

Conceptualizing Learning Evaluation in Elementary Education: An Integrative Review of Evaluation Objects, Evaluators, and Functions

Konseptualisasi Evaluasi Pembelajaran di Pendidikan Dasar: Kajian Literatur Integratif tentang Objek Evaluasi, Evaluator, dan Fungsi Evaluasi

^{*1)} Khodijah, ²⁾ Eva Rizkiya Safitri, ³⁾ Ruby Tri Maharani
^{1,2,3,)} Universitas Islam Negeri Sultan Maulana Hasanuddin, Banten, Indonesia

^{*}Corresponding author: khodyyya@gmail.com.

ABSTRACT

This study aimed to analyze and synthesize conceptualizations of the objects and subjects/evaluators of learning evaluation in elementary education and to identify the relationships among evaluation objects, evaluators, evaluation functions, and the use of evaluation results. An integrative literature review was conducted on publications issued between 2015 and 2025. Searches of Google Scholar, Scopus, and ERIC, conducted from 10 to 15 January 2026, identified 452 records; following duplicate removal and screening, 26 studies met the eligibility criteria and were included in the final *thematic synthesis*. Five themes emerged: an emphasis on student achievement (17/26; 65.4%), limited use of evaluation for instructional reflection (14/26; 53.8%), underrepresentation of affective and psychomotor dimensions (13/26; 50.0%), movement toward authentic and participatory evaluation (11/26; 42.3%), and ambiguity among evaluators, *evaluatees*, and objects of evaluation (9/26; 34.6%). The findings indicate that the reviewed literature, rather than the present study directly assessing teachers, more frequently portrays evaluation as oriented toward student outcomes than toward reflective and instructional-improvement functions. Conceptually, the synthesis highlights the need to distinguish the evaluator as the actor conducting the evaluation, the *evaluatee* as the individual or unit being evaluated, the object of evaluation as the aspect being assessed, and the evaluation function as encompassing judgment, feedback, reflection, and decision-making. The findings also suggest an emerging movement toward authentic and participatory evaluation, although this orientation was not dominant across the corpus. The review is limited by its coverage of three databases, the inclusion of only Indonesian- and English-language publications, and terminological heterogeneity across sources; therefore, future studies should empirically validate the proposed framework through teacher interviews, classroom observations, or surveys.

Key words : Learning Evaluation; Evaluation Objects; Evaluation Subjects; Elementary Education; Authentic Assessment

ABSTRAK

Penelitian ini bertujuan menganalisis dan mensintesis konseptualisasi objek dan subjek/evaluator evaluasi pembelajaran dalam literatur pendidikan dasar serta mengidentifikasi hubungan antara objek, evaluator, fungsi evaluasi, dan penggunaan hasil evaluasi. Penelitian menggunakan integrative literature review terhadap publikasi tahun 2015–2025. Penelusuran melalui Google Scholar, Scopus, dan ERIC pada 10–15 Januari 2026 mengidentifikasi 452 rekaman; setelah penghapusan duplikasi dan proses penyaringan, 26 studi memenuhi kriteria kelayakan dan dianalisis melalui thematic synthesis. Sintesis menghasilkan lima tema: orientasi evaluasi pada capaian siswa (17/26; 65,4%), terbatasnya penggunaan evaluasi untuk refleksi instruksional (14/26; 53,8%), kurang terwakilinya dimensi afektif dan psikomotorik (13/26; 50,0%), pergeseran menuju evaluasi autentik dan partisipatif (11/26; 42,3%), serta ambiguitas antara evaluator, evaluatee, dan objek evaluasi (9/26; 34,6%). Temuan tersebut menunjukkan bahwa literatur yang ditelaah, bukan penelitian ini secara langsung terhadap guru, lebih sering menggambarkan evaluasi sebagai berorientasi pada hasil siswa dibandingkan pada fungsi reflektif dan perbaikan pembelajaran. Secara konseptual, sintesis menegaskan perlunya membedakan evaluator sebagai pihak yang melakukan evaluasi, evaluatee sebagai pihak atau unit yang dievaluasi, objek evaluasi sebagai aspek yang dinilai, serta fungsi evaluasi sebagai proses yang mendukung judgment, umpan balik, refleksi, dan pengambilan keputusan. Temuan juga menunjukkan kecenderungan menuju evaluasi autentik dan partisipatif, tetapi kecenderungan tersebut belum dominan dalam corpus. Kajian ini dibatasi oleh cakupan tiga database, publikasi berbahasa Indonesia dan Inggris, serta heterogenitas terminologi antarsumber; oleh karena itu, validasi empiris melalui wawancara, observasi, atau survei guru diperlukan pada penelitian selanjutnya.

Kata Kunci : *Evaluasi Pembelajaran; Objek Evaluasi; Subjek Evaluasi; Pendidikan Dasar; Authentic Assessment*

INTRODUCTION

Learning evaluation is an integral component of the educational process because it provides a basis for teachers and other stakeholders to understand learning achievement, provide feedback, make instructional decisions, and improve the quality of teaching and learning. In elementary education, these functions are particularly important because students' development involves not only the acquisition of knowledge but also the development of attitudes, skills, and learning experiences throughout the instructional process. Contemporary literature on *assessment literacy* emphasizes that teachers' evaluative competence should extend beyond the ability to administer tests and generate scores; it should also include the capacity to select relevant evidence of learning, interpret results appropriately, provide meaningful feedback, and use assessment information to support instructional decision-making (Xu & Brown, 2016; DeLuca et al., 2016). Evaluation should therefore be understood as a process that supports learning and educational decision-making rather than merely as an administrative activity conducted at the end of instruction.

