



Data-Driven Analytic Hierarchy Process Model of Teaching Evaluation System

Ying Mao¹, Fenghong Wang¹, Yilin Wang², Renjie Li¹, Qi Chen¹ and Yong Zhang*¹

¹School of Information Engineering, Tianjin University of Commerce, Tianjin 300134, China

²School of Electrical and Electronic Engineering, Nanyang Technological University, 637616 Singapore

maoying@tjcu.edu.cn, wangfenghong101@163.com, WANG2117@e.ntu.edu.sg,
18326653378@163.com, chenqi@tjcu.edu.cn, zhangyong@tjcu.edu.cn

Corresponding author: Yong Zhang, zhangyong@tjcu.edu.cn

Abstract. In order to meet the higher requirements of knowledge innovation and educational innovation for teacher development, adapting to the educational evaluation reform and institutional innovation in the new era, a data-driven analytic hierarchy process (AHP) teaching evaluation system model is constructed in this paper. Firstly, the main content of the developmental teacher teaching evaluation system is constructed, the evaluation subject is determined by the fourth-generation educational evaluation theory, and the evaluation indicators and evaluation standards are determined by combining the developmental evaluation concept and the Likert scale. Secondly, combined with the data-driven idea and the analytic hierarchy process, the analytic hierarchy process is carried out to analyze the evaluation indexes, the weights of the evaluation indicators are calculated and determined step by step, the consistency test of the judgment matrix is carried out, so as to construct a scientific, reasonable and operable teacher teaching evaluation system. Finally, the experimental analysis of the evaluation system combined with the teacher evaluation data of a university proves the feasibility and accuracy of the evaluation system proposed in this paper, which brings the guiding significance for the follow-up auxiliary teachers to identify the problems of education and teaching quality and formulate effective reform methods.

Keywords: Data-driven; Analytic hierarchy process; Teaching evaluation

DOI: <https://doi.org/10.14733/cadaps.2025.S2.101-116>

1 INTRODUCTION

Teachers are the most valuable educational capital, and they are a decisive force in the education of students, the reproduction of productive forces, and the prestige of universities. Therefore, in order to ensure the smooth development of college education, it is of great significance to carry out high-quality and high-level teacher construction and comprehensively improve the professional quality of teachers. This is the key condition for the development of higher education, and it is also

the goal of the school's personnel management system and the improvement of the scientific evaluation system and mechanism of college teachers is a very important means to achieve this goal[1].

Colleges and universities should conduct regular or irregular evaluations of their teachers' work skills and work levels and use this as a basis to motivate teachers' self-development and ultimately promote the realization of the overall goals of the organization[2]. At present, as an effective means of internal management of schools, teacher work evaluation has received more and more attention, and more and more colleges and universities hope to improve the quality of education and promote the scientific management of teachers by carrying out teacher evaluation activities. Therefore, the establishment and improvement of a scientific, fair and efficient teacher performance evaluation system and mechanism has become an important topic for colleges and universities. Schools of different foundations, types, and levels have introduced and operated corresponding assessment and evaluation methods with their own characteristics, and are all beginning to establish a teacher work evaluation system to arouse teachers' sense of competition and promote the reasonable mobility of teachers[3,4].

Due to the influence of comprehensive factors such as regional economic underdevelopment, information asymmetry, and imperfect system, various problems in teacher evaluation have become increasingly prominent, such as the contradiction between quantitative platform and qualitative height, the contradiction between categorical assessment and comprehensive balance, the contradiction between rigid assessment and flexible tasks, and the contradiction between attaching importance to results and diluting the process[5,6]. At present, the evaluation concepts, evaluation indicators, evaluation models, evaluation methods and application of evaluation results of teachers in colleges and universities in China are not perfect, which directly affects the selection, appointment, motivation and development of teachers in colleges and universities. The existing problems are mainly manifested in: First, the concept of teacher performance evaluation is relatively confusing. Either they pay too much attention to management goals and ignore the particularity and subjective initiative of individual teachers; Either a one-sided understanding of people-oriented and a scientific evaluation of teacher performance cannot be carried out. Second, the content of the teacher performance evaluation is too simple, the evaluation criteria are relatively simple, and the evaluation focuses too much on scientific research performance and ignores teaching achievements. Most of them are limited to the quantitative assessment of teachers' performance in teaching, scientific research and talent training, and lack qualitative analysis of teachers' comprehensive quality and potential, such as ideological and political performance, humanistic spirit, and innovation ability. Third, teacher performance evaluation is a mere formality. Some schools have not even formulated a basic quantitative evaluation index system, have not given full play to the basic function of teacher performance evaluation, and cannot effectively evaluate teachers' actual performance, resulting in the workflow of teacher performance evaluation becoming a formality. In some schools, due to the lack of institutional safeguards and the improvement of management mechanisms, the evaluation results are distorted or the rewards and punishments for the evaluation results are unclear. Fourth, teacher performance evaluation places too much emphasis on the short-term benefits of schools. Some colleges and universities have a certain degree of short-sightedness, lack of strategic scientific planning, lack of policy continuity, and ignore the characteristics of teachers' labor and occupation. Even if there is a teacher evaluation system, it ignores the integration of teachers' career planning into the overall development goals of the school, relaxes the capacity building of the teaching team, and fails to establish a long-term mechanism for the teacher management system[7,8].

With the deepening of the reform and the continuous development of artificial intelligence data mining technology, the introduction of digital intelligence technology into the model construction of teacher teaching evaluation system has become the focus and difficulty in the reform of the personnel allocation system of colleges and universities. Therefore, this project intends to construct an objective, fair and comprehensive teaching evaluation system based on data-driven artificial intelligence technology and hierarchical statistical analysis method to meet the needs of teachers' individual development and the realization of college goals.

2 RELATED WORK

Gomez [9] proposed a critical review method in which the evaluation of university organizations should serve the improvement of teaching practice and should tend to the constructivist model so that different subjects participating in self-evaluation, co-evaluation, and heterosexual evaluation have clear standards and indicators to promote the realization of teacher performance evaluation. In their report, Peck et al. of the University of Washington examined the use of Instructional Performance Assessment (TPA) as a learning, program evaluation, and improvement educational resource, described the conceptual framework and defining characteristics of TPA, and discussed the conditions under which TPA can be used as a decision-making tool related to teacher qualification and program improvement[10].

