).

Qualitative data were collected through document studies and semi-structured interviews, which were subsequently analyzed thematically using ATLAS.ti software (Naeem et al., 2023). Conversely, quantitative data were gathered using a school cultural readiness questionnaire. While initially conceptualized based on the three normative criteria from the government guidelines (Open-mindedness, Eagerness to Learn, and Collaboration), a preliminary psychometric validation study empirically reconstructed these constructs into 17 valid items across three precise dimensions: (1) Mental Resistance and Professional Egoism (nine items), (2) Synergy and Collective Commitment (five items), and (3) Self-Directed Learning Agency (three items). The instrument's construct validity was confirmed using confirmatory factor analysis (CFA), with AVE values exceeding 0.50 for all factors. Excellent internal consistency reliability was also demonstrated, with McDonald's omega (ω) coefficients of 0.91, 0.84, and 0.75, respectively.

The criteria for program success were evaluated based on the revised 2024 P5 guidelines and specific school modules. Quantitative achievement standards were determined using ideal statistical criteria. Raw scores from the instrument were converted into precision percentages based on the formula proposed by Arikunto & Jabar (2018):

$$\text{Percentage} = \frac{\text{Actual Score} - \text{Minimum Score}}{\text{Score Range}} \times 100\%$$

To integrate both data types, the researcher applied a joint display analysis technique within the Stake matrix framework. Qualitative data were extensively used to clarify and explore the meaning of the quantitative findings, culminating in the final evaluative judgment (Table 2)

Table 2

General Success Criteria by Percentage

Interval	Criteria
$X > 75\%$	Very Good
$58.3\% < X \leq 75\%$	Good
$41.7\% < X \leq 58.3\%$	Fairly Good
$25\% < X \leq 41.7\%$	Not Very Good
$X \leq 25\%$	Not Good

Source: Author's Data Analysis

This study acknowledges several methodological limitations. *First*, the geographical and institutional scope was confined to public elementary schools designated as *Sekolah Penggerak* in Yogyakarta, and the sample selection relied partially on convenience sampling, which limits the generalizability of the findings to broader educational contexts. *Second*, data collection primarily depended on self-reported data from interviews with school personnel and accessible document analysis. The absence of direct field observations (due to the scheduling conflicts with the schools' final examinations) and restricted access to highly confidential internal documents may introduce response bias and limit the depth of the contextual analysis. *Finally*, due to methodological constraints, student outcomes were evaluated solely through secondary data (report cards and communication books) instead of direct measurements from participants. This limitation hindered a more in-depth exploration of the broader socio-cultural factors that influence character development.

This study was conducted in full compliance with research ethics, ensuring all participants provided informed consent before data collection. Data anonymity and confidentiality were rigorously upheld, and the research protocol was carried out with approval from the relevant institutional ethics committee.

Research Findings

This section presents the evaluation findings organized according to Stake's three-stage Countenance framework: antecedents, transactions, and outcomes. Each stage is analyzed by integrating quantitative and qualitative data.

Achievement of the Antecedent Stage

The antecedents stage evaluates the fundamental conditions before program implementation, encompassing school cultural readiness and the completeness of planning documents. The aggregate achievements at this stage are presented in Table 3.

Table 3

Antecedent Stage Results

Indicator	Score	%	Category
School Cultural Readiness	70.1	78.1	Very Good
Document	32.6	57.1	Fairly Good

Source: Author's Data Analysis

Based on Table 3, a significant readiness disparity (gap) is identified at the antecedents stage. The cultural readiness of the school ecosystem has reached the Very Good category overall (78.1%), indicating highly adequate educators' psychosocial readiness. Nevertheless, this cultural capital does not correspond to managerial and administrative readiness, where the program planning documents merely fall into the Fairly Good category (57.1%).

a. Documents in the Antecedent Stage

Document readiness focuses on fulfilling instructional design prerequisites, with detailed achievements presented in Table 4, which exposes a chronic administrative imbalance. Schools demonstrate a Very Good managerial capacity (83.3%) in formalizing team formation and a Good capacity (71.1%) in mapping elements and themes. However, performance declines in the operational domain. Program scheduling only achieves a Fairly Good category (53.3%), and most notably, the formulation of objectives, activities, and assessments plummets to the Not Very Good category (40%).

This low percentage stems from the absence of assessment designs (both diagnostic and formative) in almost all school modules. For schools unable to provide evidence of assessment documents, the methodological assessment assigned a minimum score. The lack of detailed duties in the Team Decree and the absence of assessment designs indicate that documents at the antecedents stage tend to be compiled merely to fulfill the administrative requirement to establish an implementation team, rather than serving as a robust instructional navigation instrument.

Table 4

Document at the Antecedents Stage

Indicator	Sub Indicators	Score	%	Category
Document	Team formation	7	83.3	Very Good
	Elements, themes, and programs	9.4	71.1	Good
	Programme schedule	5.2	53.3	Fairly Good
	Objectives, activities, and assessments	11	40	Not Very Good
	Overall	32.6	57.2	Fairly Good

Source: Author's Data Analysis

b. School Culture Readiness

To precisely measure psychosocial conditions, the cultural readiness indicator was evaluated through three empirical dimensions operationalizing the P5 guidelines. Referring to Table 5, the Synergy and Collective Commitment dimension is identified as the highest-achieving dimension (81.7% - Very Good). This proves that *Sekolah Penggerak* has successfully reduced sectoral ego barriers, corroborated by the educators' very low resistance to innovation (Mental Resistance reached 77.6% on a reverse scale - Very Good). However, an anomaly is found in the Self-Directed Learning Agency dimension, which lags in the Good category (73.8%). Evaluatively, this implies that the solid team cohesiveness at the beginning of the program yields a side effect of "collective dependency," where teachers' individual initiative to explore references independently becomes suboptimal due to an over-reliance on team guidance.

