ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 2 - 2026

Construction and application exploration of multi-dimensional dynamic evaluation Model of teaching effectiveness under the background of digital transformation of Higher Education

Meng Lu[1] *and Yaru Feng*[2,*]

[1] Qufu Normal University, Qufu, 276800, Shandong, China
[2] Neusoft Institute Guangdong, Foshan, 528225, Guangdong, China

SUMMARY: *In order to improve the accuracy, continuity and interpretation ability of teaching effectiveness evaluation under the background of digital transformation of higher education, this paper proposed a multi-dimensional dynamic evaluation model construction and application method. Based on the logs of learning management platform, classroom interaction records, homework performance, stage test results and text feedback, the multi-source teaching data were standardized, feature extraction and correlation modeling, and clustering analysis and machine learning classification were combined to identify teaching status, so as to alleviate the deviation caused by static, single and unbalanced samples in traditional evaluation. On this basis, the evaluation index system was constructed around teaching design, process implementation, learning participation, result achievement and feedback support. The fuzzy comprehensive evaluation and dynamic weight adjustment mechanism were used to realize the phased measurement and continuous update of teaching effectiveness. The application results showed that the comprehensive scores of course A and course B increased from 71.8 points and 70.9 points to 87.4 points and 85.3 points respectively, and the single evaluation time was controlled at 1.84 minutes and 1.91 minutes respectively, which was significantly lower than 3.96-4.37 minutes of the control method. The evaluation errors are 4.2% and 4.5%, respectively, which are also significantly lower than the control methods. The results show that the method in this paper can clearly reflect the change of teaching status, and has better performance in evaluation efficiency, stability and reliability of results, which can provide computational support for curriculum diagnosis, teaching improvement and digital teaching governance in colleges and universities.*

KEYWORDS: *Digital transformation of higher education; Teaching effectiveness evaluation; Multidimensional dynamic model; Machine learning*

1 Introduction

Under the background of the continuous advancement of the digital transformation of higher education, teaching activities are no longer limited to the single scene of classroom teaching, and multiple types of data such as course platform access, online tests, discussion and interaction, assignment submission, learning trajectory stay, text feedback and classroom behavior records are continuously accumulated, which makes the evaluation of teaching effectiveness gradually shift from empirical judgment to data support [1-5]. As an important basis for measuring the quality of curriculum implementation and the effect of talent training,

*yarufeng1120@163.com
https://doi.org/10.65102/is2026746
Published: 30 April, 2026

teaching effectiveness is not only related to the direction of teachers 'teaching improvement, but also directly affects the course optimization, resource allocation and the fine level of teaching governance [6-9]. However, from the reality, there are still some problems in the existing teaching effect evaluation in colleges and universities, such as scattered evaluation caliber, fragmented index dimension, insufficient utilization of process data, and weak interpretation of results. Especially under the condition of the coexistence of multi-source heterogeneous data, different evaluation subjects have different understandings of the importance of indicators, and some teaching performances have obvious stage volatility and context dependence, which makes the evaluation results easy to present static, fuzzy and one-sided characteristics [10, 11].

The existing research has carried out a lot of exploration around the intelligent teaching evaluation. Starting from the driving factors of digital transformation and the structure of teaching ability, some studies have constructed the digital evaluation framework of university education, which provides a macro perspective for the identification of teaching quality [12-15]. Some studies have also combined learning analysis, online formative evaluation and prediction models, and used platform log variables to identify students' learning status and teaching support effect, and achieved good results in feedback efficiency and early warning ability [16-19]. At the same time, the application research of artificial intelligence in higher education shows that machine learning, multimodal analysis and visual dashboards can improve the automation level and process traceability of teaching evaluation to a certain extent [20]. However, on the whole, the existing methods still focus on a single course result, a single time point score or local behavior characteristics, and the discussion on the dynamic changes in the teaching process, the coupling relationship between dimensions, and the mechanism of adaptive adjustment of evaluation weights with teaching status is still insufficient. When the amount of data is enlarged, the evaluation noise, feature redundancy and timing distortion will further weaken the stability and credibility of the results.

Based on the above understanding, this paper proposes a multi-dimensional dynamic evaluation model of teaching effectiveness for the digital transformation scenario of higher education. Based on the teaching process data, learning behavior data, interactive feedback data and result performance data, the model constructs a comprehensive evaluation framework that can reflect the continuous changes of teaching activities through the methods of data standardization, feature correlation modeling, teaching state recognition, machine learning classification and dynamic weight adjustment. The research goal is not only to obtain a static score, but also to try to describe the formation process and fluctuation law of teaching effectiveness on the basis of multi-dimensional data fusion, so as to provide a more computationally supported analysis path for curriculum diagnosis, teaching improvement and digital governance in colleges and universities.

2 Intelligent processing of teaching effectiveness evaluation data under the background of digital transformation of higher education

With the continuous advancement of the digital transformation of higher education, the data sources of teaching effect evaluation have expanded from the traditional final questionnaire and manual lecture records to multiple dimensions such as learning management platform logs, classroom interactive terminals, online homework systems, teaching video platforms, course forums and text feedback systems. The data from different sources are not consistent in sampling frequency, structure type and semantic granularity, including structured information such as attendance, test scores, and resource access times, as well as semi-structured or unstructured information such as classroom discussion texts, course feedback, and learning reflection. If a single outcome index is still used to judge the teaching effectiveness, it is not only difficult to identify the stage fluctuations in the teaching process, but also difficult to depict the linkage relationship between teacher engagement, student participation and platform support. Based on this, this paper takes multi-source data collection, standardization processing, feature correlation modeling and teaching state clustering analysis as the pre-links of dynamic evaluation of teaching effect, so as to form a unified data basis for subsequent model calls. Its processing flow is shown in Figure 1.

Figure 1: Flow chart of data collection and fusion of multi-source teaching effect evaluation

2.1 Multi-source data collection and fusion for teaching effect evaluation

The fusion of multi-source teaching data is not a simple concatenation of different tables, but a calculation process involving time alignment, field mapping, noise filtering and semantic unification. Let the set of course samples be:

$$\mathcal{D} = \{d_1, d_2, \dots, d_n\} \tag{1}$$

Here, d_i denotes the ith teaching unit or course sample. For any teaching sample, its original observation can be expressed as:

$$d_i = [x_i^{(l)}, x_i^{(c)}, x_i^{(a)}, x_i^{(t)}] \tag{2}$$

where, $x_i^{(l)}$ represents the learning platform log characteristics, $x_i^{(c)}$ represents the classroom interaction characteristics, $x_i^{(a)}$ represents the assignment and test characteristics, and $x_i^{(t)}$ represents the text feedback characteristics. Because the recording scales of the four types of data are different, the platform log emphasizes more frequency and length of stay, the classroom interaction emphasizes more immediate response, and the text feedback carries strong semantic information. Therefore, the reconstruction needs to be completed under a unified identifier and a unified time window. In this paper, the data is organized with the three-level time granularity of "week-units-course", and the teacher's teaching behavior, student participation behavior and result performance behavior are mapped to the same evaluation object, so as to reduce information distortion caused by cross-system sampling bias.

2.1.1 Data standardization in the teaching process

After obtaining the original data, it is necessary to standardize the teaching process data to avoid dimensional differences affecting the subsequent analysis results. Let the original feature matrix be:

$$X = [x_{ij}]_{n\times m} \tag{3}$$

Here, n is the number of samples, m is the feature dimension, and x_{ij} represents the observation value of the ith sample on the JTH index. For continuous features, we use mean-standard deviation normalization:

$$z_{ij} = \frac{x_{ij} - \mu_j}{\sigma_j} \tag{4}$$

Here, μ_j and σ_j represent the mean and standard deviation of the JTH feature, respectively. Considering that some teaching behavior variables have obvious boundary constraints, such as class attendance rate, completion rate, interactive response rate, etc., after standardization, they are mapped to [0,1] interval by interval compression:

$$\hat{x}_{ij} = \frac{z_{ij} - \min(z_j)}{\max(z_j) - \min(z_j)} \tag{5}$$

For missing values, if the missing ratio is low, the mean value of adjacent time Windows is used to fill in. If the missing has the characteristics of phased aggregation, the local imputation is performed by combining the course weekly and the similar sample distribution. After word segmentation, stop words cleaning and sentiment polarity coding, the text data are transformed into numerical variables such as text participation, feedback positivity and question focus. After the above processing, the teaching process data from different sources are mapped into a unified feature space, which lays the foundation for subsequent correlation analysis and pattern recognition.

