



Formative Assessment Practices and Students' Self-Regulation and Social-Emotional Development: An Interpretive Phenomenological Study with Historical Analysis

Syamsul Falaq^{1*}, Ival Syaputra¹, Sulaiman¹, Desi Aramiko¹, Pratiwi Syabila¹

Samudra University¹, Langsa, Indonesia

Email: syamsulfalaq030@gmail.com^{*1}

Article History

Submitted: 2026-07-28

Revised: 2026-08-01

Accepted: 2026-08-15

Published: 2026-08-15

Keywords:

educational history; formative assessment; primary school; self-regulation; social-emotional development; interpretive phenomenology

Abstract

This study explores how elementary school teachers and students experience formative assessment practices in relation to self-regulation and social-emotional development, situated within the historically shaped assessment culture of Indonesian primary education. Using a qualitative approach, this study employed an interpretive phenomenological design complemented by historical document analysis. Data were collected through semi-structured interviews with 8 classroom teachers and 12 students in grades IV and V, non-participant classroom observations across 20 learning sessions (10 sessions in each of two primary schools), and documentation of learning artifacts and curriculum policy documents. Interview and observational data were analyzed using hybrid inductive–deductive thematic analysis, while policy documents were analyzed through historical method stages. The findings reveal four main themes: (1) feedback as a mirror of learning, (2) self-assessment as a builder of autonomy, (3) social interaction in the assessment process, and (4) structural and cultural challenges. The findings suggest that formative assessment practices were perceived to support students' self-regulatory capacities and social-emotional competencies, particularly when feedback was process-oriented and when peer interaction was structured. However, implementation was constrained by curriculum density, time limitations, and an examination-oriented assessment culture that participants' experiences may reflect within a longer historical continuity. This research extends the application of Self-Regulated Learning and Self-Determination Theory by illustrating how the assessment process may function as a context for developing social-emotional skills in upper-primary classrooms. Practical implications include the need for differentiated assessment literacy programs and context-sensitive policy adaptation.



© 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution ShareAlike 4.0 International License (CC-BY-SA-4.0)

INTRODUCTION

Formative assessment remains one of the most promising yet persistently challenging pedagogical innovations in Indonesian primary education. Despite curriculum policies since the 2004 Competency-Based Curriculum (KBK) and the 2006 School-Based Curriculum (KTSP) that emphasize process-oriented assessment, classroom practice continues to exhibit significant gaps. Teachers frequently rely on written tasks and daily tests as primary indicators of student achievement, while opportunities for self-assessment, peer feedback, and reflective practice remain limited (Yolanda & Rodliyah, 2024). This disjuncture between policy intent and classroom reality raises fundamental questions about the cultural, structural, and historical conditions that shape assessment practices in Indonesian elementary schools.

The significance of this issue extends beyond technical measurement. Assessment practices are not neutral; they carry implicit messages about the nature of knowledge, the role of the learner, and the purpose of schooling. When assessment is oriented exclusively toward final outcomes, students are positioned as passive recipients of external judgment rather than active agents in their own learning.

Conversely, when formative assessment is implemented with fidelity, it creates spaces for students to monitor their progress, adjust their strategies, and develop ownership over the learning process capabilities that lie at the heart of self-regulated learning (Zimmerman, 2000) and intrinsic motivation (Ryan & Deci, 2020).

This study integrates two complementary theoretical lenses. Self-Regulated Learning (SRL) theory, as articulated by Zimmerman (2000), conceptualizes learning as a cyclical process involving three phases: forethought (planning and goal setting), performance (monitoring and strategy use), and self-reflection (evaluation and attribution). Formative assessment is theorized to support this cycle by providing information that enables students to monitor their progress against goals and adjust their learning strategies accordingly (Li & Gu, 2024).

Self-Determination Theory (SDT) (Ryan & Deci, 2020) complements SRL by emphasizing three basic psychological needs: autonomy (the need to experience volition and self-endorsement), competence (the need to feel effective in one's activities), and relatedness (the need to feel connected to and cared for by others). Formative assessment practices particularly those involving feedback, self-assessment, and peer interaction may satisfy these needs when they provide students with meaningful choice, information about their growing capabilities, and opportunities for social connection. The integration of these theories suggests the following conceptual logic: formative assessment generates process-oriented information → students engage in monitoring and reflection (SRL) → perceived satisfaction of autonomy, competence, and relatedness (SDT) → enhanced engagement and social-emotional functioning.

In this study, social-emotional development is defined specifically along five dimensions drawn from established social-emotional learning frameworks: (a) emotional regulation (the ability to manage emotional responses), (b) empathy (the capacity to understand others' perspectives), (c) cooperation (the ability to work collaboratively toward shared goals), (d) interpersonal communication (skills in expressing and receiving ideas respectfully), and (e) relatedness (the experience of positive connection with peers and teachers). This delimitation is necessary because "social-emotional development" is a broad construct, and precise operationalization strengthens analytical focus.

Internationally, research on formative assessment has predominantly focused on cognitive and academic outcomes. The seminal synthesis by Black and Wiliam (1998) established that formative assessment can produce learning gains across diverse contexts, with subsequent meta-analyses confirming medium-to-large effect sizes on academic achievement (Li & Gu, 2024). More recent work has begun examining formative assessment in relation to self-regulated learning, showing that teacher professional development in formative assessment literacy can improve students' planning, monitoring, and reflection capabilities (Dignath & Veenman, 2021; Li & Gu, 2024).

