International Journal for Housing Science and Its Applications
Publish August 4, 2025. Volume 46, Issue 3 Pages 1521-1531

https://doi.org/10.70517/ijhsa463117

Evaluation model of physical education teaching based on stochastic simulation algorithm

Gaoxun Zhang[1,*]
[1] Beihua Aerospace University, Langfang, Hebei, 065000, China
Corresponding authors: (e-mail: 13931578398@163.com).

Abstract With the deepening development of digital technology, the evaluation of physical education teaching also changes with the development of teaching theory and mode. Based on the relevant theoretical foundation, this paper proposes an evaluation method based on stochastic simulation. The algorithm utilizes the autonomous dominance matrix and dominance power vector calculation to calculate the superiority matrix and obtain the comprehensive evaluation vector and the best ranking of evaluation objects. Taking four sports colleges as an example, the proposed evaluation method is applied to solve the results of the ranking of sports colleges to verify the feasibility of the method. Through the consistency test of different evaluation models, it can be found that the first and last consistency rate of this paper's model reaches 50%, which exceeds the required 20%. Therefore, it has high credibility, good interpretability, small amount of information loss and good consistency. The results of the applied case study show that the rankings of the teaching evaluation values of the four institutions are: institution 4 (2.2-2.4) > institution 1 (2.2) > institution 2 (2.0-2.2) > institution 3 (1.8) at each number of stochastic simulations, and the model is statistically stable.

Index Terms stochastic simulation, superiority matrix, autonomous superiority measure matrix, superiority weight vector, evaluation methodology

I. Introduction

Physical education is no less important than other courses designed to cultivate students' morality, intelligence and aesthetics, and is a planned, purposeful and organized teaching activity in colleges and universities [1]. In the development of modern physical education teaching, affected by the implementation of the new curriculum standards, the traditional teaching mode, teaching concepts have been unable to meet the needs of the development of modern education and teaching, and the overall teaching content is not perfect, and the monolithic evaluation methods can not reflect the importance of physical education teaching evaluation [2]-[4]. In this regard, it is necessary to comprehensively reform the physical education teaching curriculum in colleges and universities, physical education teaching evaluation innovation, so as to meet the learning needs of students.

The construction of physical education teaching evaluation system in colleges and universities is not only the management of teachers, but also provides important evaluation standards for teachers [5]. Teachers in the process of practical teaching, according to the students' personal situation, combined with the analysis of the teaching evaluation system, the teaching mode, teaching content appropriate adjustments, so as to meet the learning needs of students, so that the teaching evaluation system in college sports teaching, give full play to its own important role and value [6]-[8]. The construction of the teaching evaluation system is based on a comprehensive analysis of the needs of sports teaching in colleges and universities, so that the evaluation of sports teaching in colleges and universities is diversified, focusing on highlighting the student's main position, which not only conforms to the teaching concept of the new standards, but also improves the overall teaching efficiency and quality [9]-[11].

In previous studies, most scholars in China mostly use gray correlation analysis, hierarchical analysis and fuzzy comprehensive evaluation method to carry out physical education teaching quality evaluation. For example, Wu, H. et al. established a physical education teaching evaluation model based on gray cluster analysis, which was applied to educational practice activities, and the validity of the evaluation results could still be ensured for the small sample data set of physical education courses [12]. Han, K. explored the specific factors affecting the quality of physical education teaching in colleges and universities by using hierarchical analysis (AHP) and constructed an index system by combining the gray system theory, and the proposed evaluation model of physical education teaching played an important role in developing the physical education curriculum in colleges and universities [13]. Chen, J. and Huang, H., on the other hand, combined the knowledge of fuzzy mathematics with the hierarchical analysis method to establish a fuzzy hierarchical analysis (FAHP) model oriented to the assessment of the quality of physical education teaching in colleges and universities, whose assessment results can accurately guide the physical education teaching of colleges and universities to be carried out in a scientific way [14]. Gui, Y. incorporates the multi-criteria decision making (MCDM) algorithm based on the fuzzy hierarchical analysis method in order to form a hybrid evaluation model, which utilizes a complex aggregation operator to deal with the fuzzy and uncertain factors in the teaching data, and then improves the accuracy of the evaluation model of physical education teaching [15]. Tang, M. et al. generated an evaluation index system for high-quality physical education with the help of expert evaluation, and further constructed the Fuzzy Comprehensive Evaluation (FCE) model, which is capable of comprehensively quantifying the results of physical education teaching and realizing the reasonable assessment of specific influencing factors [16]. However, the above evaluation methods may produce different evaluation conclusions on the same evaluation problem, and cannot avoid the problem of “non-consistency of multiple evaluation conclusions”. The use of probabilistic stochastic simulation algorithms can be used to judge the advantages of evaluation objects by calculating the depth of each evaluation object, and enhance the interpretability of the evaluation process of physical education.

