

Research on the Construction of Real-Time Monitoring and Dynamic Evaluation Index System for Teaching Process Supported by Big Data Technology

Liao Guo

Information and Future Technology College, The Open University of Shannxi, Xi'an 710119, Shaanxi, China

Corresponding author: Liao Guo (e-mail: guoliao0021@163.com)

ABSTRACT Current educational evaluation systems often suffer from fragmented data collection, single-dimensional assessment, and delayed feedback, making it difficult to perceive learning processes in real time or provide timely intervention. This study develops a real-time monitoring and dynamic evaluation index system for teaching processes supported by big data technology. The Delphi method is used to select indicators, and the Analytic Hierarchy Process determines weights for a multidimensional model covering learning behavior, cognitive development, affective attitude, and social interaction, with 23 secondary indicators. A Hadoop distributed storage architecture, Spark real-time computing engine, and machine-learning models are used to process large-scale teaching data and identify learning states. Empirical research across three universities and 1,247 students shows that the system achieves 97.3% data collection accuracy under high-concurrency conditions, with video behavior data reaching 98.6% accuracy and text input reaching 95.1%. The system provides precise support for teaching intervention and is applicable to wireless sensing networks, smart classrooms, and electromagnetic-compatible educational monitoring environments.

INDEX TERMS big data technology, teaching monitoring, dynamic evaluation, wireless sensing, smart classroom

I. INTRODUCTION

THE increased rate of informatization of the educational sector is pushing the paradigm shift in the teaching model, and the sheer volumes of data collected by online teaching platforms are offering new prospects in teaching evaluation. Nevertheless, current educational assessment systems suffer from fragmented data collection, one-dimensional evaluation, and significant feedback delays. Data are fragmented in various systems, including attendance systems, homework systems, and online classrooms, without a single combine mechanism, so teachers can hardly have an overall understanding of the learning status of students [1,2]. Assessment systems are overly dependent on mid-term and final examination scores, overlooking behavioral changes, emotional variability and the course of ability development in the learning process. The resulting feedback loop often spans several weeks or even an entire semester. Consequently, by the time instructors identify specific issues, students may have already encountered significant learning obstacles, causing them to miss optimal

windows for intervention. The above issues have a severe negative impact on the specific enhancement of the quality of education and the efficient provision of individualized instruction. The above-mentioned challenges are overcome with the assistance of the maturity of the big data technology. Teaching data at petabyte scales can be processed with distributed storage and real-time computing technologies; machine learning algorithms can be used to learn the patterns of learning in the complex data; cloud computing platforms can be used to provide access to information at large scales and respond to queries with second-level response. This paper will create a real-time tracking and dynamic appraisal index structure of the teaching process using the big data technology, form a scientific multi-dimensional evaluation model, design an effective data processing framework, design a smart warning system, and test the efficiency of the system using empirical research [3,4]. The findings of the research will serve as a foundation of decision making by the education administrators, assist the teachers to modify teaching strategies in a timely manner,

assist students to receive individualized learning directed, and encourage education evaluation based on experience judging to assessment over data, outcome evaluation to process evaluation, and single-dimensional assessment to multi-dimensional assessment.

II. RELATED WORK

The evolution of educational evaluation theory can be traced back to behaviorism, which was then followed by constructivism, where the purpose of evaluation has changed to memorizing knowledge to developing abilities, and the evaluation techniques have evolved whenever it used to be summative to both formative and diagnostic assessments. The assessment models of teaching quality can be input-output models, process-oriented models, and value-added models, which describe the process of teaching effectiveness formation in terms of various perspectives [5,6]. Approaches to building indicator systems include key performance indicators, balanced scorecards, and management by objectives, which offer a theoretical foundation of quantitative assessment. The use of big data technology in education includes learning analytics, educational data mining, and learning behavior modeling. Massive log data produced by MOOC platforms are processed with distributed computing frameworks, behavior tracking in online learning systems is performed with real-time stream processing technology, and unstructured discussion texts and multimedia resources are stored in NoSQL databases. Machine learning algorithms have advanced to predict student performance, learning styles, and learning path, with deep learning models predicting student focus in classroom videos and emotional states.

