
FORMATIVE AND SUMMATIVE EVALUATION: A COMPARATIVE ANALYSIS OF PURPOSE, PRACTICE, AND PEDAGOGICAL IMPACT

***Dr. Aadil Hussain Mir, Dr. Anmol Singh, Dr. Durga Prasad**

Assistant Professor, Department of Education, Guru Kashi University, Bathinda, Punjab.

Article Received: 15 July 2026

*Corresponding Author: Dr. Aadil Hussain Mir

Article Revised: 04 August 2026

Assistant Professor, Department of Education, Guru Kashi University, Bathinda, Punjab.

Published on: 24 August 2026

DOI: <https://doi-doi.org/101555/ijrpa.5653>

ABSTRACT

Formative and summative evaluation are the two fundamental paradigms by which educators assess learning; yet, they serve different pedagogical purposes and are commonly confused in practice. The formative-summative distinction, first introduced by Scriven (1967) in the context of curriculum evaluation and then extended to student assessment by Bloom, Hastings, and Madaus (1971), distinguishes between evaluations undertaken to enhance an ongoing process and evaluations conducted to appraise a completed one. This paper examines the conceptual history of these categories, beginning with Scriven's initial framework and progressing through Black and Wiliams' (1998) empirically grounded theory of classroom formative assessment to contemporary technology-mediated assessment systems. Drawing from Black and Wiliam's seminal review, which reported effect sizes of 0.4 to 0.7 for well-executed formative assessment, it explores the distinguishing features, instructional purposes, and evidentiary foundation of each technique. The washback effects of high-stakes summative testing on classroom practice, training limitations that restrict the quality of formative assessments, and the challenge of balancing assessment for learning with assessment of learning within accountability systems are just a few of the enduring conflicts between the two paradigms that are examined in this paper. It comes to the conclusion that formative and summative assessment are complementary elements of a cohesive assessment architecture rather than antagonistic alternatives, and it provides suggestions for a balanced integration in institutional and classroom practice.

KEYWORDS: Formative Evaluation, Summative Evaluation, Formative Assessment, Assessment For Learning, Scriven, Black And Wiliam, Educational Measurement, Feedback.

1. INTRODUCTION

Evaluation is at the center of every purposeful educational activity, serving as both a mirror reflecting what has been learnt and a compass guiding what should be taught next. The distinction between formative and summative evaluation has proven to be more enduring and consequential for classroom practice than any other in the broad field of educational assessment. The former is orientated toward the improvement of an ongoing process, providing feedback while a course, curriculum, or piece of student work is still capable of being revised; the latter is orientated toward the evaluation of a completed process, rendering a final verdict on the quality or effectiveness of what has already occurred. (ScienceDirect, 2024; EvalCommunity, 2025).

Although the phrases are now almost widely employed in educational systems around the world, their conceptual origins are in program assessment rather than classroom instruction. Michael Scriven first articulated the distinction in 1967 when evaluating educational curricula, and Benjamin Bloom and his colleagues only later applied it to individual student assessment in their influential 1971 handbook (ScienceDirect, 2024; NN/G, 2024). This genealogy is important because it clarifies a lot of the definitional ambiguity that still surrounds the two concepts. Over the course of more than 50 years, what started out as a distinction between two roles of program evaluation has been layered with additional distinctions of timing, purpose, audience, and stakes, creating a body of theory that is both fundamental and contentious (EERA, 2026).

This study conducts a thorough examination of formative and summative evaluation as they are understood and used in modern education. It traces the conceptual and historical development of the two constructs, examines each in turn in terms of defining features and practical implementation, provides a comparative analysis of the two paradigms, reviews the empirical evidence for their effectiveness, and critically evaluates the challenges associated with their translation from theory to classroom practice. The presentation concludes with a discussion of current technical advances and recommendations for developing balanced, cohesive assessment systems.