In order to solve the problem of "non-consistency of multiple evaluation conclusions" caused by the traditional teaching evaluation model, this paper establishes a new evaluation model, namely the stochastic simulation evaluation model, based on the theory of stochastic simulation evaluation, combined with the relative evaluation problem of physical education teaching evaluation, and takes six factors, including syllabus, experimental teaching plan, pre-class preparation, teaching content, teaching method and teaching effect, as evaluation indicators. The stochastic simulation algorithm utilizes the autonomous superiority matrix and the superiority power vector calculation method to obtain the superiority matrix of the superiority-disadvantage relationship between the evaluated objects, and its comprehensive evaluation vector, which gives the optimal ranking among the evaluated objects. In terms of consistency, the stochastic simulation-based evaluation model is compared and verified with other traditional evaluation models to judge the rationality and advantages and disadvantages of the evaluation method proposed in this paper. It is applied to the practice of physical education teaching quality evaluation to discuss the effectiveness of the method.

II. Evaluation of physical education

Teaching evaluation mainly refers to the nature of education, guidelines and policies based on the premise of the establishment of the teaching objectives of the use of effective teaching methods and tools, the implementation of a variety of teaching activities in the implementation of the effect of the degree of student learning needs to make value judgments; is for the students in the process of learning to achieve the goals of the teaching of the act of a systematic quantitative and qualitative description, and ultimately give a certain realistic value of the process; but also refers to the objective evaluation criteria through the collection of various information on teaching activities, based on the information collected on the overall teaching activities and the effect of relatively objective measurement and scientific judgment. The process; at the same time also refers to the objective evaluation criteria, through a variety of teaching activities to collect information, based on the information collected on the entire teaching activities and the effect of relatively objective measurement and scientific judgment process. In short, teaching evaluation is a very important part of teaching, can judge the teaching effect, for the effective development of subsequent teaching activities to provide a basis [17].

Evaluation of physical education teaching is mainly guided by educational concepts and teaching ideas, according to the established standards, the value of each point in the classroom teaching judgment. A comprehensive and scientific evaluation of classroom teaching is the key to improving the quality of physical education teaching and promoting further reform of teaching. At the same time, it is also an important part of the educational evaluation system. Evaluation index system refers to the integration of multiple interlinked and independent indicators into a group of indicators based on a specific research purpose, which is mainly composed of various evaluation indicators, the weight of each indicator and the evaluation criteria. Evaluation of the indicators indicate the characteristics of the object to be evaluated. The weights of the indicators indicate the degree of importance of the indicators in the whole system of evaluation indicators, and are reflected by the corresponding numerical values. Evaluation standards are mainly used to measure the evaluation of the evaluated to achieve the standard required by the evaluation index items, the degree of achievement is commonly used in grades or quantitative scores.

In this paper, following the principles of scientific, goal-oriented and practicable construction of the index system, combined with the literature and the analysis of the current situation of the quality of physical education teaching in secondary schools, the preliminary development of the indicators for the evaluation of the quality of physical education teaching in schools. The teaching environment (syllabus, laboratory lesson plan, pre-course preparation), the teaching process (teaching content, teaching methods, teaching effect) is established as the second and third level indicators. The evaluation system is shown in Table 1.

Table 1: Evaluation system

Primary indicator	Secondary indicator	Tertiary index
Sports teaching evaluation system	Teaching environment	Syllabus
		Experimental teaching
		Pre-class preparation
	Teaching process	Teaching content
		Teaching method

III. Stochastic simulation-based evaluation model

III. A. Stochastic simulation-based integrated evaluation model

The stochastic simulation-based integrated evaluation model is oriented to the relative evaluation problem, but is still an information processing method based on the traditional integrated evaluation method, thus the stochastic simulation-based integrated evaluation model can be viewed as a structured methodological framework that consists of two parts, one of which is the randomized setup of the traditional evaluation method [18]. The other part is the solution of the relative evaluation problem based on the stochastic simulation algorithm. The relative evaluation problem, only through a single evaluation run can not directly give the conclusion of the program's advantages and disadvantages, need to be based on large-scale comparisons of the program's advantages and disadvantages of the relationship between the statistical analysis, so as to give the possibility of reflecting the relative advantages and disadvantages of the program between the ranking conclusions. Thus, first of all, the traditional evaluation method needs to be set up for randomization, so that each evaluation run becomes a large-scale run of random sampling, so that under the premise of sufficient number of simulations, the probability of comparing the advantages and disadvantages of the programs can be obtained in a relatively stable manner. From the above analysis, it can be seen that the operation of the traditional evaluation method is actually an information processing of the relative evaluation problem, thus the randomized setting of the traditional evaluation method should be reflected in the setting of the relevant parameters in the evaluation process. The most common of these is the indicator parameter, which, as mentioned above, is not usually fixed for relative evaluation questions, providing the possibility of “random sampling” of indicator values. In addition, the parameters that can be set randomly are also common in the indicator weights, expert authority (group evaluation), information variation processing (method nature measurement) and other aspects. This paper focuses on the main parameters involved in the various aspects of comprehensive evaluation, and gives a general idea of the transformation of traditional evaluation methods into a stochastic simulation-based evaluation model.

