How Teacher-Mediated Transformation Happens: An Action Research on Generative AI-Supported Process Assessment in Primary School Writing Instruction

Jin Ziwei

Liangxiang Primary School, Fangshan, Beijing 102401, China.

DOI: 10.64649/yh.eg.issn3078-4794.20260511

Abstract: Process assessment in primary school writing instruction has long been constrained by delayed feedback, opaque criteria, and students' revision remaining at the word level. Although generative artificial intelligence (AIGC) can provide instant feedback, there is no automatic channel between technological supply and learning gains. This study asks how, under real classroom conditions, teachers transform assessment information generated by AIGC into assessment support that students can understand and act upon, and at which nodes such transformation produces gains or losses. Adopting an action research design, the study conducted three rounds of practice in a natural fifth-grade class at a primary school in Beijing, organizing data around four nodes—AI generation, teacher filtering and rewriting, student understanding, and revision behavior—to form an analytical framework of a transformation chain. The study finds that transformation gains derive mainly from two types of teacher processing operations, goal alignment and cognitive downgrading, while quantity control and contextual embedding determine whether transformed information can actually be enacted. Transformation loss concentrates at two nodes: excessive information load in AI-generated comments and decontextualized assessment criteria, the former increasing teachers' rewriting burden. Effects also differ markedly across students: weaker students experience loss at the "understanding" node and never reach the revision stage. By operationalizing teacher mediation—from the broad formulation of "teachers using technology" into an observable, codable four-node transformation chain—the study argues that the value boundary of AIGC in writing assessment depends on the weakest link in the transformation chain rather than on the functional ceiling of the tool itself.

Keywords: teacher-mediated transformation; generative AI; process assessment; writing instruction; feedback literacy; action research

0. Introduction

The Compulsory Education Chinese Language Curriculum Standards (2022 Edition) calls on teachers to establish an awareness of the integration of teaching, learning, and assessment, and to make assessment run through the entire teaching process. In actual practice, however, teachers' guidance tends to concentrate before and after writing, leaving students without necessary support during independent writing; assessment is dominated by teacher grading with long feedback cycles, so that by the time students receive comments the cognitive context of writing has often faded; and assessment criteria remain opaque to students, who rarely adjust their expression against criteria during writing. As a result, writing instruction struggles to form a virtuous cycle of "assessment for learning."

Generative artificial intelligence (AIGC), with its capabilities in learning diagnosis, resource generation, and instant feedback, offers a new technological pathway for addressing these difficulties. Existing studies have preliminarily verified the feasibility of intelligent grading systems in terms of scoring consistency and feedback generation, and research on human–AI collaborative writing instruction has shown the positive effects of AI on instructional efficiency and learning

motivation. Yet the ability to provide support does not mean technology can automatically solve problems. One often overlooked fact is that assessment information generated by AI is not automatically received and used by students. Between machine output and student revision lies a zone of transformation that requires teacher intervention. Whether comments focus on the instructional objective of the round, whether the number of suggestions falls within students´ processing capacity, and whether suggestions are anchored to specific locations in students´ compositions—none of these are decided by the tool itself, but depend on how teachers process and transmit the information.

This study therefore does not ask "whether AIGC can empower writing instruction," but takes "teacher-mediated transformation" as its core analytical concept and asks three questions. First, what features do assessment messages generated by AIGC possess, and how do these features affect the possibility of subsequent transformation? Second, what types of processing do teachers perform on AI assessment messages, and how do different types of processing alter the comprehensibility and enactability of the information? Third, under what conditions does transformed assessment information translate into students´ revision behavior, and in which student groups does transformation suffer loss?

This shift in questioning is substantive. It moves technology from the position of agent to that of condition, places teachers´ instructional judgment at the center of analysis, and narrows the research domain from "full-process coverage" to "information transformation in process assessment." The purpose is not to prove that AIGC is superior to traditional grading, but to reveal the conditions under which transformation occurs and the nodes at which loss arises, thereby providing mechanism-level evidence for the regulated use of AIGC in writing assessment.

1. Literature and Theoretical Framework

The theoretical dialogue of this study involves three lines of literature.

The first is research on formative assessment and feedback literacy. The core of formative assessment lies not in grading but in using feedback to drive learning improvement . Existing studies point out that whether feedback works depends on whether learners can receive, interpret, and act upon it, a capacity summarized as "feedback literacy". Feedback literacy research shifts attention from "how teachers give feedback" to "how students use feedback," revealing that ineffective feedback is often not a matter of content but of feedback not being effectively received by learners.

