Feedback That Lands: Bringing Formative Assessment Back into the 40-minute Chinese University Lecture

Liu Boran [1], Tan Lingzhi [2], and Dong Haiying [3,*]

[1] School of Marxism, Taizhou University, Taizhou, China
[2] Jiaojiang District Education Bureau, Taizhou, China
[3] The Affiliated School of Taizhou University, Taizhou, China
Email: brilliantz126@126.com (L.B.)
*Corresponding author

***Abstract*—Chinese universities have invested considerably in automated grading platforms that deliver instant, scalable feedback to large classes. Yet a student who receives a precise error report at midnight may still be unable, the next morning, to explain why the error occurred. This paper argues that what decides whether feedback works is less a matter of speed or precision than of timing and place: the brief window inside the 40-minute lecture when a student's reasoning is still visible and can still be adjusted. Drawing on recent evidence on formative assessment, feedback processing, and the documented limits of Artificial Intelligence (AI) grading, the paper identifies three in-class feedback mechanisms (the teacher follow-up, peer explanation, and visible thinking) and translates them into three reusable moves for lectures of eighty to one hundred and thirty students. It proposes repositioning automated grading as an out-of-class screening tool, while the feedback that most directly shapes learning is designed into the 40-minute classroom itself.**

***Keywords*—formative assessment, feedback timing, classroom interaction, large lecture, undergraduate teaching, automated grading**

I. Introduction

The dashboard did its job overnight. Before eight o'clock, the learning platform of a Chinese university had graded 130 multiple-choice quizzes, flagged the three questions that more than sixty percent of the class had answered wrongly, and arranged the results in tidy colored charts. At nine, in the tiered hall holding those 130 students, the teacher spends the first fifteen minutes on exactly those questions: the correct answers read aloud, the reasoning explained, the slides advancing on schedule. Then comes the usual question, "Is that clear now?", and the usual silence. Nobody disagrees; nobody quite confirms, either. The lesson moves on, and by the middle of the period a fair share of the room is elsewhere.

Everything in this scene worked as intended. The technology performed; the teacher followed up faithfully; the dashboard says the gap has been addressed. And yet, if the teacher re-asked one of those flagged questions a week later, a substantial share of the room would very likely pick the same wrong option again. The student now knows the correct answer is B. What the student still does not know is why their own reasoning produced C.

The standard explanation is that feedback was not fast enough or individualized enough, and the standard remedy is a better algorithm. This paper argues that both the diagnosis and the remedy look in the wrong direction. The greater difficulty lies in where the feedback happens: outside the classroom, after the piece of reasoning it should have corrected is over, and with no one present to ask, "Why did you think that?"

This paper asks a practical question: what would feedback look like if its decisive moments were moved from the platform back into the room? To answer it, the paper treats formative assessment as a set of in-class events—measured in seconds rather than days—inside a 40-minute Chinese university lecture of eighty to one hundred and thirty students. The argument runs in three steps. Automated grading has a proper but bounded role, and its limits are structural rather than algorithmic. The feedback literature, read carefully, points to something the automated turn tends to obscure: students rarely need more information about an error; they need a moment in which the reasoning behind the error becomes visible and can be adjusted while it is still fresh. And such moments can be designed through three low-cost moves that require no technology and no extra class time.

Section II reviews the automated feedback turn and the blind spot it has left. Section III sets out an analytical framework of three in-class feedback mechanisms, summarized in a comparison table. Section IV translates the framework into three reusable moves for large lectures. Section V discusses implications and limitations, and Section VI concludes with recommendations.

Manuscript received October 2, 2026; accepted November 7, 2026; published November 26, 2026.

II. Literature Review

A. The Automated Feedback Turn and Its Blind Spot

Over the past decade, Chinese universities have changed markedly in how feedback is produced and delivered. Learning management systems grade objective items the moment a student submits; Artificial Intelligence (AI)-assisted tools generate individualized error analyses; dashboards sort a whole class by weakness. The appeal is easy to see: a single teacher cannot write individual comments for 130 students every week, whereas an algorithm can. Automated grading is instant, consistent, and almost effortlessly scalable.

The blind spot is not inaccuracy, though inaccuracy is real and documented. In one study of an AI short-answer grader, errors that wrongly endorsed an incorrect answer harmed subsequent learning, largely because students who had been told they were right stopped attending to the feedback at all [1]. A systematic review of automated grading and feedback tools in programming education points to the same boundary from the technical side: across 121 studies, the feedback such tools produce is largely confined to whether a submission passes its tests, to the gap between expected and actual output, or to its distance from a reference solution [2]. The deeper blind spot is structural. Automated grading answers the question "which option is correct?" Left to itself, it does not ask "why did you think that?" A student who chooses C may be working from a half-remembered formula, a misread question stem, a genuine misconception, or a guess. The platform records the choice; the reasoning stays invisible. Feedback that cannot see the reasoning can tell a student what to change, but not what to unlearn.

