Establishment and Practice of Physical Education Evaluation Using Grey Cluster Analysis Under the Data Background

Jinxin Jiang, Bozhou University, China*

https://orcid.org/0000-0001-7106-3242

Sang Keon Yoo, Sangmyung University, North Korea

ABSTRACT

This article uses scientific methods and means to evaluate the value, elements, and processes of physical education, consistent with preset evaluation indicators through sample calculation, and then derives the characteristics of decision-making, the objectivity of indicators, the order, and other characteristics of the process. The authors have analyzed the main problems in current physical-education evaluation and its future reform and development trends under the requirements of quality education.

KEYWORDS

Assessment, Computer, Gray Clustering, Model

INTRODUCTION

Physical-education evaluation refers to the process in which teachers aim to assess students’ progress in physical education using the scientific method to make judgments about the quality of learning and level of achievement (Greaves et al., 2011). Learning-outcome assessment is an important part of students’ learning and is the basis for organization in the teaching process (Flöel et al., 2010). By evaluating the knowledge each student has mastered, teachers can provide different types of remediation for each student, thereby improving their learning (Cardona-Morrell et al., 2010).

Learning-outcome assessment has a remarkable effect on improving learning. Its meaning and effect are reflected mainly in feedback regulation and monitoring, learning support, optimizing education, and other aspects (Tremblay et al., 2016). Through evaluation, teachers can constantly discover problems, improve teaching and learning methods, and promote the all-around development of students (Zhang et al., 2021).

Scientifically evaluating students’ learning outcomes is an essential element of effective teaching and serves as a standard for determining the effectiveness of instruction (Tremblay et al., 2017). The concept of quality education should focus on students’ learning, the evaluation of the learning process, and the changes in their emotions and attitudes during the learning experience (Henson et al., 2013). Comprehensive evaluation is an effective way for students to assess their overall progress during the learning process (Ceulemans et al., 2015). Through learning evaluation, students can gain...
a more specific and detailed understanding of their learning progress, allowing them to adjust their learning methods accordingly (Papastergiou, 2009).

Because the connotation of learning outcome in quality education differs from that of traditional learning, traditional learning-outcome assessment primarily uses quantitative evaluation and test scores to obtain assessment results (Wang, 2018). In addition to evaluating students’ test scores, learning-outcome assessment in quality education should also include evaluation of their emotions, learning attitudes and overall learning quality during the learning process. Learning-outcome assessment should include qualitative evaluation criteria to comprehensively assess students’ learning outcomes (Zhang et al., 2018).

Gray clustering is a method of dividing some indexes and observation objects into several definable categories based on the correlation matrix or gray whitening weight function. A cluster can be seen as a collection of observation objects belonging to the same class (Liu et al., 2021). Some scholars have proposed a gray weighted clustering evaluation method based on triangular models (Kong & Guo, 2022). In order to make the clustering results more reasonable, some scholars have proposed the gray optimal clustering theory model (Feng, 2020). Due to its ease of understanding and programming, the gray clustering evaluation method has become a hot research topic, widely used in economics, environmental-quality assessment, remanufacturing evaluation, calculation, and transportation. Some researchers have applied the gray clustering method to calculate the weight and whiten multi-indicator data, achieving scientific weighting of multiple evaluation indexes and classification ranking of libraries (Martos et al., 2023).

Applying the gray system in the comprehensive evaluation of physical education teaching can utilize its advantage in dealing with uncertain and fuzzy information. This approach can effectively solve the gray information associated with evaluating learning outcomes while avoiding the use of precise mathematics to process fuzzy information, which may result in deviations in results. The application of gray system theory in the evaluation of learning outcomes not only effectively addresses the problem of a small amount of evaluation data in the learning process, but also avoids the deviation in evaluation results caused by the limited amount of data available.

