Received: 19 Nov. 2026 | Revised: 20 Nov. 2026 | Accepted: 24 Nov. 2026 | Published online: 25 Nov. 2026
RESEARCH ARTICLE

Contemporary Education and Teaching Research
2026, Vol. 7(11)435-441
DOI: 10.61360/BoniCETR262021231111

AI-Assisted Formative Classroom Assessment: A Multidimensional Framework with Human – AI Collaboration

Shurui Hu[1*], Jiajing Tang[1], Jingjing Jiang[1]
[1]Shanghai Jiao Tong University, P.R.China

Abstract: Recent advances in digital recording, speech recognition, and multimodal analysis have made classroom data easier to collect and analyze. This creates new opportunities for artificial intelligence (AI) in classroom assessment. AI can support preliminary analysis by organizing classroom events, identifying teacher-student interactions, and extracting evidence of student learning. However, identifying classroom activities is not the same as understanding teaching quality or student learning. A clear framework is therefore needed to guide what should be analyzed, what AI can do, and what should be reviewed by people. To meet this need, this paper proposes an AI-assisted formative assessment framework for a single lesson based on multisource classroom data. The framework has three parts: five evaluation dimensions with 14 sub-indicators, a table that links each indicator to classroom evidence, AI tasks, and human review, and a workflow for human–AI collaboration. The five dimensions are instructional structure and organization, instructional content and cognitive demand, student engagement, learning objectives and evidence of achievement, and overall analysis and formative feedback. The framework positions AI as a tool for preliminary detection, evidence organization, and initial feedback drafting, while final educational judgments are left to teachers and evaluators. As a conceptual framework, it provides a basis for future research and practical exploration.
Keywords: artificial intelligence, formative assessment, human–AI collaboration

1. Introduction

Classroom assessment helps teachers understand how a lesson is taught, how students participate in learning, and what may need improvement. Traditional observation requires considerable time and is difficult to sustain across many lessons. Recent advances in speech recognition, natural language processing, and multimodal analysis have made it possible to organize classroom recordings, transcripts, and student responses more systematically. Artificial Intelligence (AI) can help locate classroom episodes, organize teacher–student dialogue, identify question types, describe participation patterns, and compare classroom activities with learning objectives (Long et al., 2024; Wang et al., 2025; Chen et al., 2026).

However, the main issue is not merely whether AI can recognize what is happening in the classroom, but whether this information can support reasonable evaluation decisions. AI may identify how often students speak, but speaking turns do not show whether students understand the content. Existing observation tools are designed mainly for human observers and rating, whereas AI-assisted assessment requires explicit decisions about which aspects of a lesson to assess, what data to use, what AI can do with those data, and which conclusions require human review. A defensible approach therefore needs a framework that connects assessment dimensions, observable evidence, and AI-supported tasks, and that specifies where educators must interpret evidence in light of teaching context. In this sense, what is needed is not an AI evaluator, but a human–AI collaborative assessment model.

This paper proposes such a framework for AI-assisted formative assessment of a single lesson. The framework consists of three connected parts: five assessment dimensions with 14 sub-indicators, a table that links each indicator to classroom evidence, AI tasks, and human review, and a workflow for human–AI collaboration. Its aim is to support teacher reflection and instructional improvement.

2 Theoretical Background and Framework Development

2.1. Scope and data sources

This paper focuses on formative assessment at the level of a single lesson. The purpose is to use the materials available from one lesson to help teachers and evaluators understand classroom practice and

Corresponding Author: Shurui Hu
Shanghai Jiao Tong University, P.R.China
©The Author(s) 2026. Published by BONI FUTURE DIGITAL PUBLISHING CO., LIMITED This is an open access article under the CC BY License(https://creativecommons.org/licenses/by/4.0/).

identify possible areas for improvement.

The framework considers three types of information. The first type is basic classroom data, including classroom audio and video, time-stamped speech transcripts, and the text in teaching slides. These sources provide the main evidence for the five assessment dimensions. The second type is supplementary teaching and learning data, such as lesson plans, learning objectives, classroom tasks, students' written or online responses, and short quizzes. These materials help evaluators examine whether classroom activities are related to the stated objectives and whether student performance provides evidence of learning. The third type is contextual information, including student characteristics, class size, and students' prior knowledge. This information is mainly used by teachers or evaluators to interpret the analysis and the classroom evidence.