Based on the randomization setting of the traditional evaluation methods, it is necessary to compile a computer simulation program for large-scale simulation. The random information extracted each time is aggregated and solved according to the selected traditional comprehensive evaluation mode. Under the condition of sufficient simulation, the number of wins between each pair of observation schemes (evaluation objects) is observed. Form the probability judgment information of "superior" or "inferior" among the schemes. For multiple comparison scenarios an information matrix (called the superiority matrix) will be obtained. The superiority matrix contains the comparison information of each scheme, which needs to be further analyzed to obtain a probability ranking with probability information. Usually there are multiple possible orders implied in the superiority matrix, and how to derive an optimal order from the superiority matrix becomes the key to the problem. In this study, a method of defining an ideal order is given, and then an ordering among the solutions is derived in a way that is “closest to the ideal order”. In addition, we can also consider various solution ideas, such as "analyzing the basic features of the ideal ranking of superiority, and determining the reference ranking sequence according to the closeness to the ranking features of the ideal superiority, and then combining the pre-optimal ratio ranking conclusions to form the most stable ranking conclusions" or "obtaining the final sorting conclusion by selecting multiple sorting (chains) above a certain credible threshold and optimizing the synthesis of multiple chains". The algorithm for solving the possibility ranking will be discussed in depth with the basic rule of “further ensuring the stability of the ranking chain on the basis of increasing the possibility of the occurrence of the ranking chain”. Based on the above analysis, the framework of the stochastic simulation-based comprehensive evaluation model can be summarized as shown in Figure 1.

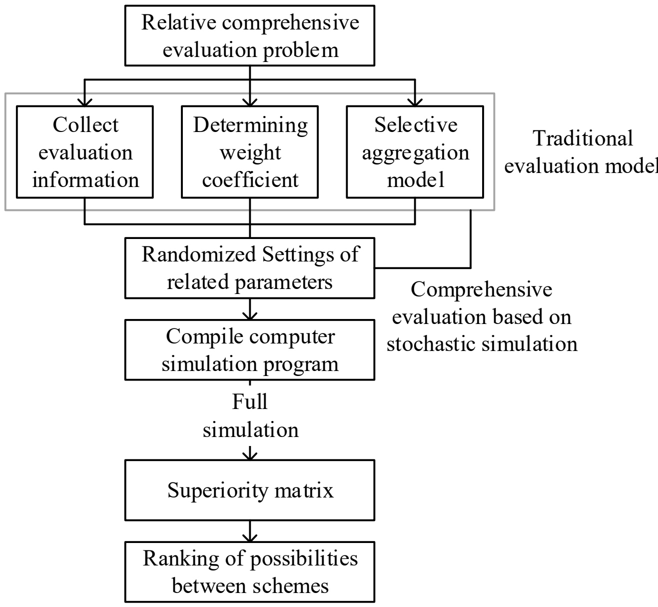

Figure 1: A framework for the comprehensive evaluation of stochastic modeling

III. B. Stochastic simulation algorithm

III. B. 1) Basic description of evaluation problems in physical education

There are n evaluated objects $u_1,u_2,\cdots,u_n$.The multi-indicator evaluation system $x_1,x_2,\cdots,x_m$ consisting of m indicators $x_{ij}=x_j(x_i)(i=1,2,\cdots,n;j=1,2,\cdots,m)$ is the evaluated object u_i is the observed value of the evaluated object x_j with respect to the indicator x_j. The evaluation data matrix (decision matrix) can be expressed as:

$$A=\left[x_{ij}\right]_{n\times m}=\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix} \tag{1}$$

where $m,n\geq 3$ and the data in A are preprocessed normalized data Physical education teaching evaluation process we describe as a general transformation:

$$y_i=f(x_{i1},x_{i2},\cdots,x_{in}),i\in N \tag{2}$$

where f is the positive transformation function; y_i is the comprehensive evaluation value of the evaluated object u_i, and $u_1,u_2,\cdots,u_n$ are ranked according to the value of $y_1,y_2,\cdots,y_n$ in ascending order. The comparison of the advantages and disadvantages of $u_1,u_2,\cdots,u_n$ can be accomplished by sorting the values from largest to smallest.

III. B. 2) Problem description

Let $\sigma_1,\sigma_2,\cdots,\sigma_n$ be the evaluated urban road objects, and $x,x_2,\cdots,x_m$ be the evaluation index system of the emergency evacuation capacity of urban roads. $x_{(ij)}(i=1,2,\cdots,n;j=1,2,\cdots,m)$ is the observed value of the evaluated road section σ_1 with respect to the indicator x_j. Denote $N=\{i=1,2,\cdots,n\},M=\{j=1,2,\cdots,m\}$, and the normalized data matrix is expressed as:

$$A=\left[x_{ij}\right]=\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix} \tag{3}$$

III. B. 3) Autonomous dominance matrix calculation

Define the autonomous dominance quantity matrix as:

$$B=\left[\lambda_{ij}\right]=\left[\mu\alpha_{ij}+\eta\beta_{ij}\right] \tag{4}$$

where $\alpha_{ij} = \sum_{k \subset N} (x_{ij} - x_{kj}) / (n-1), i \in N, j \in M$, reflecting the strength difference between the j th indicator of the evaluated roadway σ_i and the strengths of other $n-1$ evaluated roadways overall. strength difference between the evaluated road section as a whole. $\beta_{ij} = \sum_{p \in M, p \neq j} (x_{ij} - x_{ip}) / (n-1), i \in N, j \in M$, reflecting the difference in strength between the j th indicator of the evaluated roadway σ_i and other $m-1$ indicators as a whole. Indicator overall superiority difference; μ and η are competitive and developmental target favoritism coefficients respectively, $\mu, \eta \in [0,1], \mu + \eta = 1$.

