

Uneven Institutionalization of Formative Assessment Strategies under Indonesia's Merdeka Curriculum: A Collective Case Study in Primary Education

Umil Muhsinin, Kiki Fatmawati, Tarisa Jayanti

^{1,2,3}Universitas Sultban Thaba Saifuddin Jambi, Indonesia

e-mail: umilmuhsinin@uinjambi.ac.id, kikifatmawati86@uinjambi.ac.id,
tarisa.jayanti2002@gmail.com

Submitted: 08-02-2026

Revised: 10-07-2026

Accepted: 01-08-2026

ABSTRACT. Although formative assessment is positioned as a central component of Indonesia's Merdeka Curriculum, little is known about whether Black and Wiliam's (2009) five formative assessment strategies become institutionalized with comparable consistency and depth in classroom practice. Addressing this gap, this study examines the uneven institutionalization of formative assessment strategies in primary education and explores how curriculum reform interacts with teacher practices, school-level support, and student agency. Using a qualitative collective case study design, the study was conducted in five elementary schools and madrasah ibtidaiyah implementing the Merdeka Curriculum. Data were collected through non-participant classroom observations, semi-structured interviews with principals, teachers, and students, and document analysis, and were analyzed through thematic cross-case comparison. The findings reveal that formative assessment strategies do not develop in parallel. Strategies characterized by high institutional visibility and compatibility with established teacher routines, particularly learning intentions, evidence elicitation, and feedback, were more consistently institutionalized because they were reinforced through curricular guidance, documentation requirements, and academic supervision. In contrast, peer assessment demonstrated uneven pedagogical depth, while self-assessment remained weakly embedded due to limited assessment literacy, time constraints, and the continued dominance of summative assessment practices. Consequently, formative assessment predominantly functioned as assessment for learning rather than as assessment for learning. The study contributes to theory by extending Black and Wiliam's framework through the concept of uneven institutionalization, demonstrating that formative assessment strategies follow differentiated developmental trajectories during curriculum reform in a developing-country context. The findings further suggest that successful learner-centered assessment reform requires not only policy adoption but also systematic support for teacher assessment literacy, evidence-informed academic supervision, and the deliberate development of students' self-regulation and evaluative agency.

Keywords: *Formative assessment, Primary Education, Merdeka Curriculum, Assessment Literacy, Student Agency*

 <https://doi.org/10.32678/tarbawi.v12i02.12663s>

How to Cite Muhsinin, U., Fatmawati, K., & Jayanti, T. (2026). Uneven Institutionalization of Formative Assessment Strategies under Indonesia's Merdeka Curriculum: A Collective Case Study in Primary Education. *Tarbawi: Jurnal Keilmuan Manajemen Pendidikan*, 12(02), 337–356.
<https://doi.org/10.32678/tarbawi.v12i02.12663>

INTRODUCTION

Indonesia's Merdeka Curriculum represents an ambitious attempt to reposition assessment from a predominantly evaluative and reporting function toward an integral component of learner-centered, adaptive, and continuous learning. Within this reform agenda, teachers are expected to use

assessment evidence not merely to document achievement but to identify students' current understanding, diagnose learning needs, and determine appropriate next steps in teaching and learning (Aditomo, 2024; Fatmawati, 2024; Maleki, 2025; Zikriati et al., 2024). Formative assessment is therefore central to the curriculum's broader aspiration to make learning more responsive to student progress and to strengthen learners' active participation in their own development. Nevertheless, this policy aspiration generates a fundamental paradox. A curriculum explicitly oriented toward learner agency may still be enacted through assessment routines dominated by teacher control, procedural documentation, and administrative compliance. Under such conditions, formative assessment can become highly visible as a policy requirement without becoming deeply embedded as a pedagogical process that uses evidence responsively or progressively redistributes responsibility for learning. This tension between learner-centered reform aspirations and teacher-centered assessment enactment constitutes a critical problem not only for Indonesia but also for wider international efforts to transform classroom assessment.

International scholarship increasingly indicates that the formal adoption of formative assessment does not automatically transform the pedagogical logic of classroom practice. Effective formative assessment requires more than the visible presence of learning objectives, questions, tasks, feedback, or assessment instruments; it depends on the capacity to elicit relevant evidence, interpret that evidence in context, and translate it into subsequent action by teachers and learners (Almutawa, 2025; Anderson et al., 2001; Aust et al., 2025; Arvatz et al., 2025; Goertzen et al., 2025; Owen, 2016; Welsh, 2012). This distinction is consequential because evidence does not become formative simply by being collected. Its formative function depends on whether it informs instructional adaptation, supports meaningful learner action, or changes what happens next in the learning process. The challenge is therefore partly one of assessment literacy: teachers must be able not only to generate information about learning but also to judge its significance and respond pedagogically. Assessment reform may otherwise remain procedural, with teachers adopting recognizable techniques while the underlying relationship among evidence, interpretation, and instructional decision-making remains weak.

Recent research in primary education reinforces this problem. In dashboard-supported mathematics lessons, van Kessel et al. (2026) found that teachers' feedback based on real-time student data consisted mainly of task- and process-level feedback and varied substantially in when it was enacted across lesson phases. This evidence is particularly instructive because it shows that even when teachers have immediate access to visualized information about student performance, the translation of evidence into pedagogical action remains variable. The issue is therefore not simply a lack of assessment data. Rather, it concerns how evidence is interpreted, when it is acted upon, and whether it generates context-sensitive responses. Such findings sharpen a central concern in formative-assessment research: the presence of sophisticated tools, frequent assessment activities, or visible feedback practices cannot by itself demonstrate that assessment has become institutionally embedded as an evidence-responsive pedagogical process.

A parallel shift is evident in contemporary scholarship on feedback. Feedback has traditionally been conceptualized as information provided by teachers to improve student performance, but more recent perspectives emphasize learners' active role in interpreting, evaluating, and using feedback (Lui & Andrade, 2022; McCarthy et al., 2025). From this perspective, feedback is not formative merely because it is frequent or improvement-oriented; its value depends on whether learners understand the information, connect it to quality criteria, and use it to regulate subsequent action. The distinction is important because teacher-dominated feedback can improve immediate performance while leaving the learner's evaluative role largely unchanged. Recent work on teacher and peer feedback similarly indicates that feedback processes interact in complex ways with students' perceived ownership and responsibility for learning (Gwen et al., 2026). Although conducted in a collaborative-learning context, this research underscores a broader conceptual issue: increasing external feedback does not necessarily strengthen learner agency. Formative assessment must

therefore be examined not only in terms of the amount or source of information provided, but also in terms of the distribution of evaluative agency: who interprets the evidence, who judges quality, who determines the next step, and who ultimately assumes responsibility for improvement.

This issue becomes particularly salient when formative assessment is connected to self-regulated learning and metacognition. Learner-centered assessment presupposes that students progressively develop the capacity to understand goals, monitor their progress, evaluate the quality of their work, interpret feedback, and determine appropriate subsequent actions. However, these capacities cannot be assumed to emerge automatically from participation in nominally learner-centered classrooms. Research on self-regulated learning conceptualizes learners as active participants who plan, monitor, evaluate, and adapt their learning, while also emphasizing that regulatory capacity is shaped by instructional and contextual support (Baissane & Zaid, 2025; Nieminen & Boud, 2025). Recent evidence further suggests that metacognitive strategies such as planning, monitoring, self-assessment, reflection, and strategy adjustment require systematic integration and scaffolding rather than occasional use (Ramdhani & Isom, 2022). This produces a second paradox within learner-centered reform: learner-centered policy does not necessarily produce learner self-regulation. Where teachers remain the principal interpreters of assessment evidence and controllers of feedback, students may participate in formative activities without developing the capacity to evaluate and regulate their own learning.