Karel Carrel et al. [11] of the University of Chicago found that teachers with higher evaluation scores paid too much attention to students' short-term learning achievements and had problems such as test-oriented education or pleasing students, and students were more concerned about their current test scores, and most of them were in the shallow learning stage. According to a study by Lyde et al. [12] of Illinois State University, the multi-source assessment method (MME), which consists of three main data sources: student evaluation, teacher reflection, and formative external review, is more of a summative product than a formative process, and teachers should be encouraged to engage in continuous reflection and strengthen accountability and senior teachers' guidance responsibilities. Paufler et al. [13] of Clemson University in the United States reviewed the teacher evaluation and support system in Texas and raised issues such as accountability, evaluation trust, and evaluation supervision. Hoben Hoben et al. [14] of Memorial University of Canada argue that the Curriculum Assessment Questionnaire (CEQ) positions teachers as neoliberal knowledge workers and does not treat students as community members with shared responsibilities, which weakens the status and autonomy of teachers, and that alternative methods to assess teaching excellence and promote community inquiry should be explored.

In order to improve the teaching evaluation methods of teachers, Luis et al. [15] proposed an evaluation tool with the Behavior Anchored Rating Scale (BARS) to replace the Likert scale traditionally used for student teaching evaluation (SET), which eliminates ambiguity in the interpretation of results and improves the objectivity of evaluation. It also adds a step to the traditional program, which significantly reduces the loss of information in the scale-building process and helps evaluate the application of formative functions. Byrne of Morgan State University and Donlan Donlan of the University of Maryland, College Park [16] identified midterm evaluations of university teaching as a promising, low-cost solution that can provide teachers with timely feedback to improve teaching practices. MSECT-O is an effective tool for teachers to gather formative feedback on multiple aspects of their teaching. Curry Courey et al. [17] at the University of California, Irvine, used a model-based approach and a Bayesian method to examine the scale effect in student evaluation and found that students tended to rate higher when using alphabetic scales than when using numerical scales.

A holistic consideration of multiple stakeholders in higher education is suggested in the literature [18], where modified DSET based on belief dispersion and belief entropy is used for conflict management. An integrated MCGDM approach is proposed to assess the sustainable service quality of higher education institutions. The IF-TODIM approach addresses the issues of information uncertainty and individual-bounded rationality. The integrated system of assessment criteria embeds sustainability and service quality in higher education.

Translated with DeepL.com (free version)The university handbook should specifically describe the requirements for faculty re-appointment, tenure, and promotion, vague evaluation criteria can lead to a lack of validity and reliability of evaluations and the right decision can only be made by reviewing the qualifications of evaluators at each level, Adela et al. [19] of the University of Life Sciences in Prague, Czech Republic, showed that variables related to the concept, content, and outcomes of a course had a greater impact on the overall rating of the course than teacher evaluation and that students intuitively perceived the influence of the following factors on a course: whether it was beneficial or not, Whether it is interesting, easy to understand, whether it is structured logically,

whether the format of the lesson and the content of the exercises are appropriate. Dr. Boswell of St. Dao University in the United States argues that evaluation positivity affects students' course decisions and self-efficacy, but does not affect their seriousness in feedback[20].

It is not difficult to find that the current teacher evaluation system is mostly focused on classroom teaching quality and scientific research evaluation, or more of the countermeasures for the problems existing in the current evaluation system, but the research on how to construct a systematic and complete evaluation system for college teachers is not sufficient and needs to be further studied. Therefore, this paper provides ideas for the construction of a teaching evaluation system suitable for colleges and universities, based on the method of combining quantitative and qualitative analysis of the proposed evaluation indicators by analytic hierarchy process, and at the same time, in order to fully consider the influence of each index, the weight coefficient of each evaluation index is assigned, and the consistency test is carried out to enhance the reliability. Finally, based on the survey data, the feasibility and accuracy of the proposed system for teacher evaluation are proved, and guidance for teachers to improve teaching indicators in the future is provided.

3 CONSTRUCTION OF A DATA-DRIVEN TEACHER EVALUATION SYSTEM

3.1 Determination of the Subject of Evaluation

The fourth generation of educational evaluation theory emphasizes that students, teachers, peers, supervisors, and leaders should all become the main body of teaching evaluation, and teachers and students are the service objects of the evaluation subject. Students directly benefit from the teaching effect, and the effective learning of students is the ultimate foothold of evaluation, and the students' evaluation of teachers can truly reflect the classroom teaching effect of teachers and be directly used to improve the quality of teaching. At the same time, teachers can change teaching ideas or teaching methods in time according to the problems reported by students so as to improve students' enthusiasm for learning. As the main person in charge of completing teaching, teachers' participation in evaluation can enhance teachers' sense of ownership, analyze themselves objectively, and humbly accept criticism and guidance from the outside world. Teachers' teaching is based on self-evaluation, and the feedback of teaching evaluation information and evaluation results has a certain guiding and auxiliary role in teachers' teaching and finally makes the evaluation results more scientific and reasonable. The evaluation of teachers in schools is far less important than teachers' self-evaluation, and external evaluation cannot replace internal evaluation, and the evaluation of others cannot replace teachers' self-evaluation. Peer evaluation can play a radiating role in excellent teachers, eliminate the sense of isolation among peers, and promote the professional development of all teachers, which is in line with the evaluation concept of mutual learning and common growth advocated by developmental teacher teaching evaluation. Peer evaluation can not only emphasize the equal status of the evaluated teachers and the evaluators, promote equal communication, but also fully mobilize the enthusiasm of teachers and allow teachers to fully participate in all aspects of evaluation. In the evaluation process, the supervisor absorbs the opinions of teachers and students, gives corresponding solutions to specific problems, adjusts the relationship between teachers and students at the organizational level, and coordinates the work by the teachers' ethics and norms and the requirements of the state and the teaching school. As the person in charge of the teaching unit, the leader bears the responsibility and obligation to supervise the teaching quality of teachers, and plays a role in macro-control in the entire evaluation process, so the evaluation of teachers' teaching by leaders has become the proper meaning of quality evaluation.