There are major gaps in existing research. The evaluation index system is not systematic and the majority of the studies are aimed at considering individual dimensions like learning behavior or level of cognition and do not create a complete framework that includes emotional attitudes and social interactions. The design of the technical architecture is inclined to offline analysis, and the ability of the early warning is constrained by the inadequate real-time performance. The algorithm models are weak in interpretability and teachers are unable to comprehend the logical foundation of the prediction results which has an influence on the actual application effect of the system. The studies on the data privacy protection mechanisms are behind schedule and students have not been given enough attention on accepting behavioral data collection [7,8]. This paper will fill this gap by designing a multi-dimensional evaluation index system, creating an architecture of real-time data processing, and creating an intelligent analysis model that is interpretable to offer a systematic answer to the evaluation of education.

III. METHODS

A. CONSTRUCTION OF A MULTIDIMENSIONAL EVALUATION INDEX SYSTEM

All 35 invited participants (15 educational measurement experts with over 10 years of experience, 12 frontline

teachers from key universities, and 8 educational technology researchers) completed all three rounds of the survey, representing a 100% retention rate. The panel consisted of 19 males and 16 females, all holding senior professional titles. The group carried out three sets of questionnaire surveys as a means of selecting indicators. The initial round involved the use of an open-ended questionnaire to gather 62 candidate indicators, with certain dimensions, like classroom participation, homework completion quality, and speed of knowledge acquiring. A five-point Likert scale was used to rank the importance of each indicator in the second round, eliminating those whose mean was less than 3.8 and the coefficient of variation was above 0.25, leaving behind 38 indicators. The third round was used to compute the Kendall coefficient of harmony which was 0.847, which resulted in a consensus among the experts and eventually gave 23 secondary indicators. The weighting calculation used AHP. The judgment matrix for the primary dimensions (Learning Behavior B_1 , Cognitive Development B_2 , Emotional Attitude B_3 , Social Interaction B_4) was constructed as follows: $a_{12} = 1.5$, $a_{13} = 2$, $a_{14} = 2.5$, $a_{23} = 1.5$, $a_{24} = 2$, $a_{34} = 1.5$. Through this matrix, the weights were calculated as 0.356, 0.287, 0.213, and 0.144 respectively. The consistency test demonstrated a maximum eigenvalue $\lambda_{\max} = 4.098$ and a CR value of 0.037, which is within the requirements. The video viewing completion rate, test accuracy rate, and learning engagement time were some of the secondary indicators with weights of 0.089, 0.076, and 0.068 respectively. All indicators were operationalized according to whether the data could be collected; thus, the ability to transfer knowledge was turned into a score rate of the cross-chapter integrated questions, and the level of the learning anxiety was measured as the length of the pauses/number of revisions to submit the assignments.

B. BIG DATA TECHNOLOGY ARCHITECTURE DESIGN

The teaching data stream is processed with the help of a three-tier distributed architecture. The data acquisition layer is implemented as a tracking SDK on the online learning platform, which records data of student behavior (mouse movements, keyboard inputs and page dwell time per second). It is also compatible with the video player API to get operation logs like dragging the progress bar, pausing, and rewinding. The OpenCV library is used to obtain the values of face expression features by the camera, and speech, which is captured by a microphone and converted to text, is used to analyze sentiment polarity. All data is timestamped and labeled with student IDs and sent in JSON format via a Kafka message queue, with the highest throughput of 80,000 records per second.

The storage layer: a 5 node Hadoop cluster is implemented, with a three replica mechanism of HDFS to guarantee the reliability of data. The nodes have a 32-core CPU and 256GB of memory. The data is sorted by semester and stored in an HBase columnar database with composite index of student ID, course ID and a timestamp and query

response time is under 200 milliseconds. The computing layer uses Spark streaming to compute the real-time data streams to 5-second micro-batches, and invokes the random forest model in the MLlib library to determine the learning status [9,10]. Redis holds the latest behavior feature vectors of students, MongoDB receives and holds alert records and push logs, and Elasticsearch is created to create a full-text search engine to assist in querying multi-dimensional data.

C. INTELLIGENT ANALYSIS AND EARLY WARNING MECHANISM

The system forms a multi-model fusion learning state recognition system. There are 23 evaluation indicators that are converted to feature vectors that are input to the model. An 8360-labeled historical learning records are used to train a random forest classifier using 200 trees and a maximum depth of 15 layers. Using 10-fold cross-validation, an average accuracy of 91.4% is attained. It uses a Long Short-Term Memory (LSTM) network, with a batch size of 64, 128 hidden units, and a learning rate of 0.001 to learn the behavioral patterns of students over seven consecutive days.

