Biomedical Testing & Sensing
2025, VOL. 4, NO. 1, 39-45
DOI: 10.47297/wspbtsWSP2755-952108.20250401

Construction and Management Application of a Big Data-Based Quality Evaluation System for Experimental Teaching in Science Education

Zhongwei Zhu, Zhe Chen, Zhongliang Zhu

Yangzhou Polytechnic College, Yangzhou, Jiangsu, 225009, China

ABSTRACT

Against the backdrop of the digital transformation of education, and aiming to address issues such as strong subjectivity and lack of process data in traditional science experiment teaching evaluation, this study constructs a quality evaluation system for science experiment teaching integrating four dimensions—students, teachers, resources, and management—from the perspective of educational management, based on big data technology. The system adheres to three principles: integration of process and outcome orientation, balance of objectivity and practicality, and dynamic adaptability, while clarifying the definition, data source, and quantification method of indicators for each dimension. Data are collected through experiment teaching platforms, smart devices, etc., and after undergoing a four-step preprocessing procedure, a layered and progressive analysis strategy (descriptive statistics, correlation and causation analysis, prediction and early warning) is adopted to explore data value. The evaluation results are applied relying on a three-tier feedback-optimization closed loop. A pilot application in chemistry experiment teaching at a university shows that the system increases device utilization rate by 15% and student operation compliance rate by 20%, effectively solving the pain points of traditional evaluation. Meanwhile, the study identifies challenges such as data governance and teachers' technical barriers, and puts forward optimization suggestions, providing decision support for educational administrators.

KEYWORDS

Science experiment teaching; Teaching quality evaluation; Big data technology; Evaluation system; Educational digitalization

1 Introduction

Amid the digital transformation of education, big data technology has provided innovative tools for evaluating the quality of science experiment teaching. Traditional evaluation models, which rely on manual scoring and emphasize outcome assessment, suffer from limitations such as high subjectivity and a lack of process data, making it difficult to meet the demands of personalized education and precise management. China's Ministry of Education launched the "Digital Education Strategy Initiative" in 2022, explicitly advocating for data-driven optimization of educational resource allocation and improved governance efficiency, thereby providing policy support for the construction of a new evaluation system [1].

Existing research shows that certain progress has been made in the application of educational big data both domestically and internationally. For instance, the UK Open University regulates data application through its *Ethical Use of Student Data for Learning Analytics Policy* [2], while the University of Washington in the United States emphasizes the core goal of using data to serve teaching in its academic data analysis policy. Domestic studies have mostly focused on theoretical frameworks, lacking practical evaluation models specifically tailored to science experiment teaching. In the field of science experiments, Patel's study on chemistry experiments confirmed that accuracy-based assessment can significantly enhance students' operational standardization, yet a multi-dimensional systematic evaluation framework has not been established [3].

From the perspective of educational management, this study aims to construct an evaluation system integrating four dimensions—students, teachers, resources, and management—and explores the application of big data technology in diagnosing and optimizing the quality of experiment teaching. It seeks to provide educational administrators with actionable decision-support tools.

2 Design of a big data-based quality evaluation system for science experiment teaching

2.1 Principles of system construction

The evaluation system follows three core principles, each with specific implementation pathways tailored to the characteristics of science experiment teaching:

2.1.1 Integration of process and outcome orientation

Moving beyond the traditional focus on "report scores only," the system assigns a 6:4 weight to *process data* and *outcome data*. Process data includes experiment operation sequence (e.g., compliance duration of reagent addition order) and device interaction frequency (e.g., number of pH meter calibrations). Outcome data, in addition to experiment report scores, incorporates "experiment outcome conversion rate" (e.g., yield of target product in synthesis experiments, pass rate of data validation in analytical experiments). This enables dual assessment of *dynamic process tracking* and *final outcome verification*.

