Journal of Computers Vol. 37 No. 5, October 2026, pp. 153-163
https://doi.org/10.63367/19911599202610370501O

Construction and Empirical Research on Classroom Teaching Evaluation Model of Vocational Education Based on Artificial Intelligence Multimodal Data Fusion

Chunxia Lan*

School of Teacher Training, Ningbo Polytechnic University, Ningbo, 315800, China
1001656@nbpu.edu.cn

Received 1 October 2025; Revised 30 January 2026; Accepted 22 February 2026

Abstract: Against the backdrop of policies promoting quality enhancement and excellence in vocational education alongside educational evaluation reforms, traditional classroom teaching assessments struggle to comprehensively reflect teaching quality due to limitations such as single-dimensional data and insufficient technological empowerment. To address this, this study constructs an AI-driven multimodal data fusion model for classroom teaching evaluation. This model encompasses a five-layer technical architecture—data collection, preprocessing, feature extraction, metric calculation, and result feedback—integrating three types of multi-source heterogeneous information: teaching behavior data (based on DBSCAN clustering analysis), physiological state data (combined with ResNet-18 emotion recognition), and dynamic effect data (using TransE knowledge transfer evaluation). This enables comprehensive and precise diagnosis of the classroom teaching process. Empirical validation was conducted in smart classroom environments using S-T behavioral analysis, engagement curves, and the Rt-Ch model. Results demonstrate the model's effectiveness in identifying critical issues such as imbalanced teaching pacing and fluctuating student engagement, significantly enhancing the comprehensiveness and objectivity of evaluations. This research not only provides a replicable technical pathway for diagnosing vocational education classroom quality but also offers theoretical support and practical examples for the deep integration of artificial intelligence and educational assessment.

Keywords: artificial intelligence, multimodal data fusion, classroom teaching evaluation, vocational education, empirical study

1 Introduction

In 2020, the Central Committee of the Communist Party of China (CPC) and the State Council issued the "Overall Program for Deepening Educational Evaluation Reform in the New Era", which explicitly stated that the process of evaluation should be 'strengthened' and emphasized that "the use of modern information technology, such as artificial intelligence and big data, improves the scientific, professional, and objective nature of educational evaluation." [1]. The Action Plan for Improving the Quality and Excellence of Vocational Education (2020-2023) further calls for the improvement of the "learner-centered" evaluation system and the strengthening of process and comprehensive evaluation [2]. As the "baton" of teaching reform and practice, vocational education classroom evaluation plays a key role in promoting the development of high quality education.

However, the current classroom teaching evaluation is still generally characterized by the problems of "single evaluation dimension" and "insufficient technical empowerment". Specifically, there is a single source of evaluation data, mostly relying on teachers' experience to judge students' participation, ignoring the hidden state of emotion and cognition, which makes it difficult to evaluate the results comprehensively and objectively; at the same time, the evaluation methods are biased towards traditional methods such as homework and tests, which can easily lead to the "streetlight effect" [3], and it is difficult to accurately reflect the real quality of teaching.

With the gradual application of new-generation information technologies such as artificial intelligence, big data, and blockchain in educational assessment, it has become technically feasible to collect and analyze multimodal data—including students' cognitive, affective, and behavioral responses—in real time during classroom instruction. This also opens new pathways to overcome the limitations of traditional evaluation

* Corresponding Author

methods. However, existing research has yet to address the integration of multimodal data in vocational education classroom assessment, particularly lacking comprehensive evaluation models tailored to the practice-oriented, theory-practice integrated nature of vocational education. Therefore, constructing a multidimensional, comprehensive, and process-oriented classroom evaluation system based on intelligent technologies has become a trend in vocational education development, with profound research value. This study aims to address the limitations of traditional evaluation methods in vocational education classrooms—such as single-dimensionality, high subjectivity, and delayed feedback—by developing a multimodal data evaluation model that integrates teaching behaviors, physiological states, and dynamic outcomes. It further seeks to provide a reference for achieving intelligent and precise classroom diagnostics.

2 Related Research

2.1 A Review of Multimodal Learning Analytics and Artificial Intelligence in Educational Assessment

The advancement of artificial intelligence and learning analytics technologies has introduced new paradigms for educational assessment. Multimodal learning analytics achieves deep insights into the learning process by integrating data from diverse channels, such as behavioral logs, physiological signals, and audio-visual recordings [4, 12]. For instance, the multimodal data-driven smart classroom learning analytics model developed by Qianwen Tang et al. (2024) integrates teacher and student behavioral and emotional data, offering novel approaches for evaluating classroom interaction quality [8]. Rui Zhang et al. (2025) proposed the human-centered artificial intelligence (eAI) paradigm, emphasizing technology's role in understanding learner states [4]. Additionally, Jie Jiang et al. (2024) analyzed learning behaviors and multimodal data in smart classrooms, constructing a learning behavior analysis framework based on multimodal data to provide scientific support for developing learning evaluation systems [12].