Terminological clarity is particularly important in this study because the terms measurement, assessment, and evaluation are often used interchangeably despite representing different functions. *Measurement* primarily concerns the assignment of numerical

values or scores to particular attributes; *assessment* involves the collection and interpretation of evidence regarding learning; whereas *evaluation* entails making judgments or decisions based on evidence and predetermined criteria (AERA, APA, & NCME, 2014). Based on these distinctions, this study defines the object of evaluation as *what is being evaluated*, including learning outcomes, learning processes, instructional strategies, assessment practices, and curriculum or program components. In contrast, the subject/evaluator refers to *who conducts the evaluation*, including teachers, students through self-assessment, peers through *peer assessment*, or educational institutions. Individuals or components that are the targets of evaluation may be referred to as evaluatees. This distinction is essential for avoiding *category confusion*. Curriculum, infrastructure, and instructional strategies, for example, are more appropriately classified as objects or components being evaluated rather than as subjects or evaluators.

Within the Indonesian elementary education literature reviewed in the manuscript, evaluation-related issues are frequently associated with an emphasis on learning outcomes and the cognitive domain. Ahmadi et al. (2025) reported various challenges encountered by teachers in implementing learning evaluation in elementary schools, while Zainudin and Ubabuddin (2023) emphasized the importance of addressing cognitive, affective, and **psychomotor** domains in evaluating student learning outcomes. Mutiara et al. (2025) similarly demonstrated that identifying the object of evaluation is closely related to determining which aspects or components of learning are to be assessed. Meanwhile, Fernanda et al. (2025) positioned authentic assessment as an approach that provides broader opportunities to evaluate competencies and learning processes. Taken together, these studies suggest that evaluation in elementary education involves not only technical issues related to assessment instruments but also broader conceptual questions concerning the scope and functions of evaluation.

Contemporary assessment theory has also progressed beyond the classical cognitive–affective–psychomotor classification. Assessment is increasingly positioned as an integral part of learning through concepts such as formative assessment, assessment for learning, assessment as learning, self-assessment, peer assessment, and the use of feedback to support students’ self-regulated learning. Panadero et al. (2017) demonstrated that *self-assessment* is associated with self-regulated learning and self-efficacy, while Andrade (2019) emphasized that self-assessment can function as a learning process rather than merely as students assigning grades to themselves. Similarly, meta-analytic evidence on *peer assessment* indicates that peer involvement can contribute to both learning processes and outcomes when supported by explicit criteria (Double et al., 2020). These perspectives broaden the concept of the “subject of evaluation,” as students may serve not only as

individuals being evaluated but also as active agents in generating, interpreting, and using evaluative information.

Nevertheless, the sources underpinning the present manuscript suggest that studies of learning evaluation are often fragmented across different areas of emphasis. Some studies focus on the development and quality of evaluation instruments (Sudrajat, 2018; Saputra, 2025), others examine the role of evaluation in the learning process (Ansya et al., 2024; Septiani et al., 2023), while still others emphasize learning domains or authentic assessment practices (Zainudin & Ubabuddin, 2023; Fernanda et al., 2025). These strands of literature do not automatically produce a conceptual framework that explicitly distinguishes and connects the evaluator, object/evaluatee, function of evaluation, and use of evaluation results. Accordingly, the research gap addressed in this study is not framed as a claim that the objects and subjects of evaluation have “never been studied.” Rather, it concerns the need to integratively synthesize how these concepts are defined, differentiated, and interconnected within the elementary education literature. Such a formulation is more defensible because it avoids claiming a scarcity of research before the complete literature identification and selection process has been established.

Based on this gap, the conceptual contribution of the study is directed toward establishing clearer relationships among the subject/evaluator, object or evaluatee, function of evaluation, and purpose for which evaluation results are used. Within this framework, teachers, students, peers, and institutions may function as evaluators depending on the context, whereas learning outcomes, learning processes, instructional strategies, assessment practices, curriculum, and program components may constitute objects of evaluation. The resulting information may subsequently serve functions related to measurement, feedback, reflection, decision-making, and instructional improvement. This perspective is consistent with contemporary conceptions of *teacher assessment literacy*, which position assessment practice as a contextual and reflective process closely connected to pedagogical decision-making (Xu & Brown, 2016; DeLuca et al., 2016). Therefore, the novelty of this study is not claimed to lie in “modifying theory” or adopting an undefined multidisciplinary approach, but rather in providing an integrative conceptual synthesis that reduces terminological ambiguity concerning who evaluates, what is evaluated, and for what purposes evaluation is conducted in elementary education.

Accordingly, this study aims to analyze and synthesize the conceptualizations of the objects and subjects of learning evaluation in the elementary education literature and to identify their relationships with the functions of evaluation and contemporary developments in assessment practice. Because the study uses published literature as its data source, its conclusions are directed toward what previous studies report, rather than toward directly assessing teachers' understanding