2.1.2 Construction of multi-dimensional teaching feature correlation matrix

The formation of teaching effect has significant coupling. The change of teachers 'teaching rhythm may affect students 'online review frequency, the density of classroom questions may change the activity of the discussion area, and the timeliness of homework feedback will further affect subsequent learning engagement. Therefore, interpreting teaching effectiveness in terms of a single indicator tends to weaken the true structure. In order to depict the correlation relationship between multi-dimensional teaching features, this paper constructs the feature correlation matrix R:

$$R = [r_{pq}]_{m\times m} \tag{6}$$

Here, r_{pq} represents the strength of association between the PTH feature and the QTH feature. Considering the existing linear relationship of teaching data and the existence of nonlinear collaboration, this paper synthetically adopts the weighted form of Pearson correlation coefficient and cosine similarity in the implementation:

$$r_{pq} = \alpha \cdot \rho_{pq} + (1-\alpha)\cdot \cos(x_p, x_q) \tag{7}$$

where, ρ_{pq} is the Pearson correlation coefficient, $\cos(x_p, x_q)$ is the cosine similarity of the feature vectors, and $\alpha \in [0,1]$ is the balance parameter. The matrix can reveal the propagation chain between resource access, classroom interaction, assignment performance and feedback text, and can also be used to identify redundant variables and key driving variables. In the subsequent dynamic evaluation model, the correlation matrix will also be an important basis for adjusting the weight of indicators, so that the evaluation results no longer stay in the sum of isolated indicators, but can reflect the internal structural linkage of teaching activities.

2.1.3 Teaching status identification and cluster analysis

After the unified coding of multi-source features is completed, this paper further identifies the teaching status of different stages through cluster analysis. Due to the fuziness and volatility of the teaching process itself, the same course may show multiple states such as "high interaction-high achievement", "high input - low feedback" or "low participation - high differentiation" in different weeks. If we directly use a fixed threshold to determine, it is easy to ignore the intermediate transition interval. Based on this, this paper uses the idea of fuzzy clustering to divide the teaching samples. Let the normalized feature vector be v_i and the cluster center be c_k, then the membership degree of the sample to the KTH class state can be expressed as follows.

$$u_{ik} = \frac{1}{\sum_{h=1}^{K} \left(\frac{||v_i - c_k||}{||v_i - c_h||} \right)^{\frac{2}{m_f - 1}}} \tag{8}$$

where K is the number of cluster categories and m_f is the fuzzy coefficient. By updating the clustering center iteratively, the dynamic distribution structure of teaching status can be obtained.

According to the clustering results, this paper preliminarily identifies the teaching samples as four kinds of states: "stable-promoting type", "process fluctuation type", "result lag type" and "insufficient support type", and combines the time window to form a dynamic transfer diagram. Figure 2 illustrates the basic structure of the clustering of teaching status. It can be seen that the teaching effect is not fixed, but constantly moving under the combined effect of resource supply, classroom organization, student response and feedback quality. This state recognition result can provide an explanation basis for the subsequent multi-dimensional dynamic evaluation, and also make the teaching improvement suggestions more targeted.

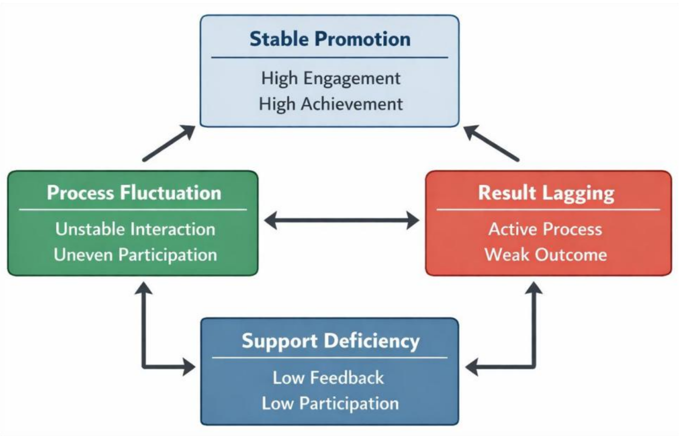

Figure 2: Schematic diagram of teaching status recognition and dynamic clustering

2.2 Modeling and analysis of teaching effect evaluation data

Due to the significant continuity, hierarchy and multi-agent coupling characteristics of the teaching process, after the completion of multi-source data collection, standardized mapping and preliminary clustering of teaching status, it is necessary to further model and analyze the teaching effect evaluation data. If the course effectiveness is judged only by a single indicator or simple weighted results, it is often difficult to identify teaching situations with mixed characteristics such as "high process investment but lagging learning achievement" and "active interaction but insufficient knowledge internalization". Especially in the digital teaching scene, course platform logs, classroom interaction records, homework behavior trajectories and text feedback information exist at the same time, the data dimension is high, the correlation between variables is complex, and the number of samples of different categories may also present unbalanced distribution. Based on this, this paper further carries out feature extraction and machine learning classification and recognition of teaching effect evaluation on the basis of the above data processing results, in order to form the computational support for multi-dimensional dynamic evaluation model.

2.2.1 Feature extraction method for teaching effect evaluation

The goal of feature extraction for teaching effect evaluation is not to mechanically compress the original data dimension, but to transform the heterogeneous signals scattered in different sources into a unified representation that is comparable, computable and input to the model while retaining the key information of the teaching process. Let the original observation of the ith course sample on the TTH teaching period be:

$$x_i^{(t)} = [l_i^{(t)}, c_i^{(t)}, a_i^{(t)}, f_i^{(t)}] \tag{9}$$

Among them, $l_i^{(t)}$ represents the characteristics of platform log, $c_i^{(t)}$ represents the characteristics of classroom interaction, $a_i^{(t)}$ represents the characteristics of assignments and tests, $f_i^{(t)}$ represents the characteristics of text feedback and evaluation. In order to reduce the interference of single time point fluctuation on the evaluation results, this paper uses the time window aggregation method to encode the phased observation:

$$\bar{x}_i = \frac{1}{T}\sum_{t=1}^{T} x_i^{(t)} \tag{10}$$

where, T is the observation time window length and $\bar{x}_i$ is the aggregated feature vector of the course sample in the current evaluation period. Considering the inconsistent representation ability of different dimensional features on teaching effectiveness, this paper introduces adaptive weight coefficient in the process of feature fusion to construct a comprehensive teaching representation vector:

$$z_i = [\alpha_1 l_i, \alpha_2 c_i, \alpha_3 a_i, \alpha_4 f_i] \tag{11}$$

Here,$\alpha_1, \alpha_2, \alpha_3, \alpha_4$ represent the fusion weights of the four types of features, respectively, and satisfy:

$$\sum_{k=1}^{4} \alpha_k = 1, \quad \alpha_k \geq 0 \tag{12}$$

The initial weight value is determined by the feature variance contribution rate and the correlation strength, and is fine-tuned according to the classification effect in the subsequent model training. The significance of this treatment is that although the frequency of platform access can reflect learning engagement, its explanatory power is not stable if it is separated from classroom response and result achievement. Textual feedback, although informative, is noisy and also needs to be co-used with structured features to avoid excessive amplification of episodic opinions.

In order to make the feature composition clearer, this paper organizes the input variables of teaching effect evaluation into four types of core feature groups as shown in Table 1. This division not only corresponds to the main data sources in the digital teaching environment, but also facilitates the hierarchical identification of subsequent models.