2.2 Current Status of Classroom Teaching Process Evaluation

The development of artificial intelligence, learning analytics, context-awareness and other information technologies has enriched the means of classroom teaching process evaluation, providing support for multi-dimensional access to teaching process information, dynamic tracking of the progress of the teaching process, multi-methods to analyze the teaching process data, and diversified embodiment of the evaluation and analysis results, so that the comprehensive evaluation, dynamic evaluation, contextual evaluation and diagnostic evaluation of classroom teaching process become possible [4]. At present, the common classroom teaching process evaluation has three aspects: classroom teaching behavior evaluation, classroom teaching status evaluation and classroom teaching effect evaluation.

Current Status of Classroom Teaching Behavior Evaluation Teachers' and students' behavior in classroom teaching is an important part of process evaluation. For example, the Flanders Interaction Analysis System explores the frequency and quality of teacher-student interactions by analyzing the active questioning, passive answering, and use of information technology tools by teachers and students during the teaching process. By analyzing classroom teaching videos, the Hias Classroom Intelligence Evaluation System is able to conduct real-time intelligent analysis of attendance, atmosphere activity, classroom participation, and type of lecture [5]. Lele Zhang et al. (2022) analyzed research on automatically identifying classroom teaching behaviors through multimodal fusion supported by artificial intelligence technology, which breaks through the limitations and developmental challenges of traditional classroom teaching behavior analysis, providing a reference for such analysis [5].

Current Status of Classroom Teaching State Evaluation The cognitive, emotional, attentional, and fatigue states of teachers and students during the implementation of classroom teaching can also provide reference for evaluating the quality of the classroom teaching process. For example, You Yang and other scholars used eye movement experiments to analyze and found that the more complex the teaching video, the more students blink and gaze, the higher the cognitive processing load. Emotional state acquisition method is very much, a variety of wearable devices, video analysis technology and other means for real-time access to teachers and students emotional emotion provides a possibility. Yiyan Wang et al. (2021) noted that research on multimodal data-driven intelligent education tends to capture learners' cognitive and emotional states with greater precision [17].

Current Status of Classroom Teaching Effectiveness Evaluation Compared with the traditional classroom teaching in which teachers use oral questions and answers, distribution of accompanying tests and other methods to evaluate and map the classroom learning effect, classroom teaching in the era of big data can help the formative evaluation of the learning effect through the technology of data statistics, speech recognition, video analysis, text analysis and so on. For example, Xunfei voice evaluation tool can score and locate the problems of reading aloud in Chinese and English, and conduct multi-dimensional evaluation on pronunciation accuracy, fluency, completeness, and phonetics and tonal patterns. FuLan Fan et al. (2025) explored a multimodal learning analysis data integration method, providing insights for integrating multi-source data to evaluate learning outcomes [6].

2.3 Current Status of Multi-Source Heterogeneous Educational Data Collection and Integration Technologies

Currently, education data collection mainly focuses on online teaching and management systems, while data acquisition in physical scenarios such as classroom teaching and outdoor learning is still highly dependent on traditional means such as manual observation, subjective scales and questionnaires, which are highly subjective and have limited coverage. Therefore, building a comprehensive and systematic view of education based on heterogeneous data from multiple sources has become an important foundation for educational big data research and application. However, due to the complexity of the teaching and learning process, there is a lack of effective data aggregation mechanisms across scenarios and systems, and modeling and analysis are mostly limited to a single data source, which makes it difficult for education administrators to make scientific decisions based on integrated data. Weifu Wang et al. (2021) emphasized that multimodal learning analytics represents a new approach to understanding and evaluating authentic learning, but its development is constrained by the depth and breadth of data integration [16].

2.4 Challenges of Traditional Classroom Teaching Evaluation

Traditional classroom evaluation mainly relies on manual observation, questionnaire survey or single platform data, which has four significant defects. First, the evaluation objectives and the teaching process are disconnected from each other, and it is difficult to reflect the real teaching effectiveness [6]. Second, the evaluation means is single, overly relying on paper-and-pencil tests and empirical judgments, favoring the quantitative analysis of externally visible behaviors and neglecting the attention to the implicit dimensions of students' cognition, emotion, and thinking [7]; the limited dimensions of data collection make it difficult to capture the dynamic changes in the complex learning context [9]. Third, the evaluation body is unbalanced, showing the tendency of "unilateralism", with teacher evaluation accounting for as high as 92%, and students' self-assessment and mutual evaluation accounting for only 8%, and most of them are in the form [8]. Fourth, insufficient technical support, evaluation means lagging behind the process of education informatization, mainly relying on manual records and empirical analysis, the lack of automated data collection tools, it is difficult to cope with the demand for massive data processing [11], resulting in inefficient evaluation, unable to adapt to the requirements of the development of intelligent education [10].

In summary, while existing research has extensively explored classroom teaching evaluation methods, it has generally overlooked the integration and application of multi-source heterogeneous data, particularly within the specific context of vocational education. Therefore, this study focuses on vocational education classrooms, concentrating on multi-source data such as teaching behaviors, learning states, and evaluation outcomes. It constructs a classroom teaching quality diagnosis and evaluation model to support precise teaching interventions and the development of smart teaching.