These international tensions are especially relevant to Indonesia's Merdeka Curriculum. The curriculum formally promotes flexible, differentiated, and learner-centered learning and positions assessment as an integral part of instruction (Aditomo, 2024). Nevertheless, emerging Indonesian research indicates that translating these principles into routine classroom practice remains challenging (Budiono & Hatip, 2023; Ihsan & Maemonah, 2023; McCallum & Milner, 2021; Tomasine, 2024; Zikriati et al., 2024). Teachers operate within institutional environments shaped by established assessment traditions, reporting requirements, academic supervision, time constraints, and uneven opportunities for sustained professional learning. These conditions may influence not merely whether formative assessment is implemented but also which dimensions of it become embedded in everyday practice. Practices that are readily observable, documentable, and compatible with established teacher routines, such as stating learning objectives, assigning tasks, recording results, or providing teacher feedback, may gain greater institutional visibility. By contrast, practices requiring students to interpret criteria, evaluate peers, monitor their own understanding, and determine subsequent learning actions may demand bigger changes in pedagogical roles and classroom relationships. The resulting problem is therefore more complex than a generic "policy-practice gap": different formative-assessment strategies may achieve different degrees of consistency, depth, and organizational embedding within the same reform environment.

Black and Wiliam's (2009) framework provides a particularly useful analytical lens for examining this possibility. The framework identifies five interconnected strategies: (1) clarifying and sharing learning intentions and criteria for success; (2) engineering effective classroom discussions, questions, and tasks that elicit evidence of learning; (3) providing feedback that moves learners forward; (4) activating students as instructional resources for one another; and (5) activating students as owners of their own learning. Although these strategies constitute an integrated architecture of formative assessment, they do not necessarily place equivalent demands on classroom roles, teacher assessment literacy, or learner agency. The first three may, at least at a procedural level, be incorporated into established teacher-led routines: teachers can communicate objectives, ask questions, assign tasks, and provide feedback while retaining substantial authority over the interpretation of evidence and subsequent decisions. By contrast, peer assessment, especially self-assessment, requires students to understand the criteria, exercise evaluative judgment, interpret evidence, provide or use feedback, and regulate subsequent action. They therefore require a more substantial redistribution of evaluative responsibility from teachers toward learners.

The theoretical problem, consequently, is not simply whether the five strategies are present or absent. It concerns whether theoretically interconnected strategies become embedded with comparable consistency and depth within the same curriculum-reform environment. A school may routinely require teachers to communicate learning objectives without using those objectives as reference points for students' evaluative judgment. Teachers may frequently ask questions while treating responses primarily as indicators of participation rather than as evidence for adapting instruction. Feedback may be recurrent but remain predominantly teacher-controlled. Peer assessment may occur as informal commenting without explicit criteria, while self-assessment remains episodic or marginal. Such configurations would suggest that formative assessment is not institutionalized as a unified pedagogical system. Instead, its constituent strategies may follow differentiated trajectories, with some becoming more deeply embedded in classroom routines and school expectations than others.

Despite growing international scholarship on formative assessment, three interrelated gaps remain. Conceptually, previous research has generated extensive evidence on the overall effects of formative assessment and on individual components such as questioning, feedback, peer assessment, self-assessment, assessment literacy, and self-regulated learning (Andersson & Palm, 2018; Beesley et al., 2018; Lui & Andrade, 2022; McCarthy et al., 2025; Nieminen & Boud, 2025; van Kessel et al., 2026). This work has substantially advanced understanding of particular practices, yet limited comparative attention has been given to whether Black and Wiliam's five interconnected strategies become embedded with comparable consistency and depth within the same institutional context. Examining strategies separately may obscure an important theoretical possibility: formative assessment can be formally promoted as an integrated approach while its constituent strategies develop unevenly because they require different configurations of teacher authority, learner participation, evaluative judgment, and institutional support.

Methodologically, substantial strands of previous research have relied on surveys, self-reported perceptions, intervention studies, or analyses of discrete assessment practices. Such approaches have generated important evidence regarding teacher conceptions, feedback, professional development, and learner regulation, but they provide more limited insight into how multiple formative-assessment strategies coexist, diverge, and become recurrently embedded in everyday school settings. The distinction matters because reported use does not necessarily indicate sustained pedagogical enactment, just as observing an isolated practice does not establish its organizational embedding. A collective case-study approach combining classroom observations, interviews with multiple participant groups, and documentary evidence can therefore illuminate not only the presence of strategies but also cross-case differences in their consistency, depth, and relationship to school-level support.

Contextually, the differentiated embedding of formative-assessment strategies remains insufficiently examined within system-level curriculum reform in primary education, particularly in settings where ambitious learner-centered policy aspirations coexist with established teacher-led, supervisory, and accountability-oriented assessment traditions. Indonesian studies have provided valuable accounts of Merdeka Curriculum implementation and specific assessment practices (Budiono & Hatip, 2023; Ihsan & Maemonah, 2023; Zikriati et al., 2024), but limited comparative evidence explains how the full configuration of Black and Wiliam's five strategies becomes embedded across different primary-school contexts. Consequently, insufficient attention has been paid to which strategies become routinized, which remain partial or episodic, and how these disparities correspond to differences between teacher-led practices and practices that require greater evaluative agency from peers and learners.

To address these gaps, this study advances uneven institutionalization of formative assessment strategies as an interpretive perspective for examining assessment reform. In this study, institutionalization refers to the extent to which a formative-assessment strategy becomes recurrently embedded in classroom routines, pedagogical tools, supervisory expectations, and the

roles enacted by teachers and students. Uneven institutionalization, therefore, refers to differences in the consistency, depth, and organizational embedding of Black and Wiliam's five strategies across school contexts. This perspective moves beyond a binary distinction between implementation and non-implementation by examining how theoretically interconnected strategies may follow differentiated trajectories of institutional embedding. It also questions the implicit expectation that the formal promotion of formative assessment will lead to parallel development across all five strategies. Instead, the study examines whether strategies compatible with established teacher routines and institutional visibility become more readily embedded, while strategies requiring a redistribution of evaluative agency to students, particularly peer and self-assessment, remain comparatively fragile. The study's conceptual novelty thus lies not in proposing an additional formative-assessment technique, but in reframing assessment reform as a process of differentiated institutional embedding across theoretically interconnected strategies.

Accordingly, this study examines how Black and Wiliam's (2009) five formative-assessment strategies are enacted and differentially institutionalized across five elementary schools and madrasah ibtidaiyah implementing Indonesia's Merdeka Curriculum. Using a collective case-study design, the analysis focuses on cross-case patterns in the consistency, depth, and organizational embedding of the five strategies, while considering teacher practices, school-level support, and students' participation in assessment processes. The study makes three contributions. Theoretically, it reframes the implementation of formative assessment from a binary question of adoption versus non-adoption toward an account of how interconnected strategies may follow uneven institutionalization trajectories. Empirically, it provides triangulated cross-case evidence from primary education within a national curriculum reform setting, which remains underrepresented in international formative assessment research. Practically, it identifies where teacher assessment literacy, evidence-responsive instructional decision-making, academic supervision, feedback processes, and deliberate scaffolding of peer and self-assessment require strengthening if learner-centered reform is to move beyond procedural visibility toward student agency and self-regulation. In this sense, the Indonesian case provides a theoretically informative setting for understanding a broader international problem: why learner-centered assessment reforms may become institutionally visible before they become pedagogically transformative.