Figure 1 below illustrates the implementation path of a data-driven teacher evaluation system. By summarizing the audio-visual images and feedback reports in the teaching process, the teacher's teaching ability is regulated as input values, and the relevant conclusions are summarized through analysis to form the methodology and improvement points of follow-up teaching, so as to achieve better teaching results.

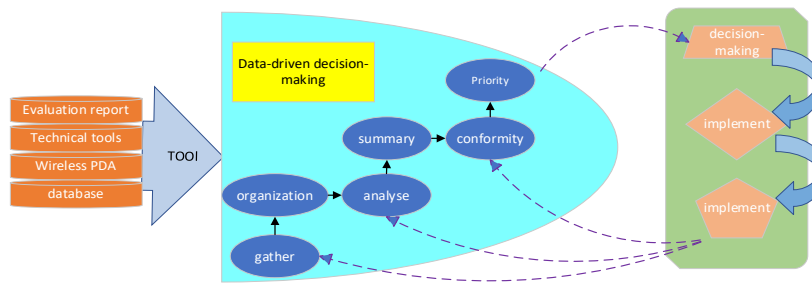


Figure 1: Data-driven evaluation decision graph.

3.2 Selection of Evaluation Indicators

The evaluation indicators should cover all aspects of teaching activities to ensure the accuracy and science of the evaluation results, not only involving the classroom teaching evaluation but also the extracurricular teaching evaluation. Different evaluation subjects should adopt different evaluation index systems, learn from each other's strengths, and give full play to their respective evaluation advantages. Classroom teaching activities are regarded as the basic items, and extracurricular teaching and research are taken as plus points, which complement each other and promote each other. Specifically, it includes pre-class preparation, in-class teaching, and after-class tutoring, and the number of indicators should not be too much, as too much will consume the patience of teachers and students. In terms of formulation, it is necessary to follow the SMART principle, i.e., specific, measurable, attainable, relevant, and time-bound. Combined with the concept of developmental evaluation, the teacher teaching evaluation index can be divided into three levels: the first-level index is a summary of the first-level index, the second-level index is subdivided, and the third-level index is a descriptive analysis of the second-level index. Specifically, it includes teacher ethics, teaching preparation, teaching attitude, teaching discipline, teaching content, teaching ability, teaching methods, teaching art, teaching effect, learning effect, classroom management, after-class guidance, course assessment, developmental evaluation, etc. The evaluation index designed in this paper draws on the existing evaluation system of colleges and universities in Harbin, which is more reliable and practical and is enriched and improved based on it, especially the developmental evaluation index involving teaching evaluation and reflection.

(1) Teacher ethics and style: In October 2020, the Central Committee of the Communist Party of China and the State Council issued the "Overall Plan for Deepening the Reform of Educational Evaluation in the New Era," proposing to "take the effectiveness of moral cultivation as the fundamental standard for testing all the work of the school, and take the teacher ethics and teacher style as the first standard for evaluating the quality of teachers." Teacher ethics is the teacher's personal moral quality, teacher style is the fashion of the teacher team, Lide Shuren to achieve the unity of "morality" and "morality teaching," based on overcoming the phenomenon of teachers emphasizing scientific research and ignoring teaching, the evaluation of teacher ethics should become a necessary premise and important content of teachers' teaching evaluation.

(2) Teaching preparation: Adequate pre-class preparation is a necessary prerequisite for the realization of effective teaching theory, and teachers can ensure the orderly progress of teaching activities by making adequate preparation before teaching. The teacher's task is not only to teach students knowledge but also to set up a good learning environment, provide students with practical supplementary materials, and increase students' opportunities for self-directed learning so as to ensure the improvement of academic performance and self-learning ability. Provide students with rich and effective extracurricular teaching resources, including online courses, audio and video resources, literature and books, forum reports, etc.

(3) Teaching attitude: Gagne proposed in his theory of the learning process that the first step of

instructional design is to set teaching goals; purposeful teaching is more effective than purposeless teaching; the teacher should let students clarify the teaching tasks and goals before starting the teaching session, and stimulate students' learning motivation in an environment of thirst for new knowledge and finding answers. The setting of goals should be in line with the discipline curriculum standards, in line with the talent training goals, and should have a corresponding relationship with the index points of students' graduation requirements: it should be conducive to the healthy development of students and the cultivation of professional emotions.

(4) Teaching discipline: Passion for teaching is an important indicator in the teaching competency model of college teachers. Only with high moral qualities can teachers train high-quality college students through words and deeds and set a good example for students. Every aspect of teachers' teaching management should conform to the teaching ethics norms, and teachers should be down-to-earth, decent, and disciplined so as to achieve teaching education, management education, and service education.

(5) Teaching content: Teachers should organically combine the ontological knowledge of the subject and the frontier knowledge of the subject and creatively develop curriculum resources. The goal of education in our country is to cultivate qualified builders and reliable successors of the socialist cause with all-round development of morality, intelligence, physical fitness, aesthetics, and labor; on the one hand, schools must strengthen ideological, political, historical, and cultural education, and on the other hand, it is necessary to supervise and evaluate the implementation of teachers, so as to ensure the correct direction of personnel training. From the perspective of pedagogical principles, the scientificity and correctness of teaching content are the primary principles of teaching, no matter how outstanding the teacher's teaching enthusiasm and teaching methods are, once the teaching content deviates from the curriculum standards, it cannot be called an excellent teacher in the evaluation results.

(6) Teaching ability: Teachers should choose appropriate methods to organize teaching and use infectious teaching language to carry out teaching, which can not only accurately and vividly convey knowledge but also enlighten students' hearts and imperceptibly make good examples. Teachers' professional knowledge should be solid and profound, and they should insist on learning cutting-edge knowledge and improving the knowledge system. Teachers can use different theories to compare and contrast in teaching, be scientifically accurate, have abundant knowledge points, be logical, clearly introduce the important views of well-known scholars or experts, and fully discuss the latest developments in the subject area.

(7) Teaching methods: On the one hand, teachers should be good at using scientific theories and research methods to solve problems, update professional knowledge in the subject area on time, and use new ideas and new insights to innovatively construct a teaching model for the coordinated development of teachers and students. On the other hand, teachers should develop information-based teaching capabilities, use modern technology and means to assist teaching, and use a variety of teaching methods such as discussion, inquiry, cooperation, case study, and project-driven to serve to teach through a variety of ways.