3 Construction of a Classroom Teaching Evaluation Model Based on Multimodal Data Fusion

3.1 Model Design Principles

Goal Orientation The classroom teaching evaluation model of multimodal data fusion needs to be oriented to the core of teaching objectives to ensure that the evaluation system forms a closed loop with curriculum design and learning activities [6]. Jiayi Wu et al. (2026) pointed out that the model needs to construct predictive indicators based on the needs of vocational positions, analyze the correlation between behavioral data and learning performance through machine learning algorithms, and ensure that the evaluation results are directly directed to the degree of achievement of teaching objectives [7]. This goal-oriented design helps to realize the virtuous circle of "promoting teaching and learning through evaluation".

Dynamic Generativity The model needs to have real-time data capture and dynamic feedback capabilities to adapt to the complexity and uncertainty of classroom teaching. Libao Wu et al. (2021) proposed that multimodal data has temporal characteristics, and students' physiological signs (e.g., heart rate, eye movement), behavioral trajectories (e.g., gestures, frequency of interactions) and environmental data (e.g., classroom layout) are captured in real time through wearable devices, smart sensors, and other technologies, forming a continuous data flow [9].

Data Multidimensionality Multimodal data integration needs to break through the limitations of single-dimensional data and integrate cognitive, emotional, behavioral and other multidimensional information in order to respond to the learning effect of students in a more scientific and comprehensive way. The classroom behavior analysis model constructed by Qianwen Tang (2024) contains three dimensions of teacher speech behavior, student activity behavior and teacher and student emotional behavior, and combines facial expression recognition and voice emotion analysis to achieve a comprehensive assessment of the quality of classroom interaction [8]. This multidimensionality design provides technical support for the evaluation of complex abilities [12].

3.2 Model Framework

In this paper, based on the principles of classroom teaching evaluation model design for multimodal data fusion, and combined with its own practice, it constructs a multimodal data classroom teaching evaluation model supported by artificial intelligence, as shown in Fig. 1.

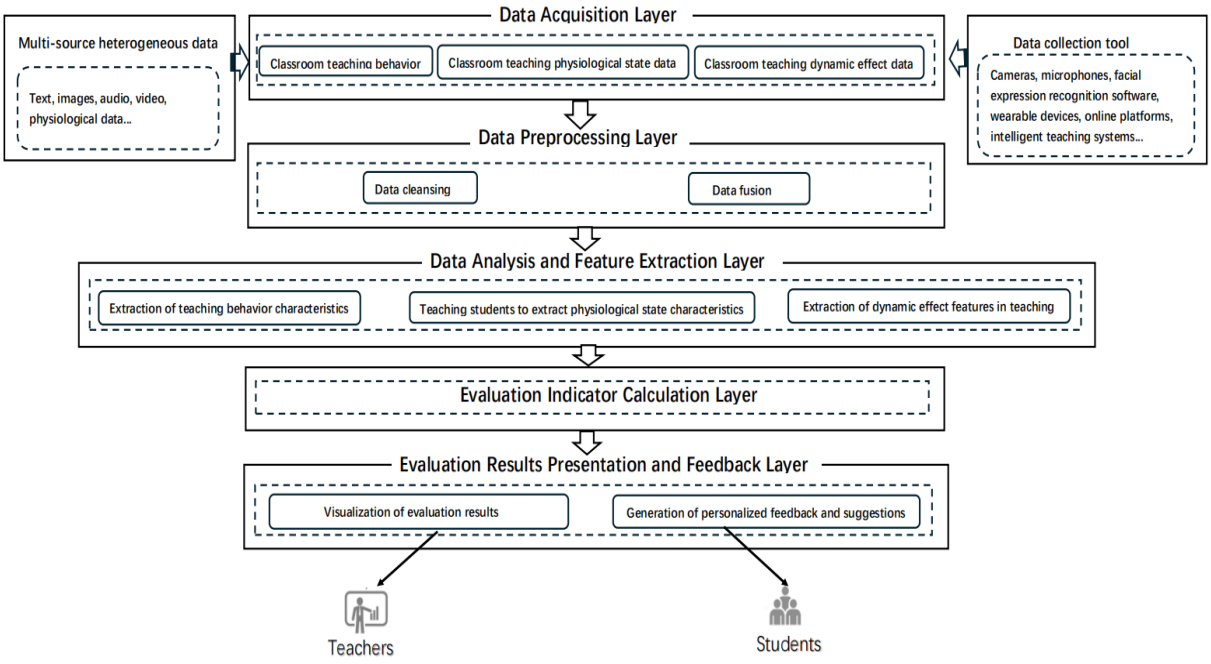


Fig. 1. AI-supported classroom teaching evaluation model based on multimodal data fusion

Data Acquisition Layer Data collection for this study was conducted in a smart classroom environment equipped with an intelligent interactive system. The specific collection process and sources are as follows:

Teaching Behavior Data: Video streams of teachers and students are captured via multiple high-definition cameras (4K ultra-high-definition recording system) installed at the front and rear of classrooms. Intelligent interactive systems (e.g., multifunctional nano smart blackboards) record interaction logs between teachers and students. Computer vision technology automatically identifies and codes 11 categories of teacher-student behaviors (Teachers: lecturing, demonstrating, writing on the board, media presentation, questioning, patrolling; Students: listening, speaking, thinking, note-taking, assignments, hands-on practice). Physiological State Data: Captures student facial video streams via front-facing high-definition cameras for subsequent ResNet-18-based expression recognition to analyze emotional states. Dynamic Effect Data: Collect student knowledge mastery data through real-time classroom feedback systems (e.g., online quizzes, interactive response devices). Evaluate skill improvement by analyzing videos of student practical exercises or project deliverables. These three data categories collectively form a comprehensive "Behavior-State-Effect" evaluation chain, providing reliable foundations for precise classroom diagnostics and intelligent feedback.

