Proceedings of the 4th International Conference on Software Engineering and Machine Learning
DOI: 10.54254/2755-2721/2026.35500

Research on Human-AI Collaborative Agent Dynamic Feedback Mechanism for Primary Chinese Large-Unit Instruction

Linghui Su

Shaanxi Normal University, Xi'an, China
lingwadesu@gmail.com

Abstract. The proposed research presents a structure of a human-AI collaboration agent to provide dynamic feedback for a Chinese large-unit instruction context. The dataset consists of eight cycles of Chinese large-unit teaching sessions in Grade 5. It includes 320 participants from 32 lessons conducted in 32 classes. Data types involved include lesson plans, dialogue between students and teachers, workbooks, reading notes, vocabulary assignments, short compositions, test scores, graded rubrics for reading and writing, and teacher's feedback notes. The project was designed with Python 3.11, MySQL 8.0, BERT Base Chinese, LightGBM, and Vue teacher dashboard. Learning states of learners can be predicted through features connected with text understanding, task accomplishment, interaction engagement, and progress monitoring. The fixed target labels include stable mastery, partial knowledge, problems with expression, and constant vulnerability. The results prove that the learning state recognition model obtains accuracy of 0.872 ± 0.024, F1 score of 0.851 ± 0.027, and weighted recall of 0.864 ± 0.025. As compared to conventional human-only feedback, the human-AI collaboration improves performance at the following task by +0.184 ± 0.036, reduces mistakes repetitions by −0.217 ± 0.041, and increases unit posttest scores by +6.84 ± 1.27.

Keywords: Human-AI collaboration, dynamic feedback, primary Chinese instruction, large-unit teaching, learning analytics

1. Introduction

Primary Chinese large-unit instruction organizes texts, language elements, reading strategies, oral expression, and writing practice around one integrated unit theme. Compared with isolated lesson teaching, large-unit instruction requires students to connect multiple texts, accumulate vocabulary and sentence patterns, understand unit-level ideas, and transfer reading comprehension into expression tasks [1].

The collaboration between humans and AI represents a practical approach to providing feedback in classrooms due to the inherent advantages each possesses. Teachers know about the classroom environment, the characteristics of students, educational goals, and pacing in instruction, whereas AI can quickly process large amounts of text written by students, transcripts of conversations, results of

DOI: 10.54254/2755-2721/2026.35500

testing, and traces of learning activity [2]. In primary Chinese learning, this collaboration can support the teacher in identifying weak textual inference, insufficient theme expression, repeated vocabulary misuse, delayed participation, and unstable task completion.

Existing AI-supported teaching research has emphasized learning analytics, adaptive feedback, teacher dashboards, and human-AI complementarity in instructional decision-making [3, 4]. However, primary Chinese large-unit instruction requires a more specific mechanism because learning states are distributed across sequential unit tasks rather than one single test. A student may understand vocabulary but fail to infer character emotion; another may participate actively in discussion but show weak written expression. Therefore, feedback must be linked to task type, learning stage, and student state.

This study develops a human-AI collaborative agent for dynamic feedback in Grade 5 Chinese large-unit instruction. The framework integrates unit-task decomposition, classroom data collection, BERT-based text analysis, LightGBM-based learning-state recognition, AI feedback recommendation, teacher dashboard selection, and learning-effect evaluation [5]. The goal is to support teacher decision-making while keeping the teacher as the final feedback actor.

2. Literature review

2.1. Large-unit instruction in primary Chinese education

Large-unit instruction in primary Chinese education connects multiple texts, language objectives, learning tasks, and expression activities around a shared unit theme. It usually includes theme introduction, text interpretation, vocabulary application, dialogue-based discussion, reading strategy training, and unit-level writing output. Feedback in this structure needs to connect local task performance with broader unit objectives. If feedback only corrects surface errors, students may improve one task but fail to build transferable reading and expression ability. Therefore, large-unit feedback should identify whether students have achieved text comprehension, language accumulation, thematic integration, and expression transfer across the unit process [6].