However, three significant gaps remain. First, most previous studies have employed quantitative or mixed-methods designs that measure outcomes without adequately capturing the subjective experiences of primary-grade students. Qualitative investigations that trace how students interpret feedback, negotiate assessment criteria, and experience peer judgment remain rare, particularly in non-Western contexts. Second, the social-emotional dimensions of formative assessment remain under-theorized. While peer assessment has been shown to provide additional perspectives on work quality, its function as a medium for developing communication, empathy, and cooperation has received limited empirical attention. Third, existing literature has largely treated formative assessment as a decontextualized technical intervention, paying insufficient attention to how historically sedimented assessment cultures constrain or enable particular practices. In Indonesia, where the basic education system has evolved through colonial, post-independence, and reform-era periods, understanding the historical layering of assessment paradigms is essential for interpreting contemporary classroom realities.

The Indonesian education system has undergone successive transformations that have left enduring traces on assessment culture. During the colonial period, education served administrative and labor needs, with selective mechanisms determining access to further schooling (Larasati & Mighwar, 2023). The post-independence 1947 Lesson Plan (Rentjana Pelajaran) sought to cultivate independent national character, yet assessment remained oriented toward final examinations. The 1975 curriculum introduced the Instructional System Development (PPSI) approach, emphasizing behavioral objectives and measurable outcomes a paradigm that, scholars note, carried centralized and particularistic political content (Damayanti & Zulkarnain, 2024). The national examination system, strengthened during

subsequent periods, reinforced an outcome-oriented culture in which examination results became the dominant indicator of educational success. Contemporary policies, including the 2013 Curriculum and the Independent Curriculum (Kurikulum Merdeka), have formally shifted toward process-based, student-centered assessment; yet field observations suggest that examination-oriented orientations persist.

This historical layering matters because it suggests that contemporary assessment practices cannot be understood solely through individual teacher competence or curriculum design. Rather, they are embedded in institutional memories, parental expectations, and administrative pressures that carry the weight of prior educational regimes. A historical lens thus provides necessary context for interpreting why process-oriented formative assessment encounters resistance despite policy support.

Given these gaps, this study aims to explore how teachers and students in Indonesian primary schools experience formative assessment practices in relation to self-regulation and social-emotional development, and how these experiences may be understood within the historical evolution of Indonesia's assessment culture. Specifically, this study addresses the following questions:

- 1) How do teachers and students experience and interpret formative assessment practices (feedback, self-assessment, and peer interaction) in upper-primary classrooms?
- 2) In what ways do these practices appear to relate to students' self-regulatory processes and social-emotional competencies from the perspectives of participants?
- 3) How do participants' experiences reflect continuities and tensions within the historical development of assessment policy and culture in Indonesian basic education?

METHOD

This study employed a qualitative approach with an interpretive phenomenological analysis (IPA) as the primary methodological framework, complemented by historical document analysis of curriculum policy materials. IPA was selected because it is specifically designed to explore how individuals make sense of their lived experiences within particular contexts (Smith et al., 2009). This approach is idiographic and interpretive, seeking to understand the detailed texture of participants' accounts rather than to test hypotheses or produce statistical generalizations.

The historical dimension was integrated as a contextualizing component rather than a mixed-methods parallel. Curriculum policy documents from key periods were analyzed to identify the evolution of assessment paradigms, which were then used to interpret the structural and cultural conditions within which participants' experiences unfolded. This design can be characterized as an interpretive phenomenological study with embedded historical analysis where phenomenological inquiry into lived experience remains central, and historical analysis provides an interpretive resource for situating those experiences within longer-term institutional trajectories.

Theoretical frameworks (SRL and SDT) functioned as sensitizing concepts (Blumer, 1954) rather than deductive codes. They oriented the researchers' attention toward processes of monitoring, reflection, autonomy, competence, and relatedness during data collection and initial analysis, while remaining open to emergent themes that fell outside these constructs.

In interpretive phenomenology, the researcher is explicitly recognized as an interpretive agent rather than a neutral observer. Rather than employing "bracketing" (epoche) in the transcendental phenomenological sense, the research team engaged in reflexive practice throughout the study. Before data collection, researchers documented their prior assumptions about formative assessment, including beliefs about its efficacy and concerns about implementation barriers in Indonesian schools. These positionality statements were revisited during analysis team meetings to examine how researchers' interpretive frames might be shaping coding decisions. Peer debriefing with two external experts one in primary education and one in educational history was conducted to challenge emerging interpretations and identify potential researcher bias.

The study was conducted in two public elementary schools one located in an urban district and one in a rural district during the even semester of the 2025/2026 academic year. These two schools were selected as contrasting cases (Yin, 2018) to capture potential variation in resources, professional development access, and community expectations. It is explicitly acknowledged that these two schools

do not represent all urban or rural schools in Indonesia; rather, they provide bounded contexts for in-depth phenomenological exploration.

Participants consisted of 8 classroom teachers (4 per school) and 12 students (6 per school, grades IV and V). Teachers were recruited through purposive sampling with the following criteria: (a) minimum three years of teaching experience, (b) current classroom teaching responsibility in grades IV or V, and (c) engagement in assessment practices that included at least one formative element (feedback, self-assessment, or peer assessment). Snowball sampling was then used to identify additional teachers recommended by initial participants based on their engagement with assessment innovation. This approach carries a positive-case selection bias: teachers recruited through reputation networks may hold more favorable orientations toward formative assessment than the general teacher population. This sampling logic was intentional, as the study aimed to understand the phenomenology of formative assessment when it is attempted rather than to survey prevalence. This limitation is acknowledged in Section 5.