METHOD

This study employed an integrative literature review to analyze and synthesize conceptualizations of the objects and subjects/evaluators of learning evaluation within the context of elementary education. This approach was selected because it enables the integration of empirical and conceptual literature to identify patterns, terminological differences, conceptual relationships, and knowledge gaps without claiming the level of comprehensiveness associated with a formal systematic literature review (Whittemore & Knafl, 2005; Torraco, 2005; Snyder, 2019). The primary review corpus was restricted to publications issued between 2015 and 2025, whereas earlier sources, such as Arikunto (2008), were used only as foundational or theoretical references and were not counted as part of the review corpus. The literature search was conducted through Google Scholar, Scopus, and ERIC using combinations of Indonesian and English search terms, including “*evaluasi pembelajaran*,” “*objek evaluasi*,” “*subjek evaluasi*,” “*learning evaluation*,” “*assessment*,” “*evaluation objects*,” “*evaluators*,” “*elementary education*,” “*assessment literacy*,” “*authentic assessment*,” “*self-assessment*,” and “*peer assessment*.” The search was conducted between 10 and 15 January 2026. A total of 452 records were initially identified. After removing 73 duplicate records, 379 records underwent title and abstract screening. Subsequently, 81 full-text articles were assessed for eligibility, and 26 studies met the eligibility criteria and were included in the final synthesis. The identification, screening, eligibility, and inclusion stages were documented to enhance transparency; nevertheless, the study remained classified as an integrative literature review and did not claim full systematic-review status (Page et al., 2021).

Study selection was guided by predefined inclusion and exclusion criteria. The inclusion criteria comprised: (1) empirical or conceptual articles published between 2015 and 2025; (2) peer-reviewed publications; (3) publications available in Indonesian or English; (4) availability of full text; (5) a focus on evaluation or assessment within elementary education; and (6) relevant information concerning evaluation objects, subjects/evaluators, evaluatees, evaluation functions, assessment literacy, authentic assessment, self-assessment, peer assessment, or the use of evaluation results. The exclusion criteria included duplicate publications, opinion pieces or non-academic sources, studies outside elementary education without a direct conceptual contribution to the review questions, studies dealing exclusively with measurement without relevance to the conceptual focus of the review, and publications for which full texts could not be accessed. The methodological quality of empirical studies was assessed using the Mixed Methods Appraisal Tool (MMAT), Version 2018, according to the methodological design of each study. Conceptual publications were evaluated separately according to their relevance to the review questions, conceptual clarity, coherence of argumentation, transparency in the use of sources, and contribution to the conceptual

synthesis (Whittemore & Knafl, 2005; Hong et al., 2018). Screening was conducted independently by two reviewers, and disagreements were resolved through discussion and consensus. When consensus could not be reached, a third reviewer served as an adjudicator. This procedure was intended to reduce subjective selection bias and enhance the consistency of inclusion decisions.

Data from each eligible source were extracted using a structured analysis form containing the author and year of publication, educational context, research design, key concepts, definitions of the object of evaluation, definitions of the subject/evaluator, evaluatee, evaluation functions, and findings relevant to the review questions. The synthesis employed a thematic synthesis approach involving repeated reading, extraction of relevant information units, generation of initial codes, grouping of related codes into categories, development of themes, cross-study comparison, and construction of an integrative conceptual synthesis (Thomas & Harden, 2008; Braun & Clarke, 2006). For example, statements positioning student learning outcomes as the principal focus of evaluation were coded as student achievement as evaluation object, grouped under the category learning-outcome orientation, and subsequently incorporated into a broader theme concerning the narrowing of evaluation to student achievement. Particular attention was given to identifying conceptual ambiguity among the evaluator, object/evaluatee, and evaluation function, thereby avoiding the automatic classification of curriculum or educational infrastructure as “evaluation subjects” when they functioned more appropriately as objects of evaluation. To strengthen credibility and dependability, two reviewers independently coded the included studies, compared their coding decisions, and resolved discrepancies through consensus discussion. Decisions concerning inclusion, exclusion, coding, category construction, and theme development were documented through an audit trail to ensure that the synthesis process remained transparent and traceable (Nowell et al., 2017; Snyder, 2019).

RESULTS AND DISCUSSION

1. Results

The literature search yielded 452 records from Google Scholar, Scopus, and ERIC. After 73 duplicate records were removed, 379 records underwent title and abstract screening. A total of 81 full-text articles were subsequently assessed for eligibility, and 26 studies met the inclusion criteria and were included in the final synthesis. Thematic analysis of the corpus generated five themes: (1) narrowing evaluation to student achievement; (2) underrepresentation of affective and psychomotor dimensions; (3) ambiguity in the roles of evaluator and evaluatee; (4) limited use of evaluation for instructional reflection; and (5) movement toward authentic and participatory evaluation.

The frequency of each theme was calculated based on the number of studies coded under that theme. Because a single study could contribute to more than one theme, the categories were not mutually exclusive and the percentages were therefore not expected to total 100%. These frequencies were used descriptively to indicate the distribution of conceptual patterns across the corpus rather than as measures of effect size or statistical significance.

a. Narrowing Evaluation to Student Achievement

The theme of narrowing evaluation to student achievement was the most frequently identified pattern, appearing in 17 of the 26 studies (65.4%). Studies assigned to this theme commonly positioned learning outcomes, attainment of instructional objectives, mastery of subject matter, and reporting of student performance as the most prominent objects of evaluation.

This pattern is reflected, for example, in Ahmadi et al. (2025), Septiani et al. (2023), and Ansyah et al. (2024), which emphasized evaluation as a means of determining the extent to which learning objectives and outcomes had been achieved. However, an emphasis on student achievement did not necessarily imply that evaluation was restricted to score production. Saputra (2025), for instance, linked evaluative information to implementation context, learner characteristics, decision-making, and instructional improvement.

The synthesis therefore revealed a dominant pattern in which student achievement remained the principal focus of evaluation, while also showing conceptual variation in how evaluation results were subsequently used. Accordingly, this finding should be interpreted as evidence of a recurrent achievement-oriented emphasis rather than as an indication that all reviewed studies reduced evaluation solely to grading.

b. Underrepresentation of Affective and Psychomotor Dimensions

The second theme was identified in 13 of the 26 studies (50.0%). This theme indicates that although the literature recognizes learning evaluation as encompassing cognitive, affective, and psychomotor domains, these dimensions were not represented with equal consistency.