Table 1: Composition and meaning of characteristics of teaching effect evaluation

Feature Category	Main Variables	Computational Meaning
Platform Behavior Features	Number of resource accesses, video replay rate, and online dwell time	Reflect learning engagement intensity and the level of resource utilization
Classroom Interaction Features	Question response rate, discussion participation, and sign-in stability	Reflect the quality of classroom participation and the state of real-time interaction
Academic Performance Features	Assignment completion rate, quiz scores, and stage achievement level	Reflect the degree of knowledge mastery and the level of outcome output
Feedback Text Features	Sentiment polarity score, problem focus degree, and feedback density	Reflect learning experience, cognitive confusion, and teaching acceptance

On this basis, this paper further uses the combination of principal component retention and correlation screening to remove redundant features and retain core variables with high discrimination for teaching effect. After processing, the input features not only retain the multi-dimensional information of the teaching process, but also reduce the interference of high-dimensional noise on the classification boundary.

2.2.2 Classification and recognition of teaching effect data based on Machine learning

Because the teaching effect data has the characteristics of high dimension, nonlinearity and fuzzy class boundary, support vector machine (SVM) is selected as the basic method of classification and recognition in this paper. This method has good generalization performance in small samples, high-dimensional features and complex boundary scenes, and can distinguish teaching status and teaching effectiveness levels more stably. Let the training sample set be:

$$\mathcal{S} = \{(z_i, y_i)\}_{i=1}^{N} \tag{13}$$

where z_i is the teaching representation vector after feature extraction, and y_i is the class label of teaching effect. In this paper, the teaching effect is divided into four categories: "significant improvement", "stable and forming", "process fluctuation" and "inefficient early warning", in order to meet the needs of hierarchical identification in digital teaching evaluation. Under the condition of linear non-separability, the SVM maps the input into a high-dimensional feature space through a kernel function, and its decision function can be expressed as:

$$f(z) = \operatorname{sgn}\left(\sum_{i=1}^{N} \lambda_i y_i K(z_i, z) + b\right) \tag{14}$$

where λ_i is the Lagrange multiplier, b is the bias term, and $K(z_i, z)$ is the kernel function. The number of "inefficient early warning" and "process fluctuation" in the teaching effect samples is usually less than that of the conventional samples. If the standard classifier is directly used, the majority class dominance phenomenon is easy to occur, which makes the identification of key abnormal samples insufficient. To this end, this paper introduces a category weighting mechanism to set differential penalty coefficients for different categories, and its optimization objective can be written as follows.

$$\min \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N} \omega_{y_i} \xi_i \tag{15}$$

Here, C is the penalty parameter, ξ_i is the slack variable, and ω_{y_i} is the class weight. After assigning higher weights to the minority classes, the model can enhance the recognition sensitivity of teaching risk states while maintaining the overall accuracy. Based on the above ideas, the teaching effect recognition process of "feature input-kernel mapping-interval optimization-category output" is formed in this paper, as shown in Figure 3.

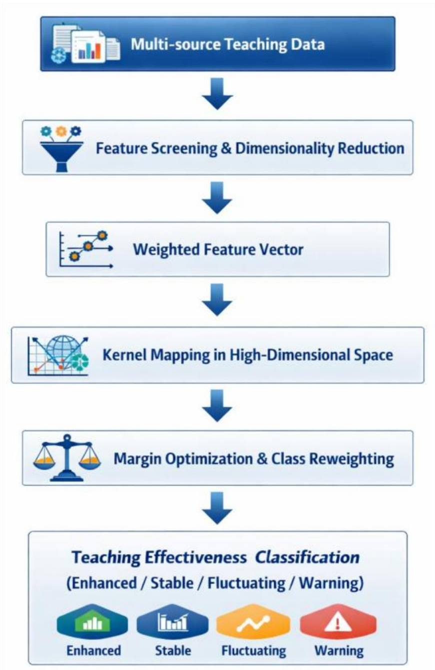

Figure 3: Flowchart of classification and recognition of teaching effect based on machine learning

In the model output stage, the classification results are no longer regarded as isolated labels, but are cross-checked with the teaching state clustering results of the previous section. If a course sample is identified as "inefficient early warning" in the feature space, and is in the "insufficient support" area for a long time in the clustering results, it means that there are persistent shortcomings in the resource supply, interactive feedback or learning support mechanism of the course. If the classification result changes from "process fluctuation type" to "stable forming type" in the adjacent period, it means that the teaching intervention has produced a positive effect. Therefore, the classification model not only undertakes the recognition task, but also becomes an important calculation module for judging the direction of teaching change in the dynamic evaluation model.

3 The design of multi-dimensional dynamic evaluation model of teaching effect under the background of digital transformation of higher education

3.1 Concept of multi-dimensional dynamic evaluation

The digital transformation of higher education does not only change the presentation way of teaching resources, but also the deeper change is that teaching activities themselves begin to be recorded and analyzed in the form of continuous data streams, and gradually enter the computable evaluation framework. In this context, the evaluation of teaching effectiveness should no longer stay in the static judgment of one-time questionnaire, final score or a small amount of manual observation, but should turn to a multi-dimensional dynamic evaluation concept that is oriented to the whole process, covers multi-subjects, and can adaptively update with the change of teaching status. This concept emphasizes that the object of evaluation is not an isolated result, but an evolutionary process composed of "instructional design, learning participation, interactive feedback, and result achievement". The goal of evaluation is not only to give high and low grades, but also to identify the formation path, fluctuation nodes and subsequent intervention links of teaching effectiveness.

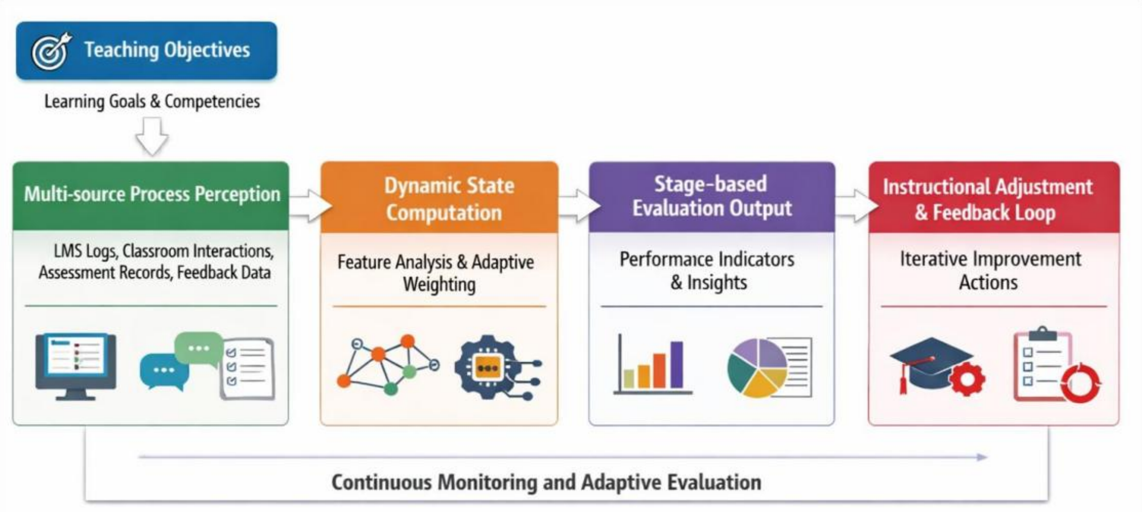

Figure 4: Framework diagram of multi-dimensional dynamic evaluation concept

From the structural point of view, the concept of multi-dimensional dynamic evaluation can be summarized into three interconnected parts, namely, goal orientation, process perception and dynamic revision, and its overall logical relationship is shown in Figure 4. The goal-oriented decision evaluation is not to summarize data without difference, but to focus on the achievement of curriculum goals, the generation of students 'ability and the effectiveness of teaching support. Process awareness emphasizes the continuous identification of teaching operation status by using multi-source data such as learning platform logs, classroom interaction records, homework completion trajectories and text feedback. The dynamic correction requires that the evaluation results can be updated continuously according to the new data in the time window, and the periodic anomalies are not misjudged as the overall conclusion, and the short-term high scores are not allowed to cover up the long-term insufficient investment. Goal orientation provides constraint boundaries for evaluation, process awareness provides data basis for evaluation, and dynamic revision makes evaluation have the ability to adapt to real teaching situations. In the level of calculation realization, this paper expresses the comprehensive evaluation value of teaching effectiveness at time t as:

$$E^{(t)} = \sum_{k=1}^{m} w_k^{(t)} x_k^{(t)} \tag{16}$$

Here, $x_k^{(t)}$ represents the observed value of the KTH evaluation index at time t, $w_k^{(t)}$ represents the dynamic weight of this index at the current stage, and is satisfied:

$$\sum_{k=1}^{m} w_k^{(t)} = 1, \quad w_k^{(t)} \geq 0 \tag{17}$$

This expression shows that teaching effectiveness is not a mechanical sum under fixed weight, but a phased comprehensive result adjusted with the change of teaching status. Teachers, students and the platform system are not separated from each other in this process: teachers provide teaching organization and feedback regulation, students generate learning behaviors and response signals, and the platform assumes the intermediary functions of data collection, state calculation and visual output. Therefore, multi-dimensional dynamic evaluation is not only the update of evaluation technology, but also an extension of digital teaching governance logic.

3.2 Design of multi-dimensional dynamic evaluation model for teaching effect

After the standardization of teaching process data, multi-dimensional feature correlation modeling, and machine learning classification and identification, the focus of evaluation has shifted from "can identify teaching differences" to "how to characterize the continuous changes of teaching effectiveness in a stable and interpretable way". Teaching activities under the background of digital transformation are not represented by one-way input and single result output, but a dynamic system composed of teacher design, classroom organization, student participation, platform support, feedback correction and learning output. If the fixed weight, static threshold and single summary score are still used as the main evaluation basis, it is difficult to reflect the stage fluctuations of teaching effectiveness, and it is more difficult to explain the differences in the formation path of different courses under the same result score. Based on this, this paper introduces a data-driven dynamic weight adjustment mechanism based on the idea of fuzzy comprehensive evaluation, and constructs a multi-dimensional dynamic evaluation model for digital teaching scenarios of higher education.

3.2.1 Principles of teaching effect evaluation

The establishment of multi-dimensional dynamic evaluation model of teaching effect needs to meet the standardization requirements of educational evaluation and the realizability requirements of computational model. Combined with the data characteristics and application goals in the digital teaching environment, this paper puts forward the following principles.

(1) Goal-oriented principle. The ultimate purpose of evaluation activities is not to form a simple ranking result, but to serve teaching improvement, curriculum optimization and digital governance. Therefore, the index design must focus on the achievement of course objectives, the generation of students 'ability and the effectiveness of teaching support, to avoid deviating from the teaching ontology and falling into data stacking.

(2) The principle of integrity. Teaching effectiveness is not determined by a certain kind of data alone, but the result of teaching design, process operation, learning participation, result output and feedback improvement. The evaluation system should cover key links, focus on process as well as results, and avoid replacing overall teaching performance with final scores.

(3) Principle of relative independence. A reasonable distinction should be maintained between different indicators to prevent the repeated inclusion of highly correlated variables such as "online time", "resource access" and "video replay rate", which will cause the evaluation results to be biased towards a certain dimension and weaken the explanatory power of the model.

(4) Principle of dynamic adaptation. In the digital teaching environment, the teaching state has obvious stages and fluctuations. The performance focus of the same course is not the same in the early stage of the course, the task concentration period and the summary consolidation period. The evaluation model must allow the index weight to be adjusted with the change of the data state, rather than staying rigid for a long time.

(5) Computability principle. Each index must have clear data sources and operable calculation methods, and can be directly obtained or indirectly derived from the learning platform, classroom system, homework system and feedback text processing module, so as to minimize the error diffusion caused by pure subjective judgment.

(6) Principle of interpretability. The results should not only provide an overall score, but also be able to trace back to specific dimensions and key indicators, explaining why the score changed, where the problem occurred, and how to intervene. For teaching governance, the evaluation score without explanation support is often difficult to form effective feedback.

3.2.2 Construction of evaluation index system

According to the above principles, this paper divides the teaching effect evaluation index system into three levels: target level, first-level dimension level and second-level indicator level. The target level was set as "the comprehensive level of teaching effectiveness under the background of digital transformation". The first-level dimension extracts five core parts from the teaching operation chain, which are teaching design preparation, teaching process implementation, learning participation state, learning results achievement and feedback support improvement. The secondary indicators are carried out according to the data that can be collected by the platform and the characteristics of the teaching situation, forming a complete multi-dimensional evaluation structure. Its overall framework is shown in Figure 5.

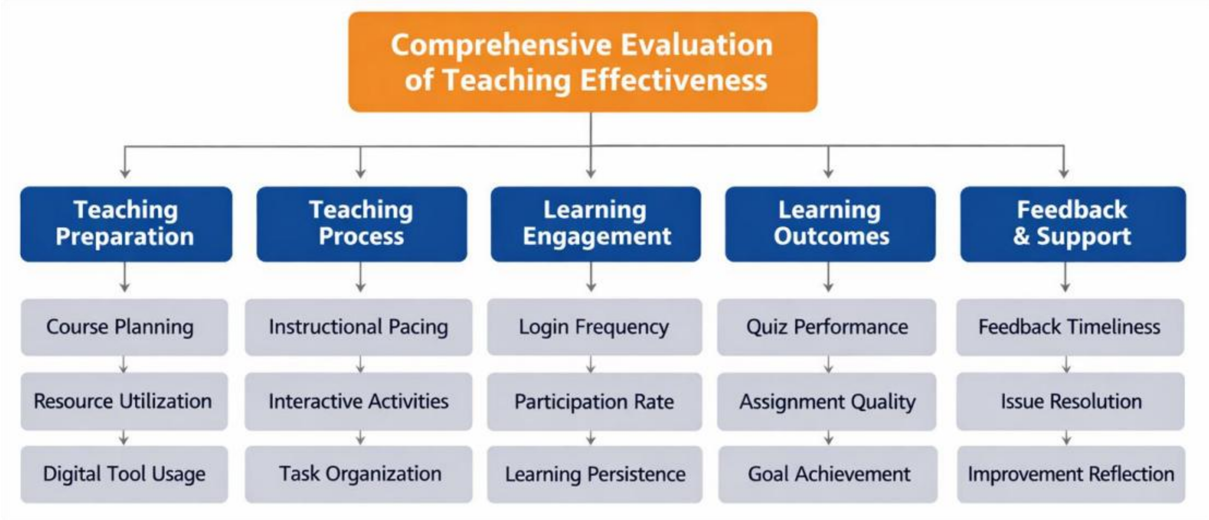

Figure 5: Architecture of multi-dimensional dynamic evaluation index for teaching effect

In the specific construction, the first-level dimensions and second-level indicators are set as shown in Table 2. Different from traditional classroom evaluation, this paper incorporates indicators such as "digital resource matching degree", "platform interactive response rate", "learning continuity" and "feedback closed-loop completion rate" into the system, so that the evaluation object extends from classroom fragments to the whole process of teaching ecology.