Table 5

School Culture Readiness Outcomes

Indicator	Dimensions	Score	%	Category
School Cultural Readiness	Mental Resistance and Professional Egoism	36.9	77.6	Very Good
	Synergy and Collective Commitment	21.3	81.7	Very Good
	Self-Directed Learning Agency	11.9	73.8	Good

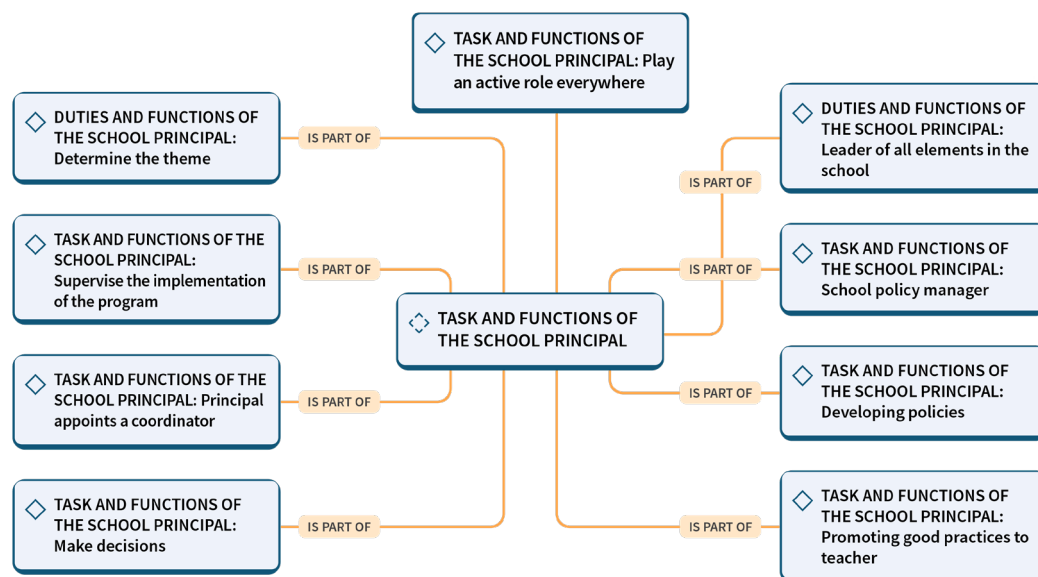
Source: Author's Data Analysis

c. Roles and Functions in the Planning Phase

The disparity between high collective commitment (Table 4) and weak assessment instruments (Table 5) does not occur in a vacuum. This phenomenon can be comprehensively explained by mapping the roles and functions of each educational actor at the antecedents stage. The thematic analysis conducted using ATLAS.ti indicates that program planning operates through a centralized (top-down) managerial command.

Figure 1

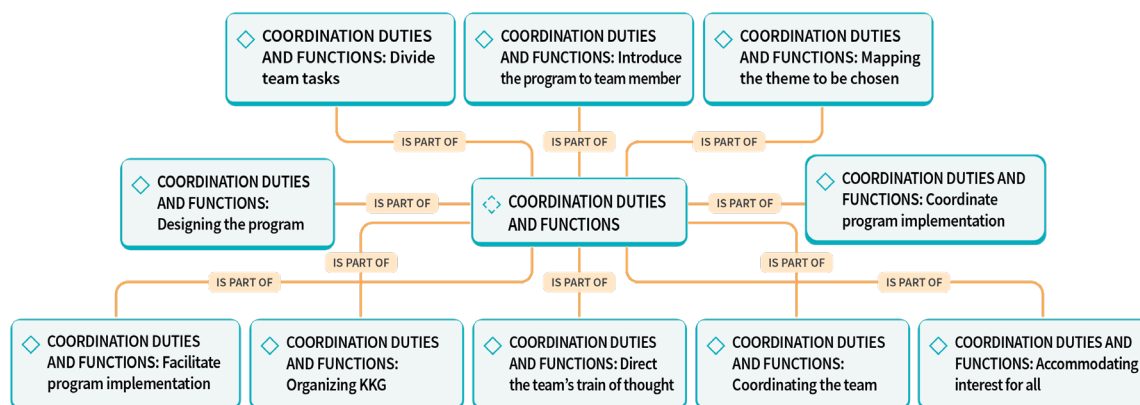
Principal Duties and Functions



Source: ATLAS.ti 25 Analysis Results

Referring to the role mapping in Figure 1, the Principal holds authoritative control at the upstream stage of planning. Qualitative findings confirm that the Principal acts not only as an administrator but as a policy initiator who definitively determines the project theme and directly appoints the committee structure. This strong managerial leadership empirically validates why the Team Formation sub-indicator in the document study achieved a Very Good score (83.3%). The principal successfully conditioned the institutional ecosystem structurally before the program began, ensuring no vacuum of responsibility at the macro level.

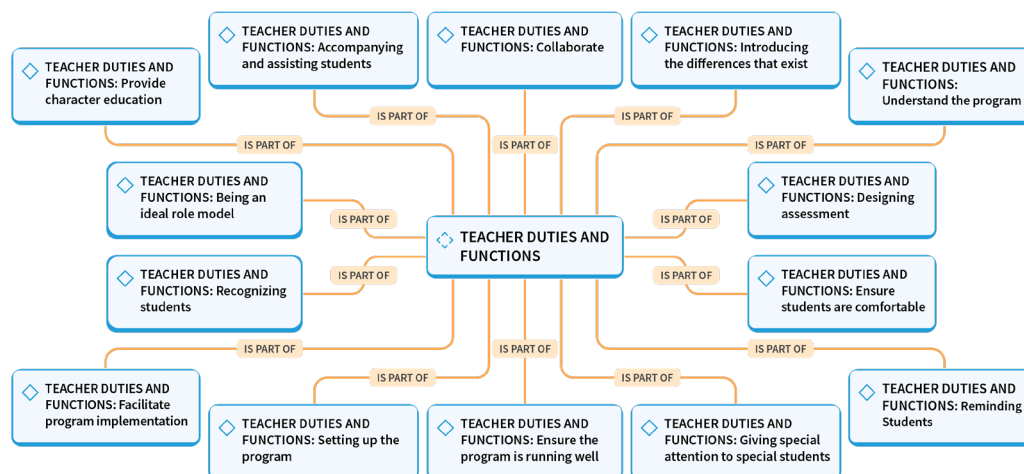
Figure 2
Coordinator Duties and Functions



Source: ATLAS.ti 25 Analysis Results

Once the macro framework is established, planning operationalization is delegated to the Program Coordinator. Thematic findings show the coordinator as the primary driving force who designs program drafts, directs the facilitators' train of thought, and consolidates the agenda through Learning Communities (KKG). This intensity of central coordination directly fosters the educators' Synergy and Collective Commitment, reaching the Very Good level (81.7%). Conversely, the coordinator's workload (being overly focused on theme synchronization and team conditioning) is suspected to create a gap in formulating technical matters. This logically explains why precise components such as the Implementation Schedule (53.3%) and Assessment formulation (40%) were overlooked and failed to mature during this preparation phase.