The second is research on teachers as curriculum mediators. In sociocultural theory, "mediation" originally refers to the regulatory role of tools between humans and the world, and subsequent research extends it to teachers´ pedagogical processing of curriculum materials and technological tools . Teachers are not passive executors of curriculum or technology, but mediators who select, transform, and reorganize materials entering the classroom according to learner conditions and objectives. This line of thinking suggests that when AIGC enters writing assessment, teachers likewise perform a mediating function, and the quality of their processing directly determines whether technological output can be converted into instructional support.

The third is research on AIGC-based writing assessment. Existing studies mostly focus on the consistency between intelligent scoring and teacher scoring, descriptions of comment features, and the construction of human–AI collaborative teaching models, and rarely examine the complete transformation process of assessment information from machine to student. The few studies attending to feedback transformation are mostly speculative discussions, lacking systematic observation and coding of teacher processing operations in real classrooms.

Integrating these three lines, this study proposes a four-node analytical framework of a "transformation chain." Node 1 is AI generation, analyzing the original form of AI assessment

messages, including dimension coverage, number of suggestions, language complexity, and degree of goal alignment. The study identifies two features: "information load," meaning that a single comment contains too many suggestions and exceeds students´ processing capacity; and "goal diffusion," meaning that comments cover multiple dimensions with equal weight, so that content directly related to the instructional objective of the round accounts for a low proportion. These two features are the structural sources of subsequent transformation loss. Node 2 is teacher filtering and rewriting, the core link of transformation. The study identifies four types of teacher processing operations: goal alignment, retaining suggestions directly related to the instructional objective while deleting or setting aside others; cognitive downgrading, rewriting directional descriptions into concrete, enactable instructions; quantity control, compressing the number of suggestions and establishing priorities; and contextual embedding, anchoring suggestions to specific locations in students´ compositions. These four operations serve both as a coding framework for teacher transformation behavior and as an analytical basis for judging transformation quality. Node 3 is student understanding, examining students´ comprehension of transformed information and their judgment behavior. The study makes understanding explicit through a "suggestion deliberation" design, in which students judge each item as agree or disagree and state reasons. Node 4 is revision behavior, using revision type and revision quality as indicators to examine the proportion of transformed information that translates into substantive revision, and to analyze differences across writing proficiency levels.

The four nodes are not linearly transmitted but involve feedback loops: students´ revision results feed back to the teacher, adjusting the transformation strategy for the next round. This loop design makes "transformation" an iterable dynamic process rather than one-off processing, and constitutes the analytical basis of the action research cycle.

2. Research Methods

The study adopts an action research paradigm, taking a natural class of 42 fifth-grade students at a primary school in Beijing as participants and conducting three rounds of teaching practice over one semester; the subject teacher is the researcher herself. The three rounds use, respectively, the character-writing task "People of All Kinds," the narrative task "That Moment, I Grew Up," and the scenery-writing task "China´s World Cultural Heritage" as vehicles, with core competence objectives of "using action and language to write characters concretely," "using process decomposition to write events concretely," and "using the combination of static and dynamic description to write scenery concretely." Each round follows the cycle of planning–action–observation–reflection, and the transformation strategy for the next round is determined by the observation data and reflection conclusions of the previous round.

Data sources include: original comments generated by the AI system, revision prompt sheets rewritten by the teacher, suggestion deliberation sheets completed by students, students´ first and revised drafts, classroom observation records, and transcribed interviews with 10 students selected by stratified sampling. For analysis, coding is used to classify teacher processing operations and student revision types. Teacher processing operations are coded according to the four operations described above; student revision types are divided into word-level, sentence-level, content-level, and structure-level, and changes in the proportion of each type are calculated; deliberation sheets are coded by "agree–disagree–uncertain" and by type of reason. Quantitative data describe trends in transformation proportions, while qualitative materials explain the conditions under which transformation occurs or is blocked. Classroom observation was recorded separately by the researcher and a Chinese language teacher from the same school and then compared to reduce single-observer bias.

3. Findings

3.1 The AI generation end: information load and goal diffusion as the structural starting point of transformation

Across the three rounds, the original comments generated by the IN Classroom system displayed highly consistent features. Taking the composition of Student B in Round 1 as an example, the system´s report included overall score, highlight appreciation, core suggestions, completion of writing requirements, detailed commentary, improvement suggestions, and writing reflections, with the number of suggestions usually ranging from two to four and covering typos, punctuation, sentence errors, structure, material selection, and descriptive methods. In terms of content, these comments are not lacking in information; the problem lies in the mismatch between how the information is organized and students´ processing capacity.