Students were nominated by their classroom teachers based on observable variation in self-regulation and social participation during regular class activities. Teachers were asked to identify students who demonstrated a range of profiles some who frequently initiated self-monitoring behaviors, some who participated actively in group work, and some who appeared more dependent on teacher direction. This nomination was based on teachers' ongoing classroom observation rather than formal psychometric screening. Students' assent and parental consent were obtained prior to participation.

Data were collected through four techniques, with source triangulation to enhance credibility:

Semi-structured interviews. Teachers participated in individual interviews (45–60 minutes) using open-ended guides (e.g., *"Tell me how you provide feedback to students after they complete an assignment"*; *"How do you support students to assess their own work?"*). Students participated in adapted interviews (20–30 minutes) using age-appropriate language and visual prompts (e.g., *"Show me how you check your work when the teacher gives it back"*; *"What do you do when a friend asks you about their work?"*). All interviews were audio-recorded and transcribed verbatim.

Non-participant classroom observation. The researcher conducted non-participant observation across 20 learning sessions (10 sessions in each school, distributed across grades IV and V classrooms). The researcher sat at the periphery of the classroom, observing without participating in instructional activities. Observation focused on: (a) teacher feedback practices (timing, specificity, tone), (b) student self-assessment activities, (c) peer interaction during assessment tasks, and (d) non-verbal emotional and behavioral responses. Detailed field notes were recorded immediately after each session.

Learning artifact documentation. Collected materials included student reflection sheets, self-assessment rubrics, peer feedback forms, and video recordings of selected peer assessment activities. Video recordings were limited to peer interaction segments and stored on encrypted devices.

Policy document study. Curriculum and assessment policy documents were collected, including the 1947 Rentjana Pelajaran, the 1975 Curriculum, KTSP 2006, the 2013 Curriculum, and the Independent Curriculum (Kurikulum Merdeka), alongside relevant ministerial regulations on student assessment.

Interview and observational data were analyzed using hybrid inductive–deductive thematic analysis following the six-phase framework of Braun and Clarke (2006), while drawing on the iterative procedures described by Miles and Huberman (1994) for data management. The analysis proceeded as follows:

Phase 1: Familiarization. Researchers read transcripts and field notes repeatedly, noting initial impressions.

Phase 2: Open coding. Initial codes were generated inductively from the data (e.g., *"student checks work against example," "teacher asks guiding question instead of giving score"*). Simultaneously, sensitizing concepts from SRL (planning, monitoring, reflection) and SDT (autonomy, competence, relatedness) were used to flag data segments potentially relevant to theoretical constructs.

Phase 3: Theme generation. Codes were grouped into potential themes, reviewed for internal coherence and distinctiveness, and refined through team discussion.

Phase 4: Theoretical interpretation. Themes were interpreted in dialogue with SRL and SDT frameworks, examining how formative assessment practices appeared to relate to self-regulatory processes and need satisfaction.

Phase 5: Member checking. Summary thematic narratives were returned to participants in simplified language. Teachers reviewed written summaries; students participated in brief verbal member-checking sessions using simplified visual maps of themes. This process confirmed that thematic representations resonated with participants' intended meanings.

Phase 6: Report production. Final themes were narrated with supporting empirical evidence.

Policy document data were analyzed using historical method stages (Kuntowijoyo, 2013): heuristics (source collection), verification (source criticism), interpretation (synthesis of facts), and historiography (narrative construction). These analyses were used to construct the historical context for interpreting participants' accounts, not as direct phenomenological evidence.

Integration occurred at the interpretation stage: historical findings were used to contextualize and deepen the analysis of contemporary experiences, while contemporary experiences were used to illustrate how historical structures are inhabited in present practice.

Table 1. Example of Hybrid Inductive–Deductive Coding Process

Raw Data Segment	Initial Code	Category	Theme	Theoretical Interpretation
"I asked the children to look back at their work before I give grades. I asked, 'Which parts do you think are good and which need improvement?'" (Teacher, Urban School)	Teacher elicits self-evaluation before external judgment	Eliciting student perspective	Feedback as a mirror of learning	Scaffolding self-monitoring (SRL: performance phase); supporting autonomy and competence (SDT)
"When I judge a friend's assignment, I have to be honest but also polite. I learned to say, 'This part is good, but maybe we could add a more detailed explanation.'" (Student, Grade V)	Student negotiates honesty and politeness in peer judgment	Managing social norms during assessment	Social interaction in assessment	Developing interpersonal communication and empathy (social-emotional dimension); relatedness (SDT)
"The curriculum is very dense. If I had to do in-depth formative assessment for every student, I would never finish." (Teacher, Urban School)	Curriculum pressure impedes individualized assessment	Time and structural constraints	Structural and cultural challenges	Institutional constraints on autonomy-supportive practice

Thematic saturation was considered achieved when no significant new codes emerged after analysis of the 18th transcript (from a total of 20 interviews), and data from the final observation sessions primarily confirmed existing patterns rather than generating new categories.

