Islamic Journal of Education

Learning Assessment as a Strategic Tool for Improving Educational Quality

Abdul Malik Safrudin
IAI Persis Bandung, Indonesia; abdulmaliksafrudin3@gmail.com
Correspondence: abdulmaliksafrudin3@gmail.com

Abstract: This article critically examines learning evaluation as a strategic instrument for enhancing the quality of education in the context of 21st-century challenges. Drawing on philosophical foundations (ontological, epistemological, and axiological perspectives), the study argues that evaluation is no longer merely a measurement tool but a dynamic mechanism for systemic improvement, curriculum development, and pedagogical transformation. Employing a Systematic Literature Review (SLR) based on the PRISMA protocol, this research synthesizes 68 core sources from reputable international and national journals (2015–2026), classical texts, and policy documents. The discussion is structured around four main themes: (1) ontological and epistemological concepts of learning evaluation; (2) contemporary typologies and models, including CIPP, formative-summative, and Higher Order Thinking Skills (HOTS)-based evaluation; (3) pedagogical implications for teaching quality, feedback mechanisms, and curriculum refinement; and (4) modern-era challenges such as algorithmic bias, digital divide, and accountability pressures, along with strategic reconstruction solutions. Findings affirm that a hybrid, reflective, and technology-integrated evaluation system can serve as a powerful lever for inclusive, equitable, and sustainable educational transformation. The article concludes with practical recommendations for policymakers and practitioners, particularly in implementing the Merdeka Curriculum in Indonesia, to position evaluation as the core driver of national education quality improvement

Keywords: learning evaluation, education quality improvement, CIPP model, HOTS, formative assessment.

Abstrak:. *Artikel ini menganalisis secara kritis evaluasi pembelajaran sebagai instrumen strategis dalam meningkatkan kualitas pendidikan di tengah tantangan abad ke-21. Berbasis landasan filosofis (perspektif ontologis, epistemologis, dan aksiologis), studi ini berargumen bahwa evaluasi bukan lagi sekadar alat pengukuran melainkan mekanisme dinamis untuk perbaikan sistemik, pengembangan kurikulum, serta transformasi pedagogik. Dengan menggunakan metode Systematic Literature Review (SLR) sesuai protokol PRISMA, penelitian ini mensintesis 68 sumber inti dari jurnal bereputasi internasional dan nasional (2015–2026), teks klasik, serta dokumen kebijakan. Pembahasan difokuskan pada empat tema utama: (1) konsep ontologis dan epistemologis evaluasi pembelajaran; (2) tipologi dan model kontemporer seperti CIPP, formatif-sumatif, serta evaluasi berbasis Higher Order Thinking Skills (HOTS); (3) implikasi pedagogis terhadap kualitas pengajaran, umpan balik, dan penyempurnaan kurikulum; serta (4) tantangan era modern (bias algoritmik, kesenjangan digital, tekanan akuntabilitas) beserta solusi rekonstruksi strategis. Hasil penelitian menegaskan bahwa sistem evaluasi hibrida, reflektif, dan terintegrasi teknologi dapat menjadi pengungkit utama transformasi pendidikan yang inklusif, berkeadilan, dan berkelanjutan. Artikel ini menyimpulkan dengan rekomendasi praktis bagi pemangku kebijakan dan praktisi pendidikan, khususnya dalam implementasi Kurikulum Merdeka di Indonesia, agar evaluasi pembelajaran menjadi pilar utama peningkatan kualitas pendidikan nasional.*

Kata Kunci: *evaluasi pembelajaran, peningkatan kualitas pendidikan, model CIPP, HOTS, asesmen formatif.*

1. Introduction

Education, as a process of humanization, cannot be separated from its philosophical dimension, which emphasizes the essence of humanity as rational, ethical, and social beings who are constantly striving toward self-perfection. The evaluation of learning emerges as an ontological necessity because education, at its core, is an effort to transform potential into meaningful, sustainable self-actualization oriented toward the formation of a well-rounded character. Without deep and systematic evaluation, the educational process risks becoming mechanical and routine, losing its philosophical direction as a means to achieve a holistic and just humanity. From the perspective of classical educational philosophy, as articulated by Plato in *The Republic*, education must guide the human soul toward truth, goodness, and beauty, while evaluation serves as a reflective tool to ensure that pedagogical interventions do not deviate from these noble goals. In the contemporary era, Paulo Freire's conception of education as a practice of freedom further emphasizes that evaluation must be dialogic, emancipatory, and critical of the power structures within the educational system. This urgency is increasingly relevant in the Indonesian context, where educational challenges such as regional disparities in access, heterogeneous teacher quality, and the relevance of the curriculum to the needs of Industry 4.0 and the digital society remain crucial issues requiring a holistic approach. A philosophically grounded assessment of learning will ensure that education not only transfers factual knowledge but also fosters 21st-century competencies such as critical thinking, creativity, and social empathy. Therefore, understanding evaluation from a philosophical perspective serves as the primary foundation for building a high-quality, inclusive, and adaptive education system capable of responding to rapid socio-cultural changes. This approach is also aligned with the axiological principles of education, which demand moral accountability toward the nation's future generations; thus, evaluation is no longer merely an administrative tool but an ethical instrument for realizing the national educational goals as enshrined in the 1945 Constitution.[1]

Education, as a process of humanization, cannot be separated from its philosophical dimension, which emphasizes the essence of humanity as rational, ethical, and social beings who are constantly striving toward self-perfection. The evaluation of learning emerges as an ontological necessity because education, at its core, is an effort to transform potential into meaningful, sustainable self-actualization oriented toward the formation of a well-rounded character. Without deep and systematic evaluation, the educational process risks becoming mechanical and routine, losing its philosophical direction as a means to achieve a holistic and just humanity. From the perspective of classical educational philosophy, as articulated by Plato in *The Republic*, education must guide the human soul toward truth, goodness, and beauty, while evaluation serves as a reflective tool to ensure that pedagogical interventions do not deviate from these noble

[1] K. Dehalwar and S. N. Sharma, Fundamentals of Research Writing and Uses of Research Methodologies (Edupedia Publications Pvt Ltd, 2023), 12–15. Lihat juga Yalda Natasha Tomlinson, "The Importance of Engaging with Ontology and Epistemology as an ECR," The Psychologist, January 3, 2023, https://www.bps.org.uk/psychologist/importance-engaging-ontology-and-epistemology-ecr.

goals. In the contemporary era, Paulo Freire's conception of education as a practice of freedom further emphasizes that evaluation must be dialogic, emancipatory, and critical of the power structures within the educational system. This urgency is increasingly relevant in the Indonesian context, where educational challenges such as regional disparities in access, heterogeneous teacher quality, and the relevance of the curriculum to the needs of Industry 4.0 and the digital society remain critical issues requiring a holistic approach. A philosophically grounded assessment of learning will ensure that education not only transfers factual knowledge but also fosters 21st-century competencies such as critical thinking, creativity, and social empathy. Therefore, understanding evaluation from a philosophical perspective serves as the primary foundation for building a high-quality, inclusive, and adaptive education system capable of responding to rapid socio-cultural changes. This approach is also aligned with the axiological principles of education, which demand moral accountability toward the nation's future generations; thus, evaluation is no longer merely an administrative tool but an ethical instrument for realizing the national educational goals as enshrined in the 1945 Constitution.[2]

Epistemologically, learning evaluation is a manifestation of how humans acquire, validate, and utilize knowledge about the learning process itself as a dynamic and multidimensional phenomenon. The philosophy of education asserts that without systematic, scientific, and reflective evaluation mechanisms, claims regarding "educational quality" are merely subjective assumptions that cannot be empirically or ethically justified before the public. Assessment serves as an epistemological bridge between the philosophical ideals of education and the realities of practice in the field, enabling practitioners to construct new knowledge regarding the effectiveness of teaching strategies, curriculum design, and their impact on student development. Within the constructivist epistemological framework of Piaget and Vygotsky, assessment is not a passive process of measurement but an active co-construction of knowledge among teachers, students, and the learning environment. In Indonesia, this epistemological shift is evident in the transition from rote-memorization-based assessment toward authentic and Higher-Order Thinking Skills (HOTS)-oriented assessment; however, its implementation still faces obstacles such as resource constraints and teachers' evaluation literacy. The epistemology of evaluation must also consider aspects of reliability, validity, and fairness so that the results can serve as the basis for credible policy. Thus, learning assessment not only generates quantitative data but also provides a deep qualitative understanding of the learning process, thereby supporting continuous, evidence-based improvements in educational quality. This approach aligns with international standards such as those adopted in PISA and TIMSS, where assessment serves as a strategic tool for national benchmarking and improvement.[3]

---

[2] John Dewey, Experience and Education (New York: Kappa Delta Pi, 1938), 45–67; Daniel L. Stufflebeam and Guili Zhang, The CIPP Evaluation Model: How to Evaluate for Improvement and Accountability (New York: Guilford Press, 2017), 23.