III. B. 4) Calculation of the dominance weight vector

If $\omega = (\omega_1, \omega_2, \cdots, \omega_m)$ is a dominance power vector and $r_k (k = 2, \cdots, m)$ is an uncertainty ratio Judgment interval $\tilde{r}_k$ on the interval follows a uniformly distributed random variable (when the upper and lower bounds of the interval r_k are equal, r_k is in exact point-valued form), then:

$$\omega_i = \left(1 + \sum_{m=2}^{i} \sum_{k=m}^{i} r_k\right)^{-1}, \omega_{i-1} = r_k \omega_i \tag{5}$$

If we call $\omega_i = (\omega_{i1}, \omega_{i2}, \cdots, \omega_{in})\ i \in N$ a vector of dominance weights for the indicator dominance ordinal group about each evaluation object σ_i, then we can According to the autonomous dominance vector $\lambda_i = (\lambda_{i1}, \lambda_{i2}, \cdots, \lambda_{im})$ each component is sorted from the largest to the smallest, and corresponds to dominance weight vector $\omega = (\omega_1, \omega_2, \cdots, \omega_m)$ of each component is obtained [19].

III. B. 5) Superiority Matrix Calculation

If we express the superiority σ_i^* of the evaluated object σ_i' over $\sigma_i^* i' \neq i^*$ in terms of $s(\sigma_i' \succ \sigma_i^*)$, we have:

$$s(\sigma_i' \succ \sigma_i^*) = P(\sum_{j=1}^{m} x_{ij}' \omega_j^* > \sum_{j=1}^{m} x_{ij}^* \omega_j^*) + 0.5P(\sum_{j=1}^{m} x_{ij}' \omega_j^* = \sum_{j=1}^{m} x_{ij}^* \omega_j^*) \tag{6}$$

where P is the event probability, $0 \leq s(\sigma_i' \succ \sigma_i^*) \leq 1, s(\sigma_i' \succ \sigma_i^*) + s(\sigma_i'\ prec \sigma_i^*) = 1$; $s(\sigma_i' - \sigma_i^*) \Leftrightarrow s(\sigma_i' \succ \sigma_i^*) = s(\sigma_i' \prec \sigma_i^*) = 0.5, s(\sigma_i' \succ \sigma_i^*) = 0.5$

Using stochastic simulation to solve the superiority matrix and superiority power vector of the evaluated objects, the specific steps are as follows;

Step 1 Arbitrarily select 2 evaluated objects σ_i', σ_i^* (if there are n evaluated objects, a total of $n^2 - n$ comparisons need to be made) [20].

Step 2 sets the simulation count variable count.

Step 3 Apply the rand() function to randomly generate uniformly distributed uncertainty ratios $r_k (k = 2, \cdots, m)$ on a given interval, and determine the dominance weight vector $\omega = (\omega_1, \omega_2, \cdots, \omega_m)$.

Step 4 Set the count variables r_s, r_e, r_f to denote "$\sigma_i' \succ \sigma_i^*$", "$\sigma_i' = \sigma_i^*$" and "$\sigma_i' \prec \sigma_i^*$", i.e. when $\sum_{i=1}^{m} x_{ij}' \omega_j^* > \sum_{i=1}^{m} x_{ij}^* \omega_j^*$, then $r_s = r_s + 1$; when $\sum_{i=1}^{m} x_{ij}' \omega_j^* = \sum_{i=1}^{m} x_{ij}^* \omega_j^*$, then $r_e = r_e + 1$; when $\sum_{i=1}^{m} x_{ij}' \omega_j^* < \sum_{i=1}^{m} x_{ij}^* \omega_j^*$, then $r_f = r_f + 1$.

Step 5 count=count+1, if count=sum(sum total number of randomized simulations), go to step 6, otherwise go to step 2.

Step 6 Use the formula $s(\sigma_i' \succ \sigma_i^*) = (r_s + 0.5 r_e) / sum$ to count $s(\sigma_i' \succ \sigma_i^*)$ of superiority randomized simulation values and save the values [21]. The superiority matrix S of n evaluated objects can be obtained by simulation:

$$S = \left[s_{ij}\right] = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1n} \\ s_{21} & s_{22} & \cdots & s_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ s_{n1} & s_{n2} & \cdots & s_{nn} \end{bmatrix} \tag{7}$$

where: the elements on the diagonal of S are all 0.5. Theoretically, $s_{ij} + s_{ji} = 1$, so we only need to get the values of $(n^2 - n)/2$ elements above (or below) the diagonal to calculate S. If $|s_{ij} + s_{ji} - 1|$ is sufficiently small, it indicates that the result is accurate and credible; otherwise, the number of stochastic simulation simulations needs to be increased by sum.In order to make the elements of S strictly satisfy the value of $s_{ij}+s_{ji}=1$, according to the formula $s_{ij}=\left[s_{ij}+(1-s_{ji})\right]/2$ to adjust it, and the adjusted superiority matrix is still noted as S.