METHOD

Research Design and Case Selection

This study employed a qualitative, collective case study design to examine the uneven institutionalization of formative assessment strategies in primary education under Indonesia's Merdeka Curriculum. The design was selected because the study sought not only to identify the presence of formative assessment practices but also to compare how Black and Wiliam's (2009) five strategies became embedded across school contexts, with varying levels of consistency, depth, and organizational support. Each school constituted a bounded case, while the enactment of formative assessment strategies represented the focal phenomenon examined within and across cases. The collective case study design, therefore, enabled both within-case interpretation and cross-case comparison of similarities, variations, and disparities in formative assessment practices.

The study was conducted over one semester at five educational institutions implementing the Merdeka Curriculum: SDN 030/XI Koto Dian, MI No. 04/E.72 Koto Baru, MIS Kurnia Kota Jambi, MIN 4 Muaro Jambi, and MIS Nurul Yaqin Muaro Jambi. The cases were selected purposively based on four considerations: implementation of the Merdeka Curriculum, the presence of classroom teachers conducting formative assessment, variation in institutional characteristics across elementary schools and madrasah ibtidaiyah, including public and private settings, and institutional willingness to provide access for repeated classroom observations, interviews, and document analysis. The inclusion of diverse school contexts was intended to support analytical

comparison of how formative assessment strategies were embedded under different organizational conditions.

Table 1: Case and participant profile

Case	Institutional context	Principal	Teachers	Students	Total	Observation
A	Public elementary school	1	2	4	7	3 meetings
B	Madrasah ibtidaiyah	1	2	4	7	3 meetings
C	Private madrasah ibtidaiyah	1	2	4	7	3 meetings
D	Public madrasah ibtidaiyah	1	2	4	7	3 meetings
E	Private madrasah ibtidaiyah	1	2	4	7	3 meetings
Total	Five cases	5	10	20	35	15 meetings

Participants

Participants were selected purposively based on their direct involvement in planning, implementation, supervision, or experience with formative assessment. Seven participants were recruited from each school, including one principal, two classroom teachers who implemented formative assessment, and four students from the classes taught by the participating teachers. In total, the study involved 35 participants: five principals, ten classroom teachers, and twenty students.

Principals were included because of their responsibilities for curriculum implementation, academic supervision, and school-level support for assessment practices. Teachers were selected because they directly planned and enacted formative assessment during classroom instruction. Students were included because they experienced the assessment process and could provide perspectives on learning objectives, questioning, feedback, peer assessment, and self-assessment. The combination of these participant groups enabled triangulation among school-level policy and supervision, teachers' classroom practices, and students' experiences of assessment.

Data Collection

Data were collected through non-participant classroom observations, semi-structured interviews, and document analysis. The use of these three techniques was intended to examine formative assessment from complementary perspectives: observed enactment, participant interpretation, and documentary evidence. The researcher acted as the primary instrument and maintained a non-participatory position during classroom observations to minimize disruption to regular learning activities.

Classroom observations were conducted in learning settings where participating teachers implemented formative assessment. In each case, observations were conducted across three meetings, with each meeting lasting 2×35 minutes. This resulted in 210 minutes of observation per case, totaling 1,050 minutes (17.5 hours) across the five cases. Observations focused on the five strategies proposed by Black & Wiliam (2009): clarification of learning intentions and success criteria; eliciting evidence through questions, discussions, and tasks; feedback that moves learning forward; peer assessment; and self-assessment. Field notes documented teacher actions, student responses, assessment interactions, feedback episodes, the use of learning evidence, and whether assessment information informed subsequent teaching or learning actions.

Semi-structured interviews were conducted face-to-face with all 35 participants. Interviews with principals explored the implementation of the Merdeka Curriculum, academic supervision, institutional expectations, and school-level support for formative assessment. Teacher interviews explored how formative assessment was planned and enacted, how evidence of learning was interpreted, how feedback was provided, and how peer and self-assessment were facilitated. Student interviews explored their understanding of learning goals, experiences of teacher feedback, participation in peer assessment, opportunities for self-reflection, and responses to formative

assessment during learning. The semi-structured format allowed the researcher to pursue common analytical issues across cases while also probing practices identified during classroom observations.

Document analysis was used to strengthen interpretation and assess whether observed practices were supported by recurring pedagogical or organizational records. The documents examined included teaching modules, lesson-related assessment notes, student worksheets, student work, feedback records, and other assessment documents available in each case. Documentary evidence was compared with observation and interview data to determine whether a practice represented an isolated classroom event or showed evidence of recurring embedding in instructional routines.

Table 2: Summary of data collection

Data source	Coverage	Amount	Main analytical focus
Classroom observations	Five cases	15 meetings; 1.050 minutes	Enactment of five formative assessment strategies
Principal interviews	Five principals	5 interviews	Supervision, policy enactment, and school support
Teacher interviews	Ten teachers	10 interviews	Planning, evidence use, feedback, peer/self-assessment
Student interviews	Twenty students	20 interviews	Learning goals, feedback, peer assessment, self-reflection
Documents	Five cases	Teaching and assessment records	Recurrence and organizational embedding of practices

Instrument Development and Expert Validation

The research instruments consisted of classroom observation guidelines, semi-structured interview guides for principals, teachers, and students, and a document analysis guide. The instruments were developed with reference to Black and Wiliam's (2009) five formative assessment strategies and aligned with the study's focus on the consistency, depth, and organizational embedding of assessment practices. Two experts in learning assessment reviewed the instruments. The validation process examined the relevance of the indicators to the formative assessment framework, the clarity of the observation items, the appropriateness of the interview questions for each participant group, and the alignment between the instruments and the research objectives. Based on the expert review, the instruments were judged suitable for use in the study. The observation guide was used to maintain consistency across classroom settings, while the semi-structured interview guides remained sufficiently flexible to allow follow-up questions in response to field conditions and case-specific practices.

Data Analysis

Data were analyzed through a framework-guided thematic cross-case procedure that combined deductive interpretation based on Black and Wiliam's (2009) five strategies with inductive attention to patterns emerging from the empirical data. The analytical process was informed by the interactive logic of data condensation, data display, and conclusion drawing described by Matthew B. Miles et al. (2014), while also incorporating explicit coding and cross-case comparison. The *first* stage involved organizing observation notes, interview data, and documents by case and data source. Data reduction then began through open coding, in which meaningful units related to formative assessment were identified and assigned provisional codes. Examples included stating learning objectives, using questions to check understanding, using tasks as evidence, verbal corrective feedback, student revision after feedback, peer comments, student reflection, assessment documentation, and supervision of assessment practice. Open coding was used to preserve empirical variation rather than forcing all data immediately into predetermined theoretical categories.

The *second* stage involved axial coding, through which related open codes were grouped, and relationships among practices were examined. The codes were then connected to the five formative assessment strategies while retaining emerging subpatterns concerning the function and depth of practice. For example, feedback was differentiated according to whether it was merely evaluative or improvement-oriented, whether students acted on it, and whether it informed revision or subsequent instructional decisions. Similarly, questioning was examined not only as a visible classroom activity but also in relation to whether student responses were interpreted and used as evidence for pedagogical adaptation. The *third* stage consisted of within-case analysis. For each school, the researchers developed an analytical matrix comparing evidence from observations, interviews, and documents across the five formative assessment strategies. This stage was used to identify the degree to which each strategy was recurrent, pedagogically substantive, supported across data sources, and connected to classroom or school routines.