2.2. Theoretical foundations

The framework draws on five groups of ideas. They guide the selection of assessment dimensions; they do not require every lesson to follow one fixed teaching pattern. Table 1 summarizes the main ideas and their contribution to the framework.

Table 1. *Theoretical foundations of the framework*

Theory or concept	Main idea	Contribution to the framework
Gagné's instructional events (Gagné et al., 1974)	Instruction may include gaining attention, informing learners of objectives, presenting content, providing guidance, eliciting practice, giving feedback, and reviewing learning.	Supports Dimension 1: lesson opening, teaching process, practice and interaction, and lesson closure.
Meaningful learning and cognitive taxonomy (Ausubel, 1963)	New knowledge is easier to understand when it is connected with prior knowledge. Questions and tasks can also be examined in terms of their cognitive demands.	Supports Dimension 2: content organization, connections between ideas, and the cognitive demands of questions and tasks.
The ICAP framework (Chi & Wylie, 2014)	Observable learning activities can be described as passive, active, constructive, or interactive.	Supports Dimension 3: the range, forms, and interaction patterns of student engagement.
Alignment of objectives and assessment (Tyler, 1949)	Learning objectives, learning experiences, and assessment should be considered in relation to one another.	Supports Dimension 4: clarity of learning objectives and evidence of student learning.
Formative evaluation and reflective thinking (Bloom, 1971; Rodgers, 2002)	Evaluation can identify issues and support reflection and later improvement.	Supports Dimension 5: summarizing strengths and suggesting possible areas for improvement.

First, Gagné's instructional events provide a reference for examining instructional structure. Gagné et al. (1974) described several events that may support learning, including gaining attention, informing learners of objectives, presenting content, providing guidance, eliciting practice, giving feedback, and reviewing learning. These ideas support the assessment of lesson opening, teaching activities, practice, interaction, and closure. However, the events should not be treated as a fixed sequence. Their value depends on the subject and classroom context.

Second, meaningful learning theory and Bloom's revised taxonomy support the analysis of instructional content and cognitive demand. Ausubel (1963) emphasized the importance of connecting new knowledge with learners' existing knowledge. This idea supports the examination of content organization, conceptual relationships, and the use of examples. Anderson and Krathwohl (2001) revised Bloom's taxonomy and provided a framework for examining different levels of cognitive demand. This taxonomy can help classify classroom questions and tasks.

Third, the ICAP framework provides a reference for examining student engagement. Chi and Wylie (2014) described four forms of engagement: passive, active, constructive, and interactive. This framework helps distinguish different types of observable learning activities. It also reminds evaluators that classroom participation involves more than speaking frequency.

Fourth, the relationship between learning objectives and assessment supports the analysis of learning evidence. Tyler (1949) emphasized the relationship between educational objectives, learning experiences, and evaluation. This principle supports the examination of whether classroom activities are connected with learning objectives and whether student performance provides evidence of progress toward those objectives.

Fifth, formative assessment and reflective thinking support the overall analysis and feedback dimension. Bloom (1971) emphasized the role of evidence in formative assessment. Shute (2008) also highlighted the importance of useful and well-designed formative feedback. Schön (1983) described how professionals reflect on their practice and use such reflection to guide improvement.

Together, these ideas support the use of classroom evidence to identify strengths and possible directions for improvement.

The five groups of ideas work together rather than separately. Gagné's instructional events provide a reference for examining how a lesson is organized. Meaningful learning theory and the revised Bloom's taxonomy support the analysis of content organization and cognitive demand. The ICAP framework helps describe student engagement beyond speaking frequency. Tyler's principle of alignment supports the examination of whether classroom activities are connected with learning objectives and whether student performance provides evidence of progress. Formative assessment and reflective practice support the final dimension, where evidence is used to identify strengths and possible directions for improvement. In each case, the theory indicates what to look for, but the judgment still depends on the subject and classroom context.