Figure 3
Teacher Duties and Functions



Source: ATLAS.ti 25 Analysis Results

As the endpoint of a centralized system, the role of teachers or facilitators during the preparation phase undergoes a functional shift. Thematic codes in ATLAS.ti indicate that teachers' perceptions of their duties are dominated by field execution functions (e.g., acting as role models, recognizing student characteristics, and assisting) rather than acting as instructional designers. This qualitative reality is the key to answering the quantitative anomaly in Table 4. The preparatory ecosystem that provides teachers with pre-developed modules and themes indirectly suppresses the teachers' Self-Directed Learning Agency (73.8%). Teachers feel no urgency to autonomously explore assessment references because they have been conditioned to remain passive and await collective instructions from the coordinator before acting.

Comprehensively, the antecedents stage evaluation concludes that the *Sekolah Penggerak* ecosystem in Yogyakarta City possesses highly robust psychosocial readiness due to centralized leadership and team synergy. However, this cohesive advantage may constrain instructional design autonomy; the dominance of upper-level managerial actors may reduce the initiative of teachers at the lower level, resulting in administrative document designs (specifically, precise schedules and assessment instruments) failing to be rigorously formulated before program execution.

Achievement of the Transactions Stage

The transactions stage focuses on the dynamics of field implementation, evaluated through two instruments: 1) analysis of operational module quality and 2) thematic analysis of instructional activities and assessment patterns. The recapitulation of the document analysis at this stage is presented in Table 6.

Table 6

Document at the Transactions Stage

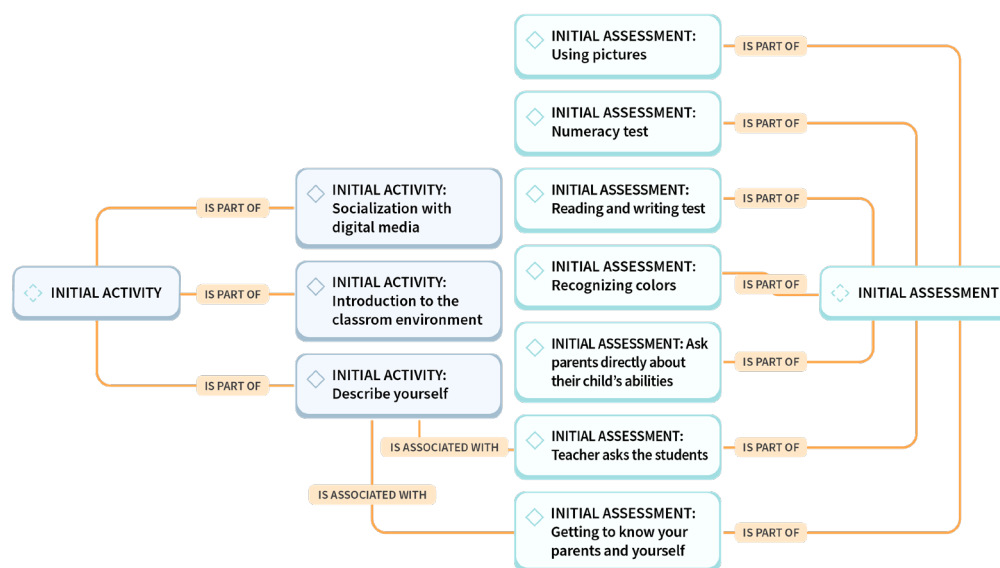
Indicator	Sub Indicators	Score	%	Category
Program Module	Starting the Activity	3.6	86.7	Very Good
	Optimizing implementation	2.6	53.3	Fairly Good
	Ending the activity	2.4	46.7	Fairly Good
	Overall	8.6	62.2	Good

Source: Author's Data Analysis

Based on Table 6, the transactions stage aggregates to a Good category (62.2%) but exhibits a linear decline in quality from the beginning to the end of the activity. A wide gap exists between the schools' capability to design program initiations (86.7%) and program closures (46.7%). The absence of reflection and follow-up descriptions in the modules across the majority of sample schools causes this low score at the end. Evaluatively, this finding suggests that P5 module planning in the field tends to be highly "project-centered," strong during initial ceremonies but weak in sustainability design and final evaluation.

a. Initial Activity and Diagnostic Assessment

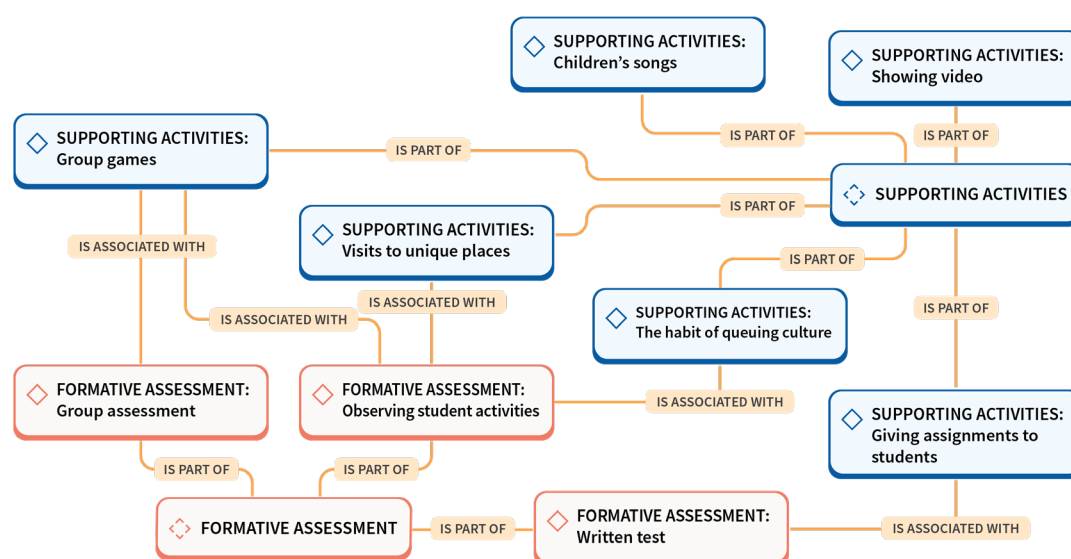
Integrating qualitative data provides a deeper understanding of how educational units compensate for document deficiencies through field improvisation. Findings in Figure 4 show that initial activities focus on building social bonding between teachers and students. The use of digital media and "self-recognition" techniques serves as a bridge to introduce the element of morality toward others. Theoretically, this aligns with the high quantitative score (86.6%) in initiating activities. Teachers consciously conduct noncognitive diagnostic assessments (recognizing emotions, colors, and parents' backgrounds) to map students' affective readiness before diving into the core character materials.