The *fourth* stage involved cross-case analysis. The five case matrices were compared to identify convergent patterns, variations, contradictions, and disparities in the institutionalization of formative assessment strategies. Cross-case interpretation focused particularly on differences between strategies that were recurrently embedded in teacher routines and supervisory expectations and strategies that required greater participation by peers and learners as evaluators of learning. The final stage involved constructing cross-case themes that explained patterns of uneven institutionalization. Themes were not derived solely from frequency. Interpretations considered the recurrence of practice, the convergence of evidence across data sources, the pedagogical depth of enactment, and the extent to which the practice was embedded in classroom routines, pedagogical tools, supervision, or participant roles. Data management was conducted using structured coding matrices, case summaries, and cross-case comparison tables. No qualitative analysis software was used.

Analytical Criteria for Institutionalization

To reduce ambiguity and strengthen transparency, the institutionalization of each formative assessment strategy was examined through four analytical dimensions—recurrence, which referred to whether a practice appeared repeatedly across the three classroom observations. Source convergence referred to whether the practice was supported by more than one type of evidence, such as observation, interview, or document data. Pedagogical depth referred to whether the strategy functioned only procedurally or contributed to evidence-responsive teaching, learner revision, evaluative judgment, or self-regulation. Organizational embedding referred to whether the practice was connected to recurring classroom routines, teaching tools, supervision, documentation, or established teacher-student roles.

The categories of consistent, fairly consistent, partial, rare, and not yet implemented were used as qualitative analytical classifications rather than numerical measurements. Their operational criteria are presented in Table 3.

Table 3: Operational criteria for levels of institutionalization

Category	Compact operational criteria
Consistent	Recurrent across nearly all observations; supported by multiple data sources; embedded in routines or tools; evidence of meaningful pedagogical use
Fairly consistent	Appears repeatedly and is supported by multiple sources, but depth, follow-up, or learner involvement varies
Partial	Appears intermittently; limited source convergence; depends mainly on individual teacher initiative; weak routine embedding.
Rare	Appears only occasionally or mainly in participant reports; little observational/documentary support; minimal follow-up.
Not yet implemented	No meaningful evidence across observations, interviews, or documents

These categories were first applied within each case and were subsequently compared across cases. Their purpose was not to quantify practice but to provide a transparent basis for distinguishing between the mere presence of a formative assessment activity and its more sustained institutional embedding.

Trustworthiness

The trustworthiness of the study was strengthened through triangulation of sources, techniques, and time. Source triangulation involved comparing perspectives from principals, teachers, and students. Method triangulation involved comparing classroom observations, semi-structured interviews, and document analysis. Time triangulation was supported through repeated observations across three classroom meetings in each case. Cross-case comparison further enabled the researchers to examine whether interpretations represented isolated events or recurring patterns across institutional contexts. Interpretive credibility was also strengthened by systematically tracing findings back to evidence from multiple sources wherever possible. Cross-case conclusions were developed only after within-case patterns had been examined, and analytical decisions were documented through coding matrices, case summaries, and category-decision records. This procedure supported an audit trail linking raw field evidence, initial codes, broader categories, within-case interpretations, and final cross-case themes.

Researcher reflexivity was addressed by recognizing the researcher's role as the primary qualitative instrument. During data collection and analysis, field notes and analytic memos were used to distinguish observed events from emerging interpretations and to record assumptions that could influence analysis. Particular attention was given to the possibility that participants might portray assessment practices positively because formative assessment is formally promoted under the Merdeka Curriculum. Interpretations were therefore compared against observed classroom practices and documentary evidence rather than relying solely on participant claims.

Methodological Integration

The overall methodological process linked purposive case selection, framework-based instrument development, expert validation, multi-source data collection, triangulation, coding, within-case analysis, and cross-case synthesis. The methodological diagram should therefore be revised to show Figure 1.

RESULT AND DISCUSSION

Result

Cross-Case Patterns of Uneven Institutionalization

The cross-case analysis showed that Black and Wiliam's (2009) five formative-assessment strategies were not embedded with comparable consistency, pedagogical depth, or organizational support across the five cases. Teacher-led practices, particularly the communication of learning

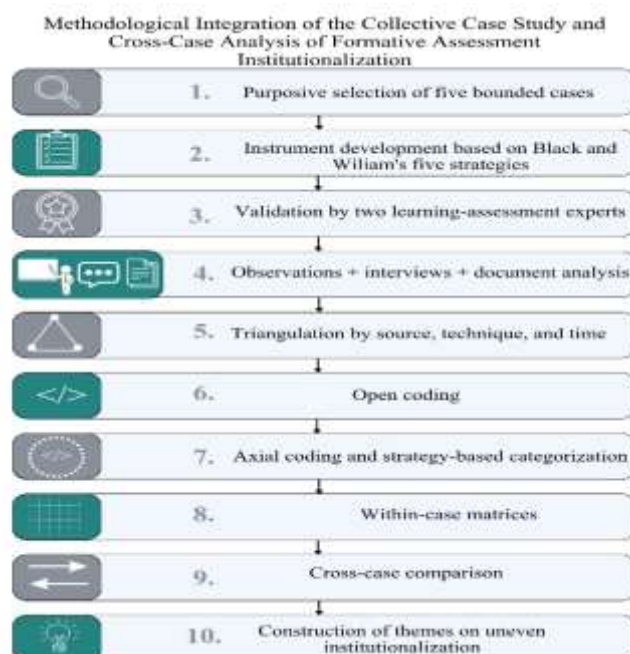


Figure 1: Methodological integration of the collective case study and cross-case analysis of formative assessment

intentions and the provision of feedback, were recurrent across observations, corroborated by multiple participant groups, and reinforced through classroom routines and academic supervision. By contrast, strategies requiring students to exercise greater evaluative responsibility, especially self-assessment, were less frequent and less well embedded. The classifications used below follow the analytical criteria described in the Method section. Consistent indicates recurrent practice supported by multiple data sources and meaningful routine embedding; fairly consistent indicates repeated enactment with variable depth or follow-up; partial refers to intermittent implementation; and rare indicates isolated or minimally supported practice. The classifications are qualitative rather than numerical.

Table 4: Cross-Case Matrix of Formative-Assessment Institutionalization

Strategy	Cross-case pattern	Classification	Analytical interpretation
Learning intentions and success criteria	Recurrent across cases; reinforced through supervision and planning documents; recognized by students	Consistent	Strong institutional visibility, but mainly teacher-communicated rather than student-used for evaluative judgment.
Eliciting evidence through questions and tasks	Open questions, discussions, and authentic tasks were recurrent; use of evidence for instructional adaptation varied.	Fairly consistent	Evidence generation is more embedded than evidence interpretation and adaptive action.
Feedback that moves learning forward	Oral, written, and symbolic feedback were recurrent; students revised their work and reported improved understanding.	Consistent, with uneven depth	Strong teacher-mediated embedding, but limited redistribution of feedback agency
Peer assessment	Present mainly in presentations and discussions; student engagement is high, but quality varies.	Fairly consistent/partial in depth	Social participation is stronger than criterion-referenced evaluative judgment.
Self-assessment	Structured reflection is uncommon; students reported rarely assessing their own learning.	Rare overall	Weakest organizational embedding and clearest limitation in learner evaluative agency

The matrix shows two forms of unevenness. First, the strategies differed in recurrence and organizational visibility. Second, they differed in pedagogical depth, because the practices they frequently enacted did not always produce evidence-responsive adaptation or greater learner responsibility.