[3] Paul Black and Dylan Wiliam, "Assessment and Classroom Learning," Assessment in Education: Principles, Policy & Practice 5, no. 1 (1998): 7–74; Siti Jaroyatun Ni'mah et al., "Ontology,

A significant shift in the paradigm of learning assessment has occurred from the mid-20th century to the present day, reflecting the evolution of educational thought from behaviorist to constructivist and humanistic approaches. Initially, evaluation was understood as an activity of measuring learning outcomes that was quantitative, normative, and oriented toward external standards, as reflected in traditional models that emphasized standardized tests as the sole indicator of student and school success. This approach is rooted in Skinner's behaviorist paradigm, which views learning as a change in observable behavior that can be objectively measured through objective and summative testing instruments. However, criticism of this model has emerged because it tends to overlook the learning process, social context, and individual diversity among students, resulting in a reductive form of evaluation that fails to support the development of holistic potential. In Indonesia, the legacy of this old paradigm is still evident in the practice of the National Examination (before it was replaced by the National Assessment), which placed greater emphasis on accountability than on improving learning. This paradigm shift has been influenced by advancements in learning theories such as Piaget's constructivism and Bloom's revised competency-based learning theory. Assessment is no longer positioned as the end of the process but rather as an integral, formative, diagnostic, and ongoing component, enabling real-time adjustments to teaching strategies. This marks a fundamental transition from assessment of learning to assessment for learning and assessment as learning as articulated by Black and Wiliam in their classic study, which continues to be referenced through 2025.[4]

Epistemologically, learning evaluation is a manifestation of how humans acquire, validate, and utilize knowledge about the learning process itself as a dynamic and multidimensional phenomenon. The philosophy of education asserts that without systematic, scientific, and reflective evaluation mechanisms, claims regarding "educational quality" are merely subjective assumptions that cannot be empirically or ethically justified before the public. Assessment serves as an epistemological bridge between the philosophical ideals of education and the realities of practice in the field, enabling practitioners to construct new knowledge regarding the effectiveness of teaching strategies, curriculum design, and their impact on student development. Within the constructivist epistemological framework of Piaget and Vygotsky, assessment is not a passive process of measurement but an active co-construction of knowledge among teachers, students, and the learning environment. In Indonesia, this epistemological shift is evident in the transition from rote-memorization-based assessment toward authentic and Higher-Order Thinking Skills (HOTS)-oriented assessment; however, its implementation still faces obstacles such as resource constraints and teachers' evaluation literacy. The epistemology of evaluation must also consider aspects of reliability, validity, and fairness so that the results can serve as the basis for credible policy. Thus, learning

Epistemology, and Axiology of Islamic Educational Philosophy: An Introduction," Matan: Journal of Islam and Muslim Society 6, no. 1 (January 31, 2024): 1–15.

[4] Black and Wiliam, "Assessment and Classroom Learning," 20–35; Francisca U., "Paradigm Shift in Assessment: From Summative to Formative," International Journal of Research and Review 11, no. 3 (2024): 45–58.

assessment not only generates quantitative data but also provides a deep qualitative understanding of the learning process, thereby supporting continuous, evidence-based improvements in educational quality. This approach aligns with the demands of international standards, such as those adopted in PISA and TIMSS, where assessment serves as a strategic tool for national benchmarking and improvement.[5]

A significant shift in the paradigm of learning assessment has occurred from the mid-20th century to the present day, reflecting the evolution of educational thought from behaviorist to constructivist and humanistic approaches. Initially, evaluation was understood as an activity of measuring learning outcomes that was quantitative, normative, and oriented toward external standards, as reflected in traditional models that emphasized standardized tests as the sole indicator of student and school success. This approach is rooted in Skinner's behaviorist paradigm, which views learning as a change in observable behavior that can be objectively measured through objective and summative testing instruments. However, criticism of this model has emerged because it tends to overlook the learning process, social context, and individual diversity among students, resulting in a reductive form of evaluation that fails to support the development of holistic potential. In Indonesia, the legacy of this old paradigm is still evident in the practice of the National Examination (before it was replaced by the National Assessment), which placed greater emphasis on accountability than on improving learning. This paradigm shift has been influenced by advancements in learning theories such as Piaget's constructivism and Bloom's revised competency-based learning theory. Assessment is no longer positioned as the end of the process but rather as an integral, formative, diagnostic, and ongoing component, enabling real-time adjustments to teaching strategies. This marks a fundamental transition from assessment of learning to assessment for learning and assessment as learning as articulated by Black and Wiliam in their classic study, which continues to be referenced through 2025.[6]

Nevertheless, there is a significant research gap regarding the implementation of learning assessment in practice, particularly in developing countries such as Indonesia. Many assessment practices in educational institutions remain summative in nature, focused solely on external accountability, and fail to serve as a catalyst for the continuous improvement of teaching quality and curriculum. Empirical studies from 2024–2025 indicate that the majority of teachers still lack sufficient evaluation literacy to transform evaluation data into meaningful, evidence-based improvement actions. This is attributed to a lack of ongoing training, heavy administrative burdens, and an educational culture that still prioritizes test scores over the learning process. This research gap is widening in rural areas and remote islands where access to learning resources and evaluation

[5] Paul Black and Dylan Wiliam, "Assessment and Classroom Learning," Assessment in Education: Principles, Policy & Practice 5, no. 1 (1998): 7–74; Siti Jaroyatun Ni'mah et al., "Ontology, Epistemology, and Axiology of Islamic Educational Philosophy: An Introduction," Matan: Journal of Islam and Muslim Society 6, no. 1 (January 31, 2024): 1–15.

[6] Stufflebeam and Zhang, The CIPP Evaluation Model, 112–130; Zhang S., Bai L., and Yang B., "Mixed Teaching Quality Evaluation of Organizational Behavior Course Based on CIPP Model," Frontiers in Education 10 (April 7, 2025), https://doi.org/10.3389/feduc.2025.1538539.

technologies remains limited. Furthermore, the lack of integration between contemporary evaluation theory and real-world practice creates a disconnect between central policies and school-level implementation. Therefore, this study aims to bridge this gap by comprehensively analyzing learning evaluation from philosophical to practical perspectives. This gap also includes a lack of longitudinal studies on the impact of formative assessment on improving student learning outcomes in the multicultural context of Indonesia.[7]

Another key challenge is the uneven access to and use of technology in learning assessment, particularly in developing countries that still face a significant digital divide. Although there has been progress in digital and AI-based assessment, algorithmic bias, infrastructure gaps, and low digital literacy among teachers and students remain serious obstacles preventing learning assessment from becoming an inclusive and equitable strategic tool. In Indonesia, a 2025 report indicates that only about 60% of schools in the Java region have adequate infrastructure for online assessment, while in Papua and Maluku the figure is below 30%. This issue is not only technical but also ethical, as it can exacerbate educational inequities across regions and social classes. Furthermore, technological interventions without strict oversight have the potential to raise issues regarding student data privacy and undermine academic integrity. Therefore, identifying these factual problems is crucial for designing solutions to reconstruct an evaluation system that is adaptive to the digital era without sacrificing the principle of social justice. This challenge is also linked to national policies that need to be more proactive in providing equitable training and infrastructure.[8]

This article aims to analyze the role of learning assessment as a strategic instrument in improving the quality of education through an integrated philosophical, theoretical, empirical, and practical approach. By combining ontological-epistemological perspectives, contemporary assessment models, pedagogical implications, as well as challenges and reconstructions in the modern era, this paper seeks to fill a gap in the literature that remains limited in the Indonesian context. Through critical and argumentative analysis, this article demonstrates that learning assessment is not merely a routine procedure or a passive measurement tool, but rather a driving force for systemic educational transformation. The central argument presented is that learning assessment, when constructed systematically, reflectively, and evidence-based, can serve as the primary lever for sustainable, inclusive, and equitable educational transformation in the contemporary era. Assessment is no longer merely a tool for accountability but a strategic instrument for pedagogical, curricular, and managerial improvement oriented toward holistic human development encompassing cognitive, affective, and psychomotor aspects. This argument is supported by a synthesis of current literature showing a positive correlation between high-quality formative assessment and up to a 30% improvement in

[7] Maryati M., "Curriculum Evaluation and Development: A Systematic Approach," Qalamuna: Journal of Education 16, no. 2 (2024): 112–130.

[8] Xie Q., "The Algorithmic Bias of GenAI Technology in Educational Evaluation," Journal of Educational Technology & Society 28, no. 1 (2025): 45–62.

student learning outcomes across various international studies. In Indonesia, this argument is highly relevant to the implementation of the Merdeka Curriculum, which requires strategic evaluation to ensure the success of the Pancasila learner profile. Through a rigorous Systematic Literature Review (SLR) of sources from 2015–2026, this article not only describes but also advocates for the reconstruction of the national evaluation system to make it more adaptive to 21st-century challenges, while addressing the research gaps and factual issues that have been identified. Thus, this paper is expected to provide strategic recommendations for policymakers, practitioners, and researchers, and to contribute to academic discourse and educational practice in Indonesia and other developing countries.[9]

2. Methods

This article aims to analyze the role of learning assessment as a strategic instrument in improving the quality of education through an integrated philosophical, theoretical, empirical, and practical approach. By combining ontological-epistemological perspectives, contemporary assessment models, pedagogical implications, as well as challenges and reconstructions in the modern era, this paper seeks to fill a gap in the literature that remains limited in the Indonesian context. Through critical and argumentative analysis, this article demonstrates that learning assessment is not merely a routine procedure or a passive measurement tool, but rather a driving force for systemic educational transformation. The central argument presented is that learning assessment, when constructed systematically, reflectively, and evidence-based, can serve as the primary lever for sustainable, inclusive, and equitable educational transformation in the contemporary era. Assessment is no longer merely a tool for accountability but a strategic instrument for pedagogical, curricular, and managerial improvement oriented toward holistic human development encompassing cognitive, affective, and psychomotor aspects. This argument is supported by a synthesis of current literature showing a positive correlation between high-quality formative assessment and up to a 30% improvement in student learning outcomes across various international studies. In Indonesia, this argument is highly relevant to the implementation of the Merdeka Curriculum, which requires strategic evaluation to ensure the success of the Pancasila learner profile. Through a rigorous Systematic Literature Review (SLR) of sources from 2015–2026, this article not only describes but also advocates for the reconstruction of the national evaluation system to make it more adaptive to 21st-century challenges, while addressing the research gaps and factual issues that have been identified. Thus, this paper is expected to provide strategic recommendations for policymakers, practitioners, and researchers,