The first feature is information load. A single comment contains many suggestions and sections, and students facing multiple parallel suggestions cannot judge priorities. In Round 1, the teacher printed and pasted the original comments for students to read and revise on their own. Revision was highly concentrated at the word level, and content-level suggestions were difficult to implement.

The second feature is goal diffusion. Comments cover multiple dimensions with equal weight, and content directly corresponding to the round´s core objective of "action and language description" accounts for a low proportion. The system does not adjust assessment weights because the current instruction focuses on a particular competence point, so students receive general evaluation rather than feedback focused on the learning objective of the round. Together, these two features constitute the structural starting point of transformation: AI output is not designed to be used directly by students; it is oriented toward teachers and presupposes a reader with filtering capacity. This is precisely why transformation is necessary.

3.2 The teacher end: four processing operations producing different transformation gains

Across the three rounds, the teacher´s processing of AI comments evolved from absent to present and from coarse to fine, and the role of the four operations became increasingly clear through comparison.

In Round 1, the teacher did no processing and distributed original comments directly, so transformation barely occurred. In Round 2, the teacher began systematic rewriting to form "revision prompt sheets," with operations including: removing scores and dimension names; compressing suggestions to no more than two; prioritizing descriptive suggestions related to "writing that moment concretely"; and converting directional descriptions into enactable requirements. Taking Student M´s composition as an example, the rewritten prompt sheet specified "What does Mom look like when she heats the milk? What are her hands doing? Write out the action that impressed you most." In the revised draft, the student rewrote the paragraph on the mother heating milk, changing "Mom skillfully cut open a carton of milk and poured it into my glass" to "Mom gently pinched one corner of the carton, the scissors slowly tracing the crease, the milk flowing thinly into the cup," and added the inner monologue "I cut open the carton the way Mom does, thinking, does she do this for me every morning?" The number of content-level additive revisions increased markedly compared with Round 1, while the proportion of word-level revisions declined.

In Round 3, an "AI suggestion deliberation sheet" was added on the basis of rewriting, requiring students to judge each item as agree or disagree and to state reasons and planned revisions. Taking Student C´s deliberation sheet as an example, the student directly checked "write the dynamic concretely" and "write out the feeling" and wrote revision plans, but marked "make the alternation of static and dynamic clear" as "uncertain," reasoning that "Longevity Hill is static in itself; would adding cloud shadows be unrealistic?" This question was resolved in peer assessment, where

a peer responded that "the trees on the hill move and the clouds move too; it is not completely static," after which the student confirmed adoption.

Comparing the three rounds, it can be summarized that goal alignment and cognitive downgrading mainly produce transformation gains: the former turns assessment information from "general feedback" into "goal-relevant feedback," and the latter turns information from "directional description" into "enactable instruction." Quantity control and contextual embedding determine whether transformation can be enacted: the former addresses students´ "cannot process it all" problem, and the latter addresses their "cannot locate it" problem. All four operations are indispensable, but their mechanisms differ.

3.3 The student end: transformation loss concentrated at the understanding node and differentiated across groups

Transformed information is not equally effective for every student. Interview and deliberation-sheet data show that loss occurs mainly at the "understanding" node and displays clear group differences.

Students at or above the intermediate writing level could make independent judgments about suggestions on the prompt sheets, and some could even question AI suggestions. Student S3 reported that the ideation prompt question "What does Mom look like when you wake up at midnight?" immediately brought back the scene, after which the student wrote the detail of the mother´s hand on his forehead. Student S10 reported that the suggestions on the prompt sheet were easier to understand than the original system text: "Now the prompt sheet has only two items, and each tells me what to do; I understand it at a glance."

Weaker students experienced loss at the understanding node. Student S4 stated: "Sometimes I don´t know how to answer the questions on the prompt strip, for example, when it asks what I was thinking at that moment, I can´t remember right away." Student S8 noted: "Some AI suggestions don´t suit me, for example, asking me to add similes, but I´m not good at using similes." These responses indicate that although cognitive downgrading reduces the linguistic difficulty of information, it does not reduce the information´s dependence on recall, association, and expression abilities. For weaker students, even when suggestions are rewritten as enactable instructions, enactment still presupposes existing expressive capacity. These students stall at the third node of the transformation chain and never reach revision, so changes in their revised drafts are limited; in the post-test, about two tenths of students remained at "average interest" and "average level."