2.2. Human-AI collaboration in intelligent teaching systems

Human-AI collaboration in teaching systems refers to a shared decision structure in which AI analyzes learning records and teachers make final instructional judgments. AI can calculate learning-state indicators, process classroom transcripts, detect repeated errors, and recommend feedback categories. Teachers can revise, reject, or implement AI suggestions according to classroom context. This division avoids replacing teacher judgment with automated decisions. It also allows feedback to become faster, more data-supported, and more traceable. Effective collaboration requires explicit task division, synchronized data flow, dashboard-based visibility, and feedback logs that record both AI suggestions and teacher implementation [7, 8].

2.3. Dynamic feedback mechanisms in classroom learning

Dynamic feedback is generated during learning rather than after all instruction ends. In Chinese language classrooms, feedback may target reading misunderstanding, weak inference, inaccurate vocabulary transfer, insufficient sentence expansion, low discussion participation, or unstable writing revision. A dynamic mechanism requires three linked processes: learning-state recognition, feedback-category generation, and implementation tracking. When feedback is connected with the

DOI: 10.54254/2755-2721/2026.35500

teaching timeline, teachers can intervene before small misunderstandings become persistent unit-level problems. Automated feedback research also shows that feedback effectiveness depends on timing, specificity, and alignment with the learner's current task state [9, 10].

3. Research methodology

3.1. Instructional dataset construction and data schema

The dataset consists of eight large-unit teaching cycles for Grade 5 primary Chinese. Every unit includes four lessons, thus, resulting in a total of 32 lessons. There are a total of 320 students involved in the sample study who are divided into eight groups. Themes covered in the unit include reading themes, text analysis, vocabulary usage, discussion tasks, worksheet assignments, writing sentences, and writing for the unit as a whole.

All data have been stored on MySQL 8.0 in five fixed tables. These include unit_task, classroom interaction, student_work, assessment record, and feedback log. The unit_task table contains information about unit number, lesson number, task number, text snippet, vocabulary, discussion, writing assignment, and learning objective. The classroom_interaction table contains information about utterance, speaker ID, transcription, response time, and task ID. Student_work table contains student answers, reading notes, sentence writing, short compositions, and corrections made. The assessment_record table stores quiz score, reading score, writing score, vocabulary application score, and rubric dimension. The feedback_log table stores AI-suggested feedback, teacher-selected feedback, teacher revision, feedback time, implementation status, and follow-up result.

3.2. Student learning-state modeling and feature construction

All preprocessing is completed in Python 3.11 with pandas, NumPy, scikit-learn, jieba, and transformers. Student texts are cleaned by removing repeated punctuation, invalid characters, empty submissions, and duplicated answers. Chinese word segmentation is performed for vocabulary-level indicators. BERT-Base Chinese is used to encode worksheet answers, reading notes, and short compositions into semantic embeddings. Structured scores and interaction records are merged with text embeddings into one student-task matrix.

The student-state feature set contains four fixed groups: text-comprehension features, task-performance features, interaction-participation features, and learning-progression features. Text-comprehension features include keyword coverage, semantic similarity to reference answers, inference completeness score, theme-expression consistency score, character-emotion inference score, and text-evidence citation density. Task-performance features include worksheet accuracy, vocabulary application accuracy, writing score, sentence expansion score, quiz correctness rate, and rubric-based expression quality. Interaction-participation features include speaking frequency, response latency, teacher-call response count, discussion contribution length, and peer-response participation. Learning-progression features include score change across lessons, repeated error frequency, feedback uptake rate, task completion stability, and revision improvement score. The four target labels are stable mastery, partial understanding, expression difficulty, and continuous risk.

3.3. Human-AI collaborative feedback model and system configuration

The collaborative feedback system contains two fixed modules: AI feedback recommendation and teacher decision interface. The AI recommendation model is implemented in LightGBM with

DOI: 10.54254/2755-2721/2026.35500

num_leaves = 31, learning_rate = 0.05, n_estimators = 200, max_depth = 6, feature_fraction = 0.85, bagging_fraction = 0.80, and random_state = 42. For each student-task record, the model predicts one learning-state label and one feedback category. The fixed feedback categories are clarification feedback, scaffolded prompting, example-based correction, and extension feedback.