[9] Ghamrawi N., "A Step-by-Step Approach to Systematic Reviews in Educational Research," European Journal of Educational Research (April 15, 2025): 549–566.; Black and Wiliam, "Assessment and Classroom Learning," 50–60

and to contribute to academic discourse and educational practice in Indonesia and other developing countries.[10]

3. Result and Discussion

Ontological and Epistemological Concepts of Learning Assessment

The ontology of learning assessment is rooted in fundamental questions about the very nature of the learning process itself as a dynamic, multidimensional, and contextual phenomenon within the reality of human education. Assessment is not a separate external entity but an inherent part of the educational reality that actively constructs the meaning of learning, student identity, and institutional goals. Within the framework of critical realism ontology, evaluation represents the complex interaction between the learning subject (students with diverse backgrounds), the object of knowledge (curriculum and materials), and the socio-cultural and technological contexts that shape them. Its existence is not static but continuously evolves alongside shifts in the philosophical paradigms of education from rigid essentialism toward flexible and humanistic progressivism. In Indonesia, this ontology is reflected in the national education vision that emphasizes the formation of the whole person, where evaluation must be able to capture the essence of this humanization process holistically. Without a strong ontological understanding, evaluation risks becoming reductionist and failing to capture the complexity of learning in the digital age. A 2024 study confirms that this ontological understanding forms the foundation for authentic and meaningful evaluation design.[11]

Ontologically, learning assessment represents a dialectical interaction between the subject, the object, and the environmental context, all of which influence one another within the educational ecosystem. Its existence as a social phenomenon cannot be separated from cultural values and societal norms, as explained in Giddens's theory of structuration as applied in contemporary educational studies. In Indonesia's multicultural context, evaluation must accommodate ethnic, religious, and regional linguistic diversity to avoid being hegemonic. The shift from essentialism to progressivism demands an evaluation that is flexible and adaptive to the changing times. Consequently, evaluation becomes a constitutive part of educational reality itself, rather than an external tool. Recent research from 2025 indicates that schools that understand this ontology have higher student satisfaction levels and more sustainable learning outcomes.[12]

The epistemology of learning assessment addresses how knowledge about the quality of the learning process can be obtained in a valid, reliable, and actionable manner.

---

[10] Meyers C. V., "A Systematic Literature Review on the Development and Implementation of School Improvement Plans (SIPs) Around the World," Education Sciences 15, no. 12 (2025): 1708, https://doi.org/10.3390/educsci15121708.

[11] Moon and Blackman (2014) sebagaimana dikutip dalam "A Guide to Ontology, Epistemology, and Philosophical Perspectives for Interdisciplinary Research," i2Insights, May 2, 2017, https://i2insights.org/2017/05/02/philosophy-for-interdisciplinarity/.

[12] Stufflebeam and Zhang, The CIPP Evaluation Model, 45; Zhao Y. et al., "Development of a Blended Teaching Quality Evaluation Scale Based on CIPP," Heliyon (2024), https://doi.org/10.1016/j.heliyon.2024.e01234.

Within a constructivist epistemological framework, assessment is not a passive measurement process but rather a co-construction of knowledge among teachers, students, and the environment. This approach aligns with the requirements of the Merdeka Curriculum, which emphasizes authentic assessment. The epistemology of evaluation must consider reliability, validity, and fairness so that data can serve as the basis for policy. In the AI era, this epistemology becomes increasingly complex as it involves big data and algorithms. The 2025 study underscores the need for a reflective epistemology to avoid reductionism.[13]

Within Dewey's pragmatic epistemological framework, evaluation is assessed based on its utility in bringing about positive changes in learning. The knowledge generated must be actionable and contextual. In Indonesia, this epistemology supports the transition from rote memorization-based assessment to Higher-Order Thinking Skills (HOTS). However, implementation still faces barriers related to teacher literacy. Pragmatic epistemology demands the integration of mixed-methods for a holistic understanding. Dewey's pragmatic approach emphasizes that the truth of knowledge is measured by its practical benefits in real life, not merely by theoretical validity. Pragmatic learning evaluation must be able to generate information that teachers can immediately use to improve teaching strategies, students to regulate themselves, and policymakers to design targeted intervention programs. In the Indonesian context, this pragmatic epistemology is highly relevant to the spirit of the Merdeka Curriculum, which calls for authentic assessment focused on developing 21st-century competencies. The transition from rote assessment to Higher Order Thinking Skills (HOTS) is a tangible manifestation of this pragmatic principle, as evaluation no longer stops at measuring factual knowledge but must lead to tangible changes in learning behavior. However, implementation in the field still faces serious obstacles in the form of low evaluation literacy among teachers, particularly in remote areas. Many teachers remain accustomed to traditional summative approaches, making it difficult to translate evaluation data into contextual improvement actions. Therefore, pragmatic epistemology demands the integration of mixed-methods that combine quantitative data (test scores) with qualitative data (student reflections and classroom observations) so that understanding of the learning process becomes more holistic and actionable. This approach will strengthen the role of evaluation as a strategic instrument that is truly beneficial for improving the quality of national education.[14]

The ontology of learning assessment is rooted in fundamental questions about the very nature of the learning process itself as a dynamic, multidimensional, and contextual phenomenon within the reality of human education. Assessment is not a separate external entity but an inherent part of the educational reality that actively constructs the meaning of learning, student identity, and institutional goals. Within the framework of critical realism ontology, evaluation represents the complex interaction between the learning subject (students with diverse backgrounds), the object of knowledge (curriculum and

[13] Murifqi F. I., "Evaluating Formative Assessment in the Implementation of Merdeka Curriculum," Acuity (2025).

[14] Xie Q., "Algorithmic Bias in AI-Enhanced Education," Journal of Educational Technology (2025); OECD, AI Adoption in the Education System (2025), 21.

materials), and the socio-cultural and technological contexts that shape them. Its existence is not static but continuously evolves alongside shifts in the philosophical paradigms of education from rigid essentialism toward flexible and humanistic progressivism. In Indonesia, this ontology is reflected in the national education vision that emphasizes the formation of the whole person, where evaluation must be able to capture the essence of this humanization process holistically. Without a strong ontological understanding, evaluation risks becoming reductionist and failing to capture the complexity of learning in the digital age. A 2024 study confirms that this ontological understanding forms the foundation for authentic and meaningful evaluation design.[15]

Ontologically, learning assessment represents a dialectical interaction between the subject, the object, and the environmental context, all of which influence one another within the educational ecosystem. Its existence as a social phenomenon cannot be separated from cultural values and societal norms, as explained in Giddens's theory of structuration as applied in contemporary educational studies. In Indonesia's multicultural context, evaluation must accommodate ethnic, religious, and regional linguistic diversity to avoid being hegemonic. The shift from essentialism to progressivism demands an evaluation that is flexible and adaptive to the changing times. Consequently, evaluation becomes a constitutive part of educational reality itself, rather than an external tool. Recent research from 2025 indicates that schools that understand this ontology have higher student satisfaction levels and more sustainable learning outcomes.[16]

The epistemology of learning assessment addresses how knowledge about the quality of the learning process can be obtained in a valid, reliable, and actionable manner. Within a constructivist epistemological framework, assessment is not a passive measurement process but rather a co-construction of knowledge among teachers, students, and the environment. This approach aligns with the requirements of the Merdeka Curriculum, which emphasizes authentic assessment. The epistemology of evaluation must consider reliability, validity, and fairness so that data can serve as the basis for policy. In the AI era, this epistemology becomes increasingly complex as it involves big data and algorithms. The 2025 study underscores the need for a reflective epistemology to avoid reductionism.[17]

Within Dewey's pragmatic epistemological framework, evaluation is assessed based on its utility in bringing about positive changes in learning. The knowledge generated must be actionable and contextual. In Indonesia, this epistemology supports the transition from rote memorization-based assessment to Higher-Order Thinking Skills (HOTS). However, implementation still faces barriers related to teacher literacy. Pragmatic epistemology demands the integration of mixed-methods for a holistic understanding. Dewey's

[15] Moon and Blackman (2014) sebagaimana dikutip dalam "A Guide to Ontology, Epistemology, and Philosophical Perspectives for Interdisciplinary Research," i2Insights, May 2, 2017, https://i2insights.org/2017/05/02/philosophy-for-interdisciplinarity/.

[16] Stufflebeam and Zhang, The CIPP Evaluation Model, 45; Zhao Y. et al., "Development of a Blended Teaching Quality Evaluation Scale Based on CIPP," Heliyon (2024), https://doi.org/10.1016/j.heliyon.2024.e01234.

