Creative Education, 2026, 17(11), 2146-2153
https://www.scirp.org/journal/ce
ISSN Online: 2151-4771
ISSN Print: 2151-4755

Research on the Through-Integrated Teaching Evaluation Model in Higher Education in the AI Era

Miaomiao Yang*, Bin Li, Xiaoli Wang, Jingming Wang, Qian Sun, Xuemei Zhang

Faculty of Mathematics and Artificial Intelligence, Qilu University of Technology, Jinan, China
Email: *3d07@163.com

How to cite this paper: Yang, M. M., Li, B., Wang, X. L., Wang, J. M., Sun, Q., & Zhang, X. M. (2026). Research on the Through-Integrated Teaching Evaluation Model in Higher Education in the AI Era. *Creative Education, 17,* 2146-2153.
https://doi.org/10.4236/ce.2026.1711119

Received: September 12, 2026
Accepted: November 7, 2026
Published: November 10, 2026

Copyright © 2026 by author(s) and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/
Open Access

Abstract

This study investigates the through-integrated teaching evaluation model in higher education in the era of artificial intelligence. We propose an AI-empowered, through-integrated evaluation model that places students' learning outcomes at the core of assessment. Built on an architecture of "vertical continuity and horizontal integration", the model integrates multiple dimensions—students, learning, and institutions—and employs an "multi-source and multi-faceted" approach to strengthen process evaluation, explore value-added evaluation, and improve comprehensive evaluation. The proposed model offers a conceptual foundation for enhancing teaching quality, improving evaluation objectivity, and fostering lifelong learning habits. This study establishes a theoretical framework for AI-empowered through-integrated evaluation.

Keywords

Teaching Evaluation Model, Through-Integrated, AI-Empowered, Teaching Reform

1. Introduction

Education is a fundamental undertaking for the nation and the Party. The Fifth Plenary Session of the 19th CPC Central Committee proposed building China into an education powerhouse by 2035, enriching the goal of realizing basic socialist modernization by the same year. Accelerating the modernization of education to lay a solid foundation for an education powerhouse requires resolving the challenge of educational evaluation reform. In October 2020, the CPC Central Committee and the State Council issued the Overall Plan for Deepening the Reform of Educational Evaluation in the New Era (CPC Central Committee & State Council,

2020, Zhongfa [2020] No. 19) charting the course for building an educational evaluation system suited to China's realities and aligned with world-class standards. The Plan explicitly introduces the "four evaluations" concept—improving outcome-based evaluation, strengthening process evaluation, exploring value-added evaluation, and enhancing comprehensive evaluation—thereby raising new requirements for teaching evaluation reform in higher education, which is pivotal to improving institutional quality and talent cultivation. Teaching evaluation reform in higher education has long attracted scholarly attention, and related research has continued domestically and internationally. Below, we present a review of the academic history and current research trends.

2. Historical Review of Scholarship and Research Trends

In Fourth Generation Evaluation, Guba and Lincoln (Guba & Lincoln, 1989) summarized the evolution of modern educational evaluation into the "four-generation theory". From the late nineteenth to the early twentieth century, the initial "measurement era" limited evaluation to observable, quantifiable states, focusing on human intelligence and personality and employing explicit phenomena or behaviors to pursue scientific accuracy. From the 1930s to the 1950s, the "description era" emphasized judging whether objectives were achieved, evaluating the degree of goal attainment while highlighting comprehensive assessment beyond measurement; the representative product of this period was Tyler's behavioral-objective evaluation model. The "judgment era" of the 1960s to the 1980s was committed to providing evidence for improving education, emphasizing that evaluation should not merely describe states but furnish a foundation for decision-making, with an emphasis on process evaluation; Stufflebeam's CIPP model (Stufflebeam, 1971) is the most representative of this period. From the 1980s to the present, the "construction era" focuses on diagnosis and development, emphasizing that evaluation is a process of negotiation and psychological construction among multiple subjects, pursuing the coordination of pluralistic value standards and consensus-building, of which Stake's responsive evaluation model (Abma & Stake, 2001) is representative.