Learning Intentions Were Consistently Embedded but Predominantly Teacher-Directed

The clarification of learning intentions was the most recurrent practice across the cases. Teachers typically stated objectives orally, wrote them on the board, or displayed them digitally. In Case D, a teacher explained: *"I state the learning objectives orally and write them on the board so that students understand the direction of the activities."* (Teacher, Case D). The practice was also reinforced through supervision. The principal of Case A stated: *"We check lesson plans and conduct classroom supervision to ensure that teachers communicate learning objectives and achievement indicators."* (Principal, Case A). Student data confirmed that learners recognized the practice: *"Our teacher always communicates the learning objectives at the beginning, so we understand what we are going to learn that day and what we are expected to achieve in the lesson."* (Student participant). The convergence of teacher, principal, student, observation, and document evidence justified classifying the strategy as *consistent*. However, the depth of enactment remained predominantly teacher-directed. Students understood and communicated goals, but there was less evidence that they independently used success criteria to judge the quality of their own work or determine subsequent action. Thus, learning intentions were strongly institutionalized as a visible teaching routine but less deeply embedded as a tool for student evaluative judgment.

Evidence Elicitation Was Recurrent, but Adaptive Use Varied

Teachers across cases used open questions, group discussions, reflective prompts, and authentic tasks to make student understanding visible. In Case E, one teacher stated:

"I often use open-ended questions to elicit students' answers and see what they understand." (Teacher, Case E). In Case D, teachers used posters and simple reports as tangible evidence of learning. School support was also evident. The principal of Case E explained: *"We supervise instruction to ensure that teachers design activities that produce evidence of learning, such as discussions and projects."* (Principal, Case E). Observations confirmed that questioning and discussion were recurrent. However, cross-case comparison revealed a distinction between eliciting evidence and acting on evidence. In some classrooms, student responses informed subsequent teaching decisions; in others, questions mainly stimulated participation or checked whether students were following the lesson. Teacher familiarity appeared to shape this variation. A teacher in Case B noted: *"I have often used prompting questions and tasks that generate authentic evidence of learning, so I feel quite familiar with these strategies."* (Teacher, Case B). The strategy was therefore classified as *fairly consistent*: evidence-generating techniques were recurrent, but the interpretation of evidence and its translation into adaptive action were less consistently embedded. This indicates uneven institutionalization even within a single strategy.

Feedback Was Consistently Embedded but Remained Predominantly Teacher-Controlled

Feedback was one of the most recurrent practices across cases. Teachers used oral comments, written comments, and symbols, generally emphasizing improvement rather than numerical scores. In Case D, a teacher stated: *"I provide brief comments and symbols to indicate the parts that need improvement."* (Teacher, Case D). The principal of Case A similarly explained: *"Teachers always provide feedback during lessons and emphasize suggestions for improvement rather than scores."* (Principal, Case A). Observational evidence showed that students corrected or revised work after the teacher's comments. Student testimony further strengthened this finding: *"The feedback provided by the teacher helps us understand our mistakes when solving problems during the lesson and helps us improve the way we learn for the next meeting."* (Student participant). This quotation indicates that students perceived feedback as useful both for identifying current errors and improving future learning. However, the depth of feedback was constrained by time and class size. A teacher in Case A stated: *"There are many students and limited time, so it is difficult to provide in-depth feedback to each student. Often, the feedback I give is only general."* (Teacher, Case A). Feedback was therefore classified as consistent in recurrence and routine embedding but uneven in pedagogical depth. Teachers remained the principal producers and interpreters of feedback. Students used teacher feedback, but there was weaker evidence that they independently generated evaluative judgments or determined improvement strategies. Feedback was thus strongly institutionalized as a teacher-mediated mechanism for improvement, but less strongly as a mechanism for developing learner feedback agency.

Peer Assessment Increased Participation but Showed Limited Evaluative Depth

Peer assessment was most visible during group presentations and class discussions. Students were invited to comment on peers' work, and observations indicated active participation. In Case E, one student explained: *"I am happy to give feedback to my friends because I feel valued, and it makes me more confident in learning."* (Student, Case E). School leaders supported this practice, but cross-case evidence showed variation in the quality of peer comments. The principal of Case A identified classroom management and students' ability to provide objective feedback as recurring challenges.

Students were generally willing to participate, but their judgments were not consistently linked to explicit success criteria. Peer assessment, therefore, functioned more successfully as a participatory and social routine than as a structured process of criterion-referenced judgment. It was classified as fairly consistent in recurrence but partial in pedagogical depth. The main limitation was not student willingness but insufficient scaffolding for judging quality, justifying comments, and using explicit criteria.

Self-Assessment Remained the Least Embedded Strategy

Self-assessment showed the clearest evidence of weak institutionalization. Although principals and teachers recognized its importance, structured opportunities for students to evaluate their own understanding were uncommon. In Case D, a teacher had attempted to use a simple reflection sheet, whereas in Case B, self-reflection remained oral: *"My reflection is only question-and-answer; there is no structured self-assessment activity yet."* (Teacher, Case B). Student data converged with this pattern. Students reported that they were rarely asked to assess their own learning outcomes. Observations similarly showed little evidence of routines in which students compared their work with explicit criteria, identified learning gaps independently, or determined their own next steps.

The weak enactment of self-assessment did not reflect rejection of its value. Principals acknowledged its importance for student independence but also recognized that implementation remained uneven across classrooms. The strategy was therefore classified as rare overall, with only isolated partial practices. The contrast with the previous themes is analytically important. Students reported understanding teacher-communicated objectives and benefiting from teacher feedback, yet they had substantially fewer opportunities to evaluate their own learning. Thus, students were positioned more strongly as users of teacher judgment than as producers of evaluative judgment.

Contextual Conditions Shaping Uneven Institutionalization

Cross-case analysis indicated that the interaction of teacher experience, academic supervision, curricular guidance, student participation, time constraints, summative pressures, and limited specialized training shaped uneven institutionalization. Teacher experience supported routine enactment. The principal of Case A observed: *"Teachers who are already accustomed to formative assessment can implement it more easily; they know how to communicate learning objectives, provide feedback, and assess students' work without needing close supervision."* (Principal, Case A). Academic supervision also reinforced visible and documentable practices. The principal of Case A explained: *"We conduct academic supervision routinely, both by examining documents such as lesson plans or teaching modules and through direct classroom supervision."* (Principal, Case A). However, supervision aligned most readily with practices that could be observed or documented, such as learning objectives, planned activities, and teacher feedback. There was less evidence that it generated equally structured routines for learner-led self-assessment.

Student enthusiasm supported peer participation but did not ensure evaluative quality. Conversely, time pressure, large classes, and summative priorities reduced pedagogical depth. A teacher in Case B stated: *"We focus more often on summative assessment, such as the end-of-semester summative assessment. Formative assessment is conducted, but it is often not the priority because much of our time and energy is spent preparing final tests"* (Teacher, Case B). Limited specialized training further constrained deeper enactment. The principal of Case E noted: *"Although there is curriculum training, training to understand formative-assessment techniques in depth is still very limited."* (Principal, Case E). These conditions affected the strategies unevenly. Practices compatible with established teacher routines and curricular guidance were easier to sustain, whereas adaptive evidence use, criterion-referenced peer assessment, and structured self-assessment required more specialized knowledge and sustained scaffolding.