Clarification feedback targets partial understanding, especially incomplete inference, unclear theme recognition, or weak text-evidence use. Scaffolded prompting targets continuous risk and supports students through guiding questions, sentence frames, and step-by-step reading prompts. Example-based correction targets expression difficulty by providing correct sentence patterns, vocabulary-use examples, and short model responses. Extension feedback targets stable mastery by adding comparison questions, transfer tasks, and higher-level expression prompts.

The feedback recommendation function is defined as shown in Equation (1) and (2):

$$f_{i,t} = arg\max_{k \in \mathcal{K}} P\left(k \mid \hat{y}_{i,t}, \mathbf{r}_{i,t}, \mathbf{q}_t, \mathbf{e}_{i,t}\right) \tag{1}$$

$$T_{i,t} = Template\left(f_{i,t}, Error_{i,t}, Task_t, TextSeg_t\right) \tag{2}$$

where $f_{i,t}$ is the recommended feedback category, $\mathcal{K}$ contains the four feedback categories, $\mathbf{r}_{i,t}$ is the student's recent learning record, $\mathbf{q}_t$ is the current task type, $\mathbf{e}_{i,t}$ is the detected error pattern, and $T_{i,t}$ is the generated feedback text template. The Vue 3 dashboard displays student state, task evidence, AI-suggested feedback, teacher-selected feedback, revision history, and follow-up results. Teachers can confirm, edit, replace, or postpone AI suggestions.

4. Experimental design

4.1. Unit arrangement, teaching sequence, and data collection process

The experiment uses one fixed Grade 5 Chinese large unit containing theme reading, text interpretation, language accumulation, discussion, and unit writing. Each class completes the same four-lesson sequence over two weeks. In Lesson 1, students will be introduced to the theme, as well as a previewing reading activity. In Lesson 2, there will be focus on close reading skills, applying the vocabularies learned, and getting textual evidence from the lesson text. In Lesson 3, discussion and inference skills will be highlighted. During each lesson, classroom audio is recorded and transcribed. Worksheet responses, reading notes, and writing texts are digitized. Short quizzes are completed on a tablet platform. At the end of each lesson, student data are uploaded to MySQL. The AI module updates learning-state records and generates feedback suggestions for the next task. During the following lesson, the teacher dashboard displays student-level and group-level feedback suggestions linked to current unit tasks.

4.2. Data processing, model training, and feedback generation workflow

Historical lesson records from the first 6 classes are used as training data, and the remaining 2 classes are used as test data. Classroom transcripts are segmented by speaking turn and aligned with student IDs and task timestamps. Student texts are encoded with BERT-Base Chinese. Structured learning features are merged with scores and interaction logs into one student-task matrix.

DOI: 10.54254/2755-2721/2026.35500

The LightGBM model is trained to predict four learning-state labels and mapped feedback categories. During execution, each new student-task record passes through the model, and one feedback recommendation is generated and written into the feedback_log table. The teacher confirms, edits, or replaces the AI suggestion through the dashboard. The implemented feedback is recorded as the final classroom intervention.

4.3. Evaluation metrics and comparative output generation

The experiment contains two comparison settings: teacher-only feedback and human-AI collaborative dynamic feedback. Learning-state classification is evaluated by accuracy, macro-F1, weighted recall, and class-level F1. Feedback effectiveness is evaluated by next-task improvement rate, error recurrence reduction rate, writing-score gain, reading-comprehension gain, and unit post-test score. Teacher-side usability is evaluated by feedback adoption rate, feedback revision rate, and average feedback response time. Statistical analysis is completed in R 4.4, and visualization is generated with ggplot2.