Because this study involved minor participants, rigorous ethical protections were implemented:

- 1) Ethical clearance was obtained from the relevant institutional research ethics committee prior to data collection.
- 2) School permission was secured through formal letters to school principals and district education offices.
- 3) Informed consent was obtained from parents/guardians of all student participants, explaining the study's purpose, voluntary nature, data protection measures, and right to withdraw.
- 4) Child assent was obtained from each student through age-appropriate verbal explanations and signed assent forms using simple language. Students were explicitly told they could stop participating at any time without consequence.
- 5) Anonymity and confidentiality: All participant names were replaced with pseudonyms. School names and specific locations are not disclosed in this report.

- 6) Data protection: Audio recordings, video files, and transcripts were stored on password-protected, encrypted devices accessible only to the research team. Video recordings will be destroyed following institutional data retention protocols after publication.
- 7) Minimizing harm: Interview questions were screened to avoid causing distress. If students appeared uncomfortable, sessions were paused or terminated. No incentives were provided that might coerce participation.

RESULTS AND DISCUSSION

The findings are organized into four themes that describe the dynamics of formative assessment practices as experienced by teachers and students. For each theme, empirical evidence is presented first (integrating interview excerpts, observational notes, and artifact evidence), followed by interpretive analysis, theoretical linkage, and historical contextualization.

Theme 1: Feedback as a Mirror of Learning

Empirical Evidence

Teachers in both schools described feedback as a central formative practice, though its character varied. A Grade V teacher at the urban school explained:

"I always ask the children to look back at their work before I give them grades. I asked, 'Which parts do you think are good and which ones need to be improved?' Only then did I add guidance from me."

This practice was observed during a mathematics lesson (Observation Note, Urban School, Session 3): after students completed problem sets, the teacher circulated and asked individuals to identify one correct step and one step needing revision before writing any score. Students were observed re-reading their work, pointing to specific lines, and verbalizing their reasoning.

A Grade IV student at the same school reflected:

"Sometimes I thought my answer was correct, but after rereading it and comparing it with the teacher's example, I realized that something was incomplete."

This student's mathematics workbook contained evidence of self-correction: initial answers were crossed out lightly, with revised calculations written beside them, suggesting that the feedback process prompted re-engagement with the task.

In contrast, a teacher at the rural school described a more directive approach:

"I usually correct their mistakes directly. If there are many errors, I ask them to redo it. But I don't always have time to ask them to find their own mistakes first."

Observation confirmed this variation: in 6 of 10 observed sessions at the rural school, feedback was delivered as direct correction rather than guided self-evaluation.

Interpretation and Theoretical Linkage

The urban teacher's practice reflects a shift from feedback as judgment to feedback as information for self-monitoring. By asking students to evaluate their own work before receiving external evaluation, the teacher positioned students as active assessors rather than passive recipients. This aligns with the performance phase of SRL, in which learners monitor their progress against standards and adjust strategies (Zimmerman, 2000). The student's account of "realizing something was incomplete" illustrates this monitoring process in action.

From an SDT perspective, this form of feedback appears to support competence (by providing information about the gap between current and desired performance) and autonomy (by inviting students to initiate self-evaluation). The teacher's follow-up guidance then functions as scaffolding, maintaining the student's sense of agency while providing necessary support.

The observed variation between teachers suggests that feedback quality is not uniform; it depends on teachers' assessment literacy, time availability, and implicit theories of student capability. The rural teacher's more directive approach appeared linked to time pressure and large class size rather than opposition to formative principles.

Historical Contextualization

Process-oriented feedback practices observed in this study may be understood as departures from historically dominant assessment paradigms. The 1975 curriculum's PPSI approach emphasized behavioral objectives and measurable outcomes, creating institutional conditions in which teachers

functioned primarily as evaluators and students as objects of assessment (Damayanti & Zulkarnain, 2024). The one-way grading system that characterized this period left limited room for student reflection. The 1947 Rentjana Pelajaran, by contrast, had articulated goals of cultivating independent national character (Alhamuddin, 2014). The feedback practices observed in this study particularly those inviting student self-evaluation may be interpreted as partial recoveries of this earlier emphasis on learner independence, though their implementation remains constrained by structural conditions carried forward from subsequent periods.

Table 2. Historical Evidence Chain: Assessment Paradigms and Contemporary Practices

Historical Period	Assessment Policy/Practice	Dominant Paradigm	Key Source	Relevance to Current Findings
Dutch Colonial	Selective examination for administrative labor; restricted indigenous access	Selection and social control	Larasati & Mighwar (2023)	Establishes precedent for examination-oriented culture; explains persistence of high-stakes mentality
1947 Rentjana Pelajaran	Formation of independent national character; limited formal assessment guidance	Nation-building and character education	Alhamuddin (2014)	Provides early policy language for student-centered development; contrast with later technical paradigms
1975 Curriculum (PPSI)	Behavioral objectives; end-of-unit testing; teacher as controller	Behaviorism and centralized control	Damayanti & Zulkarnain (2024)	Institutionalized one-way assessment; legacy visible in directive feedback practices observed
2013 Curriculum & Kurikulum Merdeka	Process-based assessment; attitude/behavior dimensions; student-centered language	Competency and holistic development	Peraturan Menteri Pendidikan dan Kebudayaan No. 54/2015; Peraturan Menteri No. 54/2015	Policy framework supporting formative assessment; gap between intent and implementation evident in field data

It is important to note that this historical analysis does not establish a deterministic causal chain. Rather, it identifies historical continuities that provide interpretive context for understanding why certain assessment orientations persist despite policy change.