III. B. 6) Comprehensive evaluation calculations

Let z_i be the relative composite evaluation value of the evaluated object σ_1, then it can be obtained by taking the minimum value of the overall comparison function between n evaluated objects:

$$\min F(z)=\sum_{i=1}^{n}\sum_{j=1}^{n}\left[z_i-z_j-(s_{ij}-s_{ji})\right]^2 \\ s.t.\ \ z_1+z_2+\cdots z_n=c \tag{8}$$

By constructing the Lagrange function of Eq. (8) and deriving it, it can be obtained:

$$z_i=\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left[(s_{ij}-s_{ji})\right]+\frac{1}{n}\sum_{j=1}^{n}\left(s_{ij}-s_{ji}\right)+\frac{c}{n} \tag{9}$$

where c is a constant. The c value selection has no effect on the size order in z_i, and does not affect the final ordering of the evaluated objects, generally taking $c=n$.

IV. Physical education quality assessment practices

IV. A. Validation against other evaluation methods

IV. A. 1) Comparison of results

In this section, we will experimentally validate the evaluation method of stochastic simulation algorithms based on the excellence criteria in existing studies, and qualitatively and quantitatively analyze whether the proposed method is scientific and reasonable. Currently, there are three main theories to discuss the criteria of excellence in evaluating models. First, the difference theory. The principle of this theory to test the advantages and disadvantages of the comprehensive evaluation model is: the greater the difference between the comprehensive evaluation results, the better the model. The second is the consistency theory. The principle of this theory is: the comprehensive evaluation results obtained by different evaluation models should be basically consistent. Third, information loss theory. The principle of this theory is that the less effective information is lost, the better the model is. As the proposed method in this paper solves the results for the relative superiority, it is not possible to use the differentiation formula to calculate the differentiation of the model, and it is also not possible to use the information loss based on the Shannon-Spearman measure to measure the model. Therefore, this section uses the idea of comparative analysis to discuss the advantages and disadvantages of the proposed methods in terms of consistency. Specifically, this section presents general evaluation methods and then compares the evaluation results between methods. Based on the consistency, the experimental process is designed from the perspectives of consistency of results among methods, consistency of results of different objective empowerment methods, consistency of results of removing suppliers, and consistency of results of removing a small number of products, etc., and the experimental results are compared to derive the reasonableness and advantages and disadvantages of the proposed evaluation methods.

It may be useful to note the evaluation model based on stochastic simulation in the previous section as F0 and the ranking result as R0. S1~S10 are the selected 10 universities. The F1~F4 models are: entropy weighting, hierarchical analysis, fuzzy comprehensive evaluation and BP neural network The results obtained from each evaluation model are shown in Table 2.

It can be seen from the results that even when switching to a different evaluation model, the first and the last of several ranking results are consistent, which is in line with the results foreseen in the previous section.S5>S7(0.522) This college ranking pair has inconsistent ranking results in all the other methods, which is within the range foreseen in the results of this evaluation. This sorting result inconsistency phenomenon also exists in other low superiority sorting pairs. However, the absolute formal evaluation values solved by models F1~F4 cannot well explain this inconsistency in the multi-method evaluation results. On the other hand, the relative superiority probability of S2>S1 is 0.611, which can be interpreted as: the probability that the teaching quality of S2 is better than that of S1 is 0.611. And the absolute value obtained from the comparative model solution does not have interpretable meaning. For example, the evaluation value of college S2 in R2 is 0.622, and this dimensionless value has no interpretable meaning.

Table 2: Comparison of five evaluation models

Evaluation model	Sorting result	Relative superiority probability/evaluation value
F0	S2>S1>S6>S9>S10>S5>S7>S4>S3>S8	[0.611,0.552,0.584,0.530,0.578,0.522,0.740,0.569,0.70]
F1	--	--
F2	S2>S1>S10>S3>S4>S6>S7>S5>S9>S8	[0.622,0.568,0.550,0.502,0.50,0.49,0.456,0.447,0.398,0.225]
F3	S2>S6>S1>S9>S10>S7>S5>S4>S3>S8	[0.520,0.53,0.512,0.505,0.489,0.485,0.476,0.426,0.401,0.363]
F4	S2>S1>S6>S10>S9>S7>S5>S4>S3>S8	[0.536,0.498,0.495,0.487,0.480.0.477,0.476,0.425,0.43,0.378]

IV. A. 2) Consistency comparison

Based on the principle that the comprehensive evaluation results obtained from different comprehensive evaluation models should be basically consistent, this subsection tests the credibility of the evaluation results of the proposed method through the consistency of the results of different evaluation models. The Spearman rank correlation coefficients between the ranked results R0, R1, R2, R3 and R4 are calculated, and the results are shown in Figure 2.