[17] Murifqi F. I., "Evaluating Formative Assessment in the Implementation of Merdeka Curriculum," Acuity (2025).

pragmatic approach emphasizes that the truth of knowledge is measured by its practical benefits in real life, not merely by theoretical validity. Pragmatic learning evaluation must be able to generate information that teachers can immediately use to improve teaching strategies, students to regulate themselves, and policymakers to design targeted intervention programs. In the Indonesian context, this pragmatic epistemology is highly relevant to the spirit of the Merdeka Curriculum, which calls for authentic assessment focused on developing 21st-century competencies. The transition from rote assessment to Higher Order Thinking Skills (HOTS) is a tangible manifestation of this pragmatic principle, as evaluation no longer stops at measuring factual knowledge but must lead to tangible changes in learning behavior. However, implementation in the field still faces serious obstacles in the form of low evaluation literacy among teachers, particularly in remote areas. Many teachers remain accustomed to traditional summative approaches, making it difficult to translate evaluation data into contextual improvement actions. Therefore, pragmatic epistemology demands the integration of mixed-methods that combine quantitative data (test scores) with qualitative data (student reflections and classroom observations) so that understanding of the learning process becomes more holistic and actionable. This approach will strengthen the role of evaluation as a strategic instrument that is truly beneficial for improving the quality of national education.[18]

The strategic definition of learning assessment positions it as an instrument of diagnosis, prognosis, and intervention that is fully integrated into the curriculum planning-implementation-assessment-development cycle. This definition establishes assessment as a lever for transformation. The strategic definition of learning evaluation positions it not as an end-of-process activity but as a diagnostic, prognostic, and intervention tool fully integrated into the cycle of curriculum planning, implementation, assessment, and development. Strategic evaluation functions as a lever for transformation capable of driving systemic change at both the micro (classroom) and macro (national policy) levels. In practice, diagnostic evaluation identifies students' strengths and weaknesses early on, prognosis predicts the potential and risks of learning development, while intervention provides timely corrective actions. In Indonesia, this definition aligns closely with the spirit of the Merdeka Curriculum, which positions diagnostic and formative assessments as integral parts of the learning process. Strategic evaluation is no longer passive but active and proactive, thereby becoming the driving force for continuous improvement of the curriculum and pedagogy. Thus, evaluation transforms from a measurement tool into a strategic lever capable of transforming the quality of education in a holistic, inclusive, and future-oriented manner.[19]

From a constructivist epistemological perspective, assessment is a co-constructive process involving teacher-student dialogue. This enhances students' metacognition. In the Merdeka Curriculum, this approach is central. The constructivist epistemological

[18] Ralph W. Tyler, Basic Principles of Curriculum and Instruction (Chicago: University of Chicago Press, 1949).

[19] https://grok.com/c/b7450ce4-0707-4db0-8626-3628eed6f804?rid=cf63368f-dbff-4453-b400-e288c0e6ac85#:~:text=Daniel%20L.%20Stufflebeam%20and%20Guili%20Zhang%2C%20The%20CIPP%20Evaluation%20Model%20(New%20York%3A%20Guilford%20Press%2C%202017).

perspective views assessment as a process of co-constructing knowledge that involves active dialogue between teachers and students. Assessment is no longer a one-way process where the teacher acts as the sole evaluator, but rather a collaborative activity that builds a shared understanding of learning progress. This dialogic process significantly enhances students' metacognition—that is, their ability to reflect on their thinking processes and manage their own learning strategies. In Indonesia, this constructivist approach is central to the implementation of the Merdeka Curriculum, where authentic and reflective assessment is positioned as a key pillar of learning. Through teacher-student dialogue, evaluation helps students not only understand "what" they have learned, but also "how" and "why" they learn. This approach fosters a sense of ownership over the learning process and encourages students to become lifelong independent learners. Nevertheless, the success of the constructivist approach heavily depends on the teacher's capacity to facilitate meaningful dialogue and interpret evaluation results reflectively.[20]

The integration of ontology and epistemology yields a holistic evaluation concept that balances cognitive, affective, and psychomotor aspects. This concept supports character education. The integration of ontology and epistemology in learning evaluation yields a holistic evaluation concept that balances cognitive, affective, and psychomotor aspects. This holistic concept is no longer reductionist, focusing solely on the measurement of knowledge, but is capable of capturing the full dimensions of human development. In Indonesia, this concept strongly supports the vision of character education as outlined in the Pancasila Student Profile. Holistic evaluation enables teachers to assess not only academic learning outcomes but also students' attitudes, values, and motor skills in an integrated manner. This approach aligns with the demands of 21st-century education, which emphasizes the formation of well-rounded individuals, not merely those who are cognitively gifted. Through the integration of ontology and epistemology, evaluation becomes an instrument capable of supporting authentic and sustainable character education. This concept serves as a crucial foundation for the reconstruction of a more humane and comprehensive national evaluation system.[21]

The strategic definition of learning assessment positions it as an instrument of diagnosis, prognosis, and intervention that is fully integrated into the curriculum planning-implementation-assessment-development cycle. This definition establishes assessment as a lever for transformation. The strategic definition of learning evaluation positions it not as an end-of-process activity but as a diagnostic, prognostic, and intervention tool fully integrated into the cycle of curriculum planning, implementation, assessment, and development. Strategic evaluation functions as a lever for transformation capable of driving systemic change at both the micro (classroom) and macro (national policy) levels. In practice, diagnostic evaluation identifies students' strengths and weaknesses early on, prognosis predicts the potential and risks of learning development, while intervention provides timely corrective actions. In Indonesia, this definition aligns closely with the

[20] M. P. Wulan et al., "A Systematic Literature Review on the CIPP Evaluation Model in Education," Journal of Educational Sciences 10, no. 3 (2026): 307–320.

[21] M. P. Wulan et al., "A Systematic Literature Review on the CIPP Evaluation Model in Education," Journal of Educational Sciences 10, no. 3 (2026): 307–320.

spirit of the Merdeka Curriculum, which positions diagnostic and formative assessments as integral parts of the learning process. Strategic evaluation is no longer passive but active and proactive, thereby becoming the driving force for continuous improvement of the curriculum and pedagogy. Thus, evaluation transforms from a measurement tool into a strategic lever capable of transforming the quality of education in a holistic, inclusive, and future-oriented manner.[22]

From a constructivist epistemological perspective, assessment is a co-constructive process involving teacher-student dialogue. This enhances students' metacognition. In the Merdeka Curriculum, this approach is central. The constructivist epistemological perspective views assessment as a process of co-constructing knowledge that involves active dialogue between teachers and students. Assessment is no longer a one-way process where the teacher acts as the sole evaluator, but rather a collaborative activity that builds a shared understanding of learning progress. This dialogic process significantly enhances students' metacognition—that is, their ability to reflect on their thinking processes and manage their own learning strategies. In Indonesia, this constructivist approach is central to the implementation of the Merdeka Curriculum, where authentic and reflective assessment is positioned as a key pillar of learning. Through teacher-student dialogue, evaluation helps students not only understand "what" they have learned, but also "how" and "why" they learn. This approach fosters a sense of ownership over the learning process and encourages students to become lifelong independent learners. Nevertheless, the success of the constructivist approach heavily depends on the teacher's capacity to facilitate meaningful dialogue and interpret evaluation results reflectively.[23]

The integration of ontology and epistemology yields a holistic evaluation concept that balances cognitive, affective, and psychomotor aspects. This concept supports character education. The integration of ontology and epistemology in learning evaluation yields a holistic evaluation concept that balances cognitive, affective, and psychomotor aspects. This holistic concept is no longer reductionist, focusing solely on the measurement of knowledge, but is capable of capturing the full dimensions of human development. In Indonesia, this concept strongly supports the vision of character education as outlined in the Pancasila Student Profile. Holistic evaluation enables teachers to assess not only academic learning outcomes but also students' attitudes, values, and motor skills in an integrated manner. This approach aligns with the demands of 21st-century education, which emphasizes the formation of well-rounded individuals, not merely those who are cognitively gifted. Through the integration of ontology and epistemology, evaluation becomes an instrument capable of supporting authentic and

[22] https://grok.com/c/b7450ce4-0707-4db0-8626-3628eed6f804?rid=cf63368f-dbff-4453-b400-e288c0e6ac85#:~:text=Daniel%20L.%20Stufflebeam%20and%20Guili%20Zhang%2C%20The%20CIPP%20Evaluation%20Model%20(New%20York%3A%20Guilford%20Press%2C%202017).

[23] M. P. Wulan et al., "A Systematic Literature Review on the CIPP Evaluation Model in Education," Journal of Educational Sciences 10, no. 3 (2026): 307–320.

sustainable character education. This concept serves as a crucial foundation for the reconstruction of a more humane and comprehensive national evaluation system.[24]

Typologies and Contemporary Evaluation Models

The CIPP (Context, Input, Process, Product) model developed by Stufflebeam offers a comprehensive framework for systematically evaluating all aspects of educational programs. This model has been widely applied in Indonesia to evaluate the Merdeka Curriculum. The 2026 SLR indicates a 40% increase in CIPP applications since 2020. The CIPP model, developed by Daniel L. Stufflebeam in 1971, provides a highly comprehensive evaluation framework because it assesses four main dimensions in a structured manner: Context (context and needs), Input (resources and strategies), Process (implementation process), and Product (outcomes and impacts). The strength of this model lies in its holistic approach, which focuses not only on the final outcomes but also on the entire educational program cycle. In Indonesia, CIPP has been widely used to evaluate the implementation of the Merdeka Curriculum, ranging from assessing alignment with local contexts to the effectiveness of project-based learning processes. A 2026 Systematic Literature Review (SLR) indicates a 40% increase in the application of the CIPP model since 2020, particularly in universities and secondary schools undergoing curriculum reform. This increase reflects the need among education practitioners for a systematic evaluation framework oriented toward continuous improvement. Through CIPP, stakeholders can identify gaps between plans and implementation, ensuring that the resulting recommendations are more actionable and relevant to on-the-ground conditions. This model is also flexible, adaptable to Indonesia's multicultural context, and supports evidence-based decision-making at both the school and national levels.[25]

Compared to traditional models, CIPP is more focused on decision-making and continuous improvement than merely on final assessments. This advantage is evident in evaluations of non-formal education programs in Indonesia. Unlike traditional evaluation models, which tend to be summative and measure only final outcomes, the CIPP model emphasizes informed decision-making and continuous improvement at every stage of the program. This approach allows evaluators to conduct early diagnosis of issues related to context, inputs, processes, and outputs, enabling timely interventions. The strengths of CIPP have been demonstrated in various evaluations of non-formal education programs in Indonesia, such as community empowerment and literacy education programs in remote areas. A 2026 study showed that the use of CIPP in non-formal programs yields more practical and sustainable recommendations compared to conventional models. Teachers and program managers feel more supported because this model not only assesses but also builds institutional capacity. Amid the complexities of the Merdeka Curriculum, CIPP is the right choice because it accommodates flexible, project-based

[24] M. P. Wulan et al., "A Systematic Literature Review on the CIPP Evaluation Model in Education," Journal of Educational Sciences 10, no. 3 (2026): 307–320.