Table 2: Multi-dimensional dynamic evaluation index system of teaching effect

Target Layer	Primary Dimension	Secondary Indicators	Main Data Sources
Comprehensive Level of Teaching Effectiveness	D1 Teaching Design and Preparation	Clarity of teaching objectives, resource matching degree, and adaptability of digital tool usage	Course syllabus, platform resource repository, and instructional design records
	D2 Teaching Process Implementation	Rationality of pace control, intensity of classroom interaction, and coherence of task organization	Smart classroom logs, discussion records, and classroom behavior data
	D3 Learning Participation Status	Login activity, discussion participation rate, on-time assignment submission rate, and learning continuity	LMS logs, forum data, and assignment system
	D4 Learning Outcome Achievement	Quiz scores, stage goal attainment rate, and stability of knowledge mastery	Testing system, assignment grading, and course stage results
	D5 Feedback Support and Improvement	Feedback timeliness, issue response rate, degree of reflective revision, and support closed-loop completion rate	Teacher feedback records, text comments, and improvement logs

In mathematical expression, let the one-level dimension factor set of the evaluation object be $U = \{U_1, U_2, U_3, U_4, U_5\}$, where U_1 to U_5 correspond to the above five one-level dimensions respectively. Let the set of evaluation grades be $V = \{v_1, v_2, v_3, v_4, v_5\}$, whose meaning is "excellent, good, medium, weak, and low" in turn. The advantage of this hierarchical design is that it can not only give a comprehensive judgment on the total amount, but also carry out difference analysis at the dimension level, so that the model output has a traceable structural interpretation.

3.2.3 Construction of multi-dimensional dynamic evaluation model

In essence, the evaluation of teaching effect is a comprehensive judgment of multiple fuzzy variables and continuous variables. The performance of different indicators does not always directly map to a clear grade. For example, a high frequency of class discussion does not necessarily mean high quality interaction; Increased learning time does not necessarily lead to improved ability. Based on this feature, the combined model of "fuzzy comprehensive evaluation + machine learning state recognition" is used in this paper. The former is used to complete the structured summary of multi-dimensional indicators, and the latter is used to identify stage states and provide a basis for weight updating. Let the ith first-level dimension contain n_i second-level indicators to form a single-factor evaluation matrix in the t-th evaluation period:

$$R_i^{(t)} = \begin{bmatrix} r_{11}^{(t)} & r_{12}^{(t)} & \cdots & r_{15}^{(t)} \\ r_{21}^{(t)} & r_{22}^{(t)} & \cdots & r_{25}^{(t)} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n_i1}^{(t)} & r_{n_i2}^{(t)} & \cdots & r_{n_i5}^{(t)} \end{bmatrix} \tag{18}$$

Among them, $r_{pq}^{(t)}$ represents the membership degree of the PTH second-level index to the level v_q in the time period, and satisfies that the sum of the elements in each row is 1. Let the weight vector of the secondary index corresponding to this dimension be $W_i^{(t)}$, then the fuzzy synthesis result of this dimension be

$$B_i^{(t)} = W_i^{(t)} R_i^{(t)} \tag{19}$$

When all five level-1 dimensions have been calculated, the upper evaluation matrix can be formed

$$B^{(t)} = \begin{bmatrix} B_1^{(t)} \\ B_2^{(t)} \\ B_3^{(t)} \\ B_4^{(t)} \\ B_5^{(t)} \end{bmatrix} \tag{20}$$

Combined with the one-level dimension weight vector $W^{(t)}$, the overall evaluation result $C^{(t)} = W^{(t)} B^{(t)}$ is obtained. In order to transform the fuzzy results into comparable numerical output, the grade assignment vector Q=[95,85,75,65,50] is set in this paper, and the comprehensive score of the t evaluation cycle can be expressed as $E^{(t)} = C^{(t)} Q^T$. This score not only retains the ability of fuzzy evaluation to accommodate complex teaching states, but also can directly serve the course comparison, stage monitoring and follow-up empirical analysis.

At the same time, this paper embeds the teaching state recognition results in the previous article into the running process of the model. Let the course status label obtained by the machine learning classifier be $S^{(t)}$, and its results can be divided into categories such as "stability promoting type, process fluctuation type, result lag type, and insufficient support type". When $S^{(t)}$ changes, the model does not immediately reconstruct all indicators, but activates the corresponding dynamic weight adjustment rules, so as to make the comprehensive evaluation closer to the current teaching state. The overall operation process is shown in Figure 6.

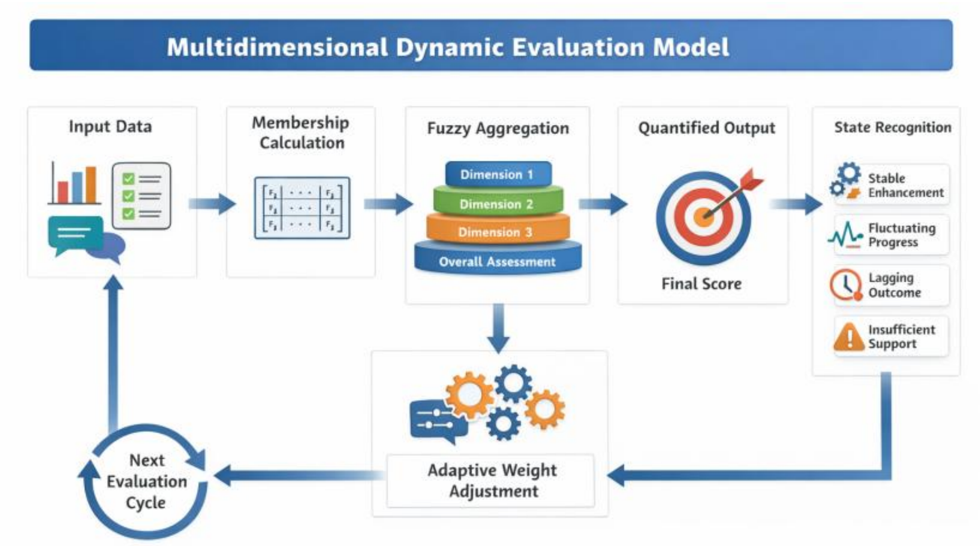

Figure 6: Operation process of multi-dimensional dynamic evaluation model of teaching effect

3.2.4 Evaluation index weight and dynamic adjustment mechanism

In the conventional comprehensive evaluation model, the weights are usually determined once at the beginning of modeling and kept unchanged for a long time. However, the course status in the digital teaching environment is time-sensitive. In some stages, more attention should be paid to the process interaction, and in some stages, more attention should be paid to the achievement of results and the closed loop of feedback. If we continue to use the fixed weight model, it will weaken the dynamic response ability of the evaluation. Based on this, this paper takes "basic weight + state correction" as the overall idea. The basic weight is determined by expert judgment, data variance contribution and correlation analysis. The status updating is completed according to the teaching status, the fluctuation degree of indicators and the importance of features in the current time window. Let the dynamic importance of the KTH first-level index at time period t be $\phi_k^{(t)}$, which is defined as

$$\phi_k^{(t)} = \alpha I_k^{(t)} + (1-\alpha)\left|x_k^{(t)} - x_k^{(t-1)}\right| \tag{21}$$

Here, $I_k^{(t)}$ represents the feature importance of the current stage extracted by the machine learning model, $\left|x_k^{(t)} - x_k^{(t-1)}\right|$ represents the fluctuation amplitude of this indicator in adjacent periods, and α is the balance parameter. Thus, the new weight update formula can be obtained

$$w_k^{(t+1)} = \lambda w_k^{(t)} + (1-\lambda)\frac{\phi_k^{(t)}}{\sum_{j=1}^{5}\phi_j^{(t)}} \tag{22}$$

Here, λ is the historical weight retention coefficient. This formula means that if the explanatory power of a dimension is significantly enhanced in the current stage, and its fluctuations have strong indicative significance for the change of teaching status, its weight will be moderately increased in the next cycle. On the contrary, the downward trend is maintained, so as to avoid a single index dominating the comprehensive evaluation results for a long time. The underlying weights of the first-level dimensions determined in this paper and their dynamic adjustment directions are shown in Table 3.