Figure 4*Initial Activity and Initial Assessment*

Source: ATLAS.ti 25 Analysis Results

b. Supporting Activities and Formative Assessments

During the core implementation stage, supporting activities are designed to maintain student engagement while serving as a medium for real-time behavioral observation. Figure 5 shows that formative assessments are not conducted as rigid written tests but are integrated into supporting activities such as group games and visits to neighboring schools. Evaluatively, this pattern demonstrates the teachers' efforts to conduct "authentic assessments." Through queuing culture and group work, teachers observe how moral elements (e.g., respecting the rights of others) are directly practiced. Although the "Optimizing Implementation" aspect in Table 6 only scores 53.3%, this qualitative finding proves that instructional interactions in the field are far more dynamic than what is rigidly prescribed within the module documents.

Figure 5*Supporting Activities and Formative Assessments*

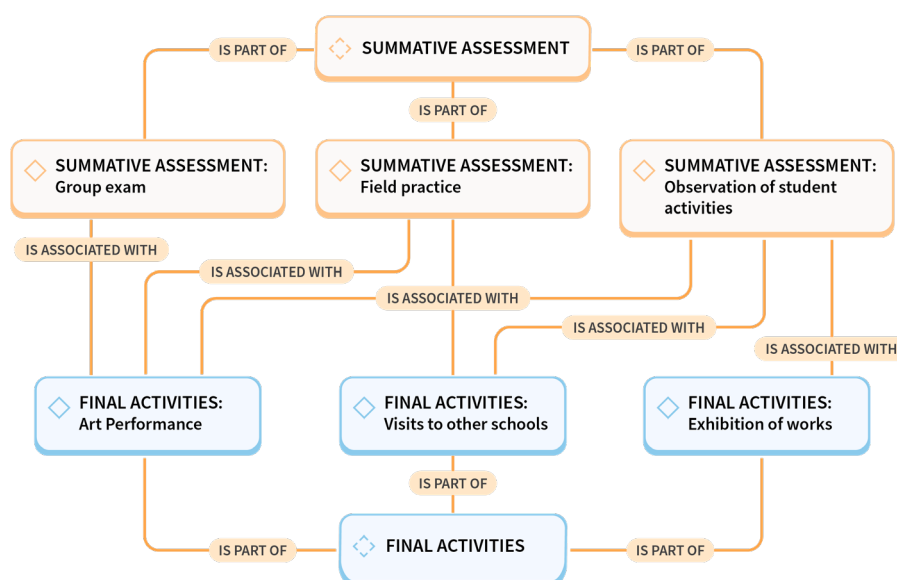
Source: ATLAS.ti 25 Analysis Results

c. Final Activity and Summative Assessment

The final section represents the most paradoxical phase in this evaluation, where inconsistencies occur between documents and practice. Based on Figure 6, the final program activities are dominated by celebrations of learning outcomes in the form of art performances, exhibitions, and inter-school visits. These activities simultaneously function as summative assessments through field practice and observation. However, the most crucial evaluative meaning here is the discovery of the “practice without a written plan” phenomenon. Although the final document design scored poorly (46.7%) due to a widespread lack of written closure plans, most schools in Yogyakarta City still hosted culminating activities and exhibitions. This suggests that the final phase of the P5 program appears to be driven by established school practices rather than by structured instructional planning.

Figure 6

Final Activity and Summative Assessment



Source: ATLAS.ti 25 Analysis Results

While the implementation stage was generally rated as 'Good,' it relied substantially on teacher improvisation in the classroom, revealing a disconnect between administrative modules and actual teaching practices. The low score in documentation design (46.7%) was effectively obscured by vibrant, collaborative art performances and exhibitions. However, this gap highlights a critical need to align written planning with classroom execution to ensure the character program maintains strong curricular accountability.

Achievement of the Outcome Stage

The outcomes stage evaluates the final products of program implementation, represented by two main outputs: administrative accountability (final report documents) and instructional impacts (students' character development).

a. Documented Outcomes

Evaluation at this stage focuses on the completeness of final P5 program reports uploaded or archived by the schools. As shown in Table 7, these final reporting documents achieved a score of 100% (Very Good), demonstrating that all sampled schools successfully produced comprehensive final reports aligned with government templates. However, this perfect score creates an evaluative anomaly when contrasted with the transactions stage, where activity closure planning was found to be notably weak (46.7%). This contradiction empirically

demonstrates a strong culture of administrative compliance within the *Sekolah Penggerak* ecosystem. Educational institutions exhibit a superior managerial capacity to produce post-activity artifacts to fulfill bureaucratic demands, even when their underlying operational designs remain incomplete.

Table 7

Document at the Outcomes Stage

Indicator	Sub Indicators	Score	%	Category
Character Education Program Final Report	Prioritizing similarities with others and respecting differences	4	100	Very Good
	Empathize with others	4	100	Very Good
	Indicators of program success	4	100	Very Good
	Overall	12	100	Very Good

Source: Author's Data Analysis

b. Student Character Development Outcomes

The instructional impact of the program is evaluated through secondary data, specifically students' P5 report card grades, focusing on the element of morality toward others. Table 8 shows that students' character outcomes have quantitatively reached the normative threshold, situated in the Developing as Expected range. The average scores of 2.73 and 2.8 indicate a positive affective development. Although this achievement appears satisfactory, a critical evaluative perspective must consider the potential of leniency bias among educators, where facilitators tend to award "safe" grades in the affective domain at the end of the semester to avoid assigning lower affective ratings. Therefore, this quantitative data must be validated with direct experiential findings from the field.