(8) The Art of Teaching: The "Dr. Fox Effect" argues that enthusiastic teachers are more likely to stimulate students' attention and interest in learning and that these interests have nothing to do with the content itself. Learning interest is the manifestation of intrinsic motivation in learning, and the sound and color of the lesson can promote students' understanding of the learning object or the tendency to approach. Teachers should have strong organizational and coordination skills, adaptability and timely evaluation skills, moderate speaking speed, vivid language, and a friendly and appealing teaching style.

(9) Teaching effect: Teaching is an interactive process between teachers' "teaching" and students' "learning.", teachers' teaching evaluation should fully consider students' actual feelings and learning effects and other internal development characteristics and strengthen the evaluation of students' knowledge, ability, values, etc., which can motivate teachers to practice the mission of teaching and educating people. In the process of teaching, teachers should create attractive learning situations, make teaching knowledge concrete and interesting, teach flexibly, and stimulate students'

enthusiasm for learning.

(10) Learning effect: According to the theory of "zone of proximal development" by the former Soviet psychologist Vygotsky, in addition to leading students to learn new knowledge on their own knowledge, teachers should also discover or organize appropriate problem situations in the area of possible development to stimulate students' intelligence. Teachers need to stimulate students to think in multiple dimensions, take students' problems as the starting point, form a dynamically generated teaching process, and cultivate students' critical and creative thinking.

(11) Classroom management: Teachers should respect the individualized differences in students' physical and mental development levels, treat students equally and teach students according to their aptitude. Students' participation in teachers' projects and production practices combines discipline construction and talent training, and teachers guide students to participate in scientific research or achievement experiments in essence, which is a process of educating people. Teachers in colleges and universities should uphold "natural" thinking, deeply integrate teaching and learning, and communicate sincerely and amicably with students to give them spiritual motivation.

(12) After-class instruction: Teacher teaching includes not only classroom teaching time but also input into teaching activities outside of the classroom. The second classroom is an important carrier for colleges and universities to cultivate new talents in the era of all-round development, and an important practical teaching link for cultivating innovative talents. Teachers guide students to participate in scientific research competitions innovation and entrepreneurship competitions and other competitions, which belong to the assessment of public service quantity and include implicit teaching work such as extracurricular tutoring, homework correction, and social practice into the evaluation content, which has an important guiding role in the teaching attractiveness of teachers.

(13) Course assessment: American psychologist Adams's fairness theory points out that fairness can make people feel satisfied and stimulate people's enthusiasm for work, while unfairness will produce negative emotions and affect work motivation. Unfair grading of student work and exams by teachers can lead to a decline in students' academic performance and moral character. The course assessment should cover all course teaching objectives and be able to provide evidence, each link participating in the evaluation should have clear and reasonable assessment standards, and the assessment scores of each item should be appropriately weighted.

Figure 2 below shows the roadmap of the data-driven teaching system proposed in this paper, firstly, the data of five evaluation subjects are collected, and secondly, the weighting functions of different subjects are assigned to different subjects through the Likert vector scale and analytic hierarchy process, and the evaluation indicators based on data-driven adjustment are applied to the dataset of relevant teaching indicators, and finally the feasibility and accuracy of the model are feedback.

3.3 Development of Evaluation Criteria

The evaluation criteria are specific provisions on the evaluation content, and the formulation of evaluation criteria can distinguish effective teaching from ineffective teaching. Establish clear criteria for different evaluation indicators, quantify them as much as possible, and use different performances of key events to anchor what cannot be quantified. Based on the positioning of developmental evaluation, the teaching evaluation system of college teachers should fully consider the subject background, age, career planning, and professional development stage of teachers and enhance the pertinence of evaluation criteria as much as possible so as to promote the development of teachers effectively. The teaching evaluation standard of teachers is composed of three parts: effectiveness, responsibility, and quality. The effectiveness standard is in the core position, and the responsibility standard and quality standard are carried out around the effectiveness standard. The performance criterion is an evaluation criterion established from the perspective of work results. Teaching evaluation mainly examines students' mastery of basic knowledge and basic skills, the development of intelligence and ability, and the formation of ideology and morality.

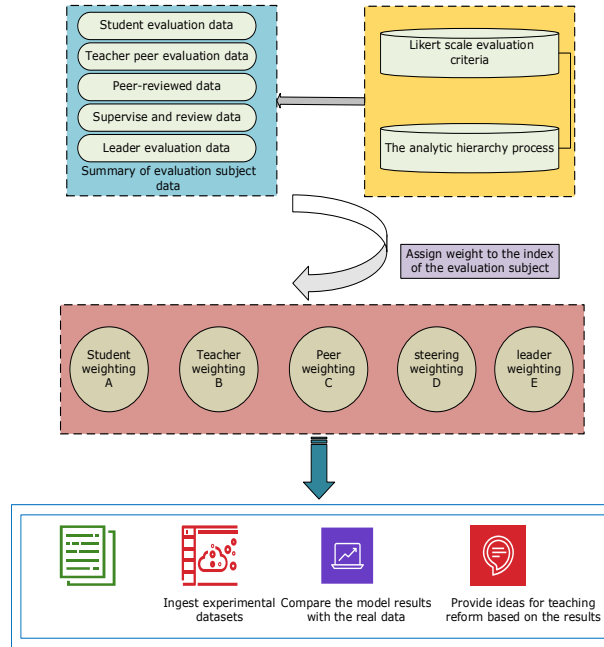


Figure 2: Flowchart of the evaluation of the main body of evaluation.