2. Conceptual Foundations and Historical Development

In a 1967 chapter that was published in an American Educational Research Association monograph series, Michael Scriven first used the word "formative evaluation" in relation to curriculum evaluation (ScienceDirect, 2024). Scriven's initial distinction focused on two roles that evaluation could play in relation to an educational program: if a program is evaluated

while it is still "malleable," that is, while it is still able to be revised in light of the evaluation's findings, the evaluation is formative in character; if, on the other hand, a mature and finalised program is evaluated in order to determine whether it should be continued, adopted more widely, or terminated, the evaluation is summative (Sabes, 2025). Scriven's framework was thus originally a distinction of function and timing within programme evaluation, not a theory of individual student assessment.

The concept acquired significant traction in education when it was adopted and expanded by Benjamin Bloom and his colleagues in their 1971 Handbook of Formative and Summative Evaluation of Student Learning. Bloom, Hastings, and Madaus used the term "summative evaluation" to describe evaluations conducted at the end of a term, course, or program for grading, certification, or curricular effectiveness research, extending Scriven's program-level distinction to individual learner assessment (ScienceDirect, 2024). Bloom's formulation also emphasised the importance of formative evaluation being more granular than summative evaluation, as its purpose was to identify precise gaps in learning that could be addressed immediately through instructional adjustment, whereas summative evaluation could operate at the level of broader, cumulative outcomes (EERA, 2026).

Subsequent theories refined and, in some cases, criticised the original distinction. Sadler (1989) shifted the emphasis away from timing and toward the quality and specificity of feedback, arguing that what distinguishes a formative assessment is whether it generates feedback that students can use to close the gap between their current and desired performance (EERA, 2026). Shepard (2005) also focused her concept on the function of improving teaching and learning rather than the assessment's calendar location within a course (EERA, 2026). In their seminal 1998 review and subsequent 2009 elaboration, Black and Wiliam advanced what may be the most empirically grounded and pedagogically actionable explanation. They defined an assessment as formative to the extent that teachers, students, or peers elicit, interpret, and use evidence about student achievement to make decisions about next instructional steps that are better, or better founded, than those that would have been made in the absence of that evidence (Wiliam, ETS; EERA, 2026).

Wiliam himself introduced a further refinement, redefining the distinction in terms of the use of information rather than its type or timing: an evaluation is formative if the information it generates is used to change what would have otherwise occurred, regardless of whether the evaluation's object is a curriculum or a student's learning (Wiliam, ETS). This "use-oriented" definition clarifies why the same assessment tool, like a classroom quiz, can serve formatively in one setting if the results are used to modify subsequent instruction, and

summatively in another if the results are merely recorded as a grade with no further instructional action.

Taras (2005) went one step further and claimed that Sadler's and Scriven's formative theories implicitly assume a prior summative judgement because it is impossible to diagnose a performance gap without an implicit standard by which current performance is evaluated. This suggests that summative judgement is, in a significant way, fundamental to formative practice rather than just its opposite (EERA, 2026).

3. Formative Evaluation: Purpose, Characteristics, and Classroom Practice

Formative evaluation, which is sometimes used synonymously with "assessment for learning," describes a variety of formal and informal methods that instructors and students employ to collect, analyse, and respond upon learning evidence while instruction is still in progress (NN/G, 2024). Its function—to mould, guide, and enhance the process it studies while it is still subject to change—is what distinguishes it rather than any specific tool or technique (ScienceDirect, 2024). It usually takes the form of continuous, low-stakes or no-stakes activities in the classroom, such as questioning during instruction, exit tickets, class discussion, draft feedback on written work, peer and self-evaluation, and diagnostic tests that are meant to find misconceptions rather than assign grades (Fairtest, 2007; NN/G, 2024).

Black and Wiliam's foundational review identified several interlocking components of high-quality formative practice, including the sharing of clear learning goals and success criteria with students, classroom questioning and dialogue that genuinely surfaces student understanding rather than merely eliciting brief right-or-wrong answers, feedback that includes explicit guidance on how to improve rather than only an evaluative judgement, and the deliberate cultivation of peer- and self-assessment skills so that students themselves become active agents in closing gaps in their own learning (Fairtest, 2007). This last element is significant: because formative evaluation is about enabling better decisions regarding subsequent learning, its effectiveness depends as much on students' capacity to interpret and act on feedback as on teachers' capacity to generate it.