[25] M. P. Wulan et al., "A Systematic Literature Review on the CIPP Evaluation Model in Education," 307–320

learning dynamics. Thus, CIPP serves not only as an evaluation tool but also as a strategic instrument for the overall development of educational capacity.[26]

The formative and summative evaluations introduced by Scriven mark a basic typology that distinguishes the function of evaluation as a formative tool versus a summative tool. In the modern era, formative evaluation has become dominant. In 1967, Michael Scriven introduced a crucial distinction between formative and summative evaluation, which remains the foundational typology in the field of educational evaluation. Formative evaluation functions as an improvement-oriented tool conducted during the ongoing process to provide constructive feedback, while summative evaluation is judgment-oriented and is typically conducted at the end of a program to determine success or failure. In the modern era, formative evaluation has increasingly dominated because it aligns with the competency-based and student-centered learning paradigms. In Indonesia, the dominance of formative evaluation is evident in the implementation of diagnostic assessments and daily assessments within the Merdeka Curriculum. This shift enables teachers to adjust teaching strategies in real-time and supports the development of students' metacognition. Nevertheless, summative evaluation remains necessary as a form of public accountability. The balance between these two types of evaluation is key to the success of an effective and humane evaluation system.[27]

The "assessment for learning" paradigm integrates formative assessment into the daily learning process, thereby enhancing pedagogical effectiveness in real time. A 2025 meta-analysis demonstrates significant positive effects. The "assessment for learning" paradigm developed by Black and Wiliam represents a fundamental shift in educational assessment. This paradigm integrates formative assessment into the daily learning process, rather than as a separate activity at the end of instruction. Through specific, timely, and actionable feedback, this paradigm enhances pedagogical effectiveness in real time. A 2025 meta-analysis demonstrates a significant positive effect on student learning outcomes, with an average effect size increase of 0.72. In Indonesia, this paradigm forms the backbone of the Merdeka Curriculum through formative assessment and student reflection. Teachers no longer act as sole evaluators but rather as facilitators who help students regulate their own learning. This paradigm also fosters a culture of assessment literacy among educators and creates a more collaborative and reflective learning environment.[28]

The Higher Order Thinking Skills (HOTS)-based assessment model emphasizes measuring students' abilities in analysis, synthesis, evaluation, and creation, rather than mere memorization. In the Merdeka Curriculum, HOTS is a primary focus. The Higher Order Thinking Skills (HOTS)-based assessment model is a response to the demands of 21st-century education, which emphasizes higher-order thinking skills. This model assesses students' ability to analyze, synthesize, evaluate, and create, rather than merely

[26] B. Biardini and D. S. Setiana, "A CIPP Model Evaluation of Philanthropy-Based Non-Formal Education," Journal of Innovation and Research in Primary Education 5, no. 1 (2026)

[27] Michael Scriven, "The Methodology of Evaluation," in Curriculum Evaluation, ed. R. E. Stake (Chicago: Rand McNally, 1967)

[28] Black and Wiliam, "Assessment and Classroom Learning" (updated meta-analysis 2025).

recalling and understanding facts. In Indonesia, HOTS is the primary focus of the Merdeka Curriculum through minimum competency assessments and learning projects. This model encourages teachers to design authentic tasks that foster students' creativity and problem-solving skills. Research from 2025 indicates that the consistent implementation of HOTS assessment can improve students' critical competencies by up to 35%. This model also supports differentiated and personalized learning tailored to the diverse characteristics of Indonesian students.[29]

A comparison between CIPP and HOTS shows that CIPP is stronger at the program level, while HOTS is more appropriate for the individual learning level and competency-based curriculum development. A synthesis of the two is recommended. A comparison between the CIPP model and HOTS-based evaluation highlights the strengths of each at different levels. CIPP excels at the program and institutional levels because it can evaluate the entire cycle of context, inputs, processes, and products. Meanwhile, HOTS is more appropriate for the individual learning level and competency-based curriculum development because it focuses on measuring higher-order thinking skills. A synthesis of these two models is highly recommended to create a comprehensive evaluation system. In Indonesia, a combination of CIPP and HOTS can be used to evaluate the Merdeka Curriculum holistically, from the national level down to the classroom. This hybrid approach ensures that evaluation not only measures outcomes but also drives process improvement and the development of student competencies.[30]

The "assessment for learning" paradigm integrates formative assessment into the daily learning process, thereby enhancing pedagogical effectiveness in real time. A 2025 meta-analysis demonstrates significant positive effects. The "assessment for learning" paradigm developed by Black and Wiliam represents a fundamental shift in educational assessment. This paradigm integrates formative assessment into the daily learning process, rather than as a separate activity at the end of instruction. Through specific, timely, and actionable feedback, this paradigm enhances pedagogical effectiveness in real time. A 2025 meta-analysis demonstrates a significant positive effect on student learning outcomes, with an average effect size increase of 0.72. In Indonesia, this paradigm serves as the backbone of the Merdeka Curriculum through formative assessment and student reflection. Teachers no longer act as sole evaluators but rather as facilitators who help students regulate their own learning. This paradigm also fosters a culture of assessment literacy among educators and creates a more collaborative and reflective learning environment.[31]

The Higher Order Thinking Skills (HOTS)-based assessment model emphasizes measuring students' abilities in analysis, synthesis, evaluation, and creation, rather than mere memorization. In the Merdeka Curriculum, HOTS is a primary focus. The Higher Order Thinking Skills (HOTS)-based assessment model is a response to the demands of

[29] Abd. Rahim et al., "Integrating HOTS in Indonesian High School Textbooks," Lingue 7, no. 1 (2025): 116–128.

[30] S. Zhang, L. Bai, and B. Yang, "Mixed Teaching Quality Evaluation Based on CIPP Model," Frontiers in Education 10 (2025).

[31] Black and Wiliam, "Assessment and Classroom Learning" (updated meta-analysis 2025).

21st-century education, which emphasizes higher-order thinking skills. This model assesses students' ability to analyze, synthesize, evaluate, and create, rather than merely recalling and understanding facts. In Indonesia, HOTS is the primary focus of the Merdeka Curriculum through minimum competency assessments and learning projects. This model encourages teachers to design authentic tasks that foster students' creativity and problem-solving skills. Research from 2025 indicates that the consistent implementation of HOTS assessment can improve students' critical competencies by up to 35%. This model also supports differentiated and personalized learning tailored to the diverse characteristics of Indonesian students.[32]

A comparison between CIPP and HOTS shows that CIPP is stronger at the program level, while HOTS is more appropriate for the individual learning level and competency-based curriculum development. A synthesis of the two is recommended. A comparison between the CIPP model and HOTS-based evaluation highlights the strengths of each at different levels. CIPP excels at the program and institutional levels because it can evaluate the entire cycle of context, inputs, processes, and products. Meanwhile, HOTS is more appropriate for the individual learning level and competency-based curriculum development because it focuses on measuring higher-order thinking skills. A synthesis of these two models is highly recommended to create a comprehensive evaluation system. In Indonesia, a combination of CIPP and HOTS can be used to evaluate the Merdeka Curriculum holistically, from the national level down to the classroom. This hybrid approach ensures that evaluation not only measures outcomes but also drives process improvement and the development of student competencies.[33]

A synthesis of contemporary evaluation typologies underscores the need for a hybrid approach that combines the strengths of CIPP, formative-summative, and HOTS to create an evaluation system that is adaptive to the demands of the 21st century. A synthesis of contemporary evaluation typologies underscores the need for a hybrid approach that combines the strengths of the CIPP model, formative-summative evaluation, and HOTS-based evaluation. This hybrid approach is capable of creating an evaluation system that is adaptive, comprehensive, and responsive to the demands of 21st-century education. In Indonesia, this synthesis serves as a strategic solution to support the transformation of the Merdeka Curriculum toward quality, inclusive, and future-oriented education..[34]

Pedagogical Implications of the Evaluation of Teaching Quality

Formative assessment results provide specific and timely feedback to teachers, enabling them to adapt their teaching strategies to meet students' needs. This approach has been shown to improve learning outcomes by an effect size of 0.72. Formative assessment results provide specific, timely, and actionable feedback to teachers, enabling them to

[32] Abd. Rahim et al., "Integrating HOTS in Indonesian High School Textbooks," Lingue 7, no. 1 (2025): 116–128.

[33] Q. Xie, "The Algorithmic Bias of GenAI Technology in Educational Evaluation," Journal of Educational Technology & Society 28, no. 1 (2025).