Zainudin and Ubabuddin (2023) explicitly identified the cognitive, affective, and psychomotor domains as objects of learning-outcome evaluation. Ahmadi et al. (2025) likewise reported challenges in implementing evaluation practices capable of accommodating a broader range of learning outcomes, while Mutiara et al. (2025) emphasized the importance of accurately determining the aspects that constitute the object of evaluation.

Across the reviewed studies, cognitive outcomes were more readily represented through tests, scores, and academic achievement, whereas attitudes, skills, and learning processes required more diverse forms of evaluative evidence. Thus, the finding across 13 studies (50.0%) does not indicate the absence of affective and psychomotor assessment, but rather that these domains were less consistently represented than cognitive outcomes.

c. Ambiguity in the Roles of Evaluator and Evaluatee

A total of 9 of the 26 studies (34.6%) demonstrated conceptual ambiguity in distinguishing among the evaluator, evaluatee, and object of evaluation. Although this was the least frequent theme, it is conceptually important because it directly relates to the central focus of the review on the objects and subjects of evaluation.

In some studies, teachers were positioned as those who conducted evaluation, whereas students were positioned as those being evaluated. However, the term *subject of evaluation* was not always used consistently. In some formulations, it also referred to individuals or components that were actually being evaluated. The earlier version of the manuscript, for example, classified curriculum and educational infrastructure as “subjects of evaluation,” even though these components do not perform evaluative actions and are more appropriately understood as objects or components that may be evaluated.

The synthesis therefore supports a clearer distinction among three categories. The evaluator/subject refers to the actor conducting the evaluation, such as a teacher, a student undertaking *self-assessment*, peers engaging in *peer assessment*, or an institution. The evaluatee refers to the individual, group, or unit being evaluated, whereas the object of evaluation refers to the particular aspect being assessed, such as learning outcomes, learning processes, instructional strategies, curriculum, programs, or supporting resources. The presence of this theme in 9 studies indicates that terminological inconsistency constitutes a meaningful conceptual issue within the review corpus, even though it was not the most prevalent pattern.

d. Limited use of Evaluation for Instructional Reflection

The theme of limited use of evaluation for instructional reflection was identified in 14 of the 26 studies (53.8%), making it the second most frequent theme after the emphasis on student achievement. This theme highlights a distinction between evaluation used primarily to determine what students have achieved and evaluation used as a source of information for improving teaching and learning.

Ansyah et al. (2024) and Septiani et al. (2023) emphasized the role of evaluation in monitoring the achievement of instructional objectives. In contrast, Saputra (2025) extended this function by positioning evaluative information as a basis for reflection, decision-making, and instructional improvement. Ardana et al. (2023) also linked evaluation results to the identification of student needs and the adjustment of instructional strategies.

The synthesis across 14 studies (53.8%) revealed two broad orientations. The first can be described as evaluation of learning, in which evaluation is used to determine and document student achievement. The second involves the use of evaluation for instructional improvement, whereby evaluative evidence informs feedback and adjustments to instructional methods, materials, strategies, and decisions. The distribution of this theme suggests that

reflective uses of evaluation are present in the literature but are not represented consistently across the entire corpus.

e. Movement Toward Authentic and Participatory Evaluation

The fifth theme was identified in 11 of the 26 studies (42.3%) and reflects an emerging orientation toward more authentic, contextualized, and participatory forms of evaluation. Within studies coded under this theme, evaluation extended beyond test performance to include evidence of competence, learning processes, student engagement, and forms of assessment that were more closely aligned with meaningful learning contexts.

Fernanda et al. (2025), for example, examined authentic evaluation in elementary-level Pancasila Education. From a more participatory perspective, self-assessment and peer assessment also broaden the role of students from being solely evaluatees to becoming active participants in evaluative processes.

Because this theme appeared in 11 studies (42.3%), it is more appropriately interpreted as an emerging tendency rather than a dominant pattern. In other words, the literature indicates movement toward authentic and participatory evaluation, but this orientation remains less prevalent than approaches centered on student achievement.

Table 1. Distribution of Themes Across the 26 Studies

Theme	Studies (n)	Percentage	Main Finding	Representative Studies
T1. Narrowing evaluation to student achievement	17/26	65.4%	Student achievement was the most prominent object of evaluation	Ahmadi et al. (2025); Septiani et al. (2023); Ansya et al. (2024); Saputra (2025)
T2. Underrepresentation of affective and psychomotor dimensions	13/26	50.0%	Affective and psychomotor dimensions were less consistently represented than cognitive outcomes	Zainudin & Ubabuddin (2023); Ahmadi et al. (2025); Mutiara et al. (2025)
T3. Ambiguity in the roles of evaluator and evaluatee	9/26	34.6%	Terminological inconsistency existed in distinguishing evaluators,	Mutiara et al. (2025); Nuriyah (2016)

Theme	Studies (n)	Percentage	Main Finding	Representative Studies
			evaluatees, and evaluation objects	
T4. Limited use of evaluation for instructional reflection	14/26	53.8%	Achievement measurement was more consistently represented than reflective use of evaluation	Saputra (2025); Ardana et al. (2023); Septiani et al. (2023); Ansya et al. (2024)
T5. Movement toward authentic and participatory evaluation	11/26	42.3%	The literature showed movement toward authentic evaluation and greater student participation	Fernanda et al. (2025) and other relevant studies in the corpus