The responsibility criterion is an evaluation criterion established from the perspective of the task to be completed by the evaluation object. Teachers' responsibilities are measured by the quality of test preparation, teaching quality, homework weight, after-class tutoring, test results, etc. The quality standard is a standard from the perspective of ideology and culture that should be possessed to undertake various responsibilities. University teachers should have noble ideological character, profound academic knowledge, scientific, educational theories, good teaching quality, lofty spiritual realm, and other basic qualities. In addition, good evaluation criteria should be tested for applicability in diverse teaching contexts, the specificity of the criteria should be determined, and a balance should be struck between clear evaluation criteria and individualization so as to improve the scientificity and fairness of evaluation. The evaluation criteria in this paper refer to the Likert scale, and each statement has five responses: "strongly agree", "agree", "general", "disagree" and "strongly disagree", which are recorded as 5 points, 4 points, 3 points, 2 points, and 1 point, as shown in Table 1. The Likert scale uses the method of adding up the scores, and the total score of each item can indicate the strength of the respondent's attitude or his different states on the scale. In order to simplify the size of the score, save the evaluator's thinking time, and facilitate the evaluation subject to score the teacher's teaching evaluation, we use 1-5 points as the specific standard of evaluation instead of the traditional 100-point standard.

Grade	Standard	Score
Strongly agree	Completely up to par and very good	5
Agree	basically meets the standard and performs well	4
Average	Barely up to standard, the performance effect is average	3
Disagree	It does not meet the standard and needs to be strengthened and improved	2
Strongly	It is far from the standard and requires individual guidance	1

 disagree

Table 1: Evaluation criteria.

3.4 Classification of Indicator Weights

The weight of the index is used to reflect the relative importance of the indicator in the whole indicator system, and it is necessary to reasonably allocate the weights between the educational performance evaluation indicators and the developmental evaluation indicators to ensure the objectivity of the evaluation results. This paper uses the literature research method to sort out and analyze the existing evaluation system, and at the same time, regarding the construction of the evaluation index system of individual universities, the relative importance of indicators at all levels is determined by the analytic hierarchy process. The analytic hierarchy process can be used for decision-making analysis and comprehensive evaluation of multi-objective, multi-factor, and multi-level complex problems. The specific steps are as follows:

First, list the set of indicators that need to be weighted.

Secondly, according to the relative importance, the indicators at the same level are compared in pairs, as shown in Table 2.

Relative importance	Meaning
1	The two factors are equally important.
3	Compared to the two factors, one is slightly more important than the other.
5	Compared to the two factors, one is significantly more important than the other.
7	One is more strongly important than the other compared to the two factors.
9	Compared to the two factors, one is more important than the other.
2, 4, 6, 8	The middle value of two degrees of proximity

Table 2: Importance scores.

Thirdly, the structure of pairwise comparison in the indicators is listed; taking the teaching evaluation subjects as an example, the five subjects are: student evaluation (A), teacher self-evaluation (B), peer evaluation (C), supervision evaluation (D), and leadership evaluation (E). If A is compared with B, A is slightly more important than B, then 2 is recorded at the intersection of A and B, and 1/2 is counted at the intersection of B and A; A is compared with C, A is slightly more important than C, 3 is recorded at the intersection of A and C, and 1/3 is counted at the intersection of C and A; A is compared with D, A is equally important than D, 1 is recorded at the intersection of A and D, and 1 is counted from the bottom at the intersection of D and A; Compare A with E, A is slightly more important than E, then write 3 at the intersection of A and E, 1/3 from the bottom of E and A, compare B with C, B is slightly more important than C, 2 at the intersection of B and C, and 1/2 at the intersection of C and B; The intersection of E and B is counted as 1/3 of the reciprocal, and so on to obtain all the comparison data, as shown in Table 3.

A	B	C	D	E	M_i	\overline{W}_i	W_i
---	---	---	---	---	-------	------------------	-------

A	1	2	3	1	3	18	1.783	0.316
B	1/1	1	2	1/2	3	1.5	1.084	0.192
C	1/3	1/2	1	1/3	1	0.056	0.561	0.100
D	1	2	3	1	2	12	0.644	0.292
E	1/3	1/3	1	1/2	1	0.056	0.561	0.100

Table 3: Weight scoring table.

Finally, the weights of each indicator are calculated, and the product of each row element is calculated according to Equation (1).

$$M_i = \prod_{j=1}^n u_{ij} \quad (1)$$

At the same time, the geometric mean value of each row of elements is calculated according to Equation (2).

$$\bar{W} = \sqrt[n]{M_i} \quad (2)$$

And finally, the values are normalized according to Equation (3), and the index weight \bar{W}_i is calculated.

$$W_i = \frac{\bar{W}_i}{\sum_{i=1}^n \bar{W}_i} \quad (3)$$

In order to ensure the correctness and rationality of the weights, it is also necessary to carry out the consistency test on the judgment matrix, that is, calculate the CR value, and when the consistency ratio according to Equation (4), it indicates that the inconsistency degree of the judgment matrix is within the allowable error range, and has strong reliability. The CI of stochastic consistency indicator is shown in the table below.

$$CR = \frac{C_1}{R_1} < 0.1 \quad (4)$$

Calculating the maximum eigenvalue of the judgment matrix according to Equation (5):

$$\lambda_{\max} = \sum_{i=1}^n \frac{(AW)_i}{nW_i} = 5.096 \quad (5)$$

Calculating the consistency metrics according to Equation (6):

$$CI = \frac{\lambda_{\max} - n}{n-1} = 0.024 \quad (6)$$

n	1	2	3	4	5	6	7	8	9	10
---	---	---	---	---	---	---	---	---	---	----

RI	0	0	0.58	0.9	1.12	1.24	1.32	1.41	1.45	1.49
----	---	---	------	-----	------	------	------	------	------	------

Table 4: RI of the stochastic consistency index

$n=5$, $CR = \frac{0.024}{1.12} = 0.021 < 0.1$, The judgment matrix passes the consistency test. According to the above steps, the weights of the evaluation indicators of the five evaluation subjects were calculated respectively.

4 EXPERIMENTAL ANALYSIS

4.1 Indicator Analysis

In the third section, the index weight analysis of each evaluation subject of the teacher evaluation system is realized, and the evaluation criteria mentioned above can be classified into evaluation subjects. Combined with the real teaching evaluation data indicators collected from a university, the subject classification and weight of each index are set, as shown in the following table.