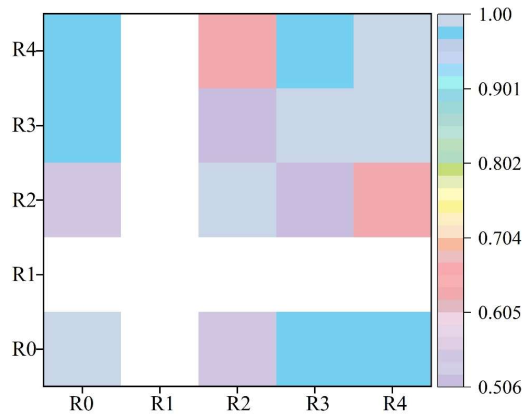

Figure 2: The correlation coefficient table between the evaluation results

The consistency of several results in descending order is R4, R0, R3, R2. The consistency of result R0 is higher, and the consistency is shown in Table 3. This result can show that the model F0, i.e., the result R obtained by the model under the stochastic simulation-based comprehensive evaluation method of college teaching quality, is credible. At the same time, the above results can pass the first and last consistency rate test, and the first and last consistency rate reaches 50%, which is more than the required 20%, which can also prove that the results of model F0 are credible.

Table 3: The average consistency of the evaluation results

Evaluation result	Average consistency with his results
R0	0.845
R1	——
R2	0.556
R3	0.822
R4	0.861

In addition, by comparing the evaluation results of different assignment methods of model F0 as shown in Table 4, the results of several objective assignment methods are consistent, which proves that several methods are better adapted to this dataset. Among them, the highest possibility coefficient of the ranking results under the CCSD assignment method indicates that the CCSD assignment method is more capable of distinguishing colleges and universities than other methods on this dataset.

Table 4: Evaluation results and related parameters

Objective method	Sorting result	Probability factor	Stability factor
CCSD	S2>S1>S6>S9>S10>S5>S7>S4>S3>S8	0.522	2.2E-07
CRITIC	S2>S1>S6>S9>S10>S5>S7>S4>S3>S8	0.514	2.5E-07
Standard deviation	S2>S1>S6>S9>S10>S5>S7>S4>S3>S8	0.500	2.8E-07
Gini coefficient method	S2>S1>S6>S9>S10>S5>S7>S4>S3>S8	0.508	2.8E-07
Entropy	S2>S1>S6>S9>S10>S5>S7>S4>S3>S8	0.514	1.7E-07

Where the objective assignment method used for each model is specified as entropy weight method. The threshold x for the number of students to be removed is set to 5. The threshold max for the number of cycles to remove i students is set to 500. The experimental data obtained is shown in Table 5. From the data in the table, it can be seen that the consistency of the ranking results of model F0 randomly removing a small number of students is better than that of other models, and according to the test principle, it can be determined that the comprehensive evaluation model of teaching quality in colleges and universities based on stochastic simulation is better. Since the model obtains the sorting results by comparing all product data, which fully measures the overall level of teaching quality in colleges and universities, removing a small number of students does not significantly affect the relative level between colleges and universities, so the model performs better.

Table 5: Remove consistency

Evaluation model	Remove consistency					Mean consistency
	i=1	i=2	i=3	i=4	i=5	
F0	0.994	0.987	0.974	0.968	0.964	0.9774
F1	0.985	0.974	0.963	0.944	0.942	0.9616
F2	0.956	0.922	0.876	0.855	0.827	0.8872
F3	0.987	0.978	0.968	0.952	0.947	0.9664
F4	0.989	0.987	0.974	0.978	0.957	0.977

IV. B. Examples of applications

The evaluation of physical education teaching in ordinary schools of higher education, what are the criteria, how to evaluate and how to set the criteria, is directly related to the basic ideas, basic concepts and guiding direction of physical education teaching in ordinary schools of higher education, the physical and mental development and progress of students, the formation of lifelong physical education lifestyles of the students, and the improvement of the teaching of teachers, and ultimately, it is directly related to the reform and realization of the objectives of the physical education teaching in ordinary schools of higher education. Ultimately, it is directly related to the reform of physical education teaching in higher education institutions and the realization of its goals. Nowadays, many problems have appeared in the evaluation mode of physical education teaching in ordinary colleges and universities.

In order to facilitate comparison, the experimental teaching evaluation data for applied research, the example includes a total of four sports colleges and universities, affecting the evaluation of the quality of physical education teaching evaluation results include a total of six indicators such as the syllabus, are the greater the better, the weight of the indicators are subjective assignment method of the weights obtained. Each evaluation index data are interval number, now using the interval number constructed in this paper stochastic simulation model for evaluation. The original evaluation index matrix of the interval number of sports colleges to be evaluated is shown in Table 6.

Table 6: The original evaluation index matrix of interval number

Evaluation university	Evaluation index					
	Syllabus	Experimental teaching	Pre-class preparation	Teaching content	Teaching method	Teaching effect
S1	[60.70]	[70.80]	[80,90]	[50,60]	[70,90]	[60,80]
S2	[80,90]	[70,80]	[70,80]	[60.80]	[60.70]	[70,80]
S3	[70,80]	[60,70]	[70,100]	[60,70]	[60,80]	[50,90]
S4	[60,70]	[70,80]	[80,90]	[70,80]	[50,90]	[70,80]

IV. B. 1) Data pre-processing results

The six evaluation indexes are all of the larger and better type evaluation indexes, the maximum value of each lower interval max(·)=100, the minimum value of min(·)=0. The maximum and minimum values of each upper interval are also 100 and 0. The data preprocessing of each interval is carried out separately, and the results are shown in Table 7.