A Support Vector Machine (SVM) is the one that uses a radial basis function kernel to manage non-linear relationships between features, the penalty parameter is set to 10 and the fault tolerance is to 0.1. A weighted voting mechanism is employed to combine the output probability values of the three models. The weight ratio (0.45:0.35:0.20) was determined through a grid search optimization process using 5-fold cross-validation on the training dataset, selecting the combination that yielded the highest weighted F1-score.

The mechanism of early warning trigger uses three-tier threshold system. Three consecutive assignments and less than 70% video completion result in a yellow warning when the accuracy of homework completed by the student is below 60 percent, and the system promotes specific learning materials. A warning on the orange level is given when classroom interaction is recorded to be zero two weeks in a row, and a decrease in test scores above 15 points, and the counselor is notified to take action. High-risk students are put on a red warning that they might fail the course; Table 1 indicates the criteria of this.

TABLE 1. Triggering Conditions and Response Strategies for Level III Early Warning

Warning Level	Homework completion rate	Changes in test scores	Platform login frequency	Interactive engagement	Response measures
Yellow Alert	60%-75%	A drop of 5-10 points	2-3 times a week	Less than 30%	Push learning resources
Orange alert	45%-60%	A drop of 10-15 points	1-2 times per week	Less than 15%	Counselor's Talk
Red Alert	Below 45%	A drop of more than 15 points	Less than once a week	Less than 5%	One-on-one tutoring by teachers
Continuous monitoring	75%-85%	Fluctuation ± 5 points	3-5 times per week	30%-60%	Regular feedback reports
Good condition	Above 85%	Stable or rising	More than 5 times a week	Above 60%	Encouraging prompts

Table 1 shows the combination of quantitative indicators corresponding to different warning levels. The system scans student data every 5 minutes. Instead of a strict ‘all-conditions’ matching logic, the system calculates a

comprehensive risk score (S) based on the weights of these four indicators. If the calculated score falls within the threshold range defined by the quantitative benchmarks in Table 1, the corresponding warning level is triggered. This prevents high-risk students with inconsistent indicators (e.g., low completion but normal login frequency) from being overlooked. Orange and red warnings will also generate detailed learning analysis reports and send them to the teacher’s workbench.

D. EMPIRICAL RESEARCH DESIGN

The research sample in this study was 1247 undergraduate students in three comprehensive universities in East China, enrolled in three typical courses: Computer Fundamentals, Advanced Mathematics, and College English. A quasi-experimental design was used where students were randomly assigned to 627 students in an experimental group and 620 students in a control group. Informed consent was obtained from all participants. No significant differences

were observed in the two groups in terms of entrance examination score ($t = 0.842, p = 0.402$), gender ratio ($\chi^2 = 0.517, p = 0.472$), or major distribution. The experimental group employed the real-time monitoring system that was developed during this research whereas the control group employment employed the traditional mid-term and final examination evaluation technique. The experimental duration was 16 weeks or the fall semester of 2024, 4 hours of classes every week.

Data collection process involved having students in the experimental group utilize the online learning platform during the whole course. The entire course videos, assignment submissions and discussions were done on the platform. A behavior data collection module was implemented to capture 628 operational data points on each student consisting of 83 adjustments of video playback speed, 147 courseware downloads and 52 forum post. Teachers put up a monitoring panel to see real time class learning dynamics and got individual teaching suggestion reports every week produced by the system. The control group had to do only the attendance in classes and paper tasks, and take a standardized midterm and final exam. The survey was done periodically every two weeks where teachers and students were asked to provide

feedback on the functionality of the system and how this could be improved. The response rate to the questionnaire was high at 94.7% with data quality being good enough to be subjected to statistical analysis.

IV. RESULTS AND DISCUSSION

A. SYSTEM PERFORMANCE EVALUATION RESULTS

The test was done in a simulated real world teaching situation. JMeter stress testing tool was implemented to imitate 1500 online learning students. All the virtual users were made to do 3-8 random operations in one second, such as video playback, submitting assignments, and responses in the discussion. The duration of the test was 72 hours and the behavioral data obtained was compared and compared with

manually labelled real data. The accuracy of data was tested, using 10,000 records, which included dimensions of video viewing time, mouse click position, keyboard input and the history of page navigation. In the case of alert response test, 50 student accounts have been configured in such a way that they are configured to go into alert conditions and the difference in time taken to achieve the threshold and the push notification received was taken. Prometheus was utilized to gather metrics like CPU usage, memory usage, and network bandwidth usage and Grafana was used to produce a visual monitoring panel. Table 2 shows the test results for key performance indicators:

TABLE 2. Statistics of System Performance Test Indicators

Performance indicators	Test values	Target value	Compliance status	Remark
Data collection accuracy	97.3%	≥95%	Meets standards	Video activity: 98.6%; Text input: 95.1%
Average response time for early warning	4 minutes and 37 seconds	≤5 minutes	Meets standards	95th percentile 4 minutes 58 seconds
System concurrent processing capability	1500 people	≥1200 people	Meets standards	Peak QPS 12000
CPU peak utilization	74%	≤80%	Meets standards	Average occupancy rate: 52%
Peak memory usage	189GB	≤200GB	Meets standards	Total memory configuration: 256GB
Data query response time	186 milliseconds	≤200 milliseconds	Meets standards	Query based on student ID index

The system maintained a data acquisition accuracy of 97.3% under high-concurrency scenarios, with video behavior data achieving the highest accuracy of 98.6%, while text input data achieved an accuracy of 95.1% due to encoding conflicts between the tracking SDK and certain third-party mobile input methods (IMEs) during high-speed typing. To address this, a debouncing algorithm and a secondary UTF-8 verification layer were implemented in the latest system iteration to ensure the integrity of discussion posts and quiz responses. The average alert response time was 4 minutes and 37 seconds, with the 95th percentile at 4 minutes and 58 seconds, meeting the requirement of a response time within 5 minutes. The server's peak CPU utilization was 74%, and peak memory usage was 189GB; the system operated stably without any downtime.

B. DATA ANALYSIS OF TEACHING EFFECTIVENESS IMPROVEMENT

The teaching effectiveness test used a pre and post test comparative design. The first test was a baseline test, which was done at the start of the semester and had three dimensions: knowledge mastery, problem-solving and application ability with a total of 120 questions. The second assessment was carried out at the semester-end with a homogeneous yet heterogeneous test paper. Both control group (620 students) and experimental group (627 students) took both tests. A paired-samples t-test was used to analyze changes in scores. Table 3 is the comparative figures of learning outcomes between the two groups and the changes that the teacher made to the learning instructions.

TABLE 3. Comparison of Teaching Effectiveness and Teachers' Teaching Adjustments

Evaluation Dimensions	Experimental group	Control group	Difference Comparison	Significance
Final average score	78.4 points	67.5 points	10.9 points	p<0.001
Improvement in grades	25.8%	9.2%	16.6% higher	p<0.001
Average time for teaching adjustment	3.2 days	5.5 days	Almost 2.3 days	p<0.01
Adjusting the ratio within a week	89.3%	47.2%	42.1% higher	p<0.001
Student satisfaction rating	6.3 points	4.8 points	1.5 points higher	p<0.01
Teacher's perception accuracy	91.7%	64.5%	27.2% higher	p<0.01

The experimental group's average final exam score was 78.4, an increase of 16.1 points from the initial 62.3, representing a 25.8% improvement. The control group's score improved from 61.8 to 67.5, a 9.2% increase. The

experimental group's improvement was 18.6 percentage points higher than the control group's. The average time for teachers to adjust their teaching was 3.2 days in the experimental group and 5.5 days in the control group, a

42.1% improvement in timeliness. 89.3% of the teaching adjustments in the experimental group occurred within one week of the problem arising, compared to only 47.2% in the control group. Student satisfaction surveys showed that the experimental group rated the effectiveness of their teaching at 6.3 points, while the control group rated it at 4.8 points.

C. PROBLEMS AND OPTIMIZATION SUGGESTIONS IN SYSTEM APPLICATION

According to system application, it was discovered that some of the teachers were over-reliant on early warning information, thus decreasing the rate of proactive observation of the students, weakening the emotional bond between teachers and students. The mechanism of data privacy protection should be enhanced as students expressed some opposition to being tracked. Humanities courses do not fit well with the algorithm model, and the quality of assessment of the discussion is less accurate compared to STEM courses. It is suggested to incorporate a teacher review procedure to balance technology and humanistic care, to optimize data anonymization policies and to create a student authorization system, and to train specialized recognizing models of various disciplines that can enhance the quality of assessment.

Furthermore, this study is subject to geographic and cultural limitations, as the sample was drawn exclusively from three universities in East China. Future research should validate the generalizability of this system across different educational levels (such as K-12) and diverse cultural backgrounds to ensure the model's robustness in varied pedagogical environments.