Evaluation subject	Evaluation indicators	Weight
student	Teaching Preparation, Teaching Attitude, Teaching Content, Teaching Discipline, Teaching Ability, Teaching Methods, Learning Effect	0.316
teacher	Teaching preparation, teaching content, teaching effectiveness, classroom management, after-class guidance	0.192
peer	Teaching content, teaching ability, teaching method, teaching art	0.100
steering	Teacher ethics, teaching discipline, teaching effectiveness, classroom management	0.292
leadership	Teacher ethics, teaching ability, teaching art, learning effect	0.100

Table 5: Weighting table.

The results of the teacher's score calculation based on the evaluation indicators of the student subject are shown in Figure 3-7 below, the purple squares represent the results of the calculation using the system proposed in this paper, and the orange dots represent the real evaluation data. It is not difficult to find that the results of the proposed method are highly approximate to the real results, and there are only a few deviations, which proves the reliability of the application of the proposed system for student subject evaluation, and the purple triangle on the right represents the difference between the evaluation value of the system and the real situation, and the errors of the two are mostly concentrated within 1 point, and the error peak is also low, which verifies the accuracy of the proposed method.

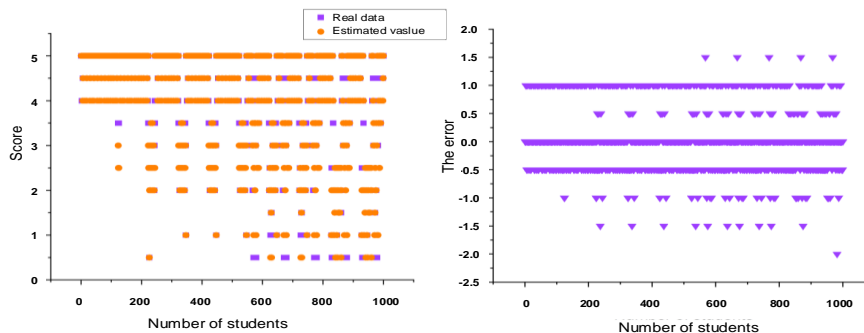


Figure 3: Chart of student evaluation indicators.

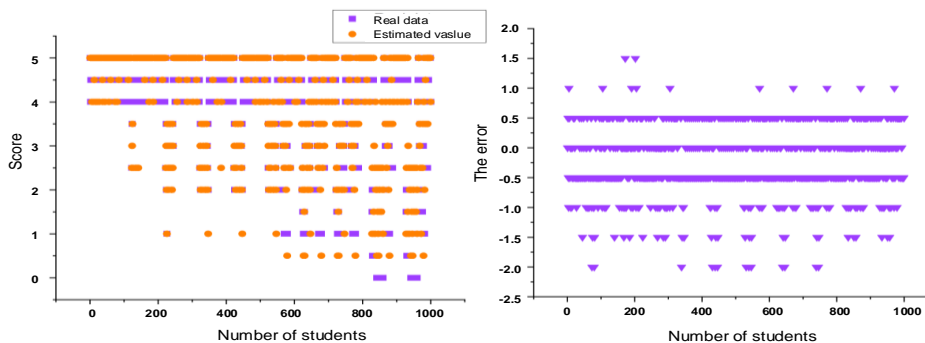


Figure 4: Diagram of teacher evaluation indicators.

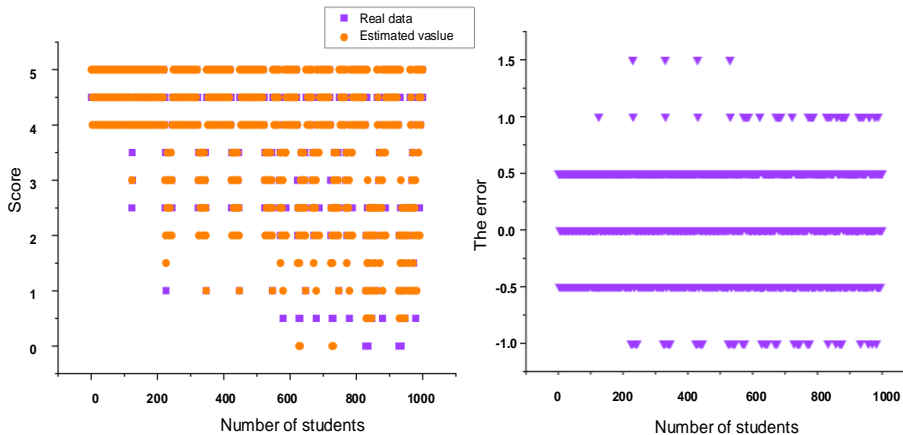


Figure 5: Peer evaluation indicator graph.

By analyzing the real evaluation data of teachers by each evaluation subject from Figure 3 to Figure 7 above and the scores calculated by the evaluation model, as shown in the following table, it can be found that the model is extremely close to the real situation of the teacher, indicating that the evaluation model has high applicability to each evaluation subject, ensures high accuracy, and can be used for the subsequent teacher evaluation system of each school.

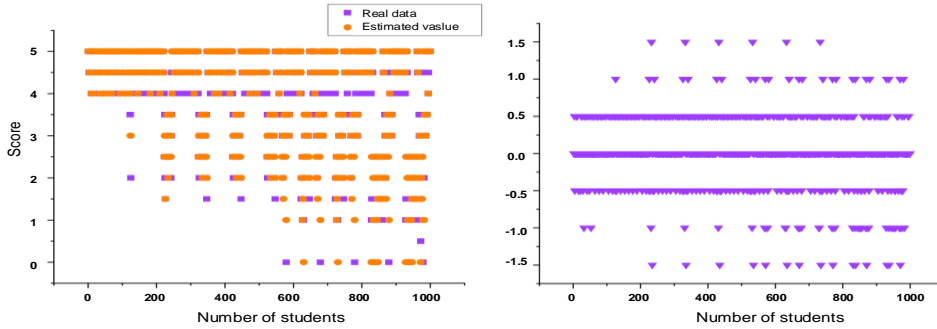


Figure 6: Diagram of supervision and evaluation indicators.