2.3. Development approach and basic principles

The framework follows a five-step process, as shown in Figure 1. First, relevant data are collected from the lesson and related teaching and learning materials. Second, AI organizes the data and carries out an initial analysis. Third, the evidence is considered through five assessment dimensions. Fourth, teacher and evaluator review the evidence and interpret it in light of the subject, curriculum goals, and students' learning needs. Finally, evidence-based formative feedback is prepared to support reflection and instructional improvement.

The second step is where AI does most of its work. Speech recognition converts classroom audio into text, and time information allows specific episodes to be located; natural language processing then organizes classroom dialogue, identifies possible questions, explanations, instructions, and feedback, and classifies questions and tasks by their likely cognitive demand. All of this is evidence work: it makes classroom activity easier to locate, compare, and summarize.

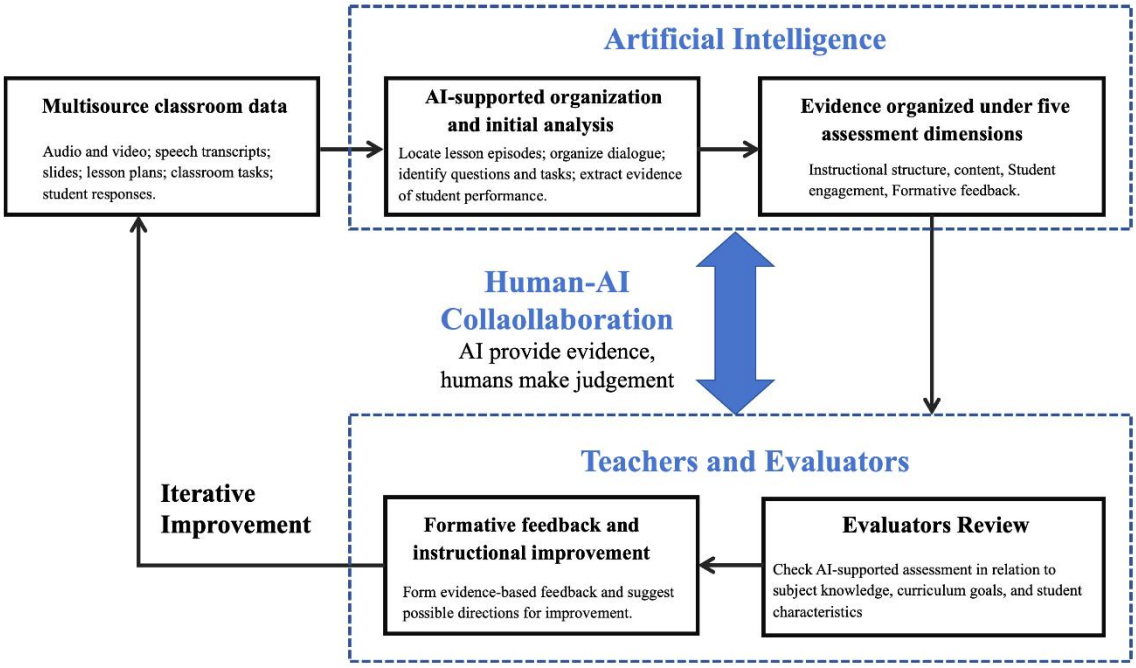


Figure 1. *The process of AI-assisted formative assessment*

3. The AI-Assisted Classroom Assessment Framework

3.1. Overview of the dimensions and indicators

Drawing on these theoretical perspectives, this paper develops an assessment framework with five dimensions and 14 sub-indicators. The first four dimensions examine the teaching process, instructional content, student engagement, and evidence of learning. The fifth dimension brings together findings from the first four dimensions and turns them into feedback for improvement.

Table 2. *Overview of the five dimensions*

Dimension	Sub-indicators	Main assessment question
1. Instructional structure and organization	1.1 Lesson opening and introduction 1.2 Main instructional events 1.3 Practice and interaction 1.4 Review and closure	How is the lesson organized? Are the main stages connected, and is the use of time appropriate?
2. Instructional content and cognitive demand	2.1 Logical organization of content 2.2 Content suitability and focus on objectives (context-dependent) 2.3 Explanation of key and difficult points 2.4 Cognitive demand of questions and tasks	Is the content clear, accurate, appropriate, well focused, and cognitively challenging?
3. Student engagement	3.1 Breadth of participation 3.2 Depth of engagement and feedback chains	How widely do students participate? Do interactions show meaningful thinking and response?
4. Learning objectives and evidence of achievement	4.1 Clarity of learning objectives 4.2 Evidence of immediate learning	Are the objectives clear? Is there evidence related to whether students are making progress toward them?
5. Overall analysis and formative feedback	5.1 Potential strengths of the lesson 5.2 Possible directions for improvement	What are the strengths of the lesson? What changes might be considered?