Theme 2: Self-Assessment as a Builder of Autonomy

Empirical Evidence

Self-assessment practices showed marked variation between the two research sites. In the urban school included in this study, teachers more frequently used simple rubrics that allowed students to assess their work against explicit criteria. During an Indonesian language lesson (Observation Note, Urban School, Session 5), students were given a three-point rubric for narrative writing (*"Does my story have a beginning, middle, and end?" "Did I use describing words?"*). Students circled their self-ratings before submitting work, and the teacher later compared student ratings with her own evaluation during individual conferences.

A Grade V student described this process:

"Before I give my book to the teacher, I check the rubric first. If I think my describing words are not enough, I add more. Then I write whether I think it's good enough."

This student's writing portfolio contained multiple drafts with self-assessment marks, showing evidence of iterative revision.

In the rural school observed in this study, self-assessment was implemented less frequently. A teacher explained:

"Children here still have a hard time judging themselves. They tend to give high marks even though their work is not up to the criteria. I'm worried that they're going to be unobjective."

Observation confirmed this teacher's concern: when asked to self-assess a mathematics quiz using a simple checklist, 4 of 6 observed students rated themselves at the highest level despite visible calculation errors. However, during a follow-up interview, one of these students revealed:

"I gave myself a good mark because I was afraid if I put a low one, the teacher would think I didn't try. But actually I know I made mistakes in number 3."

This suggests that students' inflated self-ratings may reflect social anxiety about teacher perception rather than genuine metacognitive inability a distinction with important instructional implications.

Interpretation and Theoretical Linkage

Self-assessment practices, when supported by clear criteria and a safe classroom climate, appeared to promote the self-reflection phase of SRL. The urban student's account of adding "describing words" before submission illustrates forethought and strategic adjustment. The rubric functioned as an external standard that students internalized, enabling self-monitoring.

From an SDT perspective, self-assessment directly addresses autonomy: students exercise evaluative judgment and take ownership of quality standards. However, the rural student's admission about fear of teacher perception reveals that autonomy cannot be separated from relatedness. When students perceive that honest self-evaluation might damage their relationship with the teacher or their standing in the class, they may strategically misrepresent their self-assessment, undermining its formative function.

This finding extends previous research by Dignath and Veenman (2021), who noted that teachers often underestimate students' self-regulatory capabilities. The present data suggest that students may possess more metacognitive awareness than their self-assessment products reveal, but social-emotional concerns (fear of judgment, desire for teacher approval) can mask this awareness. Effective self-assessment thus requires not only cognitive scaffolding (rubrics, examples) but also relational safety that permits honest self-appraisal.

Historical Contextualization

The disparity between the urban and rural schools in this study cannot be generalized to all urban and rural schools in Indonesia. However, it resonates with documented historical patterns of educational inequality. The colonial education system concentrated quality resources in urban centers and for select populations, while rural education received limited investment (Larasati & Mighwar, 2023). Post-independence, urban schools have generally maintained advantages in teacher professional development, instructional materials, and administrative support. These structural differences may contribute to the observed variation in self-assessment implementation, not because rural teachers are less capable, but because they may face greater resource constraints, larger class sizes, and fewer opportunities for assessment-focused professional development. This interpretation suggests that assessment reform must be differentiated rather than uniformly mandated.

Theme 3: Social Interaction in the Assessment Process

Empirical Evidence

Peer assessment and collaborative evaluation activities were observed in both schools, though more systematically in the urban setting. During a science lesson on plant classification (Observation Note, Urban School, Session 7), students worked in pairs to evaluate each other's classification diagrams using a teacher-provided checklist. The researcher observed students leaning toward each other, pointing at diagrams, and negotiating criteria:

[Student A]: *"Your mushroom is in the wrong group. It should be with the fungi, not plants."*

[Student B]: *"But it looks like a plant. It has a stem."*

[Student A]: *"Yes, but it doesn't have leaves or chlorophyll. Remember the checklist says 'plants make their own food.' Mushrooms don't."*

[Student B]: *"Oh, right. I'll move it. Thanks for telling me politely."*

This interaction was recorded in field notes and confirmed through video recording of the peer assessment segment. Both students later submitted revised diagrams with corrected classifications.

A Grade V student reflected on peer assessment experiences:

"When I judge a friend's assignment, I have to be honest but also polite. I learned to say, 'This part is good, but maybe we could add a more detailed explanation.' My friend also gave me advice. So we help each other."

This student's peer feedback form contained written comments such as "Your drawing is clear, but the label is too small. Can you make it bigger?" demonstrating the integration of positive acknowledgment with specific, actionable suggestion.

In the rural school, peer interaction during assessment was less structured but still present. During a group mathematics problem-solving activity (Observation Note, Rural School, Session 4), students were observed checking each other's calculations informally. However, without explicit peer assessment protocols, interactions sometimes focused on obtaining correct answers rather than mutual evaluation:

[Student C]: *"Is your answer 24 too?"*

[Student D]: *"Yes. Then it's correct."*

[Student C]: *"Okay, let's just write it."*

This exchange suggests that peer interaction without structured assessment criteria may limit the depth of metacognitive and social-emotional engagement.