Table 8

Program Report Card Analysis Results

Aspect	$\bar{X}$	Achievements
Global Diversity	2.73	Developing as Expected
Gotong Royong	2.8	Developing as Expected

Source: Author's Data Analysis

c. Extraction of Affective Impacts

To complement report card data, an ATLAS.ti thematic analysis was utilized to dissect how project themes genuinely transform into character development. Figure 7 exposes a crucial main finding regarding the P5 pedagogical design. The analysis reveals that the internalization of "morality toward others" is most strongly manifested through programs that, at first glance, do not directly relate to morality, such as Salted Egg Making and Traditional Food (Local Wisdom).

Table 9*Countenance Evaluation Matrix: Antecedents Stage*

Aspect	Intents / Standards	Observations / Judgments
Description	Intents: Establish a prepared school cultural ecosystem and design initial operational documents encompassing objectives, activities, and assessments.	Observations: a. Aggregate Cultural Readiness: 78.1% (Very Good). High synergy (81.7%) is driven by leadership hierarchy, while teachers' independent agency lags (73.8%). b. Aggregate Planning Documents: 57.2% (Fairly Good). Team Decrees are present (83.3%), but written assessment designs remain minimal (40%).
Judgment	Standards: a. School culture reflects uniform psychosocial readiness. b. Implementation team comprehends duties and functions. c. P5 modules contain precise schedules and explicit assessment designs prior to execution.	Judgments: Preparation Gap Identified. Highly solid cultural ecosystem (Very Good) triggers collective dependency, blunting teachers' instructional agency. Document fulfillment is limited to Team Decree formality, while technical assessment designs lack rigorous preparation (Fairly Good).

Source: Author's Data Analysis

Table 10*Countenance Evaluation Matrix: Transactions Stage*

Aspect	Intents / Standards	Observations / Judgments
Description	Intents: Implement activities aligned with operational documents, encompassing supporting activities and structured assessments.	Observations: a. Aggregate Module Execution: 62.2% (Good). Initial activity design is optimal (86.7%), but final program design weakens drastically (46.7%). b. Interviews: Absence of final guidelines compensated by high teacher improvisation (e.g., authentic assessments via art performances).
Judgment	Standards: Field implementation and assessments proceed linearly, aligning with the approved module framework.	Judgments: Instructional Decoupling Occurs. Aggregate field implementation runs well (Good), driven by facilitator improvisation rather than module compliance. Teachers' empirical adaptability masks weak final activity designs (46.7%).

Source: Author's Data Analysis

Table 11*Countenance Evaluation Matrix: Outcomes Stage*

Aspect	Intents / Standards	Observations / Judgments
Description	Intents: Generate measurable student character outcomes and document final reporting per administrative guidelines.	Observations: a. Final Reports: 100% (Very Good) across all indicators. b. Student report cards: 2,73 to 2,80 (Developing as Expected). c. Interviews: Authentic empathetic changes stimulated by experiential learning (e.g., local wisdom projects).
Judgment	Standards: a. Students identify emotions, demonstrate empathy, and respond to differences positively. b. Institutional final reports align with the standardized Kemendikbudristek format.	Judgments: Achieved with Critical Caveats. Affective outcomes and institutional reports were optimally fulfilled (Very Good). However, the absolute document score (100%) represents retrospective administrative compliance, paradoxically concealing earlier instructional design weaknesses.

Source: Author's Data Analysis

The matrix construction presented in Table 9 (Antecedents), Table 10 (Transactions), and Table 11 (Outcomes) exposes a fundamental finding: the success of the character education program in Yogyakarta City's Sekolah Penggerak is not achieved through a linear instructional flow, but rather through empirical compensation in the field. A distinct discrepancy exists between the weak administrative design in the preparation phase (Table 9) and the high achievement of reporting documents in the final phase (Table 11). The success of students' affective outcomes is proven to be sustained more by the solid synergy of teachers and the improvisational skills of facilitators during the implementation stage (Table 10), rather than originating from the precision of the planning modules.

The occurrence of instructional decoupling (i.e., a disconnect between formal planning documents and classroom implementation) and administrative compliance at the end of the program constitutes a crucial evaluative anomaly. These findings not only demonstrate how educational units navigate curriculum demands but also indicate systemic weaknesses in the operational policy design of the program. The theoretical implications and policy impacts of this phenomenon will be further dissected in the discussion section.

Discussion

Evaluating the morality toward others element in Yogyakarta's P5 program using Stake's Countenance model reveals findings that challenge traditional assumptions about linear educational planning. Theoretically, a program is assumed to be successful if its planning is flawless (Cheung et al., 2022). However, the findings of this study demonstrate that the students' affective character development does not stem from exhaustive administrative instructional planning. Instead, it is heavily sustained by the resilience and teachers' pedagogical adaptability during implementation. These findings highlight the phenomena of instructional decoupling and administrative compliance that characterize the *Sekolah Penggerak* ecosystem, revealing a substantial gap between policy design and implementation at the school level.

The Paradox of Solidarity and Instructional Agency in the Antecedents Stage

Educational-unit readiness during the preparation phase is a key predictor for program success. Previous literature consistently underscores that immature planning, ranging from inadequate curriculum design to a lack of comprehensive training, may hinder the achievement of intended outcomes (Chand et al., 2021). At the antecedents stage, the sample school ecosystem demonstrates a high level of cultural readiness overall (Very Good category). This initially appears to support the proposition that a positive school climate promotes effective uptake of educational innovations (Ayu et al., 2024).

However, a critical evaluation in this study finds that this cultural readiness conceals an institutional paradox. The high synergy and collective commitment shaped through centralized managerial leadership from principals and coordinators paradoxically fosters what may be described as collective dependency. This top-down structural dominance directly results in the reducing teachers' self-directed learning agency. As instructional designers, teachers experience diminished instructional autonomy; they position themselves more as technical executors rather than curriculum developers. This finding enriches the academic discourse on school leadership, proving that centralized command is indeed effective in ensuring rapid program compliance, yet it frequently fails to foster analytical initiative at the teacher level (Collie, 2021; Slempe et al., 2020).