Formative practice is intrinsically teacher- and context-dependent rather than instrument-dependent, according to a key implication of the use-oriented definition covered in Section 2. The same test item may or may not be formative depending on whether its results actually inform subsequent instructional decisions. This has important ramifications for teacher professional development because effective formative assessment requires teachers to develop the pedagogical judgement to interpret learning evidence in real time and modify

their instruction accordingly, rather than merely adopting a specific tool or technique (Fairtest, 2007). This idea is supported by recent research, which emphasises that formative assessment knowledge and skills are contextual and placed, shaped by the particular subject matter and learning environment in which a teacher works, rather than being a general, transferable competency (ResearchGate, 2026).

4. Summative Evaluation: Purpose, Characteristics, and Practice

In contrast, summative evaluation is the process of evaluating learning, a program, or a product at the end of an instructional time in order to determine the overall level of efficacy or achievement obtained (Wikipedia, 2026). Summative assessment is used in educational settings to certify the learning of individual participants and assess the efficacy of an instructional program, usually by comparing the results to a predefined standard or benchmark (Wikipedia, 2026). Final course grades, standardised state or national tests, end-of-term exams, and official certification evaluations are typical examples (evalcommunity, 2025).

Summative evaluation's primary purpose is retrospective rather than developmental: rather than producing feedback that alters the learning that has already taken place, it aims to summarise what has been learned or accomplished at a specific point in time, usually for grading, certification, program accountability, or research into curricular effectiveness (ScienceDirect, 2024). This does not imply that summative statistics are useless for making future decisions concerning student placement or course design. But unlike formative evaluation under Scriven's first "malleability" criterion, the assessment event itself is not intended to be modified in light of its own findings (Sabes, 2025).

Since its findings are frequently used for important choices like grade advancement, graduation, professional certification, or school and teacher accountability ratings, summative evaluation usually carries higher stakes than formative evaluation (Wikipedia, 2026). Because high-stakes summative testing frequently has a strong "washback effect" on classroom practice, influencing what and how teachers teach in ways that can either support or undermine the goals of formative, learning-oriented instruction, this elevated stakes profile has important pedagogical implications that are covered in more detail in Section 7 (Fairtest, 2007).

5. Comparative Analysis: Key Distinctions and Interrelationship

Five overlapping elements of comparison—timing, purpose, stakes, audience, and pedagogical application of results—help explain the conceptual distinctions between formative and summative evaluation as well as their interconnectedness. In terms of scheduling, summative evaluation takes place at the end of an instructional process, whereas formative evaluation takes place during it (EERA, 2026). Summative evaluation aims to evaluate a finished process, whereas formative evaluation aims to enhance an ongoing one (ScienceDirect, 2024). In terms of stakes, summative evaluation is usually high-stakes and graded, intended to guarantee a defensible, comparable judgement of performance, whereas formative evaluation is usually low-stakes or ungraded, intended to promote risk-taking and honest self-disclosure of misunderstanding. (NN/G, 2024).

In terms of audience, summative evaluation is frequently aimed at a broader range of stakeholders, such as parents, administrators, credentialing bodies, and policymakers, who need a comparable, standardised record of achievement, while formative evaluation is primarily directed at the student and teacher in the immediate instructional relationship (evalcommunity, 2025). Lastly, and perhaps most significantly from a pedagogical perspective, the two paradigms differ in how their results are applied. According to Wiliam's formulation, formative results are used to modify what would have otherwise occurred in subsequent instruction, whereas summative results are used to document, certify, or make decisions regarding what has already occurred (Wiliam, ETS).

It's crucial to avoid viewing this comparison as a strict, mutually exclusive binary. The two roles are conceptually interdependent rather than antagonistic, as Taras (2005) and later theorists have contended: formative evaluation of student work requires an implicit summative standard by which its sufficiency is assessed, and well-designed summative systems can collectively produce data that formatively influences future curricular decisions (EERA, 2026). By separating individual-level summative assessment used for certification and grading from group- or system-level summative assessment used for accountability, Black, Harrison, Lee, Marshall, and Wiliam (2004) further complicated the dichotomy. Each has unique implications for classroom practice and policy design (EERA, 2026). Thinking of the two as distinct relationships between an assessment event and subsequent action rather than as distinct test types is a helpful synthesis. The same instrument can serve either purpose depending on how its results are used, since what matters is not the instrument itself but rather the decision architecture surrounding it (Wiliam, ETS).