Perspective work on classroom-based formative assessment in the AI era reaches a similar position: AI can help feedback scale, yet the teacher's work of eliciting, hearing, and responding to student reasoning remains central, and leaning too heavily on AI-mediated feedback risks thinning out the dialogue that makes formative assessment formative [3].

B. What the Feedback Evidence Actually Says

Read carefully, the feedback literature of recent years says something rather different from what the automated turn assumes. A systematic review of feedback trials in higher education found the most encouraging evidence for low-stakes quizzing and for forms of peer and tutor feedback, while the evidence for praise, for grading on its own, and for technology-based feedback was mixed [4]. Taken together, the message is not that faster feedback is better feedback. Feedback tends to work when it stays close to a specific task, reaches a student who is still engaged with that task, and is followed by a chance to adjust.

A widely discussed multi-classroom experiment sharpened this point. Many Classes 1 tested the intuitive hypothesis that immediate feedback beats delayed feedback, running the same comparison in 38 college classes across fifteen institutions and a range of disciplines. The estimated effect of feedback timing on later performance was essentially zero (0.002, with a 95% interval from −0.05 to 0.05), and none of forty preregistered moderators showed a credible effect [5]. The authors note hints that delayed feedback may even help in certain conditions their sample underrepresented. Whatever one makes of those hints, the routine implementations they studied suggest that instant delivery, the feature most often advertised, is by itself unlikely to be the decisive ingredient. What seems to matter more is the state of the student who receives the feedback.

Theoretical work on feedback processing takes the argument one step further. On this account, feedback is not delivered and received like a parcel; it is processed—interpreted, emotionally weighed, and then acted on or set aside—by the student [6]. An automated "incorrect" read at midnight can feel like a verdict rather than information. The same error, surfaced at minute fifteen of a live lecture through a classmate's explanation and a teacher's follow-up question, stands a better chance of being experienced as an opening. There is an encouraging corollary: the student's side of the exchange is itself developable. A scoping review of feedback literacy interventions in higher education found evidence that targeted interventions improve how students manage their perceptions and attitudes and strengthen their confidence and agency in the feedback process [7].

C. Why Feedback Outside the Room Struggles

Placed side by side, these weaknesses are structural rather than incidental. There is delay: between the original reasoning and its correction, the student has closed the book, slept, and moved on. There is decontextualization: by the time the error report arrives, the student no longer remembers what they were thinking when they chose C. And there is no follow-up: the platform can say "incorrect," but on its own it does not ask what produced the answer, nor listen to the reply. None of these is a defect that a better model would obviously remove. They follow from where the feedback happens.

III. An Analytical Framework: Three In-Class Feedback Moments

This section treats feedback as a classroom event rather than a digital output, and identifies three mechanisms through which it can land inside a 40-minute lecture. Table I sets the two locations of feedback side by side.

A. The Teacher Follow-Up: "Why Do You Think That?"

When a student gives a wrong or partial answer, the teacher's instinct is to correct it. The follow-up mechanism replaces correction with inquiry: "Why do you think that?" In the next minute or so, the student puts into words the reasoning that produced the error, and that articulated reasoning is itself the feedback. The teacher need not deliver a verdict; the student has heard their own logic spoken aloud, and so has the room. Work on feedback processing suggests that students interpret feedback through their own reasoning and prior beliefs [6], so making that reasoning audible is close to a

precondition for feedback to be taken up rather than deflected.

B. Peer Explanation: Tell the Person Beside You

After a key concept has been introduced, every student is asked to explain it to a neighbor in about thirty seconds. The explaining is the feedback: a student who cannot explain has located, on the spot, the gap between recognizing an idea and understanding it. Because every pair talks at once, no one is singled out, and the teacher need not patrol the aisles. Asking two or three pairs to report what they heard keeps the exchange honest without turning it into a performance.

C. Visible Thinking: Write It Down

The third mechanism asks students to put a single step of reasoning on paper or on the board. Once thinking is externalized, an error stops being a private failure and becomes a shared object the room can examine together. The systematic review mentioned above found benefits for peer and tutor feedback in higher education, contingent on how well they are implemented [4]; visible thinking is one practical way of scaling that kind of peer diagnosis to a room of a hundred or more.