Table 7: The interval number standardization evaluation index matrix

Evaluation university	Evaluation index					
-	Syllabus	Experimental teaching	Pre-class preparation	Teaching content	Teaching method	Teaching effect
S1	[0.6,0.7]	[0.7,0.8]	[0.8,0.9]	[0.5,0.6]	[0.7,0.9]	[0.6,0.8]
S2	[0.8,0.9]	[0.7,0.8]	[0.7,0.8]	[0.6,0.8]	[0.6,0.7]	[0.7,0.8]
S3	[0.7,0.8]	[0.6,0.7]	[0.7,1.00]	[0.6,0.7]	[0.6,0.8]	[0.5,0.9]
S4	[0.6,0.7]	[0.7,0.8]	[0.8,0.9]	[0.7,0.8]	[0.5,0.9]	[0.7,0.8]

IV. B. 2) Qualitative evaluation results

Each evaluation index of each sports institution was randomly simulated for 5, 10, 50 and 100 times. Each institution after simulation was evaluated comprehensively using the weighted average method with the same subjective assignment weights W={0.16, 0.14, 0.15, 0.20, 0.20, 0.15}. The evaluation results are shown in Figure 3. The evaluation results under 5, 10, 50 and 100 random simulation times are shown respectively. As the number of simulation times increases, the evaluation value points of each institution to be evaluated become more intensive. Under each number of random simulations, the evaluation value of institution 1 fluctuates above and below 2.2, the evaluation value of institution 2 fluctuates between 2.0 and 2.2, the evaluation value of institution 3 fluctuates above and below 1.8, and the evaluation value of institution 4 fluctuates between 2.2 and 2.4. From the graph, it can be qualitatively judged that the evaluation result of institution 4 is optimal, while the evaluation result of institution 3 is the worst. However, by observing the distribution of the evaluation value points of each institution, it is found that there is a phenomenon of cross-mixing of evaluation value points between institutions, for example, there is a large amount of mixing of evaluation values between institution 1 and institution 4, which are located between 2.20 and 2.25. This phenomenon indicates that there is a large uncertainty in the evaluation value of a small number of times, and it is difficult to accurately evaluate the strengths and weaknesses of Institution 1 and Institution 4 based on the evaluation value of a small number of times only.

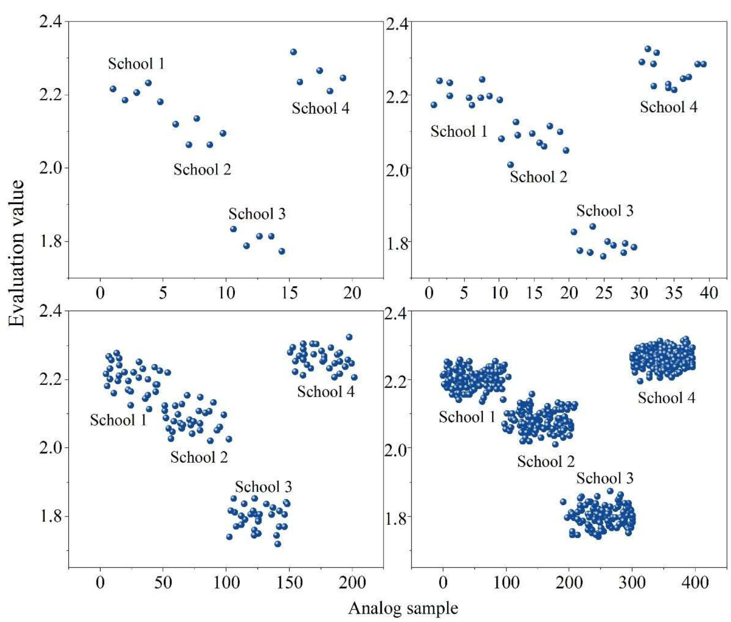

Figure 3: Teaching quality evaluation map

IV. B. 3) Quantitative evaluation results

In order to further quantitatively compare the differences in strengths and weaknesses among the institutions to be evaluated, Duncan's test was used to test the significance of the differences in the evaluation values of the institutions to be evaluated, and the results are shown in Table 8. At the 0.05 level of significance, the mean value of the evaluation results of institution 4 is statistically greater than that of institution 1, the mean value of the evaluation results of institution 1 is statistically greater than that of institution 2, and the mean value of the evaluation results of institution 2 is statistically greater than that of institution 3. The rankings of the evaluation values of the quality of teaching of the four institutions are as follows: Institution 4 > Institution 1 > Institution 2 > Institution 3, which are the same under each random simulation. This result is the same across the number of randomized simulations, which further indicates that the results of the ranked value of the 4 institutions are statistically stable. Note: Lowercase letters a, b, c, and d denote 0.05 level of significance, while uppercase letters A, B, C, and D denote 0.01 level of significance.

Table 8: Significance test results of evaluation value

Analog frequency	School 1	School 2	School 3	School 4
5	2.204±0.025bA	2.094±0.043cB	1.794±0.034dC	2.251±0.041aA
10	2.213±0.025bB	2.084±0.031cC	1.784±0.025dD	2.278±0.040aA
50	2.212±0.034bB	2.073±0.035cC	1.796±0.036dD	2.266±0.028aA
100	2.212±0.026bB	2.078±0.029cC	1.796±0.028dD	2.268±0.025aA

V. Conclusion

This paper utilizes the computational methods such as autonomous dominance amount matrix and dominance right vector, superiority matrix, comprehensive evaluation vector, etc., to establish a stochastic simulation-based evaluation method for physical education teaching. Taking four sports colleges as an example, the evaluation method of this paper is applied to evaluation practice, and the results show that.