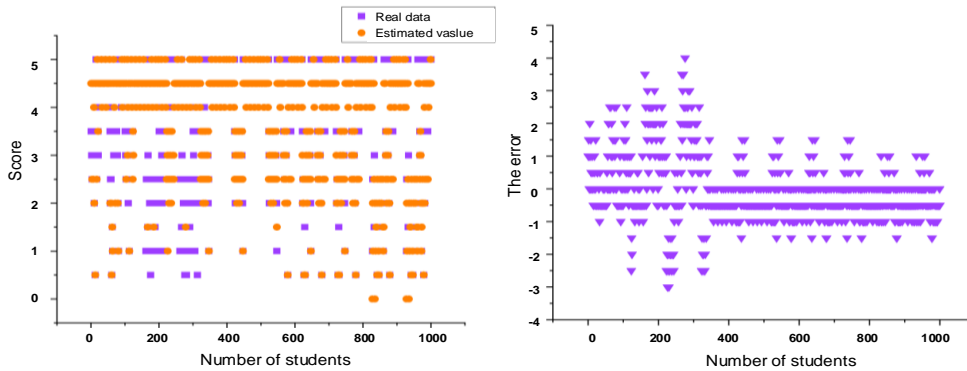


Figure 7: Diagram of leadership evaluation indicators.

	Real data	System evaluation data	Accuracy
Student Rating A	4.19	4.27	98.13%
Teacher self-assessment B	4.03	3.99	99%
Peer Review C	4.01	4.12	97.33
Supervision Evaluation D	3.92	3.97	98.74%
Leadership Evaluation E	3.71	3.83	96.61

Table 6: Accuracy of each evaluation subject.

4.2 Comprehensive Indicator Analysis

After the experimental analysis of each subject proves the feasibility and accuracy of the evaluation model, this section will conduct a comprehensive analysis of the teacher's evaluation results based on the evaluation indicators of each subject and comprehensively analyze the source of the results and the weighted average of the scores of the evaluation subjects for each evaluation index. In order to ensure the randomness of the experiment, the real data of 50 groups of evaluation indicators of each evaluation subject and the teacher were randomly selected. The new comprehensive evaluation score was calculated by using the model proposed in this paper. The average value was calculated respectively, and the median score and variance are shown in Figure 8 below. It can be found that

the proposed method is very close to the real score in three dimensions, and the discrete degree of the two is small, which proves the credibility of the proposed model.

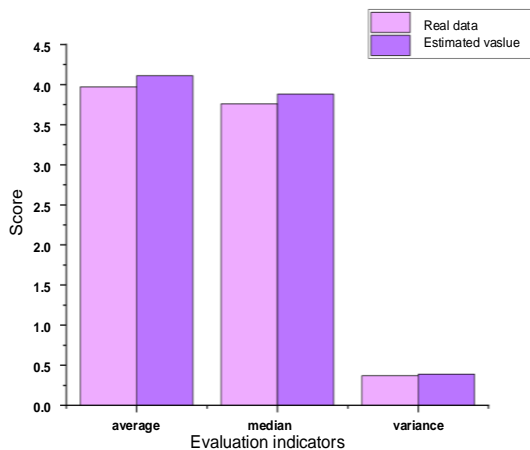


Figure 8: Comprehensive scoring index

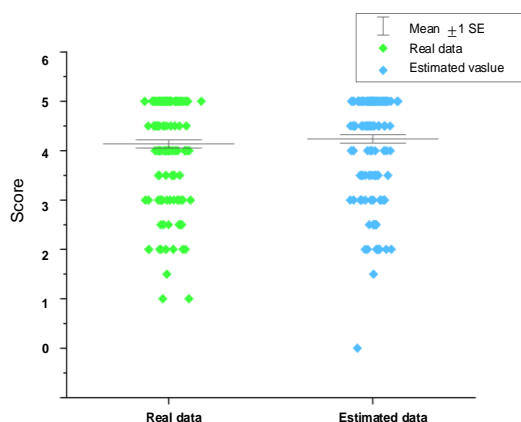


Figure 9: Discrete plot of the overall score.

Figure 9 below shows the degree of dispersion between the prediction score and the true score of the model proposed in this paper in the intercepted data, and it can be found that there is a high degree of aggregation between 3 and 5 points, which is consistent with the characteristics of the original data because the score range is 0-5 points and the interval is 0.5 points, there is a high degree of similarity between the two in the same score range, and the degree of dispersion is relatively similar. It is proved that the data-driven analytic hierarchy process weighted teacher evaluation model proposed in this paper has high credibility and feasibility under the comprehensive teacher rating.

5 CONCLUSIONS

Based on the existing teaching data, this paper constructs a model of teacher teaching evaluation system based on the combination of data-driven artificial intelligence technology and hierarchical statistical analysis. Firstly, the principle of data-driven teaching evaluation proposed in this paper is introduced, and the analytic hierarchy process is used to select different evaluation indicators for

different evaluation subjects. Besides, the weights of each index are obtained respectively, and the data are brought into the evaluation data of a university to prove that the evaluation department proposed in this paper has high applicability and credibility. Compared with the classical average weighting of each evaluation index, this paper adjusts the applicability of the weights of the five evaluation subjects and takes student feedback and supervision evaluation as the higher weights of the influencing factors, which not only ensures the initiative of students in the teaching subject but also makes up for the lack of students' professional perspective in the evaluation of teaching ability, ensuring the comprehensiveness and credibility of the system proposed in this paper. Finally, the model of this paper is fully verified under the 1000 sets of data in this experiment, which provides an effective model system for the subsequent teaching evaluation system.

ACKNOWLEDGEMENTS

This article was funded by Tianjin Undergraduate Teaching Quality and Teaching Reform Research Project(B231006901); Chinese Electronic Education Society Education and Teaching Reform Research Project(DJY23013); Tianjin Experimental Teaching and Laboratory Construction Research Project(2024); Tianjin Ideological and Political Demonstration Courses and Tianjin First-class Undergraduate Course-Power Electronics Technology(2023); Tianjin University of Commerce Research Project on Graduate Education and Teaching Reform(2023). Tianjin Municipal First-class Undergraduate Major Construction Site - Automation(2021). Tianjin University of Commerce Educational Reform Project on Undergraduate Education (TJCUJG2023057).