Source: Authors' synthesis based on the literature analyzed

The distribution shows that student-achievement orientation (65.4%) was the most frequently identified pattern, followed by limited use of evaluation for instructional reflection (53.8%), underrepresentation of affective and psychomotor dimensions (50.0%), movement toward authentic and participatory evaluation (42.3%), and ambiguity between evaluator and evaluatee (34.6%). Because multiple coding was applied, a single study could contribute to two or more themes.

f. Integrative Synthesis of the Findings

Across the five themes, the conceptualization of learning evaluation can be organized around four interrelated components:

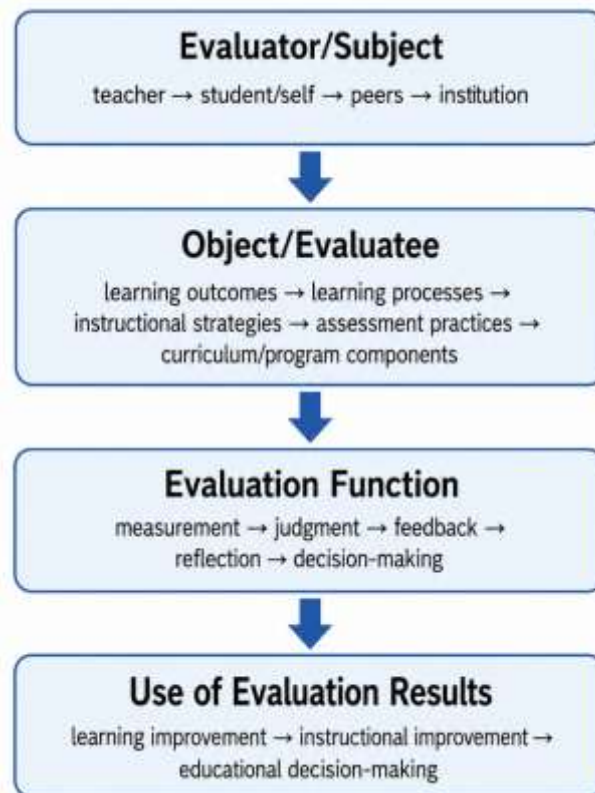


Figure 1. Integrative Synthesis of the Findings

The synthesis indicates that the objects and subjects of evaluation should not be understood as two isolated categories. Evaluators conduct evaluative processes in relation to particular individuals, units, or objects using relevant criteria and evidence, while the resulting information may be used for measurement, feedback, reflection, or decision-making. At the same time, the distribution of the five themes reveals a conceptual tension between the continued prominence of student-achievement-oriented evaluation and an emerging movement toward broader, more reflective, authentic, and participatory approaches to evaluation.

2. Discussion

The synthesis indicates that the conceptualization of learning evaluation in elementary education is currently situated between a traditional orientation emphasizing student achievement and a more contemporary orientation that views evaluation as an integral component of learning, reflection, and educational decision-making. The most frequently identified theme was the narrowing of evaluation to student achievement, which appeared in 17 of the 26 studies (65.4%), followed by the limited use of evaluation for instructional reflection in 14 studies (53.8%), the underrepresentation of affective and psychomotor dimensions in 13 studies (50.0%), movement toward authentic and participatory evaluation in 11 studies (42.3%), and ambiguity between evaluator and evaluatee in 9 studies (34.6%).

These patterns suggest that the central issue in elementary-school evaluation concerns not merely the limitations of assessment instruments, but also how the purposes, objects, actors, functions, and uses of evaluation results are conceptualized.

One of the most important theoretical implications of these findings is the need to distinguish more clearly among testing, measurement, assessment, and evaluation. Although these concepts are related, they are not synonymous. *Testing* refers to a structured procedure for eliciting samples of learners' performance or responses; *measurement* assigns numerical values or scores to attributes or performance according to specified rules; *assessment* involves the broader process of collecting and interpreting evidence about learning; and *evaluation* uses that evidence, together with explicit criteria, to make judgments and support decision-making. This distinction is consistent with contemporary principles of educational measurement, which emphasize that scores are meaningful only when their interpretations and intended uses are adequately supported by evidence (AERA, APA, & NCME, 2014). In the context of the present findings, the dominance of student-achievement orientation may reflect a tendency to reduce evaluation to the measurement and reporting of outcomes. Meaningful evaluation, however, requires the interpretation of evidence and its subsequent use in pedagogical decision-making.

The predominance of *student achievement* can also be interpreted through the concept of assessment literacy. Xu and Brown (2016) argue that teachers' assessment literacy extends beyond technical knowledge of tests to include the ability to connect assessment purposes, instructional contexts, evidence interpretation, and pedagogical decisions. Similarly, DeLuca et al. (2016) emphasize that assessment competence includes selecting appropriate assessment approaches, interpreting results, communicating assessment information, and using evidence ethically in decision-making. Thus, the finding that 65.4% of the reviewed studies emphasized student achievement should not be interpreted as suggesting that an outcome orientation is inherently problematic. Learning outcomes remain legitimate and important objects of evaluation. The problem arises when evidence of achievement is not connected to diagnosis, feedback, reflection, and instructional improvement.