2.1.2 Balance of objectivity and practicality

All indicators are based on automatically collected data to avoid subjective manual scoring. For example, "experiment operation standardization" is captured automatically by smart devices instead of relying on teachers' visual judgments. Calculation logic is simplified; e.g., "device utilization rate" is directly computed as *actual usage duration / total available duration × 100%* without complex weighting, ensuring that frontline teaching administrators can quickly understand and apply the system.

2.1.3 Dynamic adaptability

Indicator weights and definitions are adjusted according to the characteristics of different science disciplines such as chemistry and biology. For chemistry experiments, emphasis is placed on "operation precision" (e.g., pipette usage error), weighted at 40% under the student dimension. For biology experiments, "sample management standardization" (e.g., completeness of culture medium sterilization records) is prioritized, with its weight increased to 35%. A 15% "discipline-specific indicator slot" is reserved for custom additions by different specialties ("pollutant detection sensitivity" for environmental science experiments).

2.2 Evaluation indicator framework

Aligned with the entire process of science experiment teaching, each dimension is detailed with indicator definitions, data sources, and quantification methods to form an actionable indicator list:

2.2.1 Student dimension (weight: 40%)

Experiment Operation Standardization (25%): Includes "compliant operation step ratio" (number of compliant steps identified by smart video analysis system / total steps × 100%) and "key operation error rate" (titration endpoint judgment error value / standard error range × 100%). Data sources: motion capture sensors and built-in data modules of experiment devices. Thresholds are optimized with reference to Patel's quantitative methods for chemistry experiment data [3].

Learning Engagement (15%): In addition to platform login frequency and experiment step completion rate, "group collaboration contribution" is added (calculated based on the proportion of individual operation records in experiment logs and peer evaluation data). Data sources: collaboration modules of learning management platforms (ChaoXing Learning) and group task records.

2.2.2 Teacher dimension (weight: 25%)

Teaching Guidance Timeliness (15%): Core indicators include "response time to queries" (average time from student query submission to teacher response, required ⩽10 minutes) and "key step prompt coverage" (number of teacher prompts for high-risk steps / number of students × 100%). Data sources: Q&A logs of teaching platforms and classroom recording systems.

Teaching Design Adaptability (10%): Evaluated through "course content alignment" (correlation between students' pre-test scores and course difficulty, calculated using Pearson correlation coefficient) and "teaching resource update frequency" (annual update rate of experiment courseware and video resources). Data sources: student pre-test systems and teaching resource management platforms.

2.2.3 Resource dimension (weight: 20%)

Experiment Device Utilization Rate (12%): Subdivided into "core device utilization rate" and "basic device utilization rate," calculated separately and weighted at 7:3. Data source: device usage records from smart laboratory management systems.

Teaching Resource Conversion Rate (8%): Defined as "complete resource usage ratio" (e. g., duration of students' complete viewing of experiment videos / total video duration × 100%; proportion of courseware downloads opened ⩾3 times). Data source: access logs of resource platforms.

2.2.4 Management dimension (weight: 15%)

Teaching Plan Execution Rate (8%): Core indicators include "course progress compliance rate" (number of completed experiment items / planned items × 100%) and "safety inspection compliance rate" (number of monthly passed safety checks / total inspections × 100%). Data sources: progress records from teaching management systems and safety inspection logs.

Issue Resolution Efficiency (7%): Calculated as "problem resolution cycle" (average days from issue identification to resolution, required ≤7 days). Data source: quality improvement records of teaching management.

3 Implementation pathways for the evaluation system

3.1 Data collection and preprocessing

Concrete Data Sources and Technical Support: Three core data channels with specific devices and platforms are defined to ensure traceable and collectible data:

Experiment Teaching Platforms: Mainstream platforms such as *Rain Classroom* and *Experimental Space* are used to record teacher-student interaction data and student task submission data.

Smart Experiment Devices: Chemistry experiments are equipped with *smart burettes* (recording titration volume and speed) and *wireless sensor experiment kits* (transmitting real-time data such as temperature and concentration). Biology experiments use *smart incubators* (recording temperature/humidity control and sample observation logs). All devices support real-time data upload to cloud databases [4].