The consequences of this collective dependency is concretely recorded in the poor quality of operational documents, particularly in the formulation of diagnostic, formative, and summative assessment instruments, which fall into the Not Very Good category. The absence of comprehensive assessment designs in the modules suggests that teachers may perceive limited urgency to comprehensively design grading rubrics because they are accustomed to receiving macro instructions from the coordinator. This condition reinforces the arguments of Mak (2019) and Drost and Levine (2023), who assert that gaps in assessment literacy among educators constitute a primary root cause. In the context of *Sekolah Penggerak*, P5 training that is solely centralized on learning committee representatives may create a bottleneck effect and weaken knowledge transfer when transferred to classroom teachers. The failure to translate general guidelines from the central government into specific assessment instruments confirms that the preparation of instructional documents cannot rely solely on the spirit of cooperation, but rather demands technical competence and systematic curriculum design training (DeLuca et al., 2016).

Instructional Decoupling and Improvisational Capability in the Transactions Stage

Entering the implementation stage, this evaluation uncovers an anomaly that challenges conventional planning assumptions. Although the module documents in the initial phase suffer from deficits in assessment design and precision scheduling, the transactions stage was implemented adaptively in practice and falls into the Good category. A striking gap exists between the minimal documentation of final activity designs and the festive execution of program celebrations in the field. This phenomenon indicates instructional decoupling. This concept refers to a situation where actual practices in the classroom operate under a situational logic that frequently deviates from, or is completely untethered to, what is formally written in administrative documents (Hasse & Krücken, 2014).

In this context, teachers function as street-level bureaucrats, pragmatically disregarding the rigidity of official modules to adapt to the real psychological needs of students in the classroom (Hudson et al., 2019). The teachers' pedagogical agility successfully compensates for weaknesses in instructional planning documents. For instance, teachers intuitively integrate initial activities (such as self-introduction and family background) as a form of noncognitive diagnostic assessment that is not written in the module. Furthermore, supporting activities and formative assessments are not conducted through rigid written tests, but are observed through direct practice in group games, queuing culture habituation, and inter-school visits.

This approach aligns with the framework of The Jubilee Centre for Character and Virtues (2022) and the classical thoughts of Lickona (1992), which state that character development requires students' active engagement in real social interactions. This engagement-based learning strategy enables students to grapple with moral issues authentically (Black et al., 2019). The courage of educators to transform the classroom into a social ecosystem proves that the quality of teacher

capacity is the primary foundation of character education, where responsive facilitators can compensate for structural weaknesses in administrative planning documents (Sudirman, 2019).

Administrative Compliance and Experiential Learning in the Outcomes Stage

During the final assessment stage, the achievement of school reporting documents achieves a perfect score quantitatively. However, interpreting this finding demands a high degree of academic caution. Considering the weak assessment design in the initial phase, the perfection of these final reports may be interpreted as evidence of administrative compliance.

The sample schools have proven to demonstrate strong administrative capacity to prepare final reports primarily to satisfy accountability requirements, even though the instructional processes behind them are fraught with anomalies and improvisation. This phenomenon explicitly exposes a critical flaw in current national supervision and monitoring policies. Because local education authorities and school supervisors predominantly evaluate program success through rigid document-completeness checklists rather than authentic pedagogical observations, schools are structurally incentivized to prioritize document completeness (Browes, 2021; Mockler & Stacey, 2021). This confirms contemporary critiques regarding performative accountability, where schools are driven to prioritize the completeness of report formats over the substantive quality of the instructional planning itself (Holloway & Brass, 2018).

Despite being overshadowed by the appearance of bureaucratic compliance and the potential bias of educators' leniency in assigning report card grades, qualitative validation proves that students' character development still exhibits fundamental affective shifts. Interestingly, thematic analysis reveals that the internalization of human morality (such as prejudice reduction, respect for religious differences, and social empathy) is not realized through abstract moralistic lectures. These characters develop through experiential projects grounded in local wisdom, such as salted-egg-making and the introduction of traditional foods.

These activities successfully encourage meaningful social interaction and provide experiential learning spaces. As articulated by Kolb and Kolb (2018), learning through embodied, hands-on experience activates emotional regulation and deepens reflective awareness. Ego clashes when students must divide tasks in salted-egg-making groups naturally force them to practice recognizing their peers' emotions and behaving tolerantly. Grounding implementation in real-world social dynamics thus proves to be considerably more effective in reducing bullying behavior and fostering empathy than classroom instruction alone, a finding coherent with contemporary anti-violence educational intervention studies (Kang et al., 2024; Palade & Pascal, 2023; Wang et al., 2023). These findings suggest that a central strength of character education under the Merdeka Curriculum lies in providing authentic social interaction spaces, which are proven capable of rescuing the program from managerial weaknesses at the planning phase.

Conclusion and Policy Recommendations

This study reveals a structural paradox at the heart of the implementation of project-based character education: the conditions that foster institutional cohesion may simultaneously constrain the instructional agency that quality character education demands. This paradox operates across all three Countenance stages. At the antecedents stage, high team synergy driven by centralized leadership paradoxically breeds collective dependency, undermining teachers' capacity to design contextually precise assessment instruments. At the transactions stage, the resulting lack of documentations does not halt the program; instead, they are compensated by teachers' improvisational resilience, resulting in instructional decoupling between written plans and classroom practices. At the outcomes stage, perfect reporting scores may mask underlying instructional weaknesses, constituting a form of bureaucratic compliance that conceals earlier design weaknesses.

The deeper implication for the evaluation of project-based character education in Indonesian elementary schools is this: the quality of character education cannot be determined by the completeness of a school's administrative portfolio. Character internalization is fundamentally dependent on the availability of authentic social learning opportunities that teachers are empowered to create, that is, spaces that break through the rigidity of a textual curriculum.

This finding calls for a fundamental reorientation in how educational authorities conceptualize program accountability.