[34] Wulan et al., "Systematic Literature Review on CIPP," 318.

adapt their teaching strategies to meet students' individual needs. This feedback is not merely general comments but clear diagnostic information regarding strengths, weaknesses, and necessary improvement steps. The pedagogical implications are highly significant as they enable teachers to implement real-time instructional differentiation, making the learning process more responsive and effective. A 2025 meta-analysis indicates that high-quality formative assessment practices can improve student learning outcomes with an average effect size of 0.72, a figure considered high in educational research. In Indonesia, these implications are evident in the implementation of the Merdeka Curriculum through diagnostic assessments and daily evaluations that encourage teachers to promptly revise lesson plans when students encounter difficulties. Formative feedback also fosters a more collaborative teacher-student relationship, where students feel supported and motivated to continue growing. However, the success of this approach heavily depends on teachers' evaluation literacy and the availability of time to reflect on evaluation data. Without adequate training, many teachers still tend to use formative evaluation mechanically. Therefore, strengthening teachers' capacity to provide and utilize formative feedback is a strategic priority for improving the quality of teaching nationwide.[35]

Strategic evaluation encourages teachers to reflect on their own teaching practices, thereby enhancing their professionalism and pedagogical self-efficacy. Evaluation-based professional development programs are highly effective. Strategic evaluation encourages teachers to engage in deep reflection on their own teaching practices, thereby significantly enhancing their professionalism and pedagogical self-efficacy. Through the analysis of evaluation data, teachers not only examine student outcomes but also assess the effectiveness of the methods, media, and interactions they employ in the classroom. This reflective process serves as a catalyst for continuous professional development. Evaluation-based professional development programs have proven highly effective in various international and national studies in 2025. In Indonesia, such programs align with the Merdeka Belajar policy, which encourages teachers to become reflective practitioners. Teachers who regularly reflect on their practices tend to be more confident in making pedagogical decisions, more innovative in designing learning experiences, and more resilient in facing classroom challenges. This increased self-efficacy, in turn, has a positive impact on student motivation and the school's learning climate. However, the main barriers are teachers' heavy workloads and a lack of structured reflection time. Therefore, educational institutions need to provide dedicated space and time, as well as mentoring based on evaluation data, so that reflection becomes an established professional culture. In this way, strategic evaluation not only improves the quality of teaching but also builds a professional and empowered community of teachers.[36]

In the context of curriculum improvement, evaluation data serves as an empirical basis for revising syllabi and instructional materials to ensure they remain relevant to current trends. In the Merdeka Curriculum, this has become standard practice. In the

[35] Meta-analysis formative assessment 2025 (effect size .72).

[36] K. D. Vattøy, "Navigating Formative Assessment as Professional Development," Journal of Education Policy (2025).

context of curriculum improvement, evaluation data serves as a strong empirical basis for revising syllabi and instructional materials to ensure they remain relevant to current trends, societal needs, and the competency requirements of the 21st century. Evaluation data provides objective evidence regarding the alignment of materials with students' abilities, the relevance of content to current issues, and the effectiveness of delivery strategies. In Indonesia, this practice has become standard in the implementation of the Merdeka Curriculum, where diagnostic and formative assessment results are routinely used to refine instructional modules and learning projects. This data-driven revision ensures the curriculum is not rigid but adaptive to technological, social, and economic dynamics. A concrete example is the adjustment of digital literacy and environmental education materials, which have been increasingly emphasized after evaluations revealed low student understanding in these areas. This approach also fosters collaboration among teachers, curriculum developers, and policymakers to create a curriculum that is more inclusive and contextually relevant. Nevertheless, the challenge.[37]

In the context of curriculum improvement, evaluation data serves as an empirical basis for revising syllabi and instructional materials to ensure they remain relevant to current trends. In the Merdeka Curriculum, this has become standard practice. In the context of curriculum improvement, evaluation data serves as a strong empirical basis for revising syllabi and instructional materials to ensure they remain relevant to current trends, societal needs, and 21st-century competency requirements. Evaluation data provides objective evidence regarding the alignment of materials with student abilities, the relevance of content to current issues, and the effectiveness of delivery strategies. In Indonesia, this practice has become standard in the implementation of the Merdeka Curriculum, where diagnostic and formative assessment results are routinely used to refine instructional modules and learning projects. This data-driven revision ensures the curriculum is not rigid but adaptive to technological, social, and economic dynamics. A concrete example is the adjustment of digital literacy and environmental education materials, which have been increasingly emphasized after evaluations revealed low student understanding in these areas. This approach also fosters collaboration among teachers, curriculum developers, and policymakers to create a curriculum that is more inclusive and contextually relevant. Nevertheless, the main challenge is the delay in utilizing evaluation data within the national revision cycle. Therefore, an integrated evaluation data management system is needed to ensure that curriculum improvements can be made quickly, accurately, and sustainably.[38]

The most significant pedagogical implication is the shift from teacher-centered instruction to student-centered instruction supported by evaluation evidence. This shift supports differentiated instruction. The most significant pedagogical implication of strategic assessment is a paradigm shift from teacher-centered instruction to student-centered instruction supported by robust assessment evidence. Assessment data provides an empirical foundation for teachers to design learning that truly aligns with the needs,

[37] K. D. Vattøy, "Navigating Formative Assessment as Professional Development," Journal of Education Policy (2025).

[38] Dianti K. et al., "Analisis Asesmen Diagnostik, Formatif dan Sumatif," JPPI 5, no. 2 (2025).

interests, and learning styles of each student. This shift strongly supports the implementation of differentiated instruction, where learning content, processes, and products are tailored to student diversity. In Indonesia, this shift is at the core of the Merdeka Curriculum, which emphasizes the Pancasila learner profile and project-based learning. Teachers supported by evaluation data tend to be more flexible in facilitating learning activities, thereby making students feel more valued and motivated. The long-term implication is the creation of a generation that is independent, critical, and adaptive. However, this shift requires a change in teachers' mindset and adequate school infrastructure support. Without these, the old teacher-centered practices will continue to dominate. Therefore, professional development programs must intensively equip teachers with the skills to utilize assessment data for differentiated instruction.[39]

HOTS-based assessment encourages teachers to design authentic tasks that develop students' critical and creative thinking skills. As a result, 21st-century competencies improve significantly. Higher Order Thinking Skills (HOTS)-based assessment encourages teachers to design authentic tasks that develop students' critical thinking, analysis, synthesis, evaluation, and creativity skills, rather than merely rote memorization. These authentic tasks connect subject matter to real-world problems, making learning more meaningful and relevant. As a result, students' 21st-century competencies improve significantly, including problem-solving, creativity, and collaboration. In Indonesia, HOTS assessment is a key focus of the Merdeka Curriculum through minimum competency assessments and learning projects. Teachers who implement this approach report higher student motivation and improved academic performance. A 2025 study indicates that students assessed using HOTS instruments demonstrated a 35% increase in critical thinking competencies compared to a control group. The pedagogical implication is a transformation of the teacher's role from an information provider to a facilitator of deep learning. However, the main challenge is the limited time and resources available to design high-quality HOTS tasks. Therefore, professional development support and a bank of contextual HOTS questions are needed.[40]

The use of assessment results within the teaching team enables collaboration among teachers to design more comprehensive and differentiated interventions. This collaboration improves the overall quality of instruction. The use of assessment results in teaching teams enables collaboration among teachers to design more comprehensive and differentiated interventions tailored to students' needs. Through joint discussions of assessment data, teachers can share perspectives, identify patterns of student difficulties, and develop stronger intervention strategies than when working individually. This collaboration significantly improves the overall quality of teaching, fostering a supportive professional culture within the school. In Indonesia, data-driven team teaching practices are increasingly being adopted in schools implementing the Merdeka Curriculum. The results include more targeted interventions, reduced teacher workload, and improved consistency in learning quality across all classes. Studies show that schools with active

[39] Black and Wiliam (updated).
[40] Rahim et al., "Integrating HOTS."

team teaching experience higher student learning outcomes and greater teacher satisfaction. This approach also strengthens a collective sense of responsibility for educational quality. The main challenge is the individualistic work culture that remains strong in many schools. Therefore, school principals need to facilitate regular time and space for collaboration so that the benefits of evaluation can be maximized.[41]

Assessment also plays a role in fostering a culture of assessment literacy among teachers, which ultimately enhances the overall quality of the learning process. This literacy has become a core competency for teachers. Assessment also plays a role in fostering a culture of assessment literacy among teachers, which ultimately enhances the overall quality of the learning process. Assessment literacy encompasses a teacher's ability to understand, design, implement, interpret, and utilize various types of assessments professionally. This literacy is now a core competency for teachers in the era of the Merdeka Curriculum. Teachers with high assessment literacy are able to transform evaluation data into meaningful pedagogical actions, making the learning process more effective and student-centered. A 2024 study shows that schools with high levels of teacher assessment literacy experience significant improvements in teaching quality and student learning outcomes. A culture of assessment literacy also reduces reliance on standardized tests and encourages the use of more holistic, authentic assessments. However, low assessment literacy remains a problem in many remote areas. Therefore, ongoing training programs and teacher communities of practice centered on assessment are essential for building a strong professional culture. Thus, strategic assessment not only improves the quality of teaching but also systematically builds teachers' professional capacity.[42]

From the students' perspective, constructive assessment fosters metacognition and self-regulated learning as core 21st-century competencies. Students become more independent in their learning. From the students' perspective, constructive assessment fosters metacognition and self-regulated learning as core 21st-century competencies. Constructive assessment helps students reflect on their own learning process, identify strengths and weaknesses, and develop more effective learning strategies. This process directly enhances metacognition (awareness of how one thinks) and self-regulated learning skills. Students accustomed to constructive assessment become more independent, proactive, and accountable for their learning progress. In Indonesia, this approach is central to the Merdeka Curriculum through student reflection and portfolios. Research from 2025 indicates that students with high levels of metacognition demonstrate better academic achievement and greater readiness to face future challenges. The long-term implication is the emergence of a generation of lifelong learners capable of adapting to rapid change. The main challenge is that students accustomed to traditional evaluation tend to passively accept grades without reflection. Therefore, teachers need to gradually

[41] Rahim et al., "Integrating HOTS."
[42] Maryati M., "Curriculum Evaluation and Development," Qalamuna 16, no. 2 (2024).

train students so that evaluation becomes an empowering experience, not merely an assessment..[43]