Cross-Theme Synthesis: From Institutional Visibility to Uneven Evaluative Agency

The central cross-case finding was not simply that some formative-assessment strategies occurred more frequently than others. Rather, the five strategies differed systematically in recurrence, pedagogical depth, organizational embedding, and the distribution of evaluative responsibility. Learning intentions and teacher feedback showed the strongest recurrence and were reinforced by classroom routines, curricular tools, and supervisory practices. Student accounts confirmed that these practices were meaningful: students reported understanding the direction of learning when objectives were communicated and using teacher feedback to recognize errors and improve subsequent learning.

By contrast, questioning, authentic tasks, and peer assessment occupied an intermediate position. These practices were recurrent across cases, but their pedagogical depth varied. Evidence elicited through questions and tasks was not always translated into adaptive instructional action, while peer assessment generated active participation but was not consistently anchored in explicit success criteria. Self-assessment showed the weakest embedding. Students reported that they were rarely asked to evaluate their own learning, and classroom observations revealed few structured opportunities for learners to compare their performance against criteria, identify gaps independently, or determine subsequent learning actions.

Across the themes, a consistent relational pattern emerged. Students were primarily positioned as recipients of learning intentions, respondents to teacher questions, users of teacher feedback, and occasional commentators on peers' work, but only weakly as independent evaluators of their own learning. This asymmetry was evident across interviews, observations, and documentary evidence. The cross-case pattern, therefore, indicates uneven institutionalization across the five strategies. Strategies aligned with established teacher routines, curricular tools, and supervisory visibility showed stronger recurrence and organizational embedding. Strategies requiring a greater redistribution of evaluative responsibility toward students, particularly structured self-assessment, remained comparatively weak. Figure 2 summarizes this relationship among institutional visibility, pedagogical depth, and evaluative agency.

Discussion

The findings demonstrate that the implementation of formative assessment under Indonesia's Merdeka Curriculum is characterized not by the absence of formative assessment practices, but by their uneven institutionalization across the five strategies outlined by Black & Wiliam (2009). Cross-case analysis revealed that learning intentions, evidence elicitation, and teacher feedback were recurrently embedded across schools, whereas peer assessment showed variable quality and self-assessment remained marginal. This pattern suggests that formative assessment has not developed as a fully integrated pedagogical system. Instead, different strategies have followed distinct trajectories of institutionalization, reflecting variations in institutional visibility, teacher assessment literacy, and the degree of student evaluative agency required for their enactment. These findings extend the current formative assessment literature by shifting attention from whether formative assessment is implemented to how different formative assessment strategies are differentially embedded within curriculum reform contexts.

A central finding of this study is that strategies with high institutional visibility were more consistently institutionalized than strategies requiring greater learner agency. Learning intentions and teacher feedback emerged as the most stable practices across all five cases because they were strongly aligned with lesson planning requirements, classroom routines, and academic supervision. Principals reported regularly reviewing teaching modules and observing whether teachers communicated learning objectives and provided feedback during instruction. As a result, these practices became highly visible indicators of curriculum implementation and were repeatedly reinforced through organizational mechanisms.

This pattern suggests that institutionalization is shaped not only by pedagogical value but also by the extent to which practices are observable, documentable, and monitorable. The findings support previous research indicating that curriculum reform often privileges practices that can be readily incorporated into existing accountability structures (Anderson et al., 2001; Aust et al., 2025). However, the present study extends this argument by showing that visibility itself functions as a mechanism of differential institutionalization. Strategies such as communicating learning objectives and providing teacher feedback were easier to supervise and evaluate than self-assessment practices, which depend largely on students' internal evaluative processes and therefore remain less visible to organizational monitoring systems.

Consequently, the stronger institutionalization of learning intentions and feedback does not necessarily indicate a deeper transformation of assessment practices. Rather, it reflects the tendency for curriculum reforms to embed practices compatible with existing institutional routines more rapidly than those that require substantial changes in classroom roles and evaluative relationships.

The findings also reveal an important distinction between increased assessment activity and increased learner agency. Across cases, students consistently reported understanding learning objectives and benefiting from teacher feedback. Teachers regularly used questions, discussions, and authentic tasks to generate evidence of learning. Nevertheless, students rarely engaged in structured self-assessment and had limited opportunities to evaluate their learning progress independently. This pattern suggests that the Merdeka Curriculum has increased the visibility of formative assessment practices but has not yet fully transformed the distribution of evaluative responsibility within classrooms. Students were primarily positioned as recipients of learning goals, respondents to teacher questions, and users of teacher-generated feedback rather than as active evaluators of their own learning. Thus, formative assessment predominantly functioned as assessment for learning rather than as assessment for learning.

This finding is particularly significant because the Merdeka Curriculum explicitly promotes learner-centered, differentiated, and adaptive learning. However, the results indicate that learner-centered policy aspirations do not automatically translate into learner-centered assessment practices. Similar tensions have been reported internationally, where curriculum reforms often increase formative assessment activities without substantially strengthening student ownership of learning (Lui & Andrade, 2022; McCarthy et al., 2025; Nieminen & Boud, 2025; Zimmerman & J, 2010). The Indonesian case, therefore, reflects a broader international challenge: reforms may alter assessment procedures more rapidly than they alter assessment authority.

The persistence of teacher-centered assessment can be understood as a consequence of existing educational traditions. Teachers remain the primary interpreters of learning evidence, providers of feedback, and decision-makers regarding subsequent instructional actions. Although students participate in formative assessment activities, their role as evaluative agents remains comparatively weak. This finding reinforces the argument that developing learner agency requires deliberate pedagogical design rather than merely increasing the frequency of formative assessment practices. The weak institutionalization of self-assessment supports Carless's argument that learner-centered assessment requires deliberate opportunities for students to develop evaluative judgment rather than merely receive information about performance (Carless, 2015). Similarly, Carless & Boud (2018) emphasize that feedback becomes formative only when learners possess the feedback literacy needed to interpret and act upon evaluative information, a condition that remained only partially evident across the cases examined in this study.

The findings further suggest that teacher assessment literacy functions as a critical mediating mechanism between curriculum policy and classroom implementation. While teachers routinely generated evidence through questioning, discussion, and authentic tasks, the use of that evidence for instructional adaptation varied considerably across cases. This finding resonates with Heritag's (2010) conception of formative assessment as a cyclical process in which evidence must be interpreted and translated into subsequent instructional action. From this perspective, the collection of evidence alone is insufficient; its formative value depends on how teachers and learners use evidence to determine appropriate next learning steps. Similarly, feedback practices were widespread, yet time limitations, class size, and limited opportunities for individualized support constrained their depth and follow-up.

Principals and teachers repeatedly identified limited professional development related specifically to formative assessment as a significant challenge. Although teachers demonstrated familiarity with visible formative assessment practices, there was less evidence of expertise in facilitating criterion-referenced peer assessment, developing self-assessment routines, or using

assessment evidence to support adaptive instruction. This finding supports previous studies that emphasize that effective formative assessment depends not only on the collection of evidence but also on teachers' capacity to interpret and act on that evidence pedagogically (Andersson & Palm, 2018; Aust et al., 2025; Erika Bostrom & Palm, 2023; Zhang et al., 2026). It also aligns with Carles's (2015) argument that learner-centered assessment requires deliberate opportunities for students to develop evaluative judgment rather than merely receive performance information. Likewise, Carless & Boud (2018) emphasize that feedback becomes genuinely formative only when learners possess sufficient feedback literacy to interpret and use evaluative information. The limited institutionalization of self-assessment observed across the cases, therefore, suggests that students' evaluative judgment and feedback literacy remain insufficiently supported within current classroom practices. This interpretation is further supported by Panadero (2017), who found that self-assessment interventions contribute positively to students' self-regulated learning and self-efficacy. From this perspective, the limited institutionalization of self-assessment may not only constrain students' opportunities to evaluate their own learning but also restrict the development of broader self-regulatory capacities that underpin learner agency.