Teacher-Student Feedback Systems: Anonymous online questionnaires (e.g., *Wenjuanxing*) collect subjective evaluations, including 5-point Likert scale items such as "experiment resource applicability score" and "teacher guidance satisfaction," with open-ended question fields to capture non-quantifiable needs.

Standardized Preprocessing Procedures: A "four-step method" ensures data quality and compliance with educational data standards:

Data Cleaning: Python's Pandas library is used to remove outliers and delete duplicate experiment report submissions.

Missing Value Handling: Multiple imputation is applied to supplement small amount of missing data. If the missing rate exceeds 20%, data from the same student's similar experiments are used for reasonable estimation.

Standardization: Data of varying magnitudes are converted to a [0,1] range. For example, "response time to queries" is standardized using *(max response time - actual response time) / (max response time - min response time)*.

Anonymization: Student IDs are hashed, and names are desensitized, retaining only aggregated information such as "grade-major-experiment class" to avoid privacy leaks.

3.2 Application of data analysis techniques

A "layered and progressive" analysis strategy is adopted, with each layer specifying technical tools and application scenarios to ensure actionable insights for teaching management decisions:

Basic Layer: Descriptive Statistical Analysis: SPSS software generates weekly reports (e.g., "device utilization rate") and comparative charts (e.g., "student operation compliance rate by grade") to visualize teaching status. For example, a university's chemistry experiment center discovered that the utilization rate of gas chromatographs on Wednesday afternoons was only 45%, lower than the 70% at other times, providing a basis for device scheduling optimization.

Intermediate Layer: Correlation and Causation Analysis: Python's Scikit-learn library performs Pearson correlation analysis to identify key factors affecting teaching quality. For instance, a pilot study found that the correlation coefficient between "frequency of teacher's key prompts in the first 15 minutes" and "student operation compliance rate" was *r= 0.72 (p<0.01)*, indicating that early guidance significantly improves operational standardization. ANOVA revealed that "device aging" (usage years ≥5) was the main cause of increased experiment data error rates (*F=8.62, p<0.05*).

Advanced Layer: Prediction and Early Warning Models: A random forest algorithm [5] (which improved prediction accuracy by ~15% compared to traditional decision trees in the pilot) constructs a quality warning model. Input features include "student pre-test scores" and "recent device failure frequency," outputting "experiment teaching quality risk levels" (low, medium, high). When the risk level is "high," the system automatically alerts teaching managers (e.g., "Device failure risk for a certain class reaches 80%, recommend proactive maintenance"). The model's overall accuracy reached 82% in the pilot.

Visualization: Tableau builds dynamic dashboards with three modules— "teaching quality overview," "student performance tracking," and "resource usage analysis"—allowing managers to filter data by "grade," "experiment course," and "time period." All charts support export to PDF for teaching quality reporting.

3.3 Mechanisms for applying evaluation results

A "three-tier feedback-optimization closed loop"[6] is established, defining application pathways and responsibilities for different stakeholders:

Student Side: Personalized Learning Diagnostics: The system generates "personal experiment capability reports" based on evaluation results, highlighting weaknesses and recommending tailored learning resources. In the pilot, 91% of students viewed these reports, and 83% reported "clarified improvement directions."

Teacher Side: Teaching Improvement Guidance: Teachers receive "teaching behavior analysis reports, " including "guidance frequency distribution" and "student problem hotspots." For example, a chemistry teacher found that "layer judgment in extraction experiments" was a major student concern (42% of questions). By adding a 10-minute demonstration video, the question rate dropped to 15% in the next experiment. Teacher improvements are incorporated into "teaching assessment indicators," accounting for 20% of annual evaluations to incentivize result application.

Management Side: Resource and Process Optimization: Education managers make two types of decisions based on evaluation results:

(1) *Resource Allocation Optimization*: Underutilized devices are reallocated to high-demand laboratories, or new devices are purchased.

(2) *Process Improvement*: For "prolonged problem resolution cycles, " a "7-day Rectification Record" is established, requiring daily progress updates from responsible departments. Resolution rates are tied to departmental performance, reducing the average resolution cycle from 12 days to 5 days in the pilot.