6. Evidence of Impact: Research Findings on Effectiveness

Paul Black and Dylan Wiliam's seminal 1998 review, which was published as both a comprehensive scholarly article in *Assessment in Education* and the frequently cited practitioner-oriented summary "Inside the Black Box: Raising Standards Through Classroom Assessment" in *Phi Delta Kappan*, is largely responsible for the empirical case for formative evaluation's instructional value (Fairtest, 2007). Black and Wiliam came to the conclusion that well-executed classroom formative assessment yields effect sizes on standardised measures of student achievement in the range of 0.4 to 0.7, a magnitude they described as larger than that of the majority of recognised educational interventions, based on a thorough synthesis of more than 250 studies (Fairtest, 2007). This result gave formative assessment an exceptionally strong evidentiary base in comparison to rival pedagogical changes, since an effect size in this range is typically regarded as large in educational research, where conventional treatments produce effect sizes closer to 0.2 to 0.3.

Black and Wiliam's review further identified the specific instructional mechanisms through which formative assessment appeared to generate these gains, including the quality of classroom questioning and dialogue, the specificity and actionability of feedback provided to students, and the deliberate development of students' capacity for self- and peer-assessment (Fairtest, 2007). Complementary meta-analytic work by Fuchs and Fuchs (1986), examining systematic formative evaluation in special education contexts, likewise found substantial positive effects on the achievement of students with learning difficulties when teachers used frequent, structured formative assessment to adjust instructional goals and methods (Kappan Online, 2024). Notably, the same review drew a less favourable conclusion about high-stakes summative testing, cautioning that standardised examinations, when used primarily for external accountability rather than instructional adjustment, can in some circumstances harm the quality of classroom teaching and learning, a tension examined further in Section 7 (Fairtest, 2007).

These classic conclusions have been expanded upon and contextualised by more current research. Since teacher knowledge and skill in formative practice are shaped by the particular subject matter, grade level, and classroom environment in which they are applied, Brookhart (2024) emphasises that the efficacy of formative assessment cannot be assumed to generalise uniformly across contexts (ResearchGate, 2026). Formative checkpoints are most effective when explicitly aligned with the summative outcomes a course is building toward, as misalignment undermines the diagnostic value of formative feedback for students preparing for eventual summative judgement, according to recent work on aligning formative and

summative assessment, such as the framework proposed for AI-supported educational assessment (arXiv, 2025).

7. Challenges and Criticisms in Implementation

Formative evaluation has a compelling theoretical and empirical basis, but it has proven far more challenging to translate into long-term, high-quality classroom practice than early enthusiasm anticipated. The "washback effect" of high-stakes summative assessments on classroom learning is one ongoing issue. Teachers are strongly motivated to focus daily instruction on the format and content of the final summative test when accountability systems attach substantial consequences to the results of standardised summative assessments, whether for students, teachers, or schools (Fairtest, 2007). This is frequently at the expense of the open-ended questions, risk-tolerant discussion, and diagnostic feedback that define effective formative practice. Black and Wiliam specifically cautioned that a policy that places too much emphasis on "tough tests" with high stakes runs the risk of displacing classroom-level formative procedures, which their own study found to be the more effective lever for increasing achievement (Fairtest, 2007).

The training and ability limitations that restrict the effectiveness of formative assessment in practice present a second difficulty. Compared to summative assessment, which can theoretically be administered and scored in accordance with fixed, generalisable procedures, effective formative evaluation places significantly higher demands on teacher expertise because it relies on teachers' pedagogical judgement in interpreting evidence of learning and modifying instruction in real time rather than on the mechanical application of a standardised instrument (Fairtest, 2007). In situations where professional development is insufficient, formative-sounding methods like exit tickets or peer review may be adopted superficially without the significant change in questioning, feedback, and instructional responsiveness that the literature identifies as the true source of the impact of formative assessment.