China began exploring methods and systems for higher education teaching evaluation in the 1970s, relying mainly on questionnaires and student feedback for data collection. In 2003, the Ministry of Education promulgated the Measures for the Evaluation of Undergraduate Teaching Work in Regular Institutions of Higher Education, explicitly specifying evaluation indicators and methods. Since then, regions and institutions have formulated corresponding evaluation measures and actively promoted them. Professor Huang et al. (2019) published "A Study of Student-Centered Evaluation of University Course Teaching Quality" in the China Education Policy Review, one section of which is titled "'Student-Centered' Teaching Philosophy: An Interrogation", reflecting on the problems of this idea in the Chinese educational context. Building on the "learning-oriented" educational philosophy, this paper more inclusively integrates multiple teaching ele-

ments.

Recent studies have explored AI-supported assessment frameworks and learning analytics (Hayes & Trajkovski, 2022, Rajasekar et al., 2024), yet few address the integration of longitudinal tracking across multiple educational phases. This gap is addressed by our model's emphasis on vertical continuity (pre-enrollment to post-graduation) and horizontal integration (general vs. specialized courses).

3. The "Through-Integrated" Evaluation Model

3.1. Overview of the Evaluation Model

Inspired by Professor Shi Jinghuan (Huang, Zhou, & Shi, 2019) and colleagues, and guided by the Overall Plan for Deepening the Reform of Educational Evaluation in the New Era, the author conducts through-integrated research on the teaching evaluation system of higher education in the new era by integrating multiple teaching elements. Compared with existing research, the perspective proposed in this paper has distinctive scholarly and applied value.

This paper proposes a "through-integrated" evaluation model that vertically traces students' learning across "pre-enrollment—during-coursework—post-coursework—post-graduation", while horizontally comparing learning in "general education courses—specialized courses", thereby forming an evaluation model of "vertical continuity and horizontal integration" that cultivates talents with lifelong learning habits and promotes the construction of a learning society. Integrating multiple teaching elements, the model employs a "multi-source and multi-faceted" evaluation approach to strengthen process evaluation, explore value-added evaluation, and enhance comprehensive evaluation, thereby establishing a new-era teaching evaluation model for higher education. Based on big data applications, machine learning, and natural language processing, a new-era evaluation system should extract useful information from massive data to improve the scientific rigor, professionalism, and objectivity of teaching evaluation. There are some Operationalized indicators as follows:

- Student Dimension: Evaluate "knowledge" via course exam scores leveraging data from the academic affairs system; "behavior" via classroom participation rate relying on learning platforms.
- Learning Dimension: Assess "course design" using student engagement metrics from learning management systems (LMS); evaluate "teacher feedback" through response time and specificity in LMS discussions.
- Institution Dimension: Comprehensive talent cultivation is measured by graduate employment rates and industry skill surveys using data from university career center data.

This research can further improve the teaching system, enhance teaching quality, and promote the development of higher education. It relates to teaching, education management, and educational policy, and its findings can provide theoretical references for relevant disciplines, advancing educational research. Through the analysis of teaching evaluation results, it can offer strong guidance and sup-

port for teaching reform and innovation in higher education, and provide teachers with more scientific and effective teaching methods and strategies to improve their instructional level.

3.2. The Connotation and Pathways of AI-Empowered Teaching Evaluation

With the rapid development of artificial intelligence, AI is profoundly reshaping the concepts and practices of teaching evaluation in higher education. Based on big data analytics, machine learning, deep learning, and natural language processing, AI-empowered teaching evaluation can overcome the limitations of traditional evaluation in data acquisition, processing, and feedback, providing technical support for the implementation of the through-integrated evaluation model. AI-empowered teaching evaluation is mainly reflected in the following three aspects:

1) Data-driven process evaluation. Traditional evaluation tends to focus only on final examinations while overlooking each session of the teaching process. AI technology can collect in real time multimodal data produced in classroom interaction, online learning platforms, and homework assessment, incorporating all information from teacher-student participation in teaching activities into the scope of evaluation. Through data mining and learning analytics, evaluation systems can dynamically track learning trajectories, identify learning patterns, and predict learning outcomes, thereby achieving process-based, continuous evaluation of the entire learning process. Recent advances in generative AI further strengthen this capability, enabling richer formative assessment that captures nuanced evidence of student understanding beyond conventional metrics, and supporting real-time analysis of ongoing learning behaviors to sustain continuous, process-oriented evaluation.

2) Visualization and precision of value-added evaluation. Value-added evaluation emphasizes measuring students' development and progress after excluding differences in initial levels. By modeling and comparing massive amounts of data from pre-enrollment, during coursework, and even post-graduation, AI can more accurately measure students' value-added development, providing objective quantitative evidence for the "vertically continuous" dimension of through-integrated evaluation, and shifting value-added evaluation from experiential judgment to data-driven support.