TABLE I. AUTOMATED GRADING OUTSIDE CLASS VS. IN-CLASS FORMATIVE FEEDBACK: A COMPARISON SCHEMA

Dimension	Automated Grading Outside Class	In-Class Formative Feedback
Timing	Hours or days after the task	Seconds after the cognitive move
Information delivered	Right/wrong plus the correct answer	Student reasoning exposed, error located
Student's next move	Reread the solution	Adjust the very next cognitive step
Scalability	Excellent (thousands of students)	Limited per student, universal per room
Face risk	Low (private)	Managed by design (peers, simultaneous talk, safe cold-call)
What it does not do on its own	Ask "why did you think that?" and listen	—

The rows in Table I do not come from a single empirical dataset; they synthesize the formative assessment and feedback literature [1–6] reviewed in Sections II and III and are offered as an analytical baseline.

IV. DESIGNING THE MOMENTS: THREE MOVES FOR 80–130 STUDENTS

Each mechanism translates into a concrete move that fits a 40-minute lecture of eighty to one hundred and thirty students, without technology, without extra class time, and without assuming that students already know one another.

A. The 60-second Follow-up Contract

For each lecture, the teacher marks two or three points in advance at which they will not supply the answer. When a student responds incorrectly or incompletely at one of these points, the teacher asks, "Why do you think that?"—and waits. The minute that follows is not dead time; it is where the feedback happens. The student articulates the reasoning, the room hears it, the teacher or a classmate responds, and the lecture resumes. Preparing the contract beforehand frees the teacher from improvising under pressure. A note of caution from recent evidence is worth adding here: simply lengthening the time students spend on active tasks does not, by itself, track better exam performance [8]. The value of the sixty seconds lies in whether a genuine piece of reasoning surfaces, not in the seconds themselves.

B. Thirty-Second Peer Explanation

Roughly every fifteen minutes, after a concept has been presented, the teacher inserts a half-minute of peer explanation: "Turn to the person next to you and put this in one sentence". The teacher stays at the front; afterwards, two or three students are asked to report what their partner said. The move does two jobs at once. Every student receives feedback on the boundary of their own understanding, because saying an idea aloud to a peer exposes its weak points; and the teacher gains a rough sample of the room's comprehension without walking a hundred and thirty seats.

C. Safe Cold-calling

The third move is a softened cold-call. The teacher names a student; if the student stalls, the teacher neither passes over them nor treats the pause as a failure, but asks the room, "Who can help continue the thought?" The named student receives structured support, and the answer is experienced as something the room builds together rather than a solo performance. In a large lecture, where the fear of being singled out is a main reason students go quiet, this design turns the accountability problem into a feedback resource: a partial answer becomes the material from which the room constructs a fuller one.

V. DISCUSSION

Three implications follow for Chinese university teaching.

One is a matter of positioning rather than rejection. Automated grading remains a sensible tool for low-stakes factual quizzing at scale, for showing a teacher which questions the class found hard, and for taking mechanical marking off the teacher's desk. What it cannot readily take over is the part of formative assessment that depends on eliciting and responding to student reasoning [1, 3]. For universities weighing further investment, the practical question is therefore less "platform or teacher" than how the two are sequenced: the platform screens, the classroom follows up. Evidence from students who make productive use of automated feedback systems points in the same direction: in one qualitative study, the students who benefited used the systems critically rather than dependently, supported by course structure, teacher guidance, and a combination of human and automated feedback [9].

A second implication concerns professional skill. The three moves cost nothing in equipment and nothing in class time, but they ask for a discipline that platforms do

not by themselves supply: the willingness, sixty or ninety seconds at a stretch, to stop talking and let student reasoning appear. The evidence that immediate feedback shows no generalizable advantage under routine conditions [5] points in the same direction. A teacher who waits for a student to explain why they chose C is not being slower than the platform; they are doing the part of feedback that platforms, left to themselves, do not do.

Finally, this paper is best read alongside the active learning literature rather than as another entry in it. The meta-analytic case that active learning improves outcomes is well established [10], while more recent multi-classroom evidence suggests that the simple equation of more active time with more learning does not always hold [8]. What the present framework adds is a specification: what active learning can look like when it is organized around feedback rather than around engagement for its own sake. The aim is not to keep students busy, but to design the moments in which an error becomes visible and adjustable.