Definitional ambiguity presents a third, more intellectual difficulty. At least five different and only partially compatible definitional traditions have emerged since Scriven's original 1967 distinction, as the review of fifty years of formative and summative assessment theory shows. These traditions are distinguished by whether they emphasise timing, purpose, audience, the type of summative assessment being contrasted, or the fundamental relationship between formative and summative judgement (EERA, 2026). Because guidance that invokes "formative assessment" without identifying the tradition it draws from runs the danger of

causing inconsistent implementation across practitioners, this heterogeneity poses serious challenges for policy coherence.

Aligning formative and summative assessment within a single, cohesive educational sequence is a fourth obstacle that is becoming more and more prominent in modern practice. Students' trust in formative activities that don't seem to 'count' may be undermined if formative checkpoints are not in line with the final summative standard by which they will be evaluated. This is because the diagnostic feedback they produce may not adequately prepare students for summative performance (arXiv, 2025).

8. Contemporary Developments: Technology and Assessment Reform

Formative and summative assessment now face both new opportunities and challenges due to the emergence of digital learning platforms and, more recently, generative artificial intelligence. AI-based writing feedback systems and adaptive quiz platforms are examples of automated formative assessment tools that have the ability to scale personalised, real-time formative feedback far beyond what a single teacher can deliver to an entire class of pupils without assistance (arXiv, 2025). Large language models have been used in research on AI-supported writing feedback to produce thorough, rubric-aligned commentary on student drafts. This could increase the frequency and specificity of formative feedback available in between summative submissions (arXiv, 2025). However, as automated systems might not adequately capture the contextual, dialogic characteristics that Black and Wiliam highlighted as essential to successful formative practice, these innovations raise concerns regarding the validity and pedagogical authenticity of automated judgements. According to a proposed framework for constructive alignment between AI-supported formative and summative assessment, automated formative checkpoints must be purposefully created to align with summative learning outcomes in order to fulfilll their developmental purpose rather than turning into a detached activity that students ignore (arXiv, 2025).

Beyond technology, a few of educational institutions have implemented more comprehensive policy changes with the goal of rebalancing formative and summative evaluation. For instance, ongoing, teacher-mediated formative assessment integrated throughout the school year has replaced single, high-stakes end-of-year exams in the early grades as part of continuous assessment reforms implemented in some national systems in recent years. This approach reflects the internationally recognised principle of "assessment for," rather than merely "assessment of," learning (ResearchGate, 2026). More than 20 years after its first

publication, these improvements demonstrate the Black and Wiliam evidential tradition's ongoing effect on assessment policy design.

9. Recommendations for Balanced Assessment Systems

Based on the findings and problems discussed above, numerous recommendations emerge for developing assessment systems that capitalise on the combined qualities of formative and summative evaluation. First, formative and summative assessment should be treated as intentionally aligned components of a single instructional sequence, with formative checkpoints explicitly designed to build the specific knowledge and skills that will later be judged summatively, maximising both the diagnostic value of formative feedback and the validity of eventual summative judgement (arXiv, 2025). Second, since formative assessment quality is largely dependent on teachers' ability to make pedagogical decisions in real time rather than just using formative-sounding methods or tools (Fairtest, 2007; ResearchGate, 2026), ongoing investment in teacher professional development is crucial. Development should concentrate on the elements that the Black and Wiliam tradition identifies as most strongly associated with achievement gains, namely high-quality classroom questioning, actionable feedback, and the development of student self- and peer-assessment skills.

Third, in order to prevent crowding out the classroom conditions where formative assessment yields its greatest benefits, policymakers creating accountability systems should be aware of the washback risks that high-stakes summative testing poses to classroom-level formative practice. They should also take into account accountability designs that also reward evidence of effective formative practice rather than relying solely on aggregate summative test results (Fairtest, 2007). Fourth, rather than being adopted only for administrative convenience or scalability, the design of digital and AI-based assessment tools should be explicitly evaluated against the pedagogical qualities identified in the formative assessment literature, particularly the capacity to generate specific, actionable, dialogic feedback (arXiv, 2025). Lastly, institutions and policy documents that refer to "formative" or "summative" assessment should explicitly state which conceptual tradition—whether timing-based, purpose-based, or use-based—they draw from, given the ongoing definitional ambiguity documented in the academic literature. This will lower the possibility of inconsistent implementation guidance across teachers, schools, and assessment systems (EERA, 2026).