3.2. Main content of each dimension

3.2.1. Instructional structure and organization

This dimension examines how a lesson is organized and developed, including the opening, main teaching activities, practice and interaction, and closure. Gagné's instructional events provide a useful reference for identifying these activities (Gagné et al., 1974), but they do not mean that all lessons must follow the same order or use the same stages. Different subjects and teaching approaches may require different arrangements.

AI can help locate lesson episodes, organize the sequence of activities, and extract relevant content. Whether the activities are well arranged and connected still needs to be judged in relation to the lesson objectives, students' prior knowledge and the actual classroom context.

3.2.2. Instructional content and cognitive demand

This dimension examines whether the content is clearly organized, whether key ideas are explained, and what kinds of thinking are required by classroom questions and tasks. Logical organization concerns the relationships among concepts and whether the teacher uses prior knowledge, examples, or different forms of representation to support understanding.

Content suitability and focus on objectives depend on curriculum requirements and students' backgrounds, so this is a context-dependent indicator. Explanation and support for key or difficult points concern whether the teacher provides explanations, examples, and appropriate guidance.

3.2.3. Student engagement

This dimension examines the range and forms of student participation. Breadth of participation may be observed through speaking turns, student responses, classroom tasks, and available written or online work. Depth of engagement and feedback chains concern whether students ask questions, explain ideas, give reasons, respond to peers, and receive responses from the teacher.

The ICAP framework provides a useful reference for understanding different forms of engagement (Chi & Wylie, 2014). AI can organize speaking turns, identify some observable discourse actions, and summarize interaction patterns. However, classroom talk cannot fully represent students' internal thinking. Other materials can be used when possible to supplement evidence from group discussion, written activities, or participation that is not captured in recordings.

3.2.4. Learning objectives and evidence of achievement

This dimension examines whether learning objectives are clear and whether the lesson contains student performance related to those objectives. It follows the principle that objectives, learning activities, and assessment evidence should be considered in relation to one another (Tyler, 1949).

AI can extract objective statements from lesson plans, slides, and teacher explanations. It can then make an initial comparison between the objectives and materials such as student responses and exercises. However, evidence from a single lesson usually reflects students' understanding or progress at that time. It cannot show, by itself, whether students have retained knowledge over time or can transfer it to new situations.

3.2.5. Overall analysis and formative feedback

This dimension brings together evidence from the first four dimensions to identify possible strengths and areas for improvement. A summary of strengths should describe specific teaching practices and the evidence that supports them. A suggestion for improvement should identify where an issue appears

and, where possible, offer an adjustment for the teacher to consider.

This dimension is informed by formative assessment theory and research on feedback (Black & Wiliam, 2009; Shute, 2008). AI can help locate and organize evidence, summarize findings, and prepare an initial draft of feedback. These outputs are best understood as evidence-based starting points for professional reflection rather than final judgments. Teachers and evaluators interpret the findings in context and decide which feedback is useful and appropriate.

4. Classroom Data Analysis and Human–AI Collaboration

4.1. Links between indicators, classroom evidence, and AI tasks

To make the five dimensions usable in practice, each indicator is linked to observable classroom evidence. Table 3 presents the operational core of the framework. The AI tasks listed in the table describe possible forms of support.