Peer assessment partially compensated for this loss. Student S3 noted that "suggestions from classmates are sometimes easier to understand than the teacher´s, because they look at it from my perspective," and Student S6 said "suggestions from classmates are easier for me to accept." Yet peer assessment quality is constrained by students´ assessment ability: Student S7 reported that "some classmates´ opinions are not quite right," and Student S2 noted that "some classmates just say ´it´s good´ without saying what is good." This indicates that peer assessment is an effective supplementary node in the transformation chain but cannot replace teacher mediation, and itself requires guidance and regulation.

4. Discussion

The findings can be connected to three theoretical dialogue points.

First, teacher mediation is not simple information relay but a redesign involving goal judgment. Among the four processing operations, goal alignment requires teachers to select from AI output according to the instructional objective of the round, and cognitive downgrading requires teachers to reconstruct information according to students´ cognitive level; neither is merely "saying what the AI said in other words," but a second design of machine output measured against instructional

objectives. This finding corrects the simplified understanding that "teacher mediation equals teacher gatekeeping" and specifies mediation as an observable, codable set of operations.

Second, feedback literacy is not only a student capacity but also includes teachers´ capacity to reprocess machine feedback. Existing feedback literacy research mostly focuses on how students receive and use feedback; this study shows that under AIGC intervention, feedback has already undergone a round of teacher processing before students encounter it, and teachers´ feedback literacy—the capacity to judge what kind of feedback is effective for students in what form—becomes a prior variable determining feedback quality. This finding extends the object of feedback literacy research from students to both teachers and students.

Third, the value boundary of AIGC in writing assessment depends on the weakest link in the transformation chain rather than on the functional ceiling of the tool. Transformation loss concentrated at two nodes: information load and student understanding. This means that improving the effect of AIGC-based writing assessment lies not in endlessly enhancing tool functions, but in designing mediation strategies for weak nodes: at the generation end, making comments configurable and trimmable according to instructional objectives; at the teacher end, establishing transformation operation norms; and at the student end, using suggestion deliberation and peer assessment as supplementary nodes.

5. Conclusions and Recommendations

Taking "teacher-mediated transformation" as its core concept, this study examined the transformation process of AIGC assessment information in primary school writing classrooms through three rounds of action research, reaching the following conclusions.

First, teacher-mediated transformation consists of four operations: goal alignment, cognitive downgrading, quantity control, and contextual embedding. The first two produce transformation gains, turning assessment information from general feedback into goal-relevant, enactable instructions; the latter two determine whether transformation can be enacted, addressing information processing load and revision location respectively.

Second, transformation loss concentrates at two nodes. Information load and goal diffusion at the generation end constitute the structural starting point of loss, while the stalling of weaker students at the understanding end constitutes the group-level manifestation of loss. Transformation is not a smooth linear process but one that may attenuate at multiple nodes.

Third, transformation effects differ across students. Students at or above the intermediate writing level can make independent judgments about transformed information, whereas weaker students experience loss at the understanding node and never reach revision. Peer assessment can partially compensate for this loss but is constrained by students´ assessment ability and cannot replace teacher mediation.

Based on these conclusions, the study offers recommendations at three levels. At the teacher level, a transformation operation norm of "goal alignment—cognitive downgrading—quantity control—contextual embedding" should be established, students´ understanding should be checked through deliberation sheets after rewriting. At the resource level, AI comments should be made configurable and trimmable according to instructional objectives, reducing information output unrelated to the current objective and lowering transformation load from the generation end. At the student level, suggestion deliberation and peer assessment should be designed as supplementary nodes in the transformation chain: deliberation sheets make understanding explicit, and peer assessment circulates feedback among students; together they alleviate the burden of teacher mediation but require clear assessment requirements for regulation.

The limitations of this study are as follows: the single natural class yields a small sample without a control group, so generalization requires caution; the one-semester period means the stability of

transformation strategies needs longer-term testing.

About the Author

Jin Ziwei (March 1997—), female, Han nationality, from Beijing, holds a master´s degree and is a first-level teacher

References

[1] Ministry of Education of the People´s Republic of China. (2022). Compulsory Education Chinese Language Curriculum Standards (2022 Edition). Beijing: Beijing Normal University Press.

[2] Cui, Y., & Lei, H. (2015). Construction of a three-factor theoretical model of teaching-learning-assessment consistency. Journal of East China Normal University (Educational Sciences), 33(4), 15-22.

[3] Zhang, S. (2019). The origin and meaning of "teaching-learning-assessment consistency" and "teaching-learning-assessment integration." Teaching Reference of Middle School Chemistry, (13), 4-5.

[4] Shi, Y. (2021). Classroom assessment in composition instruction under "teaching-assessment consistency." Language Construction, (20), 14-18.