5. Experimental results

5.1. Descriptive statistics of large-unit instruction tasks, student learning states, and feedback records

The data set is based on 32 classes, 1,536 unit assignments, 12,842 student assignment entries, 5,926 class discussion turns, 8,714 worksheet answers, 2,560 reading annotations, and 1,280 mini-compositions. The stable mastery level is 0.342 ± 0.041 of the student assignment entries, partial comprehension is 0.284 ± 0.036, expression difficulties make up 0.211 ± 0.029, and continuous risks account for 0.163 ± 0.027. The feedback history consists of 7,436 entries from AI and 6,

5.2. Learning-state recognition performance and dynamic feedback recommendation results

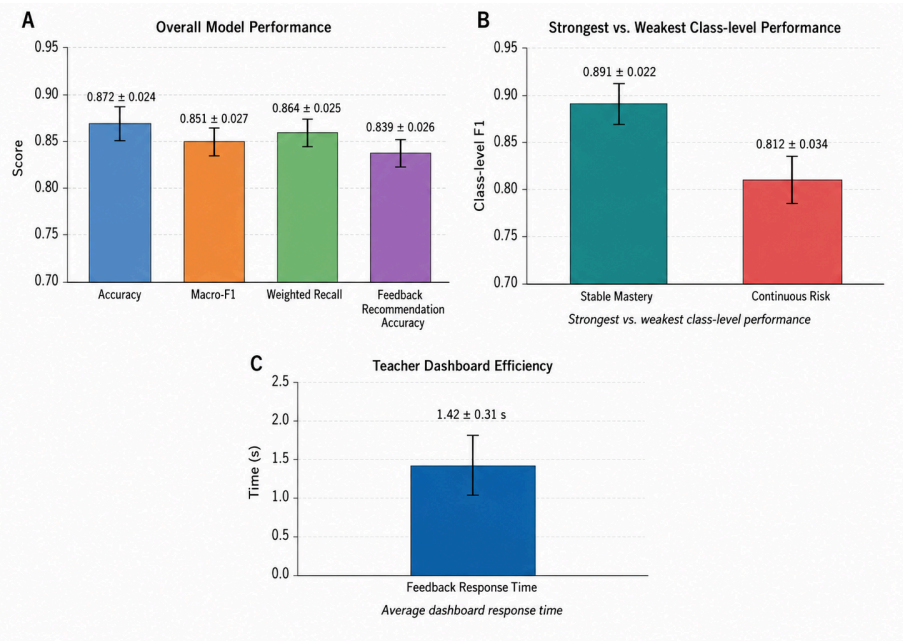

Figure 1. Learning-state recognition performance and dynamic feedback recommendation results

DOI: 10.54254/2755-2721/2026.35500

Accuracy of the proposed learning state model using LightGBM was found to be 0.872 ± 0.024, with macro-F1 score of 0.851 ± 0.027 and weighted recall rate of 0.864 ± 0.025. On a class-by-class basis, the maximum F1 score is observed for Stable Mastery class (0.891 ± 0.022), while the minimum score belongs to Continuous Risk class (0.812 ± 0.034), which is due to presence of risk states with a mixture of indicators related to low task performance, late participation, and mistakes repeated multiple times. Accuracy of feedback recommendations was equal to 0.839 ± 0.026, and the average dashboard response time amounted to 1.42 ± 0.31 sec.

5.3. Comparative effects of teacher-only feedback and Human-AI collaborative feedback on unit learning outcomes

Feedback on collaboration between humans and artificial intelligence leads to better performance improvements compared to teacher-only feedback. Next-task improvement becomes larger at 0.536 ± 0.047 to 0.720 ± 0.041***, whereas error repetition reduces from 0.438 ± 0.052 to 0.221 ± 0.039***. Gains in writing scores grow from +4.28 ± 1.04 to +7.96 ± 1.32***, whereas the unit post-test score improves from 82.14 ± 5.36 to 88.98 ± 4.72***. These results indicate that dynamic feedback works by establishing connections between student-state detection and task-oriented teacher assistance (refer to Table 1). The greatest improvements happen if the teacher uses AI recommendation after editing, not mechanically.