Table 3 *Indicators, observable classroom evidence, and possible AI support*

Indicator	Observable classroom evidence	Possible AI support
1. Instructional structure and organization		
1.1 Lesson opening and introduction	Opening questions; contextual materials; statements of learning objectives	Locate the opening episode and extract the teacher's statements and related materials.
1.2 Main instructional events	Explanations; worked examples; demonstrations; transitions between concepts	Identify and organize the sequence of teaching activities; extract representative episodes.
1.3 Practice and interaction	Questions; exercises; student responses; teacher feedback	Locate practice episodes and organize the sequence of teacher prompt, student response, and teacher feedback.
1.4 Review and closure	Summaries; closing questions; short quizzes; previews of the next lesson	Identify the review and closing stages and extract their main content.
2. Instructional content and cognitive demand		
2.1 Logical organization of content	Topic sequences; conceptual links; explanatory relationships; concrete examples	Extract key concepts and their relationships; prepare a candidate concept map or content summary.
2.2 Content suitability and focus on objectives (context-dependent)	Learning objectives; prerequisite knowledge; task difficulty	Compare objectives with lesson content and flag possible issues or differences for review.
2.3 Explanation of key and difficult points	Emphasis on key ideas; step-by-step explanations; alternative representations	Locate explanation episodes and extract examples, prompts, and forms of support.
2.4 Cognitive demand of questions and tasks	Teacher questions; classroom tasks; examples of student responses	Make an initial classification of questions or tasks by cognitive demand and organize response examples.
3. Student engagement		
3.1 Breadth of participation	Speaking turns; group contributions; written or digital responses	When the data allow, distinguish speakers and summarize participation patterns and related episodes.
3.2 Depth of engagement and feedback chains	Explanations; reasoning; peer dialogue; teacher–student feedback	Identify observable discourse actions and organize interaction chains and feedback types.
4. Learning objectives and evidence of achievement		
4.1 Clarity of learning objectives	Objective statements in lesson plans, slides, and assessment tasks	Extract objectives and compare their wording across materials; flag possible differences.
4.2 Evidence of immediate learning	Student responses; exercises; quizzes; written work	Make an initial comparison between student performance and objectives; organize correct, partly correct, and incomplete responses.
5. Overall analysis and formative feedback		
5.1 Potential strengths of the lesson	Episodes, explanations, question design, and feedback examples from Dimensions 1–4	Retrieve relevant episodes, collect evidence for possible strengths, and draft a description.
5.2 Possible directions for improvement	Participation patterns; evidence gaps; differences between objectives and activities	Connect relevant evidence and draft possible suggestions, including the conditions under which they may apply.

4.2. Main AI tasks and human review

Table 3 shows that AI-supported work is mainly evidence work: locating classroom episodes, transcribing and separating speakers, giving preliminary classifications of questions and tasks, aligning teaching content with objectives, collecting evidence, and drafting feedback. These outputs need to be checked against the source materials. Each finding in the assessment report should therefore make its basis visible: the related dimension and

indicator, the episode or time location, an extract from the original recording, transcript, or teaching material, the initial AI analysis together with any uncertainty or missing information, and the interpretation offered by teachers and evaluators.

AI can identify evidence, but it cannot explain what that evidence means for teaching and learning. Whether a lesson is well organized, whether the content is accurate and suitable for the students, and whether students are genuinely engaged are judgments that depend on subject knowledge and knowledge of the students. AI can describe what happened in a lesson, but it cannot decide whether what happened was good. That decision belongs to teachers and evaluators, who know the subject and the class.

5. Discussion

The framework links each assessment dimension to observable classroom evidence, including recordings, transcripts, slides, lesson plans, and student work, so that evaluation does not rest on a single measure such as speaking time or the number of questions. It also defines a division of labor between AI and human evaluators, treating AI output as an initial analysis rather than a final assessment. Both design choices serve the same end: findings are tied to instructional improvement rather than to an overall score, and each finding is linked to specific evidence and, where possible, to a direction for improvement.

This division of labor rests on a boundary that deserves emphasis: patterns identified from classroom data do not carry a fixed educational meaning. AI is therefore positioned as a tool for finding and organizing evidence, while educational judgment remains with teachers and evaluators. In practice, a report can organize findings under the five dimensions and attach the related episode or material to each finding, so that teachers can review both the feedback and the evidence behind it. This supports reflection and professional discussion and allows schools to move from score-based judgment towards evidence-informed feedback. Because such a report depends on AI outputs, however, those outputs need to be treated with caution: model outputs may contain classification errors or inconsistent judgments (Long et al., 2024; Wang et al., 2025), and the quality and reliability of AI-generated feedback still require careful examination (Seo et al., 2025).