3) Intelligent and personalized feedback. AI-empowered systems can generate individualized and adaptive feedback tailored to each learner's actual performance and learning trajectory, delivering timely, specific guidance that enhances both the targeting of evaluation and learning effectiveness. Generative AI can aid educators by streamlining the creation of course materials, assignments, and assessments, thereby freeing instructors to devote more time to providing support and feedback to students.

It should be noted that AI-empowered teaching evaluation does not replace teachers' judgment with technology; rather, it adheres to the principle of "human-

machine collaboration, human in the loop". In intelligent evaluation, teachers' professional judgment, value guidance, and emotional care remain dominant, thereby ensuring the fairness, scientific rigor, and educational value of evaluation.

3.3. Construction of the Evaluation Model

1) Evaluation subjects. The Chinese character xue carries rich connotations, signifying students, learning, and scholarship, as well as scholars, institutions, and academia. The teaching evaluation system of higher education in the new era should prioritize students' subjectivity while equally emphasizing other interrelated elements such as teachers and institutions—namely, the "student" as the subject of classroom instruction, the "learning" activities taking place in the classroom, and the "teachers" and "institutions" that support classroom learning, of which the role and responsibility undertaken by teachers are indispensable. Accordingly, this paper proceeds along three dimensions—"student, learning, and institution". The student dimension evaluates the "knowledge, emotion, volition, and behavior"; the learning dimension evaluates "course design" and "teacher feedback"; and the institution dimension requires a comprehensive evaluation of talent cultivation.

2) Evaluation methods. To strengthen process evaluation as required by the Overall Plan, we adopt Stufflebeam's decision-centered CIPP model—Context, Input, Process, Product (CIPP) Evaluation Model. Through this model, we aim primarily to identify the gap between "what is" and "what should be" to serve decision-making; the rationale is that evaluation should provide more comprehensive information for decisions (planning, organizing, and implementing decisions), thereby improving them. Because evaluation is conducted throughout program implementation, this model effectively realizes process evaluation.

In addition, students' (the evaluated parties') opinions are of critical importance; hence we adopt the "responsive evaluation" systematically articulated by Stake and elaborated by Abma and Stake. We maintain that all stakeholders have the right to openly discuss their claims, concerns, and issues and to demand responses, regardless of divergent value positions, emphasizing the participation of the evaluated subjects in assessment. This model redresses the managerialist tendencies, neglect of value pluralism, and over-reliance on the scientific paradigm present in previous evaluation approaches.

In through-integrated evaluation, AI-empowered methods should synergize with the above qualitative evaluation models: the CIPP model provides the decision-making framework, responsive evaluation ensures the participation of multiple subjects, and AI technology offers efficient support for data acquisition, processing, analysis, and feedback. Together, the three constitute a "framework-participation-technology" trinity of methodological systems. Here we will show an example workflow such as "Data Source-AI Task-Teacher Review-Instructional Decision":

- Data Source: Student feedback collected via LMS quizzes.

- AI Task: Natural Language Processing (NLP) analysis to identify common themes and sentiment polarity.
- Teacher Review: Faculty analyze NLP output to identify underrepresented topics.
- Instructional Decision: Redesign course modules to address identified gaps within the CIPP framework's "Process" component.

3) Evaluation framework. The proposed through-integrated teaching evaluation model for higher education aims to improve teaching quality and teachers' instructional standards, deepen pedagogical research, and ultimately cultivate innovative, lifelong-learning talents of the new era while promoting the construction of a learning society.

On the one hand, this model necessitates the formation of general education teaching teams and specialized teaching teams. The general education teams comprise four sub-teams: the "Ideological and Political Theory Course Team", the "College Mathematics Team", the "College Physics Team", and the "College English Team", enabling horizontal comparison across courses during students' tenure. Value-added evaluation is primarily employed to measure students' development more accurately, while simultaneously serving to supervise courses and provide feedback to teachers, thereby achieving horizontally integrated, comparative evaluation of students' in-school learning. Throughout this process, AI technology can standardize and correlate learning data across multiple courses, objectively revealing students' cross-curricular ability transfer and development.