Table 3: First-level dimension base weights and dynamic adjustment mechanism

Primary Dimension	Base Weight	Main Trigger Conditions	Adjustment Direction
D1 Teaching Design and Preparation	0.18	Delayed resource updates, reduced tool adaptability, and misalignment between objectives and tasks	Moderately increased to identify front-end design deficiencies
D2 Teaching Process Implementation	0.24	Increased fluctuations in classroom interaction, imbalance in pacing, and breakdowns in task organization	Significantly increased to reflect process instability
D3 Learning Participation Status	0.21	Decline in login activity, weakened continuity, and increased submission delays	Increased to strengthen participation monitoring
D4 Learning Outcome Achievement	0.22	Intensified divergence in quiz performance, unfinished stage goals, and decreased stability	Increased to highlight outcome deviation
D5 Feedback Support and Improvement	0.15	Prolonged feedback delays, insufficient issue response, and reduced closed-loop completion rate	Increased at specific stages to diagnose inadequate support

In terms of structure, D2 and D4 had slightly higher basic weights, because the implementation of teaching process and the achievement of learning results usually constituted the core support of course effect judgment. The initial weight of D5 is relatively low, but when the system identifies the state of "insufficient support type" or "result lag type", its weight will be significantly elevated to capture the impact of teacher feedback and support strategies on subsequent improvement. This design not only avoids the model's excessive dependence on one-time experience empowerment, but also enables the evaluation results to respond to the state changes in the real teaching scene.

4 Evaluation of application effect research

4.1 Scheme Design

In order to verify the application effect of the multi-dimensional dynamic evaluation model constructed in this paper in the digital transformation scenario of higher education, this study uses the process tracking scheme in the real curriculum environment to comprehensively test the stability, discrimination and teaching interpretation ability of the model output. The research object selects two undergraduate courses that implement digital teaching reform in A university, denoted as course A and Course B respectively. Both courses rely on learning management platform, smart classroom system and online assignment module to carry out teaching, and the teaching cycle is 16 weeks, which includes resource learning, classroom interaction, stage testing and feedback correction. There are 48 students in course A and 46 students in course B. The teaching content belongs to different disciplines, but the teaching organization form, platform support conditions and evaluation process are consistent, which can meet the needs of horizontal comparison and vertical tracking of models. The experimental running environment is Intel Core i7-12700 processor, 32 GB RAM, and the operating system is Ubuntu 22.04. The model training and evaluation module is implemented based on Python 3.11.

In the research design, the teaching effectiveness is not understood as a certain score fluctuation, but the continuous change of data in each stage within a semester is regarded as a complete evaluation object. In the specific implementation process, the data is organized with the three-level granularity of "teaching week - learning unit - course as a whole", and the platform access logs, classroom check-in and question and answer records, homework submission, stage test scores, discussion texts and teacher feedback records are summarized into a unified data table, and then input into the multi-dimensional dynamic evaluation model constructed in the previous section. In order to ensure the comparability of evaluation results, all original indicators are processed according to uniform rules to complete missing processing, standardized mapping and time alignment. After that, the evaluation results of the stage were output in a rolling window of two weeks, which was used to observe the state transition and effectiveness change of the course in different teaching periods.

In terms of research methods, this paper used the combination method of "model scoring + result comparison + process explanation" to carry out the application test. On the one hand, the comprehensive scores, dimension scores and status labels of each stage were calculated to analyze the model's ability to identify teaching differences. On the other hand, the model output is compared with the course stage grades, student satisfaction surveys and teachers 'teaching reflection records to test its validity and consistency. The relevant experimental Settings are shown in Table 4.

Table 4: Experimental protocol Settings for evaluating the applied study

Item	Specific Content
Research Subjects	Course A at a certain university (48 students), Course B (46 students)
Study Period	16 weeks
Data Sources	LMS logs, classroom interaction records, online assignments, stage tests, discussion texts, and teacher feedback
Evaluation Granularity	Teaching week, learning unit, and overall course
Time Window	Rolling once every 2 weeks
Core Outputs	Comprehensive score, dimension scores, and teaching status labels
Basis for Comparison	Stage grades, satisfaction questionnaires, and teacher reflection records
Analytical Tools	Python 3.11, Pandas, scikit-learn, and SPSS 27.0

4.2 Results Analysis

In order to test the practical application effect of the multi-dimensional dynamic evaluation model constructed in this paper, the evaluation results of course A and Course B were summarized and analyzed after the 16-week teaching cycle, and the method in this paper was compared with the static weighted evaluation method and BP neural network evaluation method. The analysis content included three aspects: the change of teaching effectiveness grade distribution, the trend of stage comprehensive score, evaluation efficiency and evaluation error, so as to investigate the discrimination ability, operation efficiency and result reliability of the model in real digital teaching scenarios.

From the perspective of the grade distribution of teaching effectiveness, after the implementation of the model, the number of high-level samples in the two courses increased significantly, and the number of low-level samples continued to decrease, indicating that the model could better capture the positive changes in the teaching process. Table 5 presents the grade distribution results of course A and Course B before and after the implementation of the evaluation. Compared with before implementation, the number of "excellent" and "good over" students in course A increased from 11 to 19, and that in course B increased from 9 to 15. At the same time, the combined number of "failed", "poor" and "very poor" grades fell in both courses. Such changes are not only the improvement of results, but also reflect the collaborative improvement between teaching organization, learning participation and feedback support. Especially in the middle and late stage, the platform logs showed that the discussion participation rate of the two courses increased from 62.5% and 60.9% to 81.3% and 78.2%, the on-time submission rate increased from 79.2% and 76.1% to 91.7% and 89.1%, and the resource review rate also increased from 34.6% and 32.8% to 52.4% and 49.7%.

Table 5: Changes in the distribution of teaching effectiveness grades between Course A and Course B

Course	Period	Excellent	Upper-Good	Good	Average	Pass	Fail	Poor	Very Poor
A	Before Implementation	4	7	8	12	9	5	3	0
A	After Implementation	8	11	10	9	6	3	1	0
B	Before Implementation	3	6	7	13	10	4	3	0
B	After Implementation	6	9	11	9	7	3	1	0

Figure 7 further demonstrates the changes in the combined scores of the two courses under the rolling evaluation window. It can be seen that the scores of both course A and Course B show A continuous upward trend, with Course A increasing from 71.8 points in the second week to 87.4 points in the 16th week, and Course B increasing from 70.9 points to 85.3 points. The two curves increased significantly from week 6 to 10, which coincided with the course entering the period of task-intensive and feedback reinforcement, indicating that the dynamic weight adjustment mechanism in the model could more sensitively reflect the promotion effect of process improvement on the overall effectiveness. Compared with the method of only relying on the final score for judgment, this staged output method is more suitable for revealing the time path of the formation of teaching effect.