Data Preprocessing Layer In the pre-processing stage of multimodal data, the collected raw data are first cleaned to eliminate noise such as fuzzy image frames, speech recognition error segments and abnormal physiological indicators, and fill in the missing values reasonably; and then data standardization is carried out to unify the data outlines from different sources and formats, such as converting the time of teaching behaviors into minute units and normalizing students' grades, to ensure the consistency and comparability of the data. The data standardization is then implemented to unify the data outlines from different sources and formats, such as converting the time of teaching behaviors into minutes and normalizing the students' grades, to ensure the consistency and comparability of the data, and to lay the foundation for the subsequent feature extraction and modeling.

Data Analysis and Feature Extraction Layer The in-depth analysis of classroom teaching data relies on the efficient extraction and fusion of multimodal features. In this study, we design refined algorithmic implementation paths for three types of data, namely, teaching behaviors, physiological states and dynamic effects, to ensure the accuracy and richness of feature characterization.

Feature Extraction from Teaching Behaviors Teaching method diversity features, teaching rhythm features, and IT application depth features were extracted from the teacher behavior data. For student behavior data, focus features, participation features, and learning behavior normative features are extracted.

Teaching Method Diversity Analysis, a DBSCAN density clustering algorithm was used to resolve the pattern characteristics of teacher behavior sequences. Teacher behaviors are first encoded as time-series unique heat vectors, and a joint distance matrix containing time intervals and behavioral differences is constructed:

$$\mathrm{d}(t_i - t_j) = a \cdot \left| t_i - t_j \right| + (1-a) \cdot \left\| x_i - x_j \right\|_2 . \tag{1}$$

In Formula 1, a is the weighting coefficient, and x_i represents the behavior vector at time t_i. By setting the density threshold ℓ and the minimum number of samples MinPts, high-frequency behavior patterns (such as "lecture + media presentation") are identified, and the diversity of teaching methods is quantified through the pattern occurrence frequency $F_m = N_m / N$ and duration T_m, providing a basis for optimizing teaching strategies [13].

Dynamic modeling of teaching pace, introducing the ARIMA time series model to capture the temporal characteristics of knowledge point explanations. Perform d-order differentiation on the original sequence $\{y_t\}$ until it becomes stationary. Determine the autoregressive order p and moving average order q based on the autocorrelation function (ACF) and partial autocorrelation function (PACF), and construct the $ARIMA(p, d, q)$ model:

$$\Phi(B)^d y_t = \Theta(B) \ell_t . \tag{2}$$

In Formula 2, $\Phi(B) = 1 - \varnothing_1 B - ... - \varnothing_p B^p$ is the autoregressive operator, and $\Theta(B) = 1 - \theta_1 B - ... - \theta_q B^q$ is the moving average operator. Through the model, the time allocation for teaching activities is predicted, and abnormal rhythm points (such as overly lengthy theoretical lectures leading to decreased engagement) are identified.

Teaching students to extract physiological state characteristics Based on the students' expression and emotion data, extract the positive emotion ratio features, emotion stability features and synchronization features between

emotion and teaching content. For facial expression emotion recognition, a real-time emotion classification model based on ResNet-18 is constructed. Input 32×32 grayscale face image, extract deep features by 4 residual blocks, output 7 types of emotion probability distribution $P = [p_1, p_2, \ldots, p_7]$ through global average pooling and fully connected layer, and calculate emotion time share and fluctuation standard deviation. The formula for calculating the emotion time share is:

$$\rho_k = C_k / C . \tag{3}$$

In Formula 3, ρ_k represents the percentage of time that the kth emotion (e.g., one of the seven emotion categories, such as pleasure, boredom, etc.) spends in the entire classroom video time period. C_K represents the cumulative number of frames in which the kth emotion appears in the classroom video. It reflects the frequency of the emotion in the classroom, the more frames, the more frequently the emotion occurs. C is the total number of frames in the classroom video when the emotion is counted.

Extraction of teaching dynamic effect features From the knowledge mastery and skill enhancement data, extract the knowledge mastery degree feature, skill enhancement speed feature and knowledge transfer and application feature. Based on the thinking ability development data, obtain the logical thinking clarity feature, critical thinking intensity feature and innovative thinking activity feature.

For example, for knowledge transfer ability assessment, the TransE algorithm is used to construct a knowledge point association model. Knowledge points are modeled as graph structures, and entities and relations are embedded by minimizing the energy function learning [14]:

$$L = \sum_{(h,r,t)\in T} \sum_{(h',r,t')\in T^-} \left[\gamma + d(h+r,t) - d(h'+r,t')\right]_+ . \tag{4}$$

In Formula 4, T^- is the negative sample set, $d(\cdot)$ is the L2 distance, and γ is the boundary parameter. The cosine similarity of knowledge point vectors in the extended task is calculated, and transfer behavior with a similarity > 0.7 is considered valid. The transfer frequency and accuracy are then statistically analyzed [15].