4 Case validation: application in chemistry experiment teaching at a university

4.1 Case background

A vocational university ' s College of Biological and Chemical Engineering was selected as a pilot site. The chemistry department offers 10 experiment courses (including inorganic chemistry, analytical chemistry, and organic synthesis experiments), annually covering 800 students across three grades and involving 15 experiment instructors. Prior to the pilot, three major pain points existed:

(1) *Evaluation relied on manual scoring by teachers*, with a standard deviation of 1.8 (out of 10 points) for the same experiment, indicating inconsistent standards.

(2) *Device scheduling was chaotic*: Wait times for some core devices exceeded 2 hours, while utilization rates for basic devices were only 50%.

(3) *Student feedback was delayed*: Scores and improvement suggestions were provided two weeks after experiments, hindering timely learning adjustments.

The evaluation system was implemented during the Fall 2024 semester (12-week teaching period), with five core experiment courses (e.g., analytical chemistry and organic synthesis experiments) selected for the pilot.

4.2 Application process and results

4.2.1 data collection and analysis process

Deployed 12 sets of smart experiment devices (including 6 smart burettes, 3 wireless sensor experiment kits, and 3 smart chromatographs), integrated with the university ' s "Experiment Teaching Management Platform."

Collected 2,360 sets of experiment operation data (averaging 472 sets per course), 1,280 teacher-student interaction records, and 960 student feedback questionnaires.

Data preprocessing followed the "four-step method," achieving a 95.3% valid data rate after cleaning. Hierarchical analysis was conducted using SPSS and Python, with visual dashboards pushed weekly to teachers, students, and managers.

4.2.2 Key application results

Significant Improvement in Teaching Quality Indicators:

Device utilization rate increased from 62% to 77% (e.g., HPLC utilization rose from 70% to 92%; balance utilization increased from 50% to 68%).

Student operation compliance rate (error rate <5%) improved from 68% to 88%, with the scoring standard deviation reduced to 0.7, indicating enhanced objectivity.

Average teacher response time to queries decreased from 25 minutes to 12 minutes, and the average number of experiment report revisions dropped from 3 to 1.

Enhanced Management Efficiency and Satisfaction:

A "device scheduling optimization plan" dynamically allocated device usage based on utilization rates, reducing student wait times to under 30 minutes.

An "instant feedback mechanism" provided preliminary evaluation results within 1 hour after experiments. Satisfaction surveys showed:

"Evaluation fairness" scores increased from 70 to 92 (out of 100).

"Teaching improvement effectiveness" scores rose from 68 to 89.

Experiment teaching complaint rates decreased by 40%, from 8% to 4.8%.

Validation of Key Influencing Factors:

Correlation analysis confirmed that "pre-experiment guidance duration by teachers" ($r=0.68$, $p<0.01$) and "regular device maintenance frequency" ($r=-0.59$, $p<0.05$) significantly correlated with teaching quality, providing clear directions for future improvements.

5 Practical challenges and optimization suggestions

5.1 Core challenges

Based on pilot experience and the current state of science experiment teaching in China, three core challenges are identified, each with specific manifestations:

5.1.1 Dual pressures of data governance and privacy protection

On one hand, heterogeneous data formats from different experiment devices require additional interface development for integration. In the pilot, interface development costs accounted for 30% of the total project investment.

On the other hand, strict compliance with educational data privacy regulations is required. Some students expressed concerns about "operation data being used for other purposes," with 23% initially refusing to provide complete operation data, necess repeated communication and explanation.

5.1.2 Technical application barriers among frontline teachers

Most experiment teachers excel in professional teaching but lack data analysis skills. In the early pilot stage, 67% of teachers reported "inability to understand correlation analysis results" and "difficulty using visual dashboards."

Heavy teaching workloads (averaging 12 experiment teaching hours per week) further limit time for learning data technologies, resulting in only 45% application rate of evaluation results.