10. CONCLUSION

One of the most important and empirically supported frameworks in educational theory is the distinction between formative and summative evaluation, which Michael Scriven developed more than 50 years ago in the context of curricular assessment. According to Black and Wiliam's seminal synthesis of over 250 studies (Fairtest, 2007), formative evaluation, when properly implemented through high-quality classroom questioning, specific and actionable feedback, and intentional cultivation of student self-assessment, is associated with achievement gains significantly larger than those produced by most other educational interventions. While high-stakes summative testing bears known hazards for classroom instruction when it dominates the assessment landscape, summative evaluation continues to play an essential role in certifying accomplishment and supporting system-level accountability.

The key result of this analysis is that formative and summative evaluation should not be viewed as competing paradigms for classroom time, but rather as functionally separate and mutually reinforcing components of a cohesive assessment framework. Their effective integration is determined less by the specific instruments or technologies used than by the decision architecture surrounding them: whether the information generated is truly used to improve ongoing learning, in the case of formative evaluation, or to render a defensible, comparable judgement on completed learning, in the case of summative evaluation. As education systems around the world struggle to strike the right balance between continuous, developmental assessment and periodic, high-stakes examination, and as digital and AI-mediated tools reshape the practical possibilities for both, the conceptual clarity established by Scriven, Bloom, Black, and Wiliam remains an essential foundation for sound assessment policy and practice.

REFERENCES

1. Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7-74.
2. Black, P., & Wiliam, D. (1998, October). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*. Summarised in Fairtest. <https://fairtest.org/value-formative-assessment-pdf/>
3. Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 5-31.

4. Bloom, B. S., Hastings, J. T., & Madaus, G. F. (1971). Handbook of formative and summative evaluation of student learning. McGraw-Hill.
5. EERA (European Educational Research Association). (2026). 50 years' legacy of formative and summative evaluation and assessment: A critical theoretical review of education policy and research. <https://eera-ecer.de/ecer-programmes/conference/22/contribution/41939>
6. EvalCommunity. (2025, March 16). Formative versus summative evaluation: Difference, examples. <https://www.evalcommunity.com/career-center/formative-evaluation-versus-summative-evaluation/>
7. Fairtest. (2007, August 27). The value of formative assessment. <https://fairtest.org/value-formative-assessment-pdf/>
8. Kappan Online. (2024, September 6). Inside the black box: Raising standards through classroom assessment. <https://kappanonline.org/inside-the-black-box-raising-standards-through-classroom-assessment/>
9. Mir, A. H. (2026). Continuous and comprehensive evaluation in school education: A critical qualitative review of its conceptual framework, implementation practices, and challenges. *International Journal of Economics, Business, Accounting, Agriculture, Management and Sharia Administration*, 6(2).
10. NN/G (Nielsen Norman Group). (2024, January 29). Formative vs. summative evaluations. <https://www.nngroup.com/articles/formative-vs-summative-evaluations/>
11. ResearchGate. (2026). Inside the black box: Raising standards through classroom assessment [Review with contemporary commentary]. https://www.researchgate.net/publication/44836144_Inside_the_Black_Box_Raising_Standards_through_Classroom_Assessment
12. Sabes. (2025). Formative assessment: Why, what, and whether. <https://www.sabes.org/sites/default/files/resources/ascd.org-Formative%20Assessment%20Why%20What%20and%20Whether.pdf>
13. ScienceDirect. (2024). Formative evaluation - an overview. <https://www.sciencedirect.com/topics/social-sciences/formative-evaluation>
14. Wikipedia. (2026). Summative assessment. https://en.wikipedia.org/wiki/Summative_assessment
15. Wiliam, D. (n.d.). Formative assessment: Getting the focus right. Educational Testing Service. https://discovery.ucl.ac.uk/10001128/1/EAJ_FA_special_issue_commentary_v2.pdf
16. arXiv. (2025). The impact of AI on educational assessment: A framework for constructive alignment. <https://arxiv.org/pdf/2506.23815>