The limitations of the paper should be stated plainly. It is a conceptual, practice-oriented piece, not an empirical study, and the three moves await testing across disciplines, class sizes, and student populations. Much of the evidence cited here (on feedback timing [5], on automated grading errors [1], and on feedback trials in higher education [4]) comes from Western institutions, and the review's own authors caution that rigorous feedback evidence in higher education remains limited [4]. How far these findings travel to Chinese lectures of eighty to one hundred and thirty students is an empirical question that local classroom research will need to answer.

VI. Conclusion and Recommendations

This paper has argued that in a Chinese university lecture, the feedback with the greatest leverage rarely comes from a platform. It tends to arrive inside the 40-minute class, in the seconds when a student's reasoning is exposed, heard, and answered. Automated grading can tell a teacher where the class went wrong; the question of why a student went wrong is one that in-class follow-up is far better placed to answer.

For practice, the suggestions are concrete. Teachers might prepare, for each 40-minute lecture, two or three follow-up moments at which they deliberately withhold the answer. Departments might treat the 60-second follow-up, the 30-second peer explanation, and the safe cold-call as teachable techniques in regular teaching workshops, on an equal footing with investment in grading platforms. And those who evaluate large lectures might look past platform analytics (completion rates, time on task, error counts) and ask whether, in each week, students' reasoning was made visible in the room at all.

Feedback that lands is seldom the fastest. It is the feedback a student can use while the thinking that needs it is still going on. That moment has always lived in the classroom, and it is a moment worth designing for.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Liu Boran conceived the study and wrote the manuscript. Tan Lingzhi provided writing suggestions and was responsible for literature search and translation. Dong Haiying provided writing suggestions and was responsible for language polishing, logic checking, and correspondence with the journal editors. All authors have read and approved the final manuscript.

References

[1] T. W. Li, S. Hsu, M. Fowler, Z. Zhang, C. Zilles, and K. Karahalios, "Am I wrong, or is the autograder wrong? Effects of AI grading mistakes on learning," in *Proc. 2023 ACM Conf. Int. Comput. Educ. Res.*, 2023, vol. 1, pp. 159–176. doi: 10.1145/3568813.3600124

[2] M. Messer, N. C. C. Brown, M. Kölling, and M. Shi, "Automated grading and feedback tools for programming education: A systematic review," *ACM Trans. Comput. Educ.*, vol. 24, no. 1, pp. 1–43, 2024. doi: 10.1145/3636515

[3] T. N. Hopfenbeck, Z. Zhang, S. Z. Sun, P. Robertson, and J. A. McGrane, "Challenges and opportunities for classroom-based formative assessment and AI: A perspective article," *Frontiers in Education*, vol. 8, 1270700, 2023. doi: 10.3389/feduc.2023.1270700

[4] R. Morris, T. Perry, and L. Wardle, "Formative assessment and feedback for learning in higher education: A systematic review," *Review of Education*, vol. 9, no. 3, e3292, 2021. doi: 10.1002/rev3.3292

[5] E. R. Fyfe *et al.*, "ManyClasses 1: Assessing the generalizable effect of immediate feedback versus delayed feedback across many college classes," *Advances in Methods and Practices in Psychological Science*, vol. 4, no. 3, 25152459211027575, 2021. doi: 10.1177/25152459211027575

[6] A. M. Lui and H. L. Andrade, "The next black box of formative assessment: A model of the internal mechanisms of feedback processing," *Frontiers in Education*, vol. 7, 751548, 2022. doi: 10.3389/feduc.2022.751548

[7] T. Little, P. Dawson, D. Boud, and J. Tai, "Can students' feedback literacy be improved? A scoping review of interventions," *Assessment & Evaluation in Higher Education*, vol. 49, no. 1, pp. 39–52, 2024. doi: 10.1080/02602938.2023.2177613

[8] X. Cen, R. J. Lee, C. Contreras, M. T. Owens, and J. Maloy, "Time spent on active learning activities does not necessarily correlate with student exam performance: A controlled case study," *Journal of Microbiology & Biology Education*, vol. 25, no. 3, e00073-24, 2024. doi: 10.1128/jmbe.00073-24

[9] L. Li and M. Kim, "It is like a friend to me: Critical usage of automated feedback systems by self-regulating English learners in higher education," *Australasian Journal of Educational Technology*, vol. 40, no. 1, pp. 1–18, 2024. doi: 10.14742/ajet.8821

[10] S. Freeman *et al.*, "Active learning increases student performance in science, engineering, and mathematics," in *Proc. National Academy of Sciences*, 2014, vol. 111, no. 23, pp. 8410–8415. doi: 10.1073/pnas.1319030111

Copyright © 2026 by the authors. This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited (CC BY 4.0).