Interpretation and Theoretical Linkage

Structured peer assessment appeared to function as a multidimensional developmental context. Cognitively, it provided alternative perspectives on task quality, supporting the monitoring phase of SRL. Social-emotionally, it required students to manage the tension between honesty and politeness, to perspective-take ("What will my friend feel if I say this?"), and to receive criticism without defensiveness. These processes map onto the social-emotional dimensions of empathy, interpersonal communication, and cooperation defined in this study.

From an SDT perspective, peer assessment appears particularly relevant to relatedness. The urban student's comment *"So we help each other"* suggests that assessment was experienced not as a competitive sorting mechanism but as a collaborative learning activity. This experience of mutual support may satisfy the need for positive connection, which Ryan and Deci (2020) identify as essential for intrinsic motivation and well-being.

The contrast between structured peer assessment (urban school) and informal peer checking (rural school) suggests that social-emotional benefits are not automatic byproducts of group work. They require intentional design: explicit criteria, modeled feedback language, and a classroom climate that frames assessment as mutual improvement rather than competitive ranking.

Historical Contextualization

Peer assessment practices observed in this study may be understood in relation to longstanding tensions between individualistic and collectivist educational values. Indonesian culture has historically emphasized gotong royong (mutual cooperation) and communal solidarity. The 1947 Rentjana Pelajaran explicitly valued social awareness and collective character formation (Alhamuddin, 2014). However, subsequent curriculum developments particularly the 1975 PPSI approach with its emphasis on individual behavioral objectives and measurable outcomes created institutional structures that privileged individual achievement over collaborative learning. The competitive examination culture that developed during later periods further reinforced individualistic orientations.

The structured peer assessment observed in this study may thus represent an attempt to re-integrate collectivist cultural values within formal assessment practices. Rather than positioning assessment as a private transaction between individual student and teacher, peer assessment makes evaluation a social process that requires negotiation, mutual respect, and shared responsibility. This practice does not "restore" a lost cultural essence in any simple sense; rather, it actively reconstructs collectivist values within contemporary pedagogical frameworks.

Theme 4: Structural and Cultural Challenges

Empirical Evidence

Teachers in both schools identified structural barriers to formative assessment implementation. An urban teacher stated:

"The curriculum is very dense. I had to finish a lot of material in a short amount of time. If I had to do an in-depth formative assessment for every student, I would never finish."

This concern was corroborated by observation: in 8 of 10 urban sessions, the teacher moved quickly between activities, allowing limited time for individual feedback conferences. Formative practices were often compressed into brief moments at the end of lessons.

A teacher at the rural school described a different but related pressure:

"Parents ask about test scores every semester meeting. They rarely ask whether their child can reflect on their own learning. The number on the report card is what matters."

This statement was made during an informal conversation after a classroom observation session. The researcher recorded it in field notes with the teacher's permission.

Students, too, perceived these pressures. A Grade IV student in the urban school commented:

"I like it when the teacher explains what I did wrong, but I also want a good score. Sometimes if there's no score, I feel like I didn't really do the task."

This comment reveals the internalization of outcome-oriented assessment values: even when process feedback is provided, students may still seek the validation of a numerical score.

Interpretation and Theoretical Linkage

Structural challenges (curriculum density, time constraints, large class sizes) function as contextual boundary conditions that constrain teachers' capacity to implement autonomy-supportive formative practices. These findings align with Yolanda and Rodliyah (2024), who documented similar constraints among Indonesian EFL teachers. However, the present study adds a social-emotional dimension: when teachers are pressured to prioritize coverage over depth, they may inadvertently communicate to students that learning is about task completion rather than understanding.

The parental focus on test scores reflects a cultural model of assessment that equates educational quality with measurable outcomes. From an SDT perspective, this cultural model may undermine students' intrinsic motivation by shifting the locus of evaluation externally (to parents, principals, and examinations) rather than internally (to self-monitoring and personal growth). The student's desire for a score even while valuing feedback illustrates this tension: process-oriented formative assessment must compete with deeply institutionalized outcome-oriented scripts.

Historical Contextualization

The examination-oriented culture described by participants may be understood as reflecting historical continuities in the function of assessment within Indonesian education. During the colonial period, examinations served as selection mechanisms that determined access to further education and social mobility (Larasati & Mighwar, 2023). This selective function was reinforced during the New Order period through the national examination system, which made test scores a primary indicator of school quality and teacher accountability. The political content of the 1975 curriculum, which scholars have argued aimed to foster compliance with centralized policies (Damayanti & Zulkarnain, 2024), further entrenched an assessment culture in which external validation took precedence over internal development.

The contemporary policy shift toward process-based assessment (evident in the 2013 Curriculum and Kurikulum Merdeka) represents a formal repudiation of this legacy. However, the data suggest that policy change alone is insufficient to transform classroom practice when institutional memories, parental expectations, and accountability structures continue to privilege examination outcomes. This finding implies that effective formative assessment reform requires not only teacher training but also cultural reconstruction at the community and policy levels.

Synthesis: Toward an Integrated Understanding

The four themes are not isolated findings; they describe an interconnected system. Feedback provides information for self-monitoring; self-assessment builds autonomy and reflection; social interaction develops interpersonal competence and relatedness; and structural-cultural challenges constrain the implementation of all three. Figure 1 presents a conceptual model synthesizing these relationships.