Table 1. Learning-state recognition and feedback-effect comparison results

Evaluation Indicator	Teacher-Only Feedback	Human-AI Collaborative Feedback	Change
Learning-State Accuracy	—	0.872 ± 0.024	—
Learning-State Macro-F1	—	0.851 ± 0.027	—
Weighted Recall	—	0.864 ± 0.025	—
Feedback Recommendation Accuracy	—	0.839 ± 0.026	—
Next-Task Improvement Rate	0.536 ± 0.047	0.720 ± 0.041***	+0.184 ± 0.036
Error Recurrence Frequency	0.438 ± 0.052	0.221 ± 0.039***	−0.217 ± 0.041
Writing-Score Gain	+4.28 ± 1.04	+7.96 ± 1.32***	+3.68 ± 0.84
Reading-Comprehension Gain	+3.74 ± 0.91	+6.35 ± 1.08***	+2.61 ± 0.73
Unit Post-Test Score	82.14 ± 5.36	88.98 ± 4.72***	+6.84 ± 1.27
Feedback Response Time (s)	3.86 ± 0.74	1.42 ± 0.31***	−2.44 ± 0.49

6. Conclusion

This study establishes a human-AI collaborative agent framework for dynamic feedback in primary Chinese large-unit instruction. The framework organizes unit tasks, classroom interaction records, student texts, assessment results, and feedback logs into one structured database. BERT-Base Chinese is used for student-text representation, LightGBM is used for learning-state recognition and feedback recommendation, and a Vue-based dashboard supports teacher confirmation, revision, and implementation. The results show that human-AI collaborative dynamic feedback improves both feedback efficiency and learning outcomes. The mechanism is not simple automation. AI identifies student states and recommends task-specific feedback, while teachers retain final control over classroom intervention. Collaborative feedback reduces repeated errors, improves next-task performance, strengthens reading-comprehension gains, and increases unit post-test scores. The

DOI: 10.54254/2755-2721/2026.35500

framework therefore provides a practical route for integrating learning analytics and teacher expertise into primary Chinese large-unit instruction.

References

[1] Ouyang, F., & Jiao, P. (2021). Artificial intelligence in education: The three paradigms. *Computers and Education: Artificial Intelligence, 2*, 100020.
[2] Molenaar, I. (2022). Towards hybrid human-AI learning technologies. *European Journal of Education, 57*(4), 632–645.
[3] Holstein, K., & Aleven, V. (2022). Designing for human–AI complementarity in K-12 education. *AI Magazine, 43*(2), 239–248.
[4] Cavalcanti, A. P., Barbosa, A., Carvalho, R., Freitas, F., Tsai, Y.-S., Gašević, D., & Mello, R. F. (2021). Automatic feedback in online learning environments: A systematic literature review. *Computers and Education: Artificial Intelligence, 2*, 100027.
[5] Alfredo, R., Echeverria, V., Jin, Y., Yan, L., Swiecki, Z., Gašević, D., & Martinez-Maldonado, R. (2024). Human-centred learning analytics and AI in education. *Computers and Education: Artificial Intelligence, 6*, 100215.
[6] Atchley, P., Panneton, R., Morales, G., & Lamm, C. (2024). Human and AI collaboration in the higher education environment: Opportunities and concerns. *Cognitive Research: Principles and Implications, 9*, 20.
[7] Jin, F. J. Y., Yan, L., & Gašević, D. (2025). Students' perceptions of GenAI-powered feedback in higher education. *Journal of Learning Analytics, 12*(1), 1–19.
[8] Qiao, H., Li, Y., & Zhao, X. (2023). Artificial intelligence-based language learning: Illuminating the impact on speaking skills and self-regulation in Chinese EFL context. *Frontiers in Psychology, 14*, 1255594.
[9] Shin, D. (2023). Learning and teaching with artificial intelligence in K-12 education: A systematic review. *Computers and Education: Artificial Intelligence, 5*, 100150.
[10] Xu, W., & Ouyang, F. (2022). The application of AI technologies in STEM education: A systematic review from 2011 to 2021. *International Journal of STEM Education, 9*, 59.