The results also resonate with recent research on the implementation of the Merdeka Curriculum, which highlights the increasing complexity of the assessment demands placed upon teachers. The curriculum requires teachers to conduct multiple forms of assessment while simultaneously supporting differentiated learning and student competency development. Under these conditions, teachers may prioritize formative assessment practices that are easier to implement within existing routines, while struggling to sustain practices that require more sophisticated assessment expertise. Consequently, assessment literacy emerges as a key explanatory factor for understanding why evidence generation becomes more institutionalized than evidence interpretation and adaptive pedagogical action.

The Indonesian findings contribute to broader international discussions on the implementation of formative assessment within curriculum reform. Similar challenges have been observed in other educational systems. For example, research from Ghana reported that educational assessment remains dominated by assessment-of-learning traditions, limiting the development of assessment-for-learning approaches and reducing opportunities for equitable learner progression. In contrast, the Indonesian cases demonstrate that assessment-for-learning practices have become increasingly visible within classroom routines. However, assessment-as-learning remains comparatively underdeveloped, particularly through the limited institutionalization of self-assessment.

The findings also align with recent studies emphasizing the importance of explicit criteria and structured evaluative processes. Agarwal et al. (2026) found that criterion-based assessment and feedback systems improved transparency, reliability, and students' understanding of learning expectations. Similarly, the present study found that peer assessment frequently lacked explicit success criteria, leading to participation without consistently producing high-quality evaluative judgments. This suggests that learner participation alone is insufficient; effective formative assessment requires explicit scaffolding that enables students to interpret quality criteria and use them to regulate learning.

Furthermore, recent work on Assessment for Learning instruments in science education demonstrates that formative assessment becomes more effective when deliberately designed to support higher-order thinking and student engagement. The Indonesian cases reveal that while formative assessment practices are increasingly present, their implementation remains more strongly oriented toward supporting teacher decision-making than developing students' independent evaluative capacities. Consequently, the findings contribute to international evidence indicating that expanding formative assessment practices does not automatically strengthen learner agency.

The primary theoretical contribution of this study is the development of uneven institutionalization as an interpretive lens for understanding the implementation of formative assessment under curriculum reform. Existing literature often assumes that Black and Wiliam's five strategies operate as an integrated architecture of formative assessment. However, the present findings suggest that the five strategies do not necessarily develop in parallel.

Based on the cross-case evidence, this study proposes that the institutionalization of formative assessment is shaped by the interaction of three factors: institutional pressure, teacher assessment literacy, and student evaluative agency. Strategies characterized by high institutional visibility and compatibility with established teacher routines, such as communicating learning objectives and providing feedback, are more readily reinforced through supervision, documentation, and organizational expectations. Consequently, they become institutionalized more rapidly. In contrast, strategies requiring stronger assessment literacy and greater redistribution of evaluative responsibility toward students, particularly peer and self-assessment, experience slower and more fragile institutionalization (Leenknecht et al., 2021).

This conceptual proposition extends Black and Wiliam's framework by demonstrating that formative assessment strategies may follow differentiated developmental trajectories within the same reform environment. The findings, therefore, suggest that learner-centered assessment reform should not be understood as a uniform implementation process. Rather, it is a process of uneven institutionalization in which some strategies become procedurally visible and organizationally embedded before others become pedagogically transformative. Understanding this process provides a more nuanced explanation of why formative assessment reforms frequently produce visible changes in classroom practice while generating more limited changes in student agency and self-regulation.

CONCLUSION

This study demonstrates that the implementation of formative assessment under Indonesia's Merdeka Curriculum is characterized by uneven institutionalization across Black and Wiliam's (2009) five formative assessment strategies. While learning intentions, evidence elicitation, feedback, and, to a lesser extent, peer assessment have become recurrently embedded in classroom routines, self-assessment remains weakly institutionalized. The findings indicate that strategies aligned with established teacher practices, academic supervision, and curricular documentation are more readily embedded than those that require stronger student evaluative agency. Theoretically, the study extends Black and Wiliam's framework by showing that the five strategies do not necessarily develop as a unified system during curriculum reform but instead follow differentiated trajectories of institutionalization shaped by the interaction of institutional pressure, teacher assessment literacy, and learner agency. This contribution reframes formative assessment implementation from a binary question of adoption toward an understanding of how interconnected strategies become embedded with varying levels of depth and sustainability in developing-country contexts.

The findings suggest that strengthening formative assessment under the Merdeka Curriculum requires more than increasing assessment activities; it requires systematic efforts to enhance teachers' assessment literacy, redesign academic supervision to focus on evidence-informed instructional decision-making, and integrate structured self-assessment practices into teaching materials and classroom routines. At the same time, the study is limited by its focus on five cases within a single primary-education context and observations conducted over one semester, which constrain the transferability of the findings. Future research should therefore employ longitudinal, comparative, and intervention-based designs to examine how formative assessment strategies evolve across different contexts and how targeted professional development and learner-agency interventions can support more balanced institutionalization of formative assessment practices. The

study proposes that formative assessment strategies characterized by high institutional visibility and compatibility with existing teacher routines are more rapidly institutionalized than those that require stronger learner agency and self-regulation.

ACKNOWLEDGMENT

The author would like to thank Universitas Islam Negeri Sulthan Thaha Saifuddin Jambi for providing research funding for the 2025 fiscal year through the Institute for Research and Community Service of Universitas Islam Negeri Sulthan Thaha Saifuddin Jambi, which enabled this research to be successfully conducted.

BIBLIOGRAPHY

- Aditomo, A. (2024). *Panduan Pembelajaran dan Asesmen Pendidikan Anak Usia Dini, Pendidikan Dasar, dan Pendidikan Menengah Edisi Revisi Tahun 2024*. Jakarta: Badan Standar, Kurikulum, Dan Asesmen Pendidikan Kemendikbudristek.
- Agarwal, P., Bhandari, B., Gupta, V., Agarwal, R., & Gupta, R. K. (2026). Improving consistency and feedback in essay-type assessments: evaluation of an assessment-cum-feedback checklist. *BMC Medical Education*, 26(1), 565. <https://doi.org/10.1186/s12909-026-08892-7>
- Ramdhani M, A, & Isom, M (2022). *Panduan Implementasi Kurikulum Merdeka Pada Madrasah* (1st ed.). Direktorat KSKK Madrasah Direktorat Jenderal Pendidikan Islam Kementerian Agama.
- Almutawa, H. (2025). Formative assessment of pre-service English teachers' perceptions of classroom management skills in Kuwait: a longitudinal study. *Language Testing in Asia*, 15(1), 46. <https://doi.org/10.1186/s40468-025-00382-9>
- Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., Rath, J., & Wittrock, M. C. (2001). A Taxonomy for Learning, Teaching, and Assessing—A Revision of Bloom's Taxonomy of Educational Objectives. In *Longman*. Addison Wesley Longman, Inc.
- Andersson, C., & Palm, T. (2018). Reasons for teachers' successful development of a formative assessment practice through professional development—a motivation perspective. *Assessment in Education: Principles, Policy & Practice*, 25(6), 576-597. <https://doi.org/10.1080/0969594X.2018.1430685>
- Aust, L., Linker, J. C., Eichholz, L., Schiffer, J., Nührenbörger, M., Selter, C., & Souvignier, E. (2025). Implementing Formative Assessment Into School Practice: A Matter of Structuring the Intervention? *Journal of Teacher Education*. <https://doi.org/10.1177/00224871251350680>
- Arvatz, A., Hadas, B., Waitzman, R., & Dori, Y. J. (2025). Putting self-regulated learning and teaching into practice: insights from two science teachers and their students: A. Arvatz et al. *Instructional Science*, 53(5), 973-1003. <https://doi.org/10.1007/s11251-025-09719-6>
- Baissane, O., & Zaid, H. (2025). Bridging the gap: Exploring self-regulated learning and academic achievement in Moroccan higher education. *Journal of Interdisciplinary Studies in Education*, 14(4), 139-152. <https://doi.org/10.32674/kmf9bx98>
- Beesley, A. D., Clark, T. F., Dempsey, K., & Tweed, A. (2018). Enhancing formative assessment practice and encouraging middle school mathematics engagement and persistence. *School science and mathematics*, 118(1-2), 4-16. <https://doi.org/10.1111/ssm.12255>
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 5-31. <https://doi.org/10.1007/s11092-008-9068-5>