Based on the conceptual synthesis above, this study proposes strategic recommendations distributed across three primary intervention levels. *First*, at the school level, principals should shift their management focus from merely meeting reporting deadlines to developing teachers' assessment literacy by detailing operational responsibilities in the Team Decree. Furthermore, schools need to cease spoon-feeding pre-made modules and provide teachers with greater instructional autonomy. *Second*, at the education office level, authorities should revise monitoring instruments by directing supervision toward direct classroom observations rather than document-based compliance audits. Additionally, the education office needs to facilitate technical training in assessment design that directly targets relevant teachers.

Third, future research should replicate this methodological design on other dimensions of the Pancasila Student Profile. Researchers should also employ longitudinal design approaches to test whether the affective changes achieved by students possess permanent character retention when interacting in a broader societal environment. This methodological extension will provide deeper insights into the long-term efficacy of project-based character interventions.

Acknowledgment

The authors express their gratitude to the Graduate School of Yogyakarta State University and the Center for Education Financing Services, Ministry of Primary and Secondary Education, for their financial support. The authors also thank the Yogyakarta City Education Office for granting research access and the participating teachers for their valuable time and contributions.

References

- Arikunto, S., & Jabar, C. S. A. (2018). *Evaluasi program pendidikan: Pedoman teoritis praktis bagi mahasiswa dan praktisi pendidikan* (F. Yustianti, Ed.; 2nd ed.). Bumi Aksara.
- Ayu, N. K., Astawan, I. G., & Paramita, M. V. A. (2024). Instrumen penilaian Profil Pelajar Pancasila aspek akhlak mulia dan kreativitas di sekolah dasar. *EDUKATIF: Jurnal Ilmu Pendidikan*, 6(3), 1866–1877. <https://doi.org/10.31004/edukatif.v6i3.6567>
- Black, A., Lawson, H., & Norwich, B. (2019). Lesson planning for diversity. *Journal of Research in Special Educational Needs*, 19(2), 115–125. <https://doi.org/10.1111/1471-3802.12433>
- Browes, N. (2021). Test-based accountability and perceived pressure in an autonomous education system: Does school performance affect teacher experience? *Educational Assessment, Evaluation and Accountability*, 33(3), 483–509. <https://doi.org/10.1007/s11092-021-09365-9>
- Badan Standar, Kurikulum, dan Asesmen Pendidikan. (2022). *Dimensi, elemen, dan subelemen Profil Pelajar Pancasila pada Kurikulum Merdeka*. Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi.
- Chand, S., Chaudhary, K., Prasad, A., & Chand, V. (2021). Perceived causes of students' poor performance in mathematics: A case study at Ba and Tavua secondary schools. *Frontiers in Applied Mathematics and Statistics*, 7. <https://doi.org/10.3389/fams.2021.614408>
- Cheung, W. J., Hall, A. K., Skutovich, A., Brzezina, S., Dalseg, T. R., Oswald, A., Cooke, L. J., Melle, E. Van, Hamstra, S. J., & Frank, J. R. (2022). Ready, set, go! Evaluating readiness to implement competency-based medical education. *Medical Teacher*, 44(8), 886–892. <https://doi.org/10.1080/0142159X.2022.2041585>
- Collie, R. J. (2021). COVID-19 and teachers' somatic burden, stress, and emotional exhaustion: Examining the role of principal leadership and workplace buoyancy. *AERA Open*, 7. <https://doi.org/10.1177/2332858420986187>
- DeLuca, C., Danielle, L.-M., & Luhanga, U. (2016). Teacher assessment literacy: A review of international standards and measures. *Educational Assessment, Evaluation and Accountability*, 28(3), 251–272. <https://doi.org/10.1007/s11092-015-9233-6>
- Drost, B. R., & Levine, A. C. (2023). An analysis of strategies for teaching and assessing standards-based assessment design to preservice teachers. *Journal of Education*, 203(3), 574–586. <https://doi.org/10.1177/00220574211038711>