5.1.3 Insufficient disciplinary adaptability of the evaluation system

Although the existing indicator framework reserves "discipline-specific indicator slots," cross-disciplinary adaptation remains challenging. For example, "completeness of sample observation records" in biology experiments (requiring manual judgment) is difficult to quantify automatically via smart devices, necessitating partial manual assistance.

Field experiment data in environmental science cannot be uploaded to the cloud in real time, leading to delayed data collection and reduced evaluation timeliness.

5.2 Optimization suggestions

Targeted solutions are proposed, with some suggestions drawing on domestic and international mature experience:

5.2.1 Establish a "technology + institution" dual-driven data governance system

Technical Measures: Develop a "unified interface for science experiment data" to support automatic conversion of mainstream device data formats, reducing integration costs. Adopt *blockchain technology* for data traceability, recording data collector, purpose of use, and modification history at each node to ensure traceability and tamper-resistance, enhancing students' sense of privacy security.

Institutional Measures: Formulate *Regulations for Science Experiment Teaching Data Usage*, clarifying "data collection scope" and "usage permission tiers" (e.g., administrators access full data, teachers access only their class data, students access only personal data). Drawing on the UK Open University's ethical policy [2], establish a "data usage application-approval" process to prevent misuse.

5.2.2 Create a "simplified tools + tiered training" teacher empowerment model

Tool Simplification: Develop "one-click analysis tools" that encapsulate complex algorithms into buttons like "generate teaching report" or "view student weaknesses." Teachers can obtain results without understanding technical principles (e. g., clicking "student operation analysis" automatically generates "Top 3 error-prone steps for a class" and improvement suggestions).

Tiered Training: Provide training tailored to teachers' technical backgrounds:

For beginners: "1-hour micro-training" focusing on dashboard navigation and report interpretation.

For intermediate users: "Offline workshops" (twice per semester) covering correlation analysis and basic modeling.

Incorporate "data application capability" into teachers' continuing education credits to incentivize learning, aiming to increase teacher application rates to over 80%.

5.2.3 Design a "modular + flexible" cross-disciplinary adaptation scheme

Modular Indicators: Build a "modular indicator library" categorized by discipline (e.g., chemistry: "operation precision," "reagent management"; biology: "sample management," "observation records"). Managers can directly select module combinations without redesigning indicators. For indicators not easily amenable to automatic quantification, develop "semi-automatic scoring tools" where the system first identifies key information in records, and teachers supplement with scores, reducing manual workload.

Flexible Data Collection: For special scenarios like field experiments, develop an "offline data collection APP" allowing students to record data on-site and automatically sync to the cloud upon returning to the laboratory. Set a "data supplementation window" to ensure no data omission and improve evaluation timeliness.

6 Conclusion and outlook

6.1 Research conclusions

System Value: The evaluation system constructed in this study, through its "four-dimensional indicator framework + layered implementation pathway," achieves "three transformations" in science experiment teaching quality evaluation:

Comprehensiveness of Evaluation Dimensions (covering students, teachers, resources, and management);

Automation of Data Collection (relying on smart devices and platforms to reduce manual intervention);

Closed-Loop Application of Results (forming a management closed loop from evaluation to feedback to optimization).It addresses the traditional problems of "subjectivity, fragmentation, and delayed feedback."

Practical Effectiveness: Validated through a chemistry experiment teaching pilot at a university, the system significantly improves teaching quality (student operation compliance rate increased by 20 percentage points), optimizes management efficiency (device utilization rate increased by 15 percentage points), and enhances teacher-student satisfaction (satisfaction scores improved by 22-24 points). Its operational simplicity makes it suitable for grassroots teaching administrators to promote and apply.

6.2 Innovations

(1) *Educational Management Perspective*: The system incorporates the "management dimension" into the evaluation framework, moving beyond a narrow focus on teachers, students, and resources.

(2) *Dynamic Adaptation Mechanism*: It tailors indicators to different science disciplines, overcoming the limitations of a "one-size-fits-all" approach.