(1) By comparing different evaluation models, the consistency between the first and the last of their several sorting results reaches 50%, which can prove that the results of model F0 are credible. Randomly removing a small number of students, the F0 model does not significantly affect the relative level between colleges and universities, and the consistency of the sorting results is better than other models.

(2) The ranking of the teaching quality evaluation value of 4 sports colleges and universities is as follows: institution 4>institution 1>institution 2>institution 3. Duncan statistical test on the evaluation value of each evaluation object can be found that each evaluation object is better than the other evaluation objects with significant difference.

References

[1] Spittle, M., & Spittle, S. (2016). Content of curriculum in physical education teacher education: expectations of undergraduate physical education students. Asia-pacific journal of teacher education, 44(3), 257-273.

[2] Jin, H. (2018). The Comprehensive Evaluation of the Quality of Physical Education in Colleges and Universities. Educational Sciences: Theory & Practice, 18(6).

[3] Wang, R. (2017). Research on physical education problems and management reform of chinese colleges and universities. Eurasia Journal of Mathematics, Science and Technology Education, 13(10), 7057-7069.

[4] Lan, S. (2023). The reform of physical education in colleges and universities under the background of physical education integration. International Journal of New Developments in Education, 5(26), 11.

[5] Liu, S. (2021). Research on the teaching quality evaluation of physical education with intuitionistic fuzzy TOPSIS method. Journal of Intelligent & Fuzzy Systems, 40(5), 9227-9236.

[6] Phillips, S. R., Mercier, K., & Doolittle, S. (2017). Experiences of teacher evaluation systems on high school physical education programs. Physical education and sport pedagogy, 22(4), 364-377.

[7] Zhao, Y. (2017). Research on the diversified evaluation index system and evaluation model of physical education teaching in colleges and universities. Journal of Computational and Theoretical Nanoscience, 14(1), 99-103.

[8] Zeng, Y. (2020). Evaluation of Physical Education Teaching Quality in Colleges Based on the Hybrid Technology of Data Mining and Hidden Markov Model. International Journal of Emerging Technologies in Learning, 15(1).

[9] Wang, Z. (2018). Fuzzy comprehensive evaluation of physical education based on high dimensional data mining. Journal of Intelligent & Fuzzy Systems, 35(3), 3065-3076.

[10] Xue, S., & Ding, L. (2023, August). Correlation Analysis of Physical Education Teaching Quality Assessment Based on Fuzzy Comprehensive Evaluation. In International Conference on E-Learning, E-Education, and Online Training (pp. 262-275). Cham: Springer Nature Switzerland.

[11] Wang, H., Yang, F., & Xing, X. (2022). Evaluation method of physical education teaching and training quality based on deep learning. Computational Intelligence and Neuroscience, 2022(1), 1680888.

[12] Wu, H., Li, F., & Cheng, G. (2017). Comprehensive evaluation model of physical education based on grey clustering analysis. Journal of Computational and Theoretical Nanoscience, 14(1), 84-88.

[13] Han, K. (2020). Evaluation of teaching quality of college physical education based on analytic hierarchy process. International Journal of Emerging Technologies in Learning (iJET), 15(10), 86-99.

[14] Chen, J., & Huang, H. (2018). Analytic Hierarchy Process of the Evaluation System of Physical Education Teaching Quality in Colleges and Universities. Educational Sciences: Theory & Practice, 18(6).

[15] Gui, Y. (2025). Enhancing Physical Education Analytics: An Analytic Hierarchy Process Based Approach for College Sport. IEEE Access.

[16] Tang, M., Si, X., Wang, R., & Gao, W. (2021). A new fuzzy comprehensive evaluation model for influencing factors of physical education. International Journal of Emerging Technologies in Learning (iJET), 16(17), 164-176.

[17] Song Wei & Liu Dong. (2025) .Evaluation Method of Physical Education Teaching Quality in Colleges and Universities Based on Wireless Communication Systems. Journal of Cases on Information Technology (JCIT),27(1),1-17.

[18] Haruhiko Akiyama,Kanako Komiya & Yoshiyuki Kotani. (2012) .Nested Monte-Carlo Search with simulation reduction. Knowledge-Based Systems,34,12-20.

[19] Rajneesh Sharma & Abhilash Awasthi. (2020) .An embedded element based 2D finite element model for the strength prediction of mineralized collagen fibril using Monte-Carlo type of simulations. Journal of Biomechanics,108(prepublish),109867.

[20] Vlastos Spyros I.,Xanthopoulos A. S. & Koulouriotis D. E.. (2024) .Stochastic modelling and analysis of a deteriorating serial production–inventory network. International Journal of Production Research,62(9),3084-3098.

[21] Li Hui & Tong Shaocheng. (2024) .Adaptive fuzzy consensus secure control of stochastic nonlinear multi-agent systems with false data injection attacks. International Journal of Systems Science,55(6),1191-1205.