For the "skill improvement rate" metric designed specifically for vocational education, the assessment method is as follows: By comparing indicators such as completion time, error rate, and operational fluency across different practical tasks (e.g., basic operations, comprehensive applications), the slope of skill proficiency improvement is calculated. For example, skill improvement rate V_s can be expressed as V_s = (S_post - S_pre) / T, where S_pre and S_post represent skill scores from the pre-test and post-test respectively, and T denotes the time interval between the two assessments.

Evaluation Indicator Calculation Layer *Constructing an Evaluation Indicator System Weight Allocation Model* This study employs the Analytic Hierarchy Process (AHP) to determine the weights of each evaluation dimension and specific indicators. The specific implementation process is as follows: First, a panel of evaluators comprising educational technology experts, vocational education specialists, and frontline teachers was assembled. Through pairwise comparison judgment matrices, the importance levels of each dimension (teaching behavior, physiological state, dynamic effect) and their internal indicators were determined. Subsequently, the eigenvectors of the judgment matrices were calculated to establish the weights, followed by a consistency check (CR < 0.1) to ensure the scientific rigor and objectivity of the weighting.

Calculate the comprehensive evaluation index score Based on the weighting results and extracted feature values, calculate the comprehensive evaluation index score for classroom teaching quality. The calculation formula is as follows:

$$S = \sum_{j=1}^{m} (w_j \times \sum_{i=1}^{n_j} w_{ij} \times x_{ij}) . \tag{5}$$

In Formula 5, S represents the comprehensive score for classroom teaching quality ($0 \le S \le 1$), m denotes the number of evaluation dimensions (in this model, m = 3, corresponding to teaching behavior, physiological state, and dynamic effects), w_j denotes the weight of the jth evaluation dimension ($\sum w_j = 1$), n_j denotes the number of specific indicators under the jth dimension, w_{ij} denotes the weight of the ith indicator under the jth dimension ($\sum w_j = 1$), x_{ij} denotes the standardized feature value of the i-th indicator under the j-th dimension ($0 \le x_{ij} \le 1$).

Result Presentation and Feedback Layer Present classroom evaluation results in the form of visual charts

and graphs to teachers, instructional administrators, and students. Based on the evaluation results, it generates personalized teaching improvement suggestions for teachers. Provide students with learning feedback to help them understand their performance in classroom learning. Through the above evaluation model, the quality of classroom teaching in the context of artificial intelligence can be evaluated comprehensively, objectively and accurately using multi-source heterogeneous data, and provide powerful support and guidance for teaching improvement and student learning.

4 Empirical Validation of the Model

To validate the model's effectiveness, this study selected a second-year applied chemistry class (42 students total, comprising 30 males and 12 females, with an average age of 19.5 years) at a vocational college as the research subject for a case study. It is hoped that the AI-based multimodal data classroom evaluation model can diagnose both teachers' instructional practices and students' learning outcomes, thereby providing reference for teachers to engage in instructional reflection and improve teaching practices.

4.1 Evaluation Method

In this study, the S-T (Student-Teacher) analysis method will be used to analyze the model application process. This method records the temporal distribution of teacher and student behaviors in the classroom through coding and constructs S-T curves for visual analysis, thus revealing the essential characteristics of the teaching process [18].

4.2 Evaluation Tools

This study applies an advanced smart classroom equipped with an intelligent interactive system, including a multifunctional nano-smart blackboard that supports 4K UHD display and 20-point touch control, a dual-screen interactive system, and a 4K UHD recording and broadcasting system. The front and back of the classroom are equipped with two HD resolution cameras, which can track and record the classroom teaching scene in 360 degrees in real time. The use of deep learning video analytics technology can dynamically recognize the faces and expressions of teachers and students, and finely classify the classroom teaching and learning interaction behaviors of teachers and students.

4.3 Evaluation Process

Data Collection In this study utilizes multiple high-definition cameras installed at the front and rear of classrooms to capture video and images of teachers and students during lessons. Audio data is collected via microphones, while teaching behavior data and student interaction data are gathered through an intelligent interactive system. The research primarily records 11 categories of teacher and student behaviors. Teacher behaviors include: lecturing, demonstrating, writing on the board, media presentations, questioning, and circulating among students. Student behaviors encompass: listening, speaking, thinking, note-taking, completing assignments, and practical exercises. For the three evaluation modes—teaching behavior, physiological state, and dynamic effect—the specific objective metrics adopted in this empirical study are described as follows: Teaching Behavior Mode: Evaluation metrics include teacher behavior frequency distribution (%), teacher-student behavior conversion rate (Ch value), and teacher behavior dominance rate (Rt value). Physiological State Mode: Evaluation metrics primarily comprise the proportion of overall positive emotions in the classroom (ρ_positive, see Formula 3) identified via ResNet-18, and the standard deviation of emotional fluctuations. Dynamic Effect Pattern: Evaluation metrics include knowledge transfer effectiveness (based on TransE, proportion of transfers with similarity >0.7) and skill improvement rate.