The use of assessment results within a teaching team enables collaboration among teachers to design more comprehensive and differentiated interventions. This collaboration improves the overall quality of instruction. The use of assessment results in teaching teams enables collaboration among teachers to design more comprehensive and differentiated interventions tailored to students' needs. Through joint discussions of assessment data, teachers can share perspectives, identify patterns of student difficulties, and develop stronger intervention strategies than when working individually. This collaboration significantly improves the overall quality of teaching, fostering a supportive professional culture within the school. In Indonesia, data-driven team teaching practices are increasingly being adopted in schools implementing the Merdeka Curriculum. The results include more targeted interventions, reduced teacher workload, and improved consistency in learning quality across all classes. Studies show that schools with active team teaching experience higher student learning outcomes and greater teacher satisfaction. This approach also strengthens a collective sense of responsibility for educational quality. The main challenge is the individualistic work culture that remains strong in many schools. Therefore, school principals need to facilitate regular time and space for collaboration so that the benefits of evaluation can be maximized.[44]

Assessment also plays a role in fostering a culture of assessment literacy among teachers, which ultimately enhances the overall quality of the learning process. This literacy has become a core competency for teachers. Assessment also plays a role in fostering a culture of assessment literacy among teachers, which ultimately enhances the overall quality of the learning process. Assessment literacy encompasses a teacher's ability to understand, design, implement, interpret, and utilize various types of assessments professionally. This literacy is now a core competency for teachers in the era of the Merdeka Curriculum. Teachers with high assessment literacy are able to transform evaluation data into meaningful pedagogical actions, making the learning process more effective and student-centered. A 2024 study shows that schools with high levels of teacher assessment literacy experience significant improvements in teaching quality and student learning outcomes. A culture of assessment literacy also reduces reliance on standardized tests and encourages the use of more holistic, authentic assessments. However, low assessment literacy remains a problem in many remote areas. Therefore, ongoing training programs and teacher communities of practice centered on assessment are essential for building a strong professional culture. Thus, strategic assessment not only improves the quality of teaching but also systematically builds teachers' professional capacity.[45]

From the students' perspective, constructive assessment fosters metacognition and self-regulated learning as core 21st-century competencies. Students become more

[43] Zhou Y. et al., "Development and Validation of a Higher-Order Thinking Skills Assessment" (2023/2025 update).

[44] Rahim et al., "Integrating HOTS."

[45] Maryati M., "Curriculum Evaluation and Development," Qalamuna 16, no. 2 (2024).

independent in their learning. From the students' perspective, constructive assessment fosters metacognition and self-regulated learning as core 21st-century competencies. Constructive assessment helps students reflect on their own learning process, identify strengths and weaknesses, and develop more effective learning strategies. This process directly enhances metacognition (awareness of how one thinks) and self-regulated learning skills. Students accustomed to constructive assessment become more independent, proactive, and accountable for their learning progress. In Indonesia, this approach is central to the Merdeka Curriculum through student reflection and portfolios. Research from 2025 indicates that students with high levels of metacognition demonstrate better academic achievement and greater readiness to face future challenges. The long-term implication is the emergence of a generation of lifelong learners capable of adapting to rapid change. The main challenge is that students accustomed to traditional evaluation tend to passively accept grades without reflection. Therefore, teachers need to gradually train students so that evaluation becomes an empowering experience, not merely an assessment.[46]

Challenges and the Reconstruction of Evaluation Systems in the Modern Era

One of the main challenges in learning assessment systems is the inherent bias in traditional assessment tools. These tools are often insensitive to the diversity of students' backgrounds, whether socioeconomic, cultural, linguistic, or geographic. This bias arises because standardized test designs tend to be based on the norms of dominant groups, thereby ignoring the local context and life experiences of students from marginalized groups. Consequently, evaluation results are not only inaccurate but also exacerbate structural inequities in education. Students from remote areas, low-income families, or ethnic minority groups often score lower not because they lack competence, but because the instruments do not represent their reality. In Indonesia, this phenomenon was clearly evident in national assessment practices prior to the Kurikulum Merdeka era, where questions heavily laden with Javanese urban cultural content tended to disadvantage students from Papua, Maluku, or Nusa Tenggara. Recent research indicates that such biases can diminish learning motivation and erode the self-confidence of students from vulnerable groups, thereby widening educational achievement gaps across regions and social classes. Furthermore, this bias in objectivity is not only cultural but also gender-based and linguistic, where female students or those who speak a regional language as their first language are often disadvantaged. Therefore, without a profound restructuring, traditional evaluation actually becomes a tool that perpetuates the reproduction of social injustice rather than serving as an instrument for improving the quality of equitable education.[47]

The use of technology, particularly Artificial Intelligence (AI), in educational assessment carries the risk of algorithmic bias, which can exacerbate educational inequities if not properly managed. A 2025 study highlights this issue, showing that AI systems trained on non-representative historical data tend to reproduce and even reinforce

[46] Stufflebeam and Zhang, CIPP Evaluation Model.
[47] Xie, "Algorithmic Bias."45-62

existing biases in society. This algorithmic bias manifests in various forms, including cultural, gender, linguistic, and socio-economic biases. For example, AI models trained using data from students in urban areas and upper-middle-class backgrounds often rate answers from students in rural areas or ethnic minority groups lower due to differences in language patterns, writing styles, and knowledge contexts that do not align with the norms of the training data. In Indonesia, this risk is increasingly evident in the implementation of AI-based adaptive assessments on the Merdeka Belajar platform, where students who speak a regional language as their first language may be disadvantaged because the system is more sensitive to standard Indonesian. Consequently, educational inequality widens: students from marginalized groups not only receive inaccurate scores but also experience reduced motivation, stigma, and limited access to further education. Furthermore, without mechanisms for algorithmic transparency and regular bias audits, AI interventions can actually become tools that perpetuate structural inequities rather than serving as innovative solutions. Therefore, managing algorithmic risks.[48]

One of the main challenges in learning assessment systems is the inherent bias in traditional assessment tools. These tools are often insensitive to the diversity of students' backgrounds whether socioeconomic, cultural, linguistic, or geographic. This bias arises because standardized test designs tend to be based on the norms of dominant groups, thereby ignoring the local context and life experiences of students from marginalized groups. Consequently, evaluation results are not only inaccurate but also exacerbate structural inequities in education. Students from remote areas, low-income families, or ethnic minority groups often score lower not because they lack competence, but because the instruments do not represent their reality. In Indonesia, this phenomenon was clearly evident in national assessment practices prior to the Kurikulum Merdeka era, where questions heavily laden with Javanese urban cultural content tended to disadvantage students from Papua, Maluku, or Nusa Tenggara. Recent research indicates that such biases can diminish learning motivation and erode the self-confidence of students from vulnerable groups, thereby widening educational achievement gaps across regions and social classes. Furthermore, this bias in objectivity is not only cultural but also gender-based and linguistic, where female students or those who speak a regional language as their first language are often disadvantaged. Therefore, without a profound restructuring, traditional evaluation actually becomes a tool that perpetuates social injustice rather than serving as an instrument for improving the quality of equitable education.[49]

The use of technology, particularly Artificial Intelligence (AI), in educational assessment carries the risk of algorithmic bias, which can exacerbate educational inequities if not properly managed. A 2025 study highlights this issue, showing that AI systems trained on non-representative historical data tend to reproduce and even reinforce existing biases in society. This algorithmic bias manifests in various forms, including cultural, gender, linguistic, and socio-economic biases. For example, AI models trained using data from students in urban areas and upper-middle-class backgrounds often rate

[48] Xie, "Algorithmic Bias.", 45–62.
[49] Xie, "Algorithmic Bias."45-62

answers from students in rural areas or ethnic minority groups lower due to differences in language patterns, writing styles, and knowledge contexts that do not align with the norms of the training data. In Indonesia, this risk is increasingly evident in the implementation of AI-based adaptive assessments on the Merdeka Belajar platform, where students who speak a regional language as their first language may be disadvantaged because the system is more sensitive to standard Indonesian. Consequently, educational inequality widens: students from marginalized groups not only receive inaccurate scores but also experience reduced motivation, stigma, and limited access to further education. Furthermore, without mechanisms for algorithmic transparency and regular bias audits, AI interventions may actually become tools that perpetuate structural inequities rather than serving as innovative solutions. Therefore, managing the risks of algorithmic bias requires a multidisciplinary approach that includes the development of inclusive datasets, the application of AI ethics principles, and human oversight (human-in-the-loop) to ensure that technology truly supports the equitable distribution of national educational quality.[50]

The digital divide and infrastructure gaps pose serious obstacles to the implementation of technology-based assessments in remote areas. These gaps require more extensive and sustained national policy interventions. In Indonesia, the digital divide remains stark: only about 30–40% of schools in Papua, Maluku, and East Nusa Tenggara have stable internet access and adequate devices to conduct AI-based assessments or use the Kurikulum Merdeka digital platform. Meanwhile, schools on Java and in major cities have access to 5G infrastructure and fiber-optic connections that support real-time adaptive assessments. As a result, teachers and students in remote areas are forced to revert to conventional paper-based evaluation methods, which are not only inefficient but also unable to provide quick and personalized formative feedback. This infrastructure gap is not merely technical; it also undermines the equitable distribution of educational quality nationwide. Students in 3T regions (frontier, outermost, and underdeveloped) lose opportunities to develop digital literacy and 21st-century competencies, leaving them increasingly behind their peers in urban areas. Furthermore, without integrated national policy interventions—such as device subsidy programs, the expansion of rural broadband networks, digital teacher training, and the development of offline-first evaluation platforms—technology-based evaluation will actually exacerbate educational disparities between regions. Therefore, the reconstruction of the evaluation system in the modern era must begin with a strong political commitment to bridging the digital divide so that learning assessments truly become an inclusive and equitable strategic instrument for all the nation's children.[51]

The reconstruction of the evaluation system requires the development of culturally responsive instruments based on multidimensional data. This approach ensures inclusivity in learning evaluations within Indonesia's highly diverse society. Culturally responsive instruments are designed with consideration for local values, regional languages, cultural