This article first uses a questionnaire to obtain the physical-education learning outcomes of six evaluation subjects. Then a gray clustering model was used to evaluate the academic performance of six students based on factors such as student performance, learning attitude, and quality development. This study aims to use gray clustering technology to construct a comprehensive evaluation model for physical-education learning effectiveness, explore a new evaluation method, and provide a foundation and reference for theoretical exploration in related fields. By evaluating the effectiveness of physical-education learning, students’ strengths, weaknesses, and potential in sports can be more accurately identified, providing a basis for targeted education and promoting their learning progress and comprehensive development.

**DESIGN OF SINGLE-CLASS EVALUATION INDEX OF PHYSICAL EDUCATION**

The comprehensive evaluation of physical education considers multiple indexes to produce evaluation results. To conduct a comprehensive evaluation, the evaluation of a single-class index must be carried out first. This is because the specific evaluation index can be determined by differentiating index categories (O’Connor et al., 2016). The evaluation of a single-class index provides a grade for the evaluation object on that particular index. However, the quality of an object on a single index does not necessarily represent its overall quality. The single-class index evaluation provides the grade of the evaluation object on only a particular evaluation index. However, the quality of an object on a single evaluation index does not reflect the overall quality of the evaluation object (Giese & Ruin, 2018). Therefore, to understand the overall performance of the evaluation object, a comprehensive evaluation model that combines all kinds of evaluation indexes must be established to conduct an integrated evaluation of the evaluation object. The learning outcome assessment process is illustrated in Fig. 1.
Fig. 1 shows that the evaluation process is divided into six parts: single-class evaluation index, single-class evaluation, comprehensive evaluation model, comprehensive evaluation, survey, and results analysis. Among them, the single-class evaluation index mainly determines the content and focus of the evaluation, which usually needs to be combined with the characteristics of the subject and teaching objectives to determine the types and quantities of evaluation indexes. The single-class evaluation and comprehensive-evaluation models convert the evaluation results into numerical form based on evaluation index and weights, providing data support for subsequent comprehensive evaluations. When the evaluation result is basically consistent with the teacher’s and the student’s overall evaluation of the student, which indicates the evaluation result is reasonable; otherwise, it should continue to investigate the sample values of the student for reappraisal. Result analysis is a
crucial step in the evaluation process, which requires statistical analysis of the evaluation results and targeted improvement suggestions and measures.

COMPREHENSIVE EVALUATION BASED ON GRAY CLUSTERING

A white function is the function expression for the possibility of each element’s value in the gray class and is used to describe the deviation of a gray number with the different value within its range (Goodyear & Dudley, 2015). The white function is generally designed by researchers according to the available information, and there is no fixed formula for its design (Richards et al., 2014). However, determining the turning point of the function is key to defining the white function. The white function is important in the gray system theory. Determining the white function is also crucial in the gray clustering evaluation and is the key to moving from qualitative analysis to quantitative modeling (Yin & Huang, 2015). There are four basic forms for white function that are commonly used: typical white function, upper limit measure white function, lower measurement white function, and moderate white function.

The importance of the white function is the determination of the function’s turning point, which reflects whether the determination of the white function is scientific (Tsangaridou, 2016). The typical white function \( f(x) \) is defined by its starting point and ending point. It is an increasing function on the left and a decreasing function on the right. If the white function \( f(x) \) does not have a first or second turning point, it is a lower limit measure white function. If the second and third turning points of the white function \( f(x) \) coincide, it is a moderate measure white function. If the white function \( f(x) \) does not have a third or fourth turning point, it is an upper limit measure white function.

When we need to know about the position or level of an evaluation object in an index sample, we will need to use the gray clustering technology in the gray system (Hill, 2015). Gray clustering technology is used to establish the evaluation model, and the specific expression is as follows: \( PJ = \{A, B, C\} \), wherein \( A = \{A_1, A_2, A_3, \ldots, A_m\} \) is the evaluated object set; \( B = \{B_1, B_2, B_3, \ldots, B_n\} \) is the evaluation class set; \( C = \{C_1, C_2, C_3, \ldots, C_t\} \) is the evaluation level.

For example, if the object set \( A \) consists of \( m \) students, and the single-class evaluation is conducted on the evaluation class, the result constitutes \( B = \{B_1, B_2, B_3, \ldots, B_n\} \); setting \( t \