Data Visualization The classroom analysis reports provided by the system mainly include classroom teacher and student teaching behavior analysis report (Fig. 2), classroom participation data analysis graph (Fig. 3), and S-T and Rt-Ch analysis report (Fig. 4). The classroom teacher-student teaching behavior analysis report presents the frequency distribution of 11 classroom teacher-student behaviors in both teacher and student dimensions,

and the classroom engagement data analysis graph as well as the S-T and Rt-Ch analysis report can visually reflect the trend of students' classroom participation. The case shown in this paper is an information technology application course in applied chemical engineering.

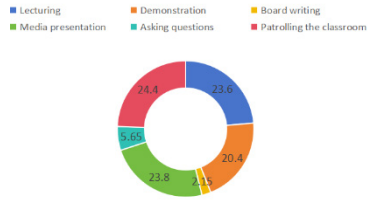
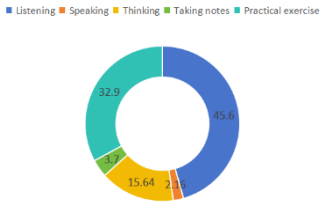
Observation dimension	Behavior	Proportion (%)	Visualization legend
Teacher behaviors	Lecturing	23.6	
	Demonstrating	20.4	
	Board writing	2.15	
	Media presentations	23.8	
	Asking questions	5.65	
	Patrolling the classroom	24.4	
Student behaviors	Listening	45.6	
	Speaking	2.16	
	Thinking	15.64	
	Taking notes	3.7	
	Practical exercises	32.9	

Fig. 2. Statistical chart of teacher and student behavior in the classroom

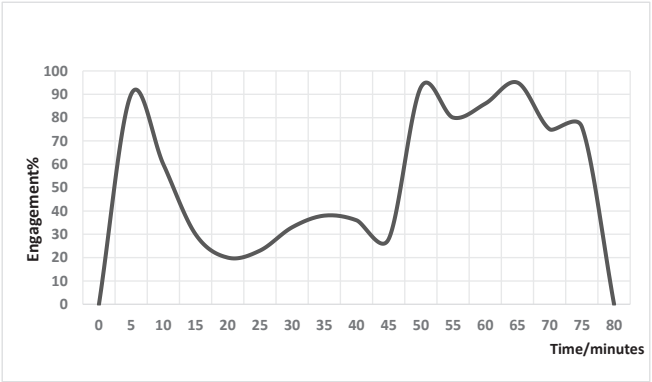


Fig. 3. Curve chart of student participation in class

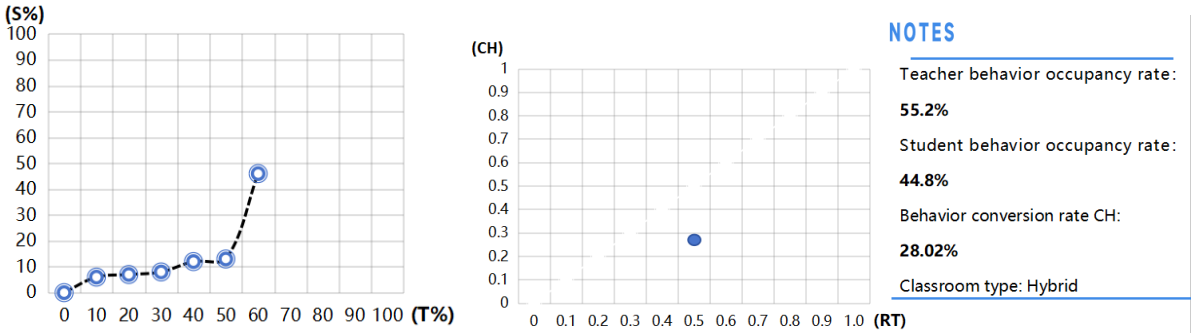


Fig. 4. S-T curve and RT-CH graph

Data Analysis and Interpretation A team of teachers with many years of rich teaching experience, consisting of school-level master teachers and school-level supervisors, analyzes and interprets the data and charts output by the machine to ensure the accuracy of data interpretation.

Fig. 2 shows that 71.8% of the teachers' behaviors in the categories of media demonstration (23.8%), inspections (24.4%), and lectures (23.6%) indicate that the teachers used the mixed teaching strategy of "demonstration + classroom management + lectures". However, 69.95% of the teacher's behavior was used for lecturing, demonstration, board and presentation, while teacher-student interaction and students' practical exercises accounted for 30.05%, indicating that the classroom did not well reflect the teaching concept of "student-centered", and that the time for teacher-student interaction and students' practical exercises was relatively small, so it is recommended to improve students' participation in the classroom and increase the time for practical exercises. Fig. 3 shows that the participation rate peaked at 90% in the start-up phase of the course, dropped to 20% in the decay phase, and then rose to 93% in the hands-on practice. It is recommended that more project-task-based teaching methods should be used in vocational education classrooms to improve students' classroom engagement. Fig. 4 is the S-T curve and RT-CH graph shows that the first half of the classroom S-T curve shows a "jagged" flat fluctuations, basically based on the teacher's lecture, demonstration, etc., student participation is not high, about 45 minutes or so the curve appears to be a clear turnaround, the students' behavior changed abruptly, corresponding to the practical session, reflecting the characteristics of the practical course. The Rt and Ch values can be used to determine the classroom teaching mode, as shown in Table 1. In the RT-CH diagram, the RT value of the teacher's behavioral occupancy rate is 55.2%, the student behavioral occupancy rate is 44.8%, the behavioral conversion rate is about Ch=28.02%, and the Rt value is between 0.5 and 0.75, which is judged to be a blended classroom teaching mode, which is basically compatible with the nature of the information technology application course. It is recommended to increase the Ch value to more than 60% to improve the interactivity of the classroom.