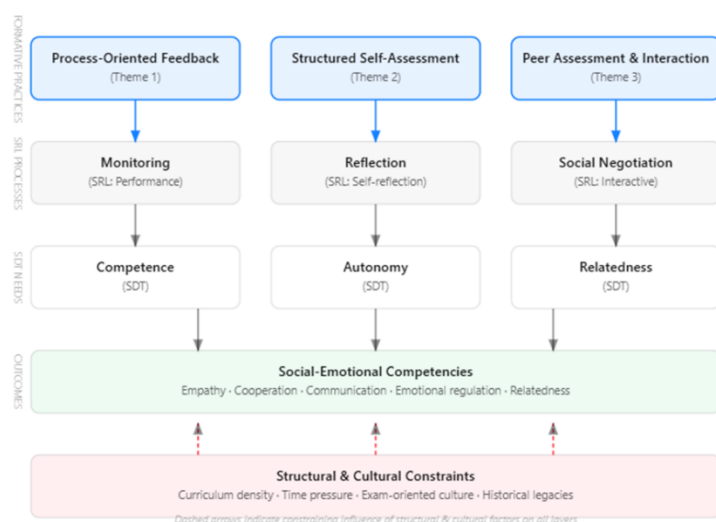


Figure 1. Conceptual Model: Formative Assessment, Self-Regulated Learning, and Social-Emotional Development

This model suggests that formative assessment functions as a multidimensional developmental context when implemented with attention to process, autonomy, and social connection. However, its transformative potential is bounded by structural and cultural conditions that carry the weight of historical assessment paradigms.

Limitations

This study has several limitations that bound the interpretive claims made. First, the qualitative, idiographic nature of IPA precludes statistical generalization to all Indonesian primary schools. The findings represent the lived experiences of participants in two specific schools and should be understood as such. Second, the purposive sampling of teachers known for innovative assessment practices introduced a positive-case selection bias: these teachers may hold more favorable orientations toward formative assessment than the broader teacher population. This sampling was intentional for exploring the phenomenology of implementation, but it means that barriers identified may be underrepresented relative to schools with less formative assessment activity. Third, the study was conducted in regular classroom contexts and did not examine adaptations for students with special needs or from markedly different cultural backgrounds. Fourth, while historical document analysis provided contextual depth, the evidentiary chain connecting historical policies to contemporary practices is interpretive rather than causal. Claims about "historical continuities" should be understood as analytical framings rather than deterministic causal statements.

CONCLUSION

This study explored how teachers and students in two Indonesian primary schools experienced formative assessment practices in relation to self-regulation and social-emotional development. The findings suggest that when formative assessment was implemented through process-oriented feedback, structured self-assessment, and peer interaction, participants perceived it as supporting students' self-monitoring, reflective awareness, and interpersonal competence. Specifically, feedback that invited students to identify their own strengths and gaps appeared to scaffold the monitoring phase of self-regulated learning. Self-assessment, when supported by clear criteria and a safe classroom climate, was associated with students' sense of ownership over learning quality. Peer assessment provided a context for developing communication, empathy, and cooperation, though these benefits required intentional structural support rather than emerging automatically from group work.

However, these practices were unevenly implemented and constrained by structural factors including curriculum density, time limitations, and an examination-oriented cultural model that participants' experiences may reflect within a longer historical continuity. Students in this study sometimes internalized outcome-oriented values, seeking numerical validation even when process feedback was available. Teachers in the rural school observed in this study faced particular challenges

related to resource access and professional development, underscoring the need for differentiated support.

This study extends the application of Self-Regulated Learning theory and Self-Determination Theory to the context of Indonesian primary education by illustrating how formative assessment practices may function as contexts for satisfying basic psychological needs. The findings suggest that the assessment process itself not merely its cognitive outcomes may serve as a medium for social-emotional skill development when it is designed to support autonomy, competence, and relatedness. The identification of relational safety as a condition for honest self-assessment adds nuance to existing SRL research, suggesting that social-emotional and metacognitive processes are interdependent rather than separate domains.

For teachers, this study highlights the importance of assessment literacy that extends beyond technical skills to include the design of safe, autonomy-supportive classroom climates. Specific practices include: (a) using feedback questions that elicit student self-evaluation before external judgment, (b) providing rubrics that make criteria transparent and actionable, (c) modeling respectful peer feedback language, and (d) creating time for individual reflection even within dense curricula.

For school leaders, the findings suggest a need to create institutional conditions that protect teachers' time for formative practices and that communicate to parents the value of process-oriented learning evidence alongside numerical scores.

For policy makers, the findings imply that assessment reform requires more than curriculum redesign. It necessitates attention to historically sedimented assessment cultures, differentiated resource allocation between schools, and accountability mechanisms that value developmental process rather than relying exclusively on examination outcomes.

Future research might productively examine: (a) longitudinal designs tracking students' self-regulatory and social-emotional development over multiple semesters; (b) intervention studies testing the efficacy of differentiated professional development programs for teachers in urban and rural contexts; (c) comparative historical studies examining assessment evolution in other post-colonial education systems; and (d) in-depth investigations of formative assessment adaptations for students with diverse learning needs and cultural backgrounds.