- Dyahningtyas, A. A. S., & Muthmainah, M. (2023). Proyek karnaval HUT Republik Indonesia sebagai penguatan Profil Pelajar Pancasila. *Jurnal Obsesi: Jurnal Pendidikan Anak Usia Dini*, 7(2), 2024–2036. <https://doi.org/10.31004/obsesi.v7i2.4075>
- Harjanti, R., Supriyati, Y., & Rahayu, W. (2019). Evaluation of learning programs at elementary school level of “Sekolah Alam Indonesia (SAI)”. Evaluative research using Stake’s Countenance Model. *American Journal of Educational Research*, 7(2), 125–132. <https://doi.org/10.12691/education-7-2-2>
- Hasse, R., & Krücken, G. (2014). Decoupling and coupling in education. In B. Holzer, F. Kastner, & T. Werron (Eds.), *From Globalization to World Society: Neo-Institutional and Systems-Theoretical Perspectives* (1st ed., pp. 197–214). Routledge. <https://doi.org/10.4324/9781315772967>
- Holloway, J., & Brass, J. (2018). Making accountable teachers: The terrors and pleasures of performativity. *Journal of Education Policy*, 33(3), 361–382. <https://doi.org/10.1080/02680939.2017.1372636>
- Hudson, B., Hunter, D., & Peckham, S. (2019). Policy failure and the policy-implementation gap: Can policy support programs help? *Policy Design and Practice*, 2(1), 1–14. <https://doi.org/10.1080/25741292.2018.1540378>
- Indah, F., & Muhson, A. (2025). The implementation of Pancasila student profile strengthening program in pioneer elementary schools. *Journal of Education Research and Evaluation*, 9(1), 106–118. <https://doi.org/10.23887/jere.v9i1.87318>
- Jannah, M. (2018). Studi komparasi akhlak terhadap sesama manusia antara siswa fullday school dengan siswa boarding school di kelas XI SMA IT Abu Bakar Yogyakarta. *Jurnal Pendidikan Agama Islam Al-Thariqah*, 3(2), 1–15. [https://doi.org/10.25299/althariqah.2018.vol3\(2\).2216](https://doi.org/10.25299/althariqah.2018.vol3(2).2216)
- Kang, K.-A., Kim, S.-J., Kang, S., & Lee, J. (2024). Effects of a violence prevention education program using empathy (VPEP-E) on fifth-grade students in South Korea. *The Journal of School Nursing*, 40(4), 361–371. <https://doi.org/10.1177/10598405221094282>
- Kolb, D. A., & Kolb, A. Y. (2018). Eight important things to know about the experiential learning cycle. *Australian Educational Leader*, 40(3), 8–14. https://www.acel.org.au/ACELWEB/Publications/AEL/2018/3/Lead_Article_1.aspx
- Lickona, T. (1992). *Educating for character: How our schools can teach respect and responsibility*. Bantam Books.
- Mahanani, P., Akbar, S., Kamaruddin, A. Y. B., & Hussin, Z. B. (2022). Educational analysis to develop character in Malaysia and Indonesia. *International Journal of Instruction*, 15(3), 377–392. <https://doi.org/10.29333/iji.2022.15321a>
- Mak, P. (2019). Impact of professional development programme on teachers’ competencies in assessment. *Journal of Education for Teaching*, 45(4), 481–485. <https://doi.org/10.1080/02607476.2019.1639266>
- Masruraini, M., Rasyid, M. N. A., & Mania, S. (2023). Evaluation of character education strengthening program through scouts extra in KKSC (Kisada Kids Scout Camp) activities using countenance evaluation at MIS Ushuluddin Singkawang. *International Journal of Multi Discipline Science (IJ-MDS)*, 6(2), 156. <https://doi.org/10.26737/ij-mds.v6i2.4700>
- Mockler, N., & Stacey, M. (2021). Evidence of teaching practice in an age of accountability: When what can be counted isn’t all that counts. *Oxford Review of Education*, 47(2), 170–188. <https://doi.org/10.1080/03054985.2020.1822794>
- Molyneux, T. M., Zeni, M., & Oberle, E. (2023). Choose your own adventure: Promoting Social and emotional development through outdoor learning. *Early Childhood Education Journal*, 51(8), 1525–1539. <https://doi.org/10.1007/s10643-022-01394-3>
- Mulyani, S., Nurmeta, I. K., & Maula, L. H. (2023). Analisis implementasi Profil Pelajar Pancasila di sekolah dasar. *Jurnal Educatio FKIP UNMA*, 9(4), 1638–1645. <https://doi.org/10.31949/educatio.v9i4.5515>
- Naem, M., Ozuem, W., Howell, K., & Ranfagni, S. (2023). A step-by-step process of thematic analysis to develop a conceptual model in qualitative research. *International Journal of Qualitative Methods*, 22. <https://doi.org/10.1177/16094069231205789>
- Palade, T., & Pascal, E. (2023). Reducing bullying through empathy training: The effect of teacher’s passive presence. *Behavioral Sciences*, 13(3), 216. <https://doi.org/10.3390/bs13030216>

- Purwati, E., & Sukirman, D. (2024). Evaluation of the implementation of the Merdeka Curriculum in internalizing serviam's character values with The Countenance Stake Model. *Proceedings of the 7th International Conference on Learning Innovation and Quality Education (ICLIQE 2023)*, 48–58. https://doi.org/10.2991/978-2-38476-301-6_6
- Rachman, A., Putro, H. Y. S., Rusandi, M. A., & Situmorang, D. D. B. (2024). The development and validation of the “Kuesioner Tema Proyek Penguatan Profil Pelajar Pancasila” (KT P5): A new tool for strengthening the Pancasila student profile in Indonesian pioneer schools. *Heliyon*, 10(16), e35912. <https://doi.org/10.1016/j.heliyon.2024.e35912>
- Raihani, Kailani, N., Kartika, D., Supraptini, A., Maulana, T., Pratiwi, I., Utama, B., Hijriani, I., Purnama, J., & Azizah, S. N. (2024). *Evaluasi proses dan konteks perubahan program sekolah penggerak: Potret transformasi pendidikan di tingkat sekolah dan daerah*. Pusat Standar dan Kebijakan Pendidikan (PSKP). <https://pskp.kemendikdasmen.go.id/analisis-kebijakan/detail/526>
- Rohana, E. (2019). Character education relation with spiritual intelligence in Islamic education perspective. *International Journal of Nusantara Islam*, 6(2), 165–174. <https://doi.org/10.15575/ijni.v6i2.4803>
- Satria, M. R., Adiprima, P., Jeanindya, M., Anggraena, Y., Anitawati, Sekarwulan, K., & Harjatanaya, T. Y. (2024). *Panduan pengembangan Proyek Penguatan Profil Pelajar Pancasila* (Revisi). Pusat Kurikulum dan Pembelajaran.
- Slemp, G. R., Field, J. G., & Cho, A. S. H. (2020). A meta-analysis of autonomous and controlled forms of teacher motivation. *Journal of Vocational Behavior*, 121, 103459. <https://doi.org/10.1016/j.jvb.2020.103459>
- Stake, R. E. (1967). The countenance of educational evaluation. *Teachers College Record: The Voice of Scholarship in Education*, 68(7), 1–15. <https://doi.org/10.1177/016146816706800707>
- Sudirman, S. (2019). The 21st-century teacher: Teacher's competence within the character education framework towards a cultural-oriented development and promoting tolerance. *International Education Studies*, 12(8), 21. <https://doi.org/10.5539/ies.v12n8p21>
- The Jubilee Centre for Character & Virtues. (2022). *The Jubilee Centre framework for character education in schools* (3rd ed.). University of Birmingham.
- Ulandari, S., & Rapita, D. D. (2023). Implementasi proyek penguatan profil pelajar Pancasila sebagai upaya menguatkan karakter peserta didik. *Jurnal Moral Kemasyarakatan*, 8(2), 116–132. <https://doi.org/10.21067/jmk.v8i2.8309>
- Wang, L., Chen, Q., Peng, Z., Ye, C., & Zhou, X. (2023). Six-week grit training course reduces school bullying: A quasi-experimental study. *Frontiers in Child and Adolescent Psychiatry*, 1. <https://doi.org/10.3389/frcha.2022.1045808>
- Zhang, Y., Larson, M., & Cook, C. R. (2025). A multi-site cluster randomized controlled trial evaluating the effect of characterstrong's purposefull people elementary program on student outcomes. *Canadian Journal of School Psychology*, 40(2), 79–101. <https://doi.org/10.1177/08295735251340750>