Table 1. Criteria for evaluating classroom teaching models

Teaching model	Rt (Teacher behavior percentage)	Ch (Teacher-Student transition rate)
Lecture-based	≥75%	≤20%
Blended	50%-75%	20%-60%
Interactive	30%-50%	60%-80%
Inquiry-based	<30%	>80%

4.4 Comparative Analysis with Mainstream Research

Compared with the current mainstream research, this study has obvious advantages in the depth of multimodal data integration, methodological innovation and adaptability to VET scenarios: we break through the limitations of unimodal analysis, and construct a three-dimensional integration framework of "teaching behavior-physiological state-dynamic effect" [16-17]; we innovatively integrate DBSCAN clustering, ResNet-18 emotion recognition and TransE knowledge migration assessment. It breaks through the limitations of unimodal analysis and constructs a three-dimensional integration framework of "teaching behavior-physiological state-dynamic effect" [16-17]; innovatively integrates DBSCAN clustering, ResNet-18 emotion recognition and TransE knowledge migration assessment, realizing the whole chain of analysis methods from behavioral sequences analysis, emotion modeling to migration ability assessment, which significantly improves the evaluation accuracy and objectivity and overcomes the coarse-grained, subjective and one-sided nature of traditional S-T analysis, questionnaires and performance evaluation [18]; and closely adheres to the "integration of science and practice" of vocational education. In line with the characteristics of vocational education, we design vocational indicators such as "speed of skill enhancement", and construct a closed loop of "collection-diagnosis-intervention" through real-time collection and dynamic feedback. Through real-time collection and dynamic feedback, we construct a "collection-diagnosis-intervention" closed-loop, effectively solving the problem of static evaluation lag [19], and realizing real-time and intelligent diagnosis of classroom teaching.

5 Conclusion

This study, grounded in the policy direction of educational assessment reform and the practical demands of vocational classroom teaching, constructs an AI-based multimodal data fusion model for classroom teaching evaluation. By incorporating algorithms such as DBSCAN, ResNet-18, and TransE, it overcomes the limitations of traditional single-dimensional assessment, achieving cross-modal integration and intelligent analysis of teaching behaviors, physiological states, and learning outcomes. By providing visual, multi-source data-driven classroom diagnostic reports (e.g., teacher-student behavior distributions, engagement curves, Rt-Ch analysis), the model helps educators precisely identify specific issues such as imbalanced teaching pacing, fluctuating student engagement, and skill acquisition gaps. This empowers evidence-based teaching reflection and targeted improvements (e.g., increasing hands-on sessions, adopting project-based learning). Empirical results demonstrate that the model effectively enhances the comprehensiveness and objectivity of evaluations in vocational education practical courses, providing technological support for the integration of teaching, learning, and assessment.

However, the study still has several limitations: First, although the data collection dimensions cover behavior, emotion, and outcomes, they do not incorporate deeper cognitive physiological data such as eye-tracking and EEG signals, resulting in an incomplete quantitative representation of teacher-student cognitive interactions [20]. Second, while the model weight allocation employs the AHP method, expert judgments still involve a degree of subjectivity. Future research aims to explore dynamic optimization strategies like deep reinforcement learning to enhance objectivity. Furthermore, this empirical study is currently limited to vocational education practical courses. The model's universality and effectiveness in theoretical courses or across different disciplinary contexts require cross-scenario validation [21]. Additionally, the model's reliance on specific hardware environments like smart classrooms somewhat restricts its widespread application. This study provides an empirical example of technology-enabled educational evaluation. Future research is anticipated to achieve new breakthroughs in data fusion depth, model generalization capabilities, and hardware universality.

Acknowledgement

This research was supported by the following project: the 2025 Key Project of the Zhejiang Provincial Education Science Planning Program, "A Study on the Evaluation of Vocational Education Classroom Instruction Based on Multi-source Heterogeneous Data in the Context of Artificial Intelligence" (2025SB094).

References

[1] Z.J. Wang, S. Long, J. Zhang, Research on a Human-AI Collaborative Hierarchical Model for Intelligent Classroom Teaching Evaluation, Journal of Distance Education 43(5)(2025) 32-40. https://doi.org/10.15881/j.cnki.cn33-1304/g4.2025.05.004

[2] Y. Lin, Y.Q. Tian, J.R. Wang, Logic, Dilemma and Realization of Digital Transformation of Classroom Teaching Assessment,Journal of Teacher Education 12(2)(2025) 39-47. https://doi.org/10.13718/j.cnki.jsjy.2025.02.005

[3] B.Q. Liu, The Core Values of AI-enabled Classroom Transformation: Wisdom Generation and Model Innovation, Open Education Research 28(4)(2022) 42-49. https://doi.org/10.13966/j.cnki.kfjyyj.2022.04.005