(3) *Technical Practicality*: All analysis techniques and tools are designed with frontline teachers' application capabilities in mind, avoiding "technology for technology's sake."

6.2 Research limitations and future directions

6.3 Two limitations exist

(1) The pilot only covered chemistry in higher education and has not yet been extended to middle school science experiment teaching or other disciplines such as biology and environmental science. The system's generalizability requires further validation.

(2) Some indicators (e.g., observation record scoring in biology experiments) still require manual assistance, and the degree of automation needs improvement.

Future research can be deepened in three directions:

(1) *Cross-Educational-Level and Cross-Disciplinary Expansion*: Apply the system to middle school science experiment teaching (e.g., "oxygen preparation experiments" in junior chemistry) and optimize indicator modules for disciplines like

biology and environmental science, forming an evaluation system that "covers primary/secondary schools to higher education and multiple science disciplines."

(2) *Integration of AI Technologies*:

Introduce computer vision to automatically recognize "sample morphology" in biology experiments (e.g., whether cells are pathological), replacing manual scoring.

Use natural language processing to analyze the "completeness of phenomenon descriptions" in student experiment reports, further enhancing evaluation automation.

(3) *Construction of a Multi-Stakeholder Collaborative Platform*: Collaborate with universities, experiment device enterprises, and education administrative departments to build a "science experiment teaching data platform" for cross-entity sharing of "device data, teaching data, and management data." For example:

Enterprises can optimize device designs based on platform data.

Education departments can formulate regional science experiment teaching policies based on data insights.

This fosters an ecosystem of "teaching-management-industry" synergistic development.

Funding

2025 Annual Philosophy and Social Sciences Research Project of Colleges and Universities in Jiangsu Province: Study on the Path of Promoting Higher Vocational College Students' Development and High-Quality Population Development under the Employment Priority Strategy (No. 2025SJSZ0979); 2025 International Exchange Research Project of Yangzhou Polytechnic University: Study on the Benchmarking Path of Vocational Undergraduate Education Internationalization Oriented by Industrial Demands (No. 2025GJ02)

About the Author

1. Zhongwei Zhu (1989-09), male, Han ethnicity, born in Gaoyou, Jiangsu Province, holds a master's degree, is an assistant researcher at the Enrollment and Employment Office of Yangzhou Polytechnic University, with research interest in education management.

2. Zhe Chen (1986-07), male, Han ethnicity, born in Huaining, Anhui Province, holds a master's degree, is a lecturer at the Normal College of Yangzhou Polytechnic University, with research interest in ideological and political education.

3. Zhongliang Zhu (1984-05), male, born in Gaoyou, Jiangsu Province, holds a bachelor's degree, is a senior secondary school teacher, and engages in teaching and research of senior high school physics.

References

[1] Tang Y, Liu Y, Xie T. The Dilemmas and Breakthroughs in Educational Data Privacy Protection: A Multidimensional Analysis of Policies from World-Class Universities[J]. Distance Education in China, 2025, 45(4): 69-84.

[2] The Open University. Policy on Ethical Use of Student Data for Learning Analytics[EB/OL]. [2025-08-15]. https://help.open.ac.uk/documents/policies/ethical-use-of-student-data/files/22/ethical-use-of-student-data-policy.pdf.

[3] Patel B A. A Chemical Analysis Laboratory Class Solely Assessed on the Accuracy and Precision of Student Data[J]. Journal of Chemical Education, 2022, 99(10): 3585-3589.

[4] Li L J, Tang M, Wang W Z, et al. Exploration of Intelligent Mobile Devices in Biochemistry Experiment Teaching[J]. Basic Medical Education, 2017, 19(6): 462-466.

[5] Yang Z K. Adhering to the "3I" New Concept to Deepen Educational Digitalization[J]. Distance Education in China, 2024, 44(12): 3-14.

[6] Dillenbourg P, Järvelä S, Fischer F. The Evolution of Research on Computer-Supported Collaborative Learning[J]. Informatics in Education, 2009, 8(1): 3-19.