The second finding, namely the underrepresentation of affective and psychomotor dimensions in 50.0% of the corpus, further supports the need to broaden the conceptualization of evaluation objects. The cognitive–affective–psychomotor framework remains relevant as a classical classification, and Zainudin and Ubabuddin (2023) explicitly identified these three domains as important objects of learning-outcome evaluation. However, contemporary evaluation should not be restricted to these domains alone. Evaluation objects may also include learning processes, the application of knowledge, self-regulation, feedback use, participation, problem solving, and performance on authentic tasks. This broader perspective is consistent with developments in formative assessment and *assessment for learning*, in

which evidence is collected not simply to determine what has been achieved, but also to identify learning gaps and inform subsequent instructional action. Meta-analytic evidence by Yao et al. (2024), covering 118 K–12 studies, indicates that formative assessment has a positive effect on academic learning, reinforcing the argument that assessment should extend beyond documenting outcomes to actively supporting learning processes.

This development is also closely related to assessment as learning, self-assessment, and peer assessment. In traditional models, students primarily function as *evaluatees*, whereas teachers act as evaluators. However, the fifth theme indicates that 11 of the 26 studies (42.3%) reflected a movement toward more authentic and participatory forms of evaluation. This trend is important because *self-assessment* enables students to compare their performance against criteria, identify gaps, and take corrective action. Panadero et al. (2017), through four meta-analyses, found positive associations between self-assessment, self-regulated learning, and self-efficacy. Andrade (2019) likewise emphasizes that formative self-assessment is not simply a process in which students assign grades to themselves, but rather a criterion-referenced reflective activity directed toward improving learning. Likewise, the meta-analysis by Double et al. (2020) indicates that peer assessment can positively contribute to academic performance. Students may therefore occupy two conceptually distinct positions: as individuals being evaluated and, under certain conditions, as active participants in the evaluative process.

The finding concerning evaluator–evaluatee ambiguity in 34.6% of the studies constitutes an important conceptual contribution. Literature that uses the term “subject of evaluation” inconsistently risks conflating three analytically distinct categories: the evaluator, the evaluatee, and the object of evaluation. Teachers, students engaging in self-assessment, peers, or institutions may act as evaluators because they perform evaluative activities. Students, teachers, classes, or program units may function as evaluatees when they are the targets of evaluation. Meanwhile, learning outcomes, instructional processes, teaching strategies, assessment practices, curriculum, programs, and educational resources constitute objects of evaluation because they are the aspects being judged. Curriculum and educational infrastructure, for example, should not be classified as evaluators because they do not perform evaluative actions. This distinction clarifies the conceptual weakness identified in the review corpus and provides a more consistent basis for understanding the relationship between the objects and subjects of evaluation.

The fourth theme also highlights an important limitation in the function of evaluation. Fourteen of the 26 studies (53.8%) indicated that the use of evaluation for instructional reflection was not consistently represented. Yet evaluation that is oriented toward learning should operate as a cycle in which evidence is collected, interpreted, compared with intended criteria or objectives, and then used to determine subsequent action. From this perspective,

feedback becomes a critical link between assessment and improvement. Carless and Boud (2018) introduced the concept of feedback literacy, referring to the capacity to understand, interpret, and productively use feedback. This concept broadens evaluation beyond score provision by emphasizing the ability of both students and teachers to act on evaluative information. Accordingly, the finding regarding limited reflective use can be interpreted as a gap between evaluation of learning and evaluation for learning and instructional improvement.

The movement toward authentic evaluation further reinforces this interpretation. Fernanda et al. (2025), one of the studies included in the corpus, examined authentic evaluation in elementary-level Pancasila Education. Conceptually, authentic assessment requires evidence of performance that is more closely connected to the application of knowledge and competence in meaningful tasks rather than the reproduction of information through conventional tests. Villarroel et al. (2018) identify authenticity, cognitive challenge, and the application of capabilities in meaningful contexts as important features of authentic assessment design. Within Indonesian elementary education, this orientation is relevant to competency-based learning and the broader direction of the *Kurikulum Merdeka*, which emphasizes process, competence, and student development. However, this relationship should be presented as a contextual implication, rather than as evidence that the present review systematically examined all assessment practices within the *Kurikulum Merdeka*.

Conceptually, the findings can be integrated through the framework presented in Figure 1 on page 5 of the Results document. The figure organizes evaluation into four sequential and interrelated components: Evaluator/Subject → Object/Evaluated → Evaluation Function → Use of Evaluation Results. This framework offers a stronger conceptual contribution than treating objects and subjects as isolated categories. Within the framework, the evaluator may be a teacher, student, peer, or institution; the object/evaluated may include learning outcomes, learning processes, instructional strategies, assessment practices, or program components; the evaluation function may include measurement, judgment, feedback, reflection, and decision-making; and the use of evaluation results may be directed toward learning improvement, instructional improvement, and educational decision-making. Thus, the relationship between objects and subjects is not static, but forms part of an evaluative process in which particular actors generate and use evidence concerning specific objects for particular purposes.

The framework also helps explain the relationship among the five themes. The dominance of student-achievement orientation indicates that the object of evaluation often remains concentrated on learning outcomes. The underrepresentation of affective and psychomotor dimensions suggests that the range of evaluative evidence may still be broadened. Evaluator–evaluated ambiguity reflects a conceptual problem at the level of actors and terminology. The limited reflective function indicates that evaluative evidence is not always translated into

instructional action. In contrast, the emergence of authentic and participatory assessment reflects movement toward a more integrated orientation. Therefore, the principal contribution of this review is not to argue that evaluation theory should be modified, but rather to suggest the need to contextualize existing evaluation frameworks so that the relationships among actors, objects, functions, and uses of results are made more explicit in elementary education.