[4] R. Zhang, Y.P. Xie, Value Representation, Logical Framework and Practical Approaches of Classroom Teaching Empowered by Digitization, Modern Education Management (4)(2025) 79-89. https://doi.org/10.16697/j.1674-5485.2025.04.007

[5] L.L. Zhang, X.Q. Gu, Analysis of Classroom Teaching Behavior Supported by Multimodal Data: Model Construction and Practical Framework, Open Education Research 28(6)(2022) 101-110. https://doi.org/10.13966/j.cnki.kfjyyj.2022.06.011

[6] F. Fan, X. Wang, L. Mei, Y. Li, X. Nie, A Research on the Impact of Emotional Feedback on Collaborative Learning Engagement Supported by Multimodal Data, Chinese Journal of Distance Education 45(11)(2025) 95-115. https://doi.org/10.13541/j.cnki.chinade.2025.11.005

[7] J.Y. Wu, T.Y. Li, Paradigm Shift and Practical Reconstruction of Higher Education Assessment Systems in the Age of Artificial Intelligence, Open Education Research 32(5)(2026) 80-88.

https://doi.org/10.13966/j.cnki.kfjyyj.2026.05.009
[8] Q.W. Tang, H. Zhang, Y.A. Wu, Modeling a Multimodal Learning Analysis in the Context of Smart Classroom, Journal of Teacher Education 11(5)(2024) 49-58.
https://doi.org/10.13718/j.cnki.jsjy.2024.05.006
[9] L.B. Wu, Y.N. Cao, Y.M. Cao, Reform and Practical Paths of Classroom Teaching Evaluation Under Artificial Intelligence, China Educational Technology (5)(2021) 94-101.
https://doi.org/10.3969/j.issn.1006-9860.2021.05.013
[10] H.T. Long, Educational Assessment Reform in AI Era: Opportunities, Challenges and Path Selection, Journal of China Examinations (11)(2021) 10-18+34.
https://doi.org/10.19360/j.cnki.11-3303/g4.2021.11.002
[11] J.L. Zhang, M. Yang, G.F. Tan, Intelligent Technology Empowers Basic Education Evaluation Reform: the Practical Dilemma and Path Choice, Chinese Journal of Distance Education 43(2)(2023) 18-27.
https://doi.org/10.13541/j.cnki.chinade.2023.02.002
[12] J. Jiang, W.T. Yu, H.Y. Wang, Research on Student Learning Behavior in Smart Classroom Based on Multimodal Data, Chinese Journal of ICT in Education 30(4)(2024) 107-117.
https://doi.org/10.3969/j.issn.1673-8454.2024.04.012
[13] B. Cheng, J. Zhu, M. Guo, MultiJAF: Multi-modal joint entity alignment framework for multi-modal knowledge graph, Neurocomputing 500(2022) 581-591.
https://doi.org/10.1016/j.neucom.2022.05.058
[14] Z.C. Wang, Q.S. Lv, X.H. Lan, Y. Zhang, Cross-lingual knowledge graph alignment via graph convolutional networks, in: Proc. 2018 the AAAI Conference on Artificial Intelligence, 2018.
https://doi.org/10.18653/v1/D18-1032
[15] S. Vashishth, S. Sanyal, V. Nitin, P. Talukdar, Composition-based Multi-Relational Graph Convolutional Networks, in: Proc. 2020 The Eighth International Conference on Learning Representations, 2020.
https://openreview.net/forum?id=BylA_C4tPr
[16] W.F. Wang, M.J. Mao, Multimodal Learning Analytics: A New Approach to Understand and Evaluate Authentic Learning, e-Education Research 42(2)(2021) 25-32.
https://doi.org/10.13811/j.cnki.eer.2021.02.004
[17] Y.Y. Wang, Y.C.X. Wang, Y.H. Zheng, Multimodal Learning Analytics: a New Research Trend in Intelligence Education Driven by Multimodality, China Educational Technology (3)(2021) 88-96.
https://doi.org/10.3969/j.issn.1006-9860.2021.03.013
[18] Y.J. Shang, Using the S-T Analytical Method to Analysis the Classroom Instruction in Educational Technology Undergraduate Courses, Modern Educational Technology (10)(2008) 29-31.
https://doi.org/10.3969/j.issn.1009-8097.2008.10.007
[19] D. Wu, Q. Guo, L.K. Wu, H. Chen, Artificial General Intelligence to Reshape the Reform of Educational Evaluation, Open Education Research 29(4)(2023) 4-10.
https://doi.org/10.13966/j.cnki.kfjyyj.2023.04.001
[20] Z.T. Zhu, Z.M. Han, C.Q. Huang, Educational Artificial Intelligence (eAI): A New Paradigm of Human-centered Artificial Intelligence, e-Education Research 42(1)(2021) 5-15.
https://doi.org/10.13811/j.cnki.eer.2021.01.001
[21] Y. Song, C.L. Xu, X.X. Mu, Research on New Classroom Teaching Evaluation and Optimization Enabled by Generative Artificial Intelligence, Modern Educational Technology 34(12)(2024) 27-36.
https://doi.org/10.3969/j.issn.1009-8097.2024.12.003