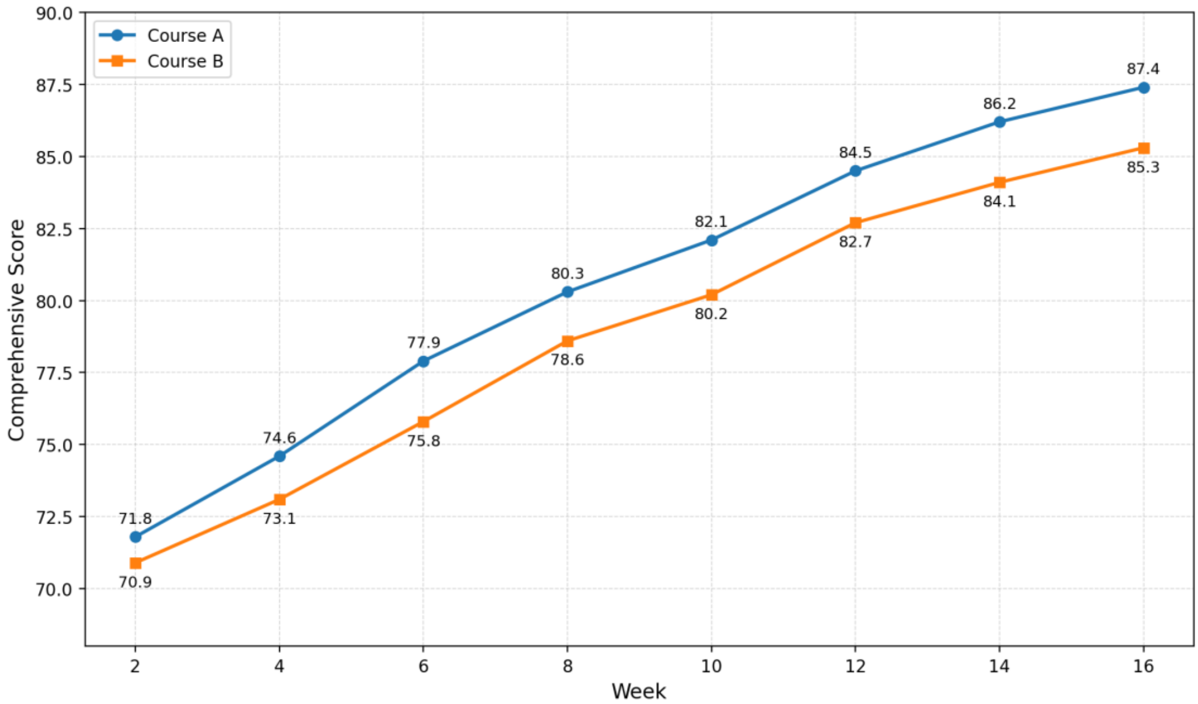

Figure 7: Changes in rolling window comprehensive scores of the two courses

At the level of application effect, this paper further compares the evaluation efficiency of different methods. Figure 8 shows that under the same hardware environment and the same data scale, the proposed method takes significantly less time to complete one course stage evaluation than the two control methods. In course A, the average time of the proposed method was 1.84 min, the static weighted evaluation method was 3.96 min, and the BP neural network evaluation method was 4.28 min. In course B, the corresponding results were 1.91 min, 4.05 min and 4.37 min, respectively. The reason for this difference is that the feature selection and state clustering of the proposed model have been completed in the input stage, and the subsequent evaluation process does not need to repeat the high-cost calculation of all the original variables, but directly updates the results based on the unified feature vector and the rolling time window, so it has higher operating efficiency. This is especially important for digital teaching scenarios that need to continuously monitor the status of the course, because it is difficult to form real effective teaching feedback if the evaluation cannot be returned in time.

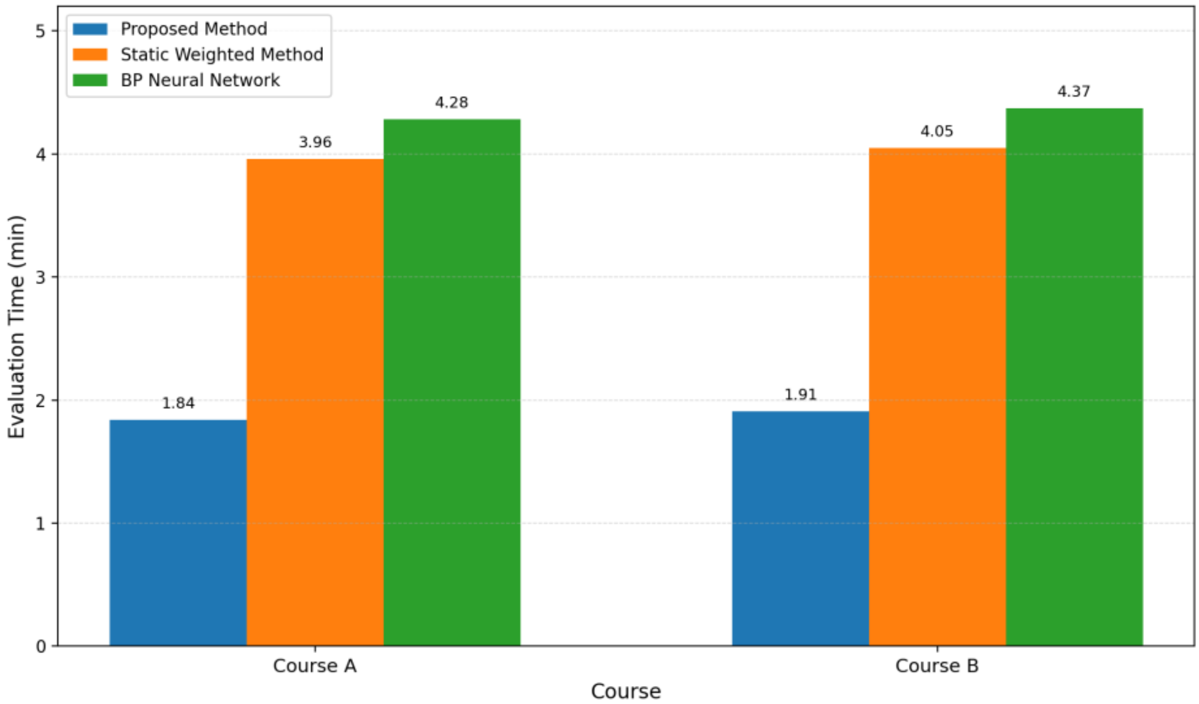

Figure 8: Comparison of evaluation efficiency of different methods

In order to further investigate the reliability of the evaluation results, the average comprehensive rating of three teachers with rich digital teaching experience was used as the reference value, and the normalized absolute error between the model output and the reference value was calculated. Figure 9 indicates that the error of the proposed method is lower than that of the control method on both courses. In course A, the error of the proposed method is 4.2%, the static weighted evaluation method is 10.3%, and the BP neural network evaluation method is 9.1%. In course B, the three are 4.5%, 10.9% and 9.6%, respectively. This result shows that the proposed model does not sacrifice the evaluation accuracy for improving the running speed. On the contrary, because it absorbs the information of multi-source data, fuzzy comprehensive calculation and dynamic weight correction at the same time, it can more accurately approximate the artificial comprehensive judgment. Especially in the "process fluctuation" and "result lag" samples, the dynamic state identification mechanism effectively reduces the common deviation accumulation problem of the static model, and makes the evaluation results more stable and explanatory.