6. Limitations and Future Research

The framework is currently at the conceptual stage and has not been examined through expert review, classroom trials, or student learning outcomes, so its validity, reliability, and practical usefulness remain to be established. Its value at this stage lies in offering a clear structure for linking assessment dimensions, classroom evidence, and AI tasks, not in demonstrating measurement properties. Expert review is the most immediate next step: specialists in teaching, instructional design, and educational assessment can judge whether the 14 indicators are clear, relevant, and sufficiently distinct.

A second limitation concerns teaching context. An indicator may look different in a discussion, a laboratory lesson, or a practical activity: engagement in discussion may be visible in spoken interaction, while engagement in individual work may be better reflected in written products. The framework therefore needs subject-specific descriptions before it can be applied across different teaching formats.

A third limitation concerns privacy. Recordings and student work contain personal information, so informed consent, access control, and secure storage are required (Gaisie et al., 2025). These requirements affect how the framework can be applied, but they do not change what the framework is designed to do.

Future research can proceed in three steps: expert review of the indicators, pilot applications across subjects and teaching formats, and comparisons between AI-supported analysis and trained human evaluators. A further question is how teachers interpret and use AI feedback in practice. Addressing these questions would strengthen the framework and clarify the conditions under which it can be used responsibly.

7. Conclusion

This paper proposed an AI-assisted formative assessment framework for a single lesson. The framework organizes classroom assessment around five dimensions, each linked to observable classroom evidence, and defines a model of human–AI collaboration in which AI locates and organizes evidence while teachers and evaluators interpret that evidence and make the final judgment. Its contribution lies in connecting assessment dimensions, classroom evidence, and AI tasks within a single structure, and in making the boundary between evidence and judgment explicit. By making this boundary explicit, the framework offers a testable starting point for further research on AI-assisted classroom assessment.

Conflict of Interest

The author(s) declare that there are no conflicts of interest to this work.

References

Anderson, L. W., & Krathwohl, D. R. (2026). *A taxonomy for learning, teaching, and assessing*. Addison Wesley Longman.

Ausubel, D. P. (1963). The psychology of meaningful

verbal learning.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability (formerly: Journal of personnel evaluation in education)*, *21*(1), 5-31.

Bloom, B. S. (1971). Handbook on formative and summative evaluation of student learning.

Chen, G., Han, G., Niu, J., & He, J. (2026). Exploring teacher-student interaction through multimodal large language models: an empirical investigation. *Scientific Reports*, *16*(1), 7602.

Chi, M. T., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational psychologist*, *49*(4), 219-243.

Gagne, R. M., & Briggs, L. J. (1974). *Principles of instructional design*. Holt, Rinehart & Winston.

Gaisie, E., Owusu-Boateng, O., Yidana, I., & Ghansah, B. (2025). Ethical considerations and data privacy in artificial intelligence. *Unlocking the potential: Artificial intelligence in education*, *58*.

Long, Y., Luo, H., & Zhang, Y. (2024). Evaluating large language models in analysing classroom dialogue. *npj Science of Learning*, *9*(1), 60.

Rodgers, C. (2002). Defining reflection: Another look at John Dewey and reflective thinking. *Teachers college record*, *104*(4), 842-866.

Schön, D. A. (1983). *The reflective practitioner: How professionals think in action*. Basic Books.

Seo, H., Hwang, T., Jung, J., Kang, H., Namgoong, H., Lee, Y., & Jung, S. (2025). Large language models as evaluators in education: Verification of feedback consistency and accuracy. *Applied Sciences*, *15*(2), 671.

Shute, V. J. (2008). Focus on formative feedback. *Review of educational research*, *78*(1), 153-189.

Tyler, R. W. (1949). *Basic principles of curriculum and instruction*. University of Chicago Press.

Wang, J., Hankour, K., Zhang, Y., LoCasale-Crouch, J., & Whitehill, J. (2025, March). Classroom Observation Evaluation with Large Language Models. In *Innovation and Responsibility in AI-Supported Education* (pp. 83-93). PMLR.

How to Cite: Hu, S., Tang, J. & Jiang, J. (2026). AI-Assisted Formative Classroom Assessment: A Multidimensional Framework with Human–AI Collaboration. *Contemporary Education and Teaching Research*, 7(11), 435-441. https://doi.org/10.61360/BoniCETR262021231111