The practical implication is that professional development should not focus solely on teachers' ability to construct test items or assign scores, but should address a broader form of assessment literacy. This includes clarifying assessment purposes, selecting appropriate evidence, using multiple sources of evidence, designing authentic tasks, providing meaningful feedback, involving students in self- and peer assessment, and using evaluative information to support instructional decisions. The present review does not directly demonstrate that elementary-school teachers have inadequate assessment competence because the unit of analysis was published literature rather than teachers themselves. A more defensible conclusion is therefore that the reviewed literature suggests a need to strengthen conceptually coherent, reflective, and participatory approaches to learning evaluation in elementary education.

Several limitations should be considered when interpreting the synthesis. First, the literature search was limited to Google Scholar, Scopus, and ERIC, meaning that publications indexed exclusively in other databases may not have been identified. Second, restricting the corpus to Indonesian- and English-language publications may have introduced publication-language bias. Third, although double screening and double coding were used, the processes of selection and interpretation remain potentially subject to *selection* and *interpretive bias*. Fourth, definitions of "object," "subject," "evaluator," and "evaluatee" varied substantially across sources, which constituted one of the findings but also limited direct comparability. Fifth, part of the corpus appears to be concentrated in the Indonesian educational context, and the transferability of the synthesis to other educational systems should therefore be considered cautiously. Finally, theme frequencies such as 65.4% and 53.8% indicate the distribution of patterns within the reviewed corpus and should not be interpreted as prevalence estimates for teachers or elementary schools more broadly. Future research may empirically examine the proposed conceptual framework through teacher interviews, classroom observations, assessment-instrument analysis, or surveys of *assessment literacy*, thereby testing how evaluator, object/evaluatee, evaluation function, and use of results operate in actual educational practice.

CONCLUSION

The synthesis of 26 studies indicates that learning evaluation in elementary education is still frequently conceptualized with a strong emphasis on student achievement. A *student-*

achievement orientation was the most dominant theme, appearing in 17 studies (65.4%), followed by the limited use of evaluation for instructional reflection in 14 studies (53.8%), the underrepresentation of affective and psychomotor dimensions in 13 studies (50.0%), movement toward authentic and participatory evaluation in 11 studies (42.3%), and ambiguity between evaluators and *evaluatees* in 9 studies (34.6%). These findings do not directly demonstrate that elementary school teachers have a limited understanding of evaluation; rather, they indicate that the reviewed literature frequently portrays evaluation practices and conceptualizations as being more strongly oriented toward student outcomes than toward reflective, participatory, and instructional-improvement functions. Conceptually, the review underscores the importance of clearly distinguishing the evaluator/subject as the actor conducting the evaluation, the *evaluatee* as the individual or unit being evaluated, the object of evaluation as the aspect being assessed, and the evaluation function as encompassing measurement, judgment, feedback, reflection, and decision-making. The contribution of this review, therefore, is not to modify evaluation theory, but to contextualize existing evaluation frameworks by making the relationships among evaluators, objects/*evaluatees*, evaluation functions, and the uses of evaluation results more explicit in relation to learning improvement, instructional improvement, and educational decision-making.

From a practical perspective, the synthesis highlights the importance of strengthening teachers' assessment literacy and reflective evaluation practices, including their capacity to clarify assessment purposes, select and interpret appropriate evidence of learning, provide meaningful feedback, use evaluation to improve instructional strategies, and involve students through self-assessment, peer assessment, and authentic forms of evaluation. These implications should, however, be interpreted in light of the study's limitations, including the restriction of the literature search to Google Scholar, Scopus, and ERIC, the inclusion of only Indonesian- and English-language publications, potential selection and interpretive bias, heterogeneous definitions of "object," "subject," "evaluator," and "evaluatee," and the concentration of part of the corpus within the Indonesian educational context. Accordingly, the findings provide a conceptual synthesis rather than direct evidence of teachers' understanding. Future research should empirically validate the proposed framework through teacher interviews, classroom observations, analysis of assessment instruments, or surveys of *assessment literacy*, thereby examining how evaluators, objects/*evaluatees*, evaluation functions, and the use of evaluation results operate in actual elementary-education practice.

REFERENCES

- Ahmadi, A. M., Raihan, D., Alawiyah, H., Martinez, M., & Kistian, A. (2025). Problematika guru dalam melaksanakan evaluasi pembelajaran siswa di kelas IV sekolah dasar. *Jurnal Binagogik*, 12(1), 33–40. <https://ejournal.uncm.ac.id/index.php/pgsd/article/view/1519>
- American Educational Research Association, American Psychological Association, & National

- Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- Andrade, H. L. (2019). A critical review of research on student self-assessment. *Frontiers in Education*, 4, Article 87. <https://doi.org/10.3389/feduc.2019.00087>
- Ansyah, Y. A. U., Alfianita, A., Syahkira, H. P., & Syahrial, S. (2024). Peran evaluasi pembelajaran pada mata pelajaran matematika kelas V sekolah dasar. *Indiktika: Jurnal Inovasi Pendidikan Matematika*, 6(2), 173–184. <https://doi.org/10.31851/indiktika.v6i2.15030>
- Ardana, L. N., Vega, N., Kirana, P., & Sunaryati, T. (2023). Peran penting evaluasi pembelajaran pada mata pelajaran Pendidikan Kewarganegaraan di sekolah dasar berbasis pendidikan karakter. *Jurnal Pendidikan Tambusai*, 7(2), 15814–15819. <https://doi.org/10.31004/jptam.v7i2.8871>
- Arikunto, S. (2008). *Dasar-dasar evaluasi pendidikan* (Cet. ke-8). Bumi Aksara.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment & Evaluation in Higher Education*, 43(8), 1315–1325. <https://doi.org/10.1080/02602938.2018.1463354>
- DeLuca, C., LaPointe-McEwan, D., & Luhanga, U. (2016). Teacher assessment literacy: A review of international standards and measures. *Educational Assessment, Evaluation and Accountability*, 28(3), 251–272. [10.1007/s11092-015-9233-6](https://doi.org/10.1007/s11092-015-9233-6)
- Double, K. S., McGrane, J. A., & Hopfenbeck, T. N. (2020). The impact of peer assessment on academic performance: A meta-analysis of control group studies. *Educational Psychology Review*, 32(2), 481–509. [10.1007/s10648-019-09510-3](https://doi.org/10.1007/s10648-019-09510-3)
- Fernanda, S. A., Fernica, V. O., & Pratama, M. B. (2025). Penerapan evaluasi autentik dalam pembelajaran Pendidikan Pancasila di tingkat sekolah dasar. *Menulis: Jurnal Penelitian Nusantara*, 1(5), 334–340. <https://doi.org/10.59435/menulis.v1i5.277>
- Hong, Q. N., Pluye, P., Fàbregues, S., Bartlett, G., Boardman, F., Cargo, M., Dagenais, P., Gagnon, M.-P., Griffiths, F., Nicolau, B., O’Cathain, A., Rousseau, M.-C., & Vedel, I. (2018). *Mixed Methods Appraisal Tool (MMAT), version 2018*. Canadian Intellectual Property Office. <https://doi.org/10.3233/EFI-180221>
- Mutiara, A., Subayu, D., Pratama, M. R., Syahputri, N., Handayani, P., Lestari, R., & Panggabean, H. S. (2025). Objek evaluasi pembelajaran Pendidikan Agama Islam. *Indonesian Research Journal on Education*, 5(1), 50–55. <https://doi.org/10.31004/irje.v5i1.1901>
- Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic analysis: Striving to meet the trustworthiness criteria. *International Journal of Qualitative Methods*, 16, 1–

13. <https://doi.org/10.1177/1609406917733847>
- Nuriyah, N. (2016). Evaluasi pembelajaran: Sebuah kajian teori. *Eduksos: Jurnal Pendidikan Sosial & Ekonomi*, 3(1). <https://doi.org/10.24235/edueksos.v3i1.327>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., . . . Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71. <https://doi.org/10.1136/bmj.n71>
- Panadero, E., Jonsson, A., & Botella, J. (2017). Effects of self-assessment on self-regulated learning and self-efficacy: Four meta-analyses. *Educational Research Review*, 22, 74–98. [10.1016/j.edurev.2017.08.004](https://doi.org/10.1016/j.edurev.2017.08.004)
- Saputra, A. (2025). Pengembangan instrumen evaluasi. *Ar-Raudah: Jurnal Pendidikan dan Keagamaan*, 2(4), 1–14. <https://doi.org/10.61891/ar-raudah.v2i4.651>
- Septiani, A. N., Pratiwi, D., & Rossy, R. (2023). Evaluasi pembelajaran dalam meningkatkan mutu pendidikan di sekolah dasar. *MASALIQ*, 3(5), 824–832. [10.58578/masaliq.v3i5.1380](https://doi.org/10.58578/masaliq.v3i5.1380)
- Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, 104, 333–339. <https://doi.org/10.1016/j.jbusres.2019.07.039>
- Sudrajat, H. (2018). *Analisis alat evaluasi pada mata pelajaran IPA di kelas V Sekolah Dasar Negeri 3 Nyerot Kecamatan Jonggat Lombok Tengah Nusa Tenggara Barat* [Doctoral dissertation, Universitas Islam Negeri Maulana Malik Ibrahim].
- Thomas, J., & Harden, A. (2008). Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Medical Research Methodology*, 8, Article 45. <https://doi.org/10.1186/1471-2288-8-45>
- Torraco, R. J. (2005). Writing integrative literature reviews: Guidelines and examples. *Human Resource Development Review*, 4(3), 356–367. <https://doi.org/10.1177/1534484305278283>
- Usri, U., Salam, A. M. I., Hasanuddin, M. I., & Badaruddin, S. (2025). Bridging Expectations and Outcomes: A Discrepancy Model Evaluation of Learning Effectiveness in an Islamic Higher Education Context. *AL-ISHLAH: Jurnal Pendidikan*, 17(2), 2058-2069. <https://doi.org/10.35445/alishlah.v17i2.7544>
- Villarroel, V., Bloxham, S., Bruna, D., Bruna, C., & Herrera-Seda, C. (2018). Authentic assessment: Creating a blueprint for course design. *Assessment & Evaluation in Higher Education*, 43(5), 840–854. <https://doi.org/10.1080/02602938.2017.1412396>
- Whittemore, R., & Knafl, K. (2005). The integrative review: Updated methodology. *Journal of Advanced Nursing*, 52(5), 546–553. <https://doi.org/10.1111/j.1365-2648.2005.03621.x>

- Xu, Y., & Brown, G. T. L. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, 58, 149–162. [10.1016/j.tate.2016.05.010](https://doi.org/10.1016/j.tate.2016.05.010)
- Yao, Y., Amos, M., Snider, K., & Brown, T. (2024). The impact of formative assessment on K–12 learning: A meta-analysis. *Educational Research and Evaluation*, 29(7–8), 452–475. <https://doi.org/10.1080/13803611.2024.2363831>
- Zainudin, Z., & Ubabuddin, U. (2023). Ranah kognitif, afektif dan psikomotorik sebagai objek evaluasi hasil belajar peserta didik. *ILJ: Islamic Learning Journal*, 1(3), 915–931. <https://doi.org/10.31004/jptam.v7i3.11169>