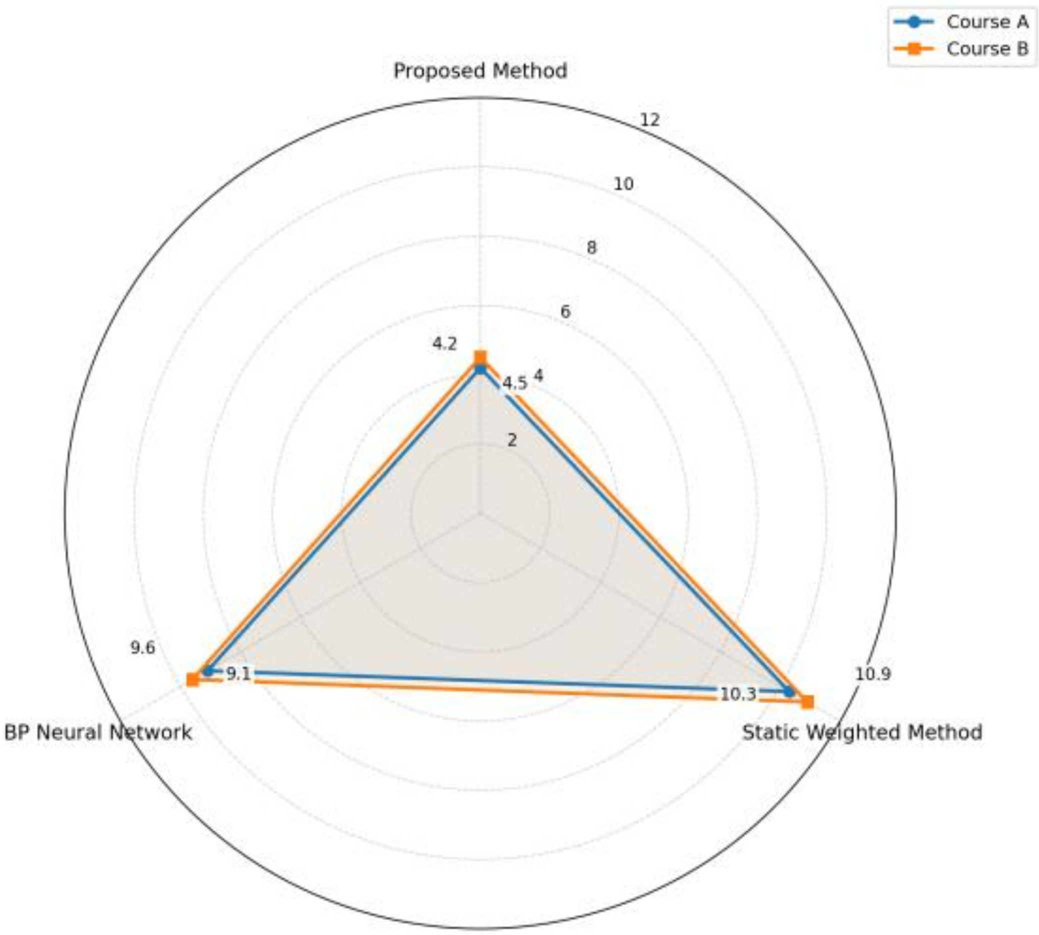

Figure 9: Comparison of evaluation errors of different methods

Synthesizing the above results, it can be seen that the multi-dimensional dynamic evaluation model constructed in this paper shows a good application effect in the real teaching environment. On the one hand, it can reflect the teaching improvement process clearly through the stage score change and grade distribution adjustment. On the other hand, it is superior to the control method in evaluation efficiency and error control, which shows that the model is not only suitable for course summative analysis, but also suitable for embedding digital platform to carry out continuous teaching diagnosis. For the digital transformation of higher education, this evaluation method with real-time, structure and interpretability can better meet the development needs of fine teaching governance than traditional static evaluation.

5 Conclusion

Under the background of the continuous advancement of the digital transformation of higher education, this paper proposes a multi-dimensional dynamic evaluation model construction and application scheme around the problems of staticization, singleness and insufficient explanation in the evaluation of teaching effectiveness. Starting from the standardized processing of teaching process data, this study integrates multi-source data such as learning platform logs, classroom interaction records, homework performance, stage test results and feedback texts into a unified analysis framework. On this basis, feature extraction, state recognition, classification modeling and fuzzy comprehensive evaluation design are completed, and a dynamic weight adjustment mechanism is introduced. So that the evaluation of teaching effectiveness can be continuously updated with the change of teaching status. The application research results show that the proposed method has good feasibility and effectiveness in real digital teaching scenarios. The model can not only clearly identify the stage differences and effectiveness changes in the course operation process, but also show a better level in evaluation efficiency, error control and result consistency. The reason is that the model does not simply understand the teaching effect as the static sum of several scores, but fully considers the multidimensionality, process and fuzziness of teaching activities, and improves the closeness of the evaluation results to the objective teaching reality with the help of machine learning and dynamic calculation mechanism. The significance of this paper is not only to provide a computable model for digital teaching evaluation in colleges and universities, but also to provide a more structurally supported analysis path for curriculum diagnosis, teacher feedback, teaching improvement and teaching governance. For education management departments, the model can provide reference for teaching quality monitoring and resource allocation optimization. For teachers, the output results can help them more accurately locate teaching shortcomings and implement targeted adjustment. In general, this paper provides a realization path with practical application value for the evaluation of higher education teaching effectiveness from experience judgment to data-driven, from static conclusion to dynamic recognition.

About the Author

Meng Lu was born in Qingdao, Shandong, China in 1989. She obtained a master’s degree from the Conservatory of Rueil- Malmaison in France. She is currently working at the School of Music, Qufu Normal University. Her main research interests are educational assessment and music education.

Yaru Feng was born in Zhibo Shandong, China in 1992. She obtained a master’s degree from the Swinburne University of Technology in Australia. She is currently working at the School of Digital Media and Design, Neusoft Institute Guangdong. Her main research interests are digital media art and design, interactive media, new media exhibition design, and digital cultural communication.

References

[1] Mukul E, Büyüközkan G. Digital transformation in education: A systematic review of education 4.0[J]. Technological forecasting and social change, 2023, 194: 122664.

[2] Wang K, Li B, Tian T, et al. Evaluate the drivers for digital transformation in higher education institutions in the era of industry 4.0 based on decision-making method[J]. Journal of Innovation & Knowledge, 2023, 8(3): 100364.

[3] Schleicher A. Toward the digital transformation in education[J]. Frontiers of Digital Education, 2024, 1(1): 4-25.

[4] Zhao B, Zhou J. Research hotspots and trends in digitalization in higher education: A bibliometric analysis[J]. Heliyon, 2024, 10(21).

[5] Chen X, Zou D, Xie H, et al. Technology-enhanced higher education: Text mining and bibliometrics[J]. Heliyon, 2024, 10(4).

[6] Alkandari F, Alsaber A, Al-Kandari A, et al. Enhancing the effectiveness of digital transformation on teaching in higher education in Kuwait[C]//Frontiers in Education. Frontiers Media SA, 2024, 9: 1417062.

[7] Moreira-Choez J S, Gómez Barzola K E, Lamus de Rodríguez T M, et al. Assessment of digital competencies in higher education faculty: A multimodal approach within the framework of artificial intelligence[C]//Frontiers in education. Frontiers Media SA, 2024, 9: 1425487.

[8] Mejías-Acosta A, Regnault D A, Vargas-Cano E, et al. Assessment of digital competencies in higher education students: Development and validation of a measurement scale[C]//Frontiers in Education. Frontiers, 2024, 9: 1497376.

[9] Pelaez-Sanchez I C, Glasserman-Morales L D, Rocha-Feregrino G. Exploring digital competencies in higher education: design and validation of instruments for the era of Industry 5.0[C]//Frontiers in education. Frontiers Media SA, 2024, 9: 1415800.

[10] Zhao L, Xu P, Chen Y, et al. A literature review of the research on students' evaluation of teaching in higher education[J]. Frontiers in Psychology, 2022, 13: 1004487.

[11] Quansah F, Cobbinah A, Asamoah-Gyimah K, et al. Validity of student evaluation of teaching in higher education: a systematic review[C]//Frontiers in Education. Frontiers Media SA, 2024, 9: 1329734.

[12] Banihashem S K, Noroozi O, Van Ginkel S, et al. A systematic review of the role of learning analytics in enhancing feedback practices in higher education[J]. Educational Research Review, 2022, 37: 100489.

[13] Pan Z, Biegley L, Taylor A, et al. A systematic review of learning analytics: Incorporated instructional interventions on learning management systems[J]. Journal of Learning Analytics, 2024, 11(2): 52-72.

[14] Márquez L, Henríquez V, Chevreux H, et al. Adoption of learning analytics in higher education institutions: A systematic literature review[J]. British Journal of Educational Technology, 2024, 55(2): 439-459.

[15] Divjak B, Svetec B, Horvat D, et al. Assessment validity and learning analytics as prerequisites for ensuring student-centred learning design[J]. British journal of educational technology, 2023, 54(1): 313-334.

[16] Wang Q, Mousavi A. Which log variables significantly predict academic achievement? A systematic review and meta-analysis[J]. British Journal of Educational Technology, 2023, 54(1): 142-191.

[17] Bulut O, Gorgun G, Yildirim-Erbasli S N, et al. Standing on the shoulders of giants: Online formative assessments as the foundation for predictive learning analytics models[J]. British Journal of Educational Technology, 2023, 54(1): 19-39.

[18] Yan L, Echeverria V, Jin Y, et al. Evidence-based multimodal learning analytics for feedback and reflection in collaborative learning[J]. British Journal of Educational Technology, 2024, 55(5): 1900-1925.

[19] Masiello I, Mohseni Z, Palma F, et al. A current overview of the use of learning analytics dashboards[J]. Education Sciences, 2024, 14(1): 82.

[20] Castillo-Martínez I M, Flores-Bueno D, Gómez-Puente S M, et al. AI in higher education: A systematic literature review[C]//Frontiers in Education. Frontiers Media SA, 2024, 9: 1391485.