- Budiono, A. N., & Hatip, M. (2023). Asesmen pembelajaran pada kurikulum merdeka. *Jurnal Axioma: Jurnal Matematika Dan Pembelajaran*, 8(1), 109-123. <https://doi.org/10.56013/axi.v8i1.2044>
- Carless, D. (2015). Exploring learning-oriented assessment processes. *Higher Education*, 69(6), 963–976. <https://doi.org/10.1007/s10734-014-9816-z>
- Carless, D., & Boud, D. (2018). The development of student feedback literacy: enabling uptake of feedback. *Assessment and Evaluation in Higher Education*, 43(8), 1315–1325. <https://doi.org/10.1080/02602938.2018.1463354>
- Erika Bostrom, & Palm, T. (2023). The Effect of a Formative Assessment Practice on Student Achievement in Mathematics. *Frontiers in Education*. <https://doi.org/10.3389/feduc.2023.1101192>
- Goertzen, L., Heeneman, S., & Schils, T. (2025). The impact of formative assessment on pupil academic achievement: An empirical study of the effects of a formative assessment practices program that was co-designed in a teacher community. *Learning and Instruction*, 99(May), 102153. <https://doi.org/10.1016/j.learninstruc.2025.102153>
- Gwen, M. C. M., Nagaratnam, S., Chew, S. Y., Hanefar, S. B. M., & Naz, F. A. (2026). Teacher and Peer Feedback in Collaborative Learning: Perceived Psychological Ownership, Perceived Quality, and Perceived Learning Outcomes. *Journal of Interdisciplinary Studies in Education*, 15(2), 291–318. <https://doi.org/10.32674/wx807j32>
- Heritage, M. (2010). Formative assessment: Making it happen in the classroom. *Formative Assessment: Making It Happen in the Classroom*, 1–144. <https://doi.org/10.4135/9781452219493>
- Ihsan, M., & Maemonah. (2023). Penilaian Formatif dalam Kurikulum Merdeka pada Pembelajaran Pendidikan Agama Islam di Kelas IV SDN Sekumpul 1 Martapura. *Tarbiyah Islamiyah: Jurnal Ilmiah Pendidikan Agama Islam*, 13(2). <https://doi.org/10.18592/jtipai.v13i2.9872>
- Fatmawati, K. (2024). The Dynamics of Merdeka Curriculum Change: Challenges and Opportunities in Private Madrasah Ibtidaiyah Management. *MUDARRISA: Jurnal Kajian Pendidikan Islam*, 16(1), 49-76. <https://doi.org/10.18326/mudarrisa.v16i1.918>
- Leenknecht, M., Wijnia, L., Köhlen, M., Fryer, L., Rikers, R., & Loyens, S. (2021). Formative assessment as practice: the role of students’ motivation. *Assessment and Evaluation in Higher Education*, 46(2), 236–255. <https://doi.org/10.1080/02602938.2020.1765228>
- Lui, A. M., & Andrade, H. L. (2022). The Next Black Box of Formative Assessment: A Model of the Internal Mechanisms of Feedback Processing. *Frontiers in Education*, 7. <https://doi.org/10.3389/feduc.2022.751548>
- Maleki, A. (2025). Voices from the virtual classroom: students’ views toward ethical challenges in the context of online formative assessment. *Discover Education*, 4(1). <https://doi.org/10.1007/s44217-025-00445-2>
- Matthew B. Miles, A. Michael Huberman, & Johnny Saldana. (2014). *Qualitative-Data-Analysis A Method Sourcebook* (Helen Salmon, Ed.; Third). SAGE Publications, Inc.
- McCallum, S., & Milner, M. M. (2021). The effectiveness of formative assessment: student views and staff reflections. *Assessment and Evaluation in Higher Education*, 46(1), 1–16. <https://doi.org/10.1080/02602938.2020.1754761>
- McCarthy, N., Neville, K., & Pope, A. (2025). Feedback and formative assessment—looking backward to move forward. *Discover Education*, 4(1). <https://doi.org/10.1007/s44217-025-00414-9>
- Nieminen, J. H., & Boud, D. (2025). Placing authenticity at the heart of student self-assessment: an integrative review. *Teaching in Higher Education*, 30(3), 640–662. <https://doi.org/10.1080/13562517.2024.2367669>
- Owen, L. (2016). The Impact of Feedback as Formative Assessment on Student Performance. *International Journal of Teaching and Learning in Higher Education*, 28(2), 168–175. <http://www.isetl.org/ijtlhe/>

- Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for research. *Frontiers in Psychology*, 8(APR), 1–28. <https://doi.org/10.3389/fpsyg.2017.00422>
- Tomasine, J. S. (2024). The multimodal organization of feedback and its documentation: multiactivity during formal, formative reading assessment. *Frontiers in Communication*, 9. <https://doi.org/10.3389/fcomm.2024.1390843>
- van Kessel, M., Knoop-van Campen, C. A. N., de Jonge, M., Molenaar, I., & Saab, N. (2026). Feedback practices with teacher dashboards in primary education: Exploring dashboard-prompted feedback across lesson phases. *Computers and Education*, 249(April 2025), 105615. <https://doi.org/10.1016/j.compedu.2026.105615>
- Welsh, M. (2012). Student perceptions of using the PebblePad e-portfolio system to support self- and peer-based formative assessment. *Technology, Pedagogy and Education*, 21(1), 57–83. <https://doi.org/10.1080/1475939X.2012.659884>
- Zhang, L. X., Yan, Z., & Wang, X. (2026). Mapping formative assessment in higher music education: A scoping review of its implementation and impact on musical intelligence. *Research Studies in Music Education*. <https://doi.org/10.1177/1321103X251400571>
- Zikriati, Z., Safitri, A., Zuhendra, D., Nurazan Mohmad, R., & Fadillah, M. R. (2024). Implementation of the Merdeka Curriculum at Madrasah Tsanawiyah in Aceh Tamiang, Indonesia. *Tarbawi: Jurnal Keilmuan Manajemen Pendidikan*, 10(02), 261–278. <https://doi.org/10.32678/tarbawi.v10i02.10690>
- Zimmerman, J.B. (2010). Becoming a Self-Regulated Learner: An Overview Becoming a Self-Regulated Learner: An Overview. *Theory Into Practice*, 5841(2002), 64–70. <https://doi.org/10.1207/s15430421tip4102>