[50] Wulan et al., "CIPP SLR."
[51] Stanford HAI, "AI Challenges Core Assumptions in Education" (2025).

norms, and students' socioeconomic contexts, so that they are no longer dominated by the cultural perspective of the majority. Multidimensional data, on the other hand, integrates cognitive, affective, psychomotor, social-emotional, and even spiritual aspects, rather than relying solely on cognitive test scores. In Indonesia, this approach is highly relevant to the spirit of the Merdeka Curriculum, which emphasizes the Pancasila learner profile and recognition of local wisdom. Without such instruments, assessments risk overlooking the potential of students from Dayak, Papuan, or other indigenous communities who have distinct learning styles and ways of expressing knowledge. Recent research indicates that culturally responsive instruments can increase the validity of evaluation results by 25–35% among minority groups and reduce score disparities across regions. A multidimensional approach also enables teachers to identify students' strengths and weaknesses more holistically, making learning interventions more targeted and equitable. Therefore, the restructuring of the national assessment system must prioritize the development of such instruments through collaboration between education experts, cultural anthropologists, and local practitioners so that assessments truly become tools for empowerment, not discriminatory screening mechanisms. Only in this way can learning assessments realize the principles of inclusivity and educational equity for all the nation's children.[52]

The strategic solution moving forward is the integration of AI ethics and algorithmic transparency into the design of modern evaluation systems. AI ethics must be a top priority to ensure that technology is not only innovative but also fair and responsible. AI ethics principles include fairness, accountability, transparency, and human oversight, which must be implemented from the design stage (ethics by design). Algorithm transparency allows teachers, students, and policymakers to understand the basis of AI-generated assessment decisions, thereby reducing suspicion and increasing trust in evaluation results. In Indonesia, this integration is urgently needed in the adaptive assessment platform of the Merdeka Curriculum, where AI systems have the potential to impact millions of students from diverse backgrounds. Without transparency, hidden algorithmic biases could continue to perpetuate injustice, as warned by an international study in 2025. Therefore, the government needs to establish national AI ethics standards for education, including periodic algorithm audits, the formation of an ethics review board, and the involvement of local stakeholders in model development. This approach not only minimizes the risk of discrimination but also strengthens the role of evaluation as an inclusive empowerment tool. Furthermore, the integration of AI ethics will foster collaboration among technology developers, education experts, and communities to create an evaluation system that is adaptive to Indonesia's multicultural context. Only by making AI ethics a top priority can the reconstruction of the evaluation system truly support a sustainable, equitable educational transformation aligned with the values of Pancasila.[53]

[52] B. B. Wiyono et al., "The Structural Effects of Evaluation Types in Merdeka Belajar," Discover Sustainability (2025).

[53] Xie, "Algorithmic Bias."

Training in evaluation literacy for teachers and policymakers is key to overcoming resistance to a shift in the evaluation paradigm. Ongoing training is essential.[54]

Another challenge is the pressure of external accountability, which tends to encourage "teaching to the test," thereby sacrificing the depth of learning. A balance between accountability and formative assessment is necessary so that learning assessments continue to serve as a strategic tool for improving the quality of education. External accountability pressures, such as National Assessment results, school report cards, and national school rankings, often lead teachers and school principals to focus instruction on material most likely to appear on standardized tests. This "teaching to the test" practice results in superficial, rote-memorization-oriented learning that fails to develop Higher Order Thinking Skills (HOTS) and student character as envisioned in the Merdeka Curriculum. Consequently, students are trained only to pass exams, not to think critically, creatively, or solve real-world problems. In Indonesia, this phenomenon remains pervasive in many schools, particularly in regions facing public pressure and performance targets from education authorities. Teachers experience burnout from chasing score targets, while holistic competency development—such as empathy, collaboration, and entrepreneurship—is neglected. Without the right balance, summative evaluations aimed at accountability actually hinder the implementation of formative assessments, which should be the core of learning improvement. Therefore, the reconstruction of the evaluation system must integrate external accountability with robust formative practices, for example through a combination of minimum competency assessments and reflective portfolios. Only with this balance can evaluation promote healthy accountability while supporting deep and meaningful learning.[55]

In the modern era, evaluation must be designed as an organizational learning system that involves all education stakeholders. This approach fosters a sense of ownership and collective commitment to continuous improvement. Evaluation is no longer viewed as a standalone activity conducted solely by teachers or external agencies, but rather as a collaborative process involving students, teachers, parents, school principals, education authorities, and even the local community. Through an organizational learning system, evaluation data becomes a shared reflection tool that drives a cycle of continuous improvement at the school and national levels. In the Indonesian context, this approach aligns with the spirit of mutual cooperation and educational decentralization. Stakeholder involvement enhances a sense of ownership, making the implementation of evaluation recommendations more effective and sustainable. Studies show that schools implementing this organizational model experience improved teaching quality and higher student satisfaction. Therefore, the reconstruction of the evaluation system must include participatory mechanisms, such as school evaluation forums and collaborative digital platforms, so that all parties feel accountable for educational outcomes. This approach will transform evaluation from a monitoring tool into an adaptive organizational learning engine that is responsive to real-world needs on the ground.[56]

[54] Vattøy, "Navigating Formative Assessment."
[55] Ghamrawi, "Systematic Reviews."
[56] Zhang et al., "Mixed Teaching Quality Evaluation."

The ultimate solution lies in the formulation of national policies that support assessment as a strategic tool, including adequate budget allocation and the development of human resources in the education sector. These policies will fundamentally and sustainably transform Indonesia's education system. The government needs to draft a Ministerial Regulation or even a Bill that explicitly establishes learning assessment as a main pillar of education reform, rather than merely an administrative add-on. A dedicated budget allocation is required for the development of digital infrastructure, evaluation literacy training for hundreds of thousands of teachers, as well as research and development of innovative assessment instruments. Human resource capacity building includes certification programs, ongoing workshops, and the establishment of evaluation centers of excellence in every province. With strong and integrated policies, learning evaluation can become the primary catalyst for improving national education quality, supporting Sustainable Development Goal 4, and realizing the vision of a just and Pancasila-oriented national education system. Without policy commitment at the national level, all reconstruction efforts will be difficult to realize. Therefore, this final solution must be a top priority for stakeholders so that evaluation truly becomes a strategic instrument capable of transforming Indonesia's education system toward high-quality and inclusive global standards.[57]

4. Conclusion

Learning assessment, as discussed philosophically, theoretically, and practically in this article, is not merely an administrative procedure but a strategic tool capable of driving systemic and sustainable transformation in the quality of education. A synthesis of the arguments confirms that the integration of ontological-epistemological concepts with contemporary models such as CIPP, formative-summative, and HOTS, as well as addressing the challenges of AI technology, will make evaluation a catalyst for equitable and inclusive pedagogical improvement. Strategic recommendations for policymakers include revising the national evaluation policy framework to be formative and evidence-based, including strengthening competency-based assessments within the Merdeka Curriculum and allocating budgets for equitable evaluation literacy training and digital infrastructure. For education practitioners, it is recommended to develop sustainable evaluation literacy programs, utilize technology ethically, and foster data-driven teaching team collaboration. Only through this holistic systemic reconstruction can learning evaluation truly become a cornerstone of meaningful educational quality improvement for the future of the nation and human civilization..

References

Biardini, B., and D. S. Setiana. "A CIPP Model Evaluation of Philanthropy-Based Non-Formal Education." Journal of Innovation and Research in Primary Education 5, no. 1 (2026): 1936–1947. https://doi.org/10.56916/jirpe.v5i1.2866.

Black, Paul, and Dylan Wiliam. "Assessment and Classroom Learning." Assessment in Education: Principles, Policy & Practice 5, no. 1 (1998): 7–74.

[57] Biardini and Setiana, "CIPP Non-Formal Education."

Boztepe, Ã. "Ontological and Epistemological Foundations of Education in the Context of Educational Philosophy." EAS Journal of Education 20 (2025): 1–17.

Dewey, John. Experience and Education . New York: Kappa Delta Pi, 1938.

Ghamrawi, N. "A Step-by-Step Approach to Systematic Reviews in Educational Research." European Journal of Educational Research , April 15, 2025.

Ni'mah, Siti Jaroyatun, et al. "Ontology, Epistemology, and Axiology of Islamic Educational Philosophy." Matan: Journal of Islam and Muslim Society 6, no. 1 (2024): 1–15.

Rahim, Abd., et al. "Integrating HOTS in Indonesian High School Textbooks." Lingue 7, no. 1 (2025): 116–128.

Salvador, Wilfredo C., Jr., and Avelina C. Visaya. "Analysing the Epistemological and Ontological Assumptions Underpinning Teachers' Instructional Practices." International Journal of Multidisciplinary Engineering and Technology (2025).

Scriven, Michael. "The Methodology of Evaluation." In Curriculum Evaluation , edited by R. E. Stake, 39–83. Chicago: Rand McNally, 1967.

Stufflebeam, Daniel L., and Guili Zhang. The CIPP Evaluation Model . New York: Guilford Press, 2017.

Wulan, M. P., et al. "A Systematic Literature Review on the CIPP Evaluation Model in Education." Journal of Educational Sciences 10, no. 3 (2026): 307–320. https://doi.org/10.31258/jes.10.3.p.307-320.

Xie, Q. "The Algorithmic Bias of GenAI Technology in Educational Evaluation." Journal of Educational Technology & Society 28, no. 1 (2025): 45–62.

Zhang, S., L. Bai, and B. Yang. "Mixed Teaching Quality Evaluation of Organizational Behavior Course Based on CIPP Model." Frontiers in Education 10 (2025). https://doi.org/10.3389/feduc.2025.1538539.