V. CONCLUSION

This research develops a system of teaching monitoring and dynamic evaluation index based on big data, that can greatly mitigate the issue of data fragmentation and slow feedback in educational evaluation. The multi-dimensional assessment model developed with the help of the Delphi technique and the analytic hierarchy process (AHP) balances the rigor and operability of science, whereas the distributed technical architecture provides the possibility of real-time processing of large volumes of data. Machine learning algorithms allow proper determination of learning states. The effectiveness and stability of the system is proven empirically, which is a technical route towards enhancing the quality of education. The further studies will be devoted to the maximization of subject-specific flexibility of the algorithm, the investigation of the balance mechanism between privacy protection and the use of the data, the further enhancement of evaluation technology and teaching practice, the innovative development of individual teaching models.

AUTHOR CONTRIBUTIONS

Liao Guo designed, collected and analyzed the data, and drafted the manuscript. Liao Guo conducted the study, critically revised the manuscript for important intellectual

content, and gave final approval of the version to be published. Liao Guo participated fully in the work, take public responsibility for appropriate portions of the content, and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

CONFLICTS OF INTEREST

The author declares no conflict of interest.

FUNDING

This research was supported by,

- 1) The Open University of Shaanxi (Shaanxi Business College) Scientific Research Project in 2024: «(Digitalization Empowering Educational Evaluation: Research on the Construction of a Growth Evaluation System for Higher Vocational Students)»(2024KY-A06)
- 2) Research Project on Teacher Education Reform and Teacher Development in Shaanxi Province in 2025: «(Research on the Construction of a Data-driven Evaluation Index System for Teachers and Classroom Teaching Quality)»(SJS2025YB065)

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

- [1] H. Yang and J. Wang, "An online teaching process monitoring method of MOOC platform based on video recognition," *International Journal of Continuing Engineering Education and Life-Long Learning*, vol. 33, no. 2/3, pp. 149, 2023, doi: 10.1504/ijceell.2023.129203.
- [2] K. C. Lin, Y. H. Lin, and M. Y. Chen, "A Realtime Classroom Assessment System for Analysis of Students' Evaluation of Teaching Through a Deep Learning and Emotional Contagion Mechanism," *International Journal of Interactive Multimedia & Artificial Intelligence*, vol. 9, no. 5, 2025, doi: 10.9781/ijimai.2025.03.002.
- [3] B. A. D. Morais, D. R. A. M. Salha, and C. A. G. Blaha, "The importance of monitoring in remote teaching: Experiences of a nursing student in the collaborative teaching-learning process in academic training," *INTERDISCIPLINARITY AND INNOVATION IN SCIENTIFIC RESEARCH*, 2023, doi: 10.56238/interdiinnovationscrese-067.
- [4] U. Iroegbu, K. Drury, F. McGill, et al., "Quality improvement project on improving the process of antimicrobial therapeutic drug monitoring in Leeds Teaching Hospitals," *Clinical Infection in Practice*, vol. 25, Art. no. 100478, 2025, doi: 10.1016/j.clinpr.2025.100478.
- [5] L. V. Shkerina, A. Gavriluk, O. Tabinova, et al., "Bi-subject monitoring of the results of mastering universal educational activities by students in 7th-9th grades in the process of teaching mathematics," 2020, doi: 10.32744/pse.2020.2.15.
- [6] M. D. A. Nogueira, Márcia Cristina Monteiro dos Reis, J. C. Barbosa, et al., "Contribution of academic monitoring to the teaching-learning process: Experience report of monitors in the discipline Nursing Care in Emergency and Trauma," *International Journal of Advanced Engineering Research and Science*, vol. 7, no. 7, pp. 92-96, 2020, doi: 10.2161/ijaers.77.11.
- [7] N. Lozovaya, "TECHNIQUE FOR COMPUTER MONITORING OF COMPETENCIES DEVELOPMENT AMONG UNIVERSITY STUDENTS SPECIALIZING IN ENGINEERING IN THE PROCESS OF TEACHING THEM MATHEMATICS," 2020, doi: 10.25146/1995-0861-2020-51-1-183.

- [8] V. Gupta, V. Viswesh, C. Cone, et al., "Qualitative Analysis of the Impact of Changes to the Student Evaluation of Teaching Process," *American Journal of Pharmaceutical Education*, vol. 84, no. 1, 2020, doi: 10.5688/ajpe7110.
- [9] H. Dekhn, "Evaluation of Teaching Process Using Capability Index," 2020, doi: 10.30684/etj.37.4C.2.
- [10] G. Jianliang, "Application of SVM Algorithm in Teaching Process Evaluation," 2021, doi: 10.1007/978-3-030-51556-0_64.