Ying Mao, <https://orcid.org/0009-0006-7173-8362>
Fenghong Wang, <https://orcid.org/0009-0007-3495-2500>
Yilin Wang, <https://orcid.org/0009-0000-9010-0117>
Renjie Li, <https://orcid.org/0009-0005-3488-1146>
Qi Chen, <https://orcid.org/0009-0003-8210-7978>
Yong Zhang, <https://orcid.org/0000-0001-8854-1319>

REFERENCES

- [1] David, C.; Chris, J.; Madison, M.; Alex, M.; Madeline, P.; Amy, M. F.: Teaching Through the Student Lens: Qualitative Exploration of Student Evaluations of Teaching, *American Journal of Pharmaceutical Education*, 88(3), 2024, 100672. <https://doi.org/10.1016/j.ajpe.2024.100672>
- [2] Li, B.; Guo, W.: Quality assessment of intelligent physical education teaching in universities based on multivariate statistical analysis and regression analysis, *Applied Mathematics and Nonlinear Sciences*, 9(1), 2024. <https://doi.org/10.2478/AMNS.2023.2.00914>
- [3] Zhang, L.; Jiang, S.; Shi, G. X.: Intelligent Learning-Driven College Physical Education Teaching Method based on Fuzzy Evaluation Theory, *Computer-Aided Design and Applications*, 21(S9), 2024, 186-204. <https://doi.org/10.14733/cadaps.2024.S9.186-204>
- [4] Zhu, C; Zhang, Y; Xu, H. P.; Chen, Q.; Fu, Q. Y.: Improved Harris Hawks Optimization algorithm based on quantum correction and Nelder-Mead simplex method, *Mathematical Biosciences and Engineering*, 19(8), 2022, 7606-7648. <https://doi.org/10.3934/mbe.2022358>
- [5] Zhao, H.; Guo, L. N.: Design of Intelligent Computer-Aided Network Teaching System Based on Web, *Computer-Aided Design and Applications*, 19(S1), 2022, 12-23. <https://doi.org/10.14733/cadaps.2022.S1.12-23>
- [6] Zhang, Y.; Li, X. Y.; Wang, L.; Fan, S. R.; Zhu, L.; Jiang, S. H.: An Autocorrelation Incremental Fuzzy Clustering Framework Based on Dynamic Conditional Scoring Model, *Information Sciences*, 648(11), 2023, 119567. <https://doi.org/10.1016/j.ins.2023.119567>

- [7] Steve, C.; Duncan, W.; Robert, W.: Performance evaluation in teaching: Dissecting student evaluations in higher education, *Studies in Educational Evaluation*, 81, 2024, 101342. <https://doi.org/10.1016/j.stueduc.2024.101342>
- [8] Zhang, Y.; Li, R. J.; Chen, Q.; Zhi, D. R.; Wang, X. L.; Feng, C. Z.; Shang, J. H.; Jiang, S. H.: An improved bicubic interpolation SLAM algorithm based on multi-sensor fusion method for rescue robot, *International Journal of Sensor Networks*, 42(2), 2023, 125-136. <https://doi.org/10.1504/ijnsnet.2023.131656>
- [9] Gomez, V.: The Evaluation of Teacher Performance in Higher Education. *Propositos Representaciones*, 7(2), 2019, 479-515. <https://doi.org/10.20511/pyr2019.v7n2.255>
- [10] Peck, C. A.; Young M. G.; Zhang W. Q.: Using Teaching Performance Assessments for Program Evaluation and Improvement in Teacher Education, National Academy of Education, 2021. <https://doi.org/10.31094/2021/3/3>
- [11] Carrrell, S. E.; West J. E.: Does Professor Quality Matter Evidence from Random Assignment of Students to Professors, *Journal of Political Economy*, 118(3), 2010, 409-432. <https://doi.org/10.1086/653808>
- [12] Lyde, A. R.; Grieshaber, D. C.: Faculty Teaching Performance: Perceptions of a Multi-Source Method for Evaluation (MME), *Journal of the Scholarship of Teaching and Learning*, 16(3), 2016, 82-94. <https://doi.org/10.14434/josotl.v16i3.1814>
- [13] Paufler, N. A.; King, K. M.: Delivering on the Promise of Support for Growth Evaluator Perceptions of a New State Teacher Evaluation System, *Journal of Educational Supervision*, 3(2), 2020, 32-50. <https://doi.org/10.31045/jes.3.2.3>
- [14] Hoben, J. L.; Badenhorst C.: Student Evaluations and the Performance of University Teaching:/Teaching to the Test, *Learning Landscapes*, 13(1), 2020, 161-172. <https://doi.org/10.36510/learnland.v13i1.1012>
- [15] Luis, M. L.; Santiago, L. G.; Luis-Miguel, D. P.: Converting Likert Scales Into Behavioral Anchored Rating Scales(Bars)For The Evaluation of Teaching Effectiveness For Formative Purposes, *Journal of University Teaching & Learning Practice*, 16(3), 2019. <https://doi.org/10.53761/1.16.3.9>
- [16] Byrne, V. L.; Donlan, A. E.: Presenting a Validated Mid-Semester Evaluation of College Teaching to Improve Online Teaching, *Online Learning Journal*, 24(2), 2020, 94-110. <https://doi.org/10.24059/olj.v24i2.2126>
- [17] Courey, K. A.; Lee, M. D.: A Model-Based Examination of Scale Effects in Student Evaluations of Teaching, *AERA Open*, 7(1), 2021, 1-13. <https://doi.org/10.1177/23328584211040083>
- [18] Bao, T. T.; Liu, Y. F.; Yang, Z. Z.; Wu, S. H.; Yan, Z. L.: Evaluating sustainable service quality in higher education from a multi-stakeholder perspective: An integrated fuzzy group decision-making method, *Socio-Economic Planning Sciences*, 92, 2024, 101849. <https://doi.org/10.1016/j.seps.2024.101849>
- [19] Fajelkova A.; Fejfarova M.: Evaluation of the Quality of Teaching from the Perspective of University Students, *Journal on Efficiency and Responsibility in Education and Science*, 12(2), 2019, 34-40. <http://doi.org/10.7160/eriesj.2019.120201>
- [20] Boswell, S. S.: Effects of Ratemyprofessors.com and university student evaluations of teaching on students' course decision-making and self-efficacy, *Higher Learning Research Communications*, 10(2), 2020, 1-20. <https://doi.org/10.18870/hlrc.v10i2.1194>