[5] Cao, H. (2024). Primary school Chinese composition classroom assessment from the perspective of teaching-learning-assessment consistency. Teaching and Management, (26).

[6] Yang, F. (2021). Establishing and implementing composition instructional objectives under teaching-learning-assessment integration. Teaching and Management, (35).

[7] Huang, T., Gong, M., Yang, H., et al. (2020). Human-AI collaborative support for primary school Chinese composition instruction. E-education Research, (2), 108-114.

[8] Xia, J., & Lai, Y. (2025). Possibilities and paths of generative AI-assisted instruction: The case of primary and secondary school Chinese composition teaching. Primary and Secondary School Management, (2), 36-40.

[9] He, Z. (2025). Making the scene present so that writing flows: Exploring large model technology in primary school Chinese composition instruction. China Educational Technology & Equipment, (1), 55-58.

[10] Liu, S., Li, Y., Yang, P., et al. (2021). A study on the effects of intelligent composition assessment. Open Education Research, 27(3), 73-84.

[11] Lu, Y., Yu, J., Chen, P., et al. (2023). Educational applications and prospects of generative AI: The case of the ChatGPT system. Chinese Journal of Distance Education, (4), 24-31.

[12] Wan, P., & Gu, X. (2024). Human-AI collaborative assessment supported by generative AI. Modern Educational Technology, 34(2), 5-14.

[13] Ye, H. (2011). A brief discussion of backward design. Contemporary Educational Science, (4), 23-26.

[14] Wen, P., & Jia, G. (2002). Constructivist theory and teaching reform: A review of constructivist learning theory. Theory and Practice of Education, 22(5), 17-22.

[15] Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. Assessment & Evaluation in Higher Education, 43(8), 1315-1325.

[16] Winstone, N. E., Nash, R. A., Parker, M., & Rowntree, J. (2017). Supporting learners´ agentic engagement with feedback: A systematic review and a taxonomy of recipience

processes. Educational Psychologist, 52(1), 17–37.

[17] Boud, D., & Molloy, E. (Eds.). (2013). Feedback in higher education: Understanding it and doing it well. Routledge.

[18] Wertsch, J. V. (1991). Voices of the mind: A sociocultural approach to mediated action. Harvard University Press.

[19] Brown, A. L., & Campione, J. C. (1996). Psychological theory and the design of innovative learning environments. In L. Schauble & R. Glaser (Eds.), Innovations in learning: New environments for education (pp. 289–325). Lawrence Erlbaum.

[20] Wiggins, G., & McTighe, J. (2005). Understanding by design (2nd ed.). ASCD.

[21] Page, E. B. (2003). Project Essay Grade: PEG. In M. D. Shermis & J. C. Burstein (Eds.), Automated essay scoring: A cross-disciplinary perspective (pp. 43–54). Lawrence Erlbaum.

[22] Humble, N., & Mozelius, P. (2022). The threat, hype, and promise of artificial intelligence in education. Discover Artificial Intelligence, 2, 22.

[23] Holmes, W., Bialik, M., & Fadel, C. (2019). Artificial intelligence in education: Promises and implications for teaching and learning. Center for Curriculum Redesign.

[24] Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). Intelligence unleashed: An argument for AI in education. Pearson.

[25] Sato, M. (2003). Curriculum and teachers (Zhong Qiquan, Trans.). Beijing: Educational Science Publishing House.

[26] Wu, Z. (2005). Promoting a virtuous cycle of teaching and learning through action research. Foreign Languages in China, (6), 4.

[27] Liu, J. (2020). Research methods in physics education. Shanghai: Shanghai Popular Science Press.

[28] Fu, Y. (2015). Classroom elements from an ecological perspective and big-data-supported teaching evaluation: A commentary on Cui Yunhuo's LICC classroom observation. Journal of Nanyang Normal University, (10), 59–61.

[29] Zhang, P. (2025). Research and practice on the effectiveness of primary school Chinese composition correction. Questions and Guidance, (13), 1–4.

[30] Tribble, C. (1996). Writing. Oxford University Press.

[31] IN Classroom Intelligent Education Platform. (2018, November 20). IN Classroom: Intelligent Chinese composition correction—Education moving toward intelligence [EB/OL].

[32] Dan, W., & Xiao, M. (2023). Construction of a teaching–learning–assessment integration system from the perspective of core competencies. Basic Education Curriculum, (9), 4–8.

[33] Shuai, F. (2019). Teaching–learning–assessment integration: Goal orientation, research impetus, and bidirectional construction. Educational Measurement and Evaluation, (10), 27–33.