In addition, vertical continuity evaluation is conducted through four stages: formulating plans, collecting information, organizing and analyzing, and reaching conclusions. Taking mathematics as an example, we track students' pre-enrollment data (e.g., Gaokao mathematics scores), during-coursework data (e.g., evaluations of the three public college mathematics courses—Advanced Mathematics, Linear Algebra, and Probability and Mathematical Statistics), post-coursework data (e.g., students' mathematics performance in junior and senior years up to postgraduate entrance examinations), and post-graduation data (e.g., the influence of mathematical thinking on personal development three to five years after graduation). Organizing and analyzing require adequate reliance on technological means to render evaluation results more objective and scientific. The philosophy of "vertical continuity" is to record students' learning continuously across the four stages—pre-enrollment, during coursework, post-coursework, and post-graduation—and, according to students' different types and actual situations, to attend to their growth and the cultivation of a lifelong-learning personality. The AI-empowered through-integrated platform can consolidate and dynamically track data across these four stages, providing sustained data support for nurturing lifelong learners.

Value-added comparisons must account for heterogeneous assessment scales (e.g., normalized scores vs. pass/fail grades) and initial ability differences (measured via pre-enrollment tests). Additionally, selective participation in postgradu-

ate exams or alumni follow-ups introduces selection bias. Therefore, observed development should be interpreted cautiously as a potential effect of teaching rather than definitive evidence.

4. Conclusion

Inspired by the "learning-oriented" philosophy proposed by Professor Shi Jinghuan and colleagues, this paper focuses on the three dimensions of "student, learning, and institution", integrates multiple teaching elements, and employs a "multi-source and multi-faceted" evaluation approach to strengthen process evaluation, explore value-added evaluation, and enhance comprehensive evaluation.

This paper innovatively proposes a through-integrated evaluation model that vertically traces students' learning across "pre-enrollment—during-coursework—post-coursework—post-graduation", while horizontally comparing learning in "general education courses—specialized courses", including public foundation courses such as "Ideological and Political Courses", "College Mathematics", and "College English", with attention to value-added evaluation across courses for horizontal comparison. This ultimately forms an evaluation model of "vertical continuity and horizontal integration", cultivating talents with lifelong learning habits and promoting the construction of a learning society.

The proposed through-integrated model offers a conceptual foundation for AI-empowered teaching evaluation. By integrating CIPP, responsive evaluation, and AI analytics, this framework provides a blueprint for future empirical research to validate its effectiveness in enhancing teaching quality and student outcomes.

Acknowledgements

This work was supported by the Shandong Province Undergraduate Teaching Reform Research Project (Grant No. Z2025233), the Shandong Province Graduate Education Teaching Reform Research Project (Grant No. SDYJSJGC2025041), the Key Teaching Research Projects of Qilu University of Technology (Grant Nos. 2024zd18 and 2024zd19), and the General Teaching Research Projects of Qilu University of Technology (Grant Nos. 2024yb27 and 2024yb28).

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

Abma, T. A., & Stake, R. E. (2001). Stake's Responsive Evaluation: Core Ideas and Evolution. *New Directions for Evaluation, 2001,* 7-22. https://doi.org/10.1002/ev.31

CPC Central Committee & State Council. (2020). Overall Plan for Deepening the Reform of Educational Evaluation in the New Era (Zhongfa [2020] No. 19) [Policy document]. https://www.gov.cn/zhengce/2020-10/13/content_5551049.htm

Guba, E. G., & Lincoln, Y. S. (1989). *Fourth Generation Evaluation*. Sage Publications.

Hayes, H., & Trajkovski, G. (2022). Identifying Structure in Program-Level Competencies

and Skills. In *Advances in Educational Technologies and Instructional Design* (193-218). IGI Global. https://doi.org/10.4018/978-1-7998-9644-9.ch010

Huang, Y., Zhou, X., & Shi, J. (2019). Research on the Student-Centered Quality Evaluation of University Course Teaching. *China Education Policy Review, No. 1,* 262-278. (In Chinese)

Rajasekar, A., Mallenahalli, S., Mirza, I., Palaboyina, P. K., Pola, S. K., Quadri, S. F. et al. (2024). AI Framework for Scalable Automated Continuous Formative Assessment. In 2024 IEEE International Conference on Teaching, Assessment and Learning for Engineering (TALE) (pp. 1-6). IEEE. https://doi.org/10.1109/tale62452.2024.10834316

Stufflebeam, D. L. (1971). The Relevance of the CIPP Evaluation Model for Educational Accountability. *Journal of Research and Development in Education, 5,* 19-25.