REFERENCES

- Alhamuddin, Alhamuddin. "Sejarah Kurikulum di Indonesia (Studi Analisis Kebijakan Pengembangan Kurikulum)." Nur El-Islam, vol. 1, no. 2, Oct. 2014, pp. 48-58. <https://media.neliti.com/media/publications/226468-sejarah-kurikulum-di-indonesia-studi-ana-bac69203.pdf>
- Birt, L., Scott, S., Cavers, D., Campbell, C., & Walter, F. (2016). Member Checking: A Tool to Enhance Trustworthiness or Merely a Nod to Validation? *Qualitative Health Research*, 26, 1802-1811. <https://doi.org/10.1177/1049732316654870>
- Black, P., & Wiliam, D. (1998). Assessment and Classroom Learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7-74. <https://doi.org/10.1080/0969595980050102>
- Blumer, H. (1954). What is Wrong with Social Theory? *American Sociological Review*, 19(1), 3-10. <https://doi.org/10.2307/2088165>
- Braun, V. and Clarke, V. (2006) Using Thematic Analysis in Psychology. *Qualitative Research in Psychology*, 3, 77-101. <https://doi.org/10.1191/1478088706qp0630a>
- Creswell, J. W., & Poth, C. N. (2024). *Qualitative inquiry and research design: Choosing among five approaches* (5th ed.). SAGE Publications.
- Damayanti, S., & Zulkarnain. (2024). Politik Orde Baru: Dampak Kurikulum 1975 Pada Mata Pelajaran Sejarah Tingkat SMA/ sederajat. *Jurnal Pendidikan Sejarah Indonesia*, 7(2), 358-368. <http://dx.doi.org/10.17977/um0330v7i2p358-367>
- Dignath, C., Veenman, M.V.J. The Role of Direct Strategy Instruction and Indirect Activation of Self-Regulated Learning—Evidence from Classroom Observation Studies. *Educ Psychol Rev* 33, 489-533 (2021). <https://doi.org/10.1007/s10648-020-09534-0>
- Freeman, M., DeMarrais, K., Preissle, J., Roulston, K., & St. Pierre, E. A. (2007). Standards of evidence in qualitative research: An incitement to discourse. *Educational Researcher*, 36(1), 25-32.

- <https://doi.org/10.3102/0013189X06298009>
- Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, 77, 81-112. <https://doi.org/10.3102/003465430298487>
- Krefting L. (1991). Rigor in qualitative research: the assessment of trustworthiness. *The American journal of occupational therapy : official publication of the American Occupational Therapy Association*, 45(3), 214–222. <https://doi.org/10.5014/ajot.45.3.214>
- Kuntowijoyo. (2013). *Pengantar ilmu sejarah*. Tiara Wacana.
- Larasati, S., & Mighwar, M. A. (2023). History of the development of education in Indonesia. *Historical: Journal of History and Social Sciences*, 2(3), 142-148. <https://doi.org/10.58355/historical.v2i3.85>
- Li, J and Gu, YP. (2024). Formative assessment for self-regulated learning: Evidence from a teacher continuing professional development programme. *System*, 125, 103414. <https://doi.org/10.1016/j.system.2024.103414>.
- Lincoln, Y.S. and Guba, E.G. (1985) *Naturalistic Inquiry*. SAGE, Thousand Oaks, 289-331. [http://dx.doi.org/10.1016/0147-1767\(85\)90062-8](http://dx.doi.org/10.1016/0147-1767(85)90062-8).
- Lopez, K. A., & Willis, D. G. (2004). Descriptive versus interpretive phenomenology: their contributions to nursing knowledge. *Qualitative health research*, 14(5), 726–735. <https://doi.org/10.1177/1049732304263638>
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Sage Publications, Inc.
- Moustakas, C. (1994). *Phenomenological Research Methods*. SAGE Publications, Inc. <https://doi.org/10.4135/9781412995658>
- Ryan, R. M., & Deci, E. L. (2020). Intrinsic and Extrinsic Motivation from a Self-Determination Theory Perspective: Definitions, Theory, Practices, and Future Directions. *Contemporary Educational Psychology*, 61, Article 101860. <https://doi.org/10.1016/j.cedpsych.2020.101860>
- Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gagné, & M. Scriven (Eds.), *Perspectives of curriculum evaluation* (pp. 39–83). Rand McNally.
- Smith, J. A., Flowers, P., & Larkin, M. (2009). *Interpretative phenomenological analysis: Theory, method and research*. SAGE Publications.
- Sugiyono. (2019). *Metode penelitian kuantitatif, kualitatif, dan R&D (Edisi Revisi)*. Alfabeta.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge MA: Harvard University Press.
- Yolanda, & Rodliyah, R. S. (2024). Exploring the challenges of formative assessment practices faced by an Indonesian EFL teacher. *TELL-US Journal*, 10(2), 521-533. <https://doi.org/10.22202/tus.2024.v10i2.8228>
- Zahavi D. (2019). Getting It Quite Wrong: Van Manen and Smith on Phenomenology. *Qualitative health research*, 29(6), 900–907. <https://doi.org/10.1177/1049732318817547>
- Zhang, H., Quan, J., Zhang, N., & Zhang, L. (2025). Investigating the effects of formative assessment on EFL students' achievement and motivation: a Self-Determination Theory perspective. *Frontiers in psychology*, 16, 1664871. <https://doi.org/10.3389/fpsyg.2025.1664871>
- Zimmerman, B. J. (2000). Attaining Self-Regulation: A Social Cognitive Perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of Self-Regulation* (pp. 13-39). San Diego, CA: Academic Press. <http://dx.doi.org/10.1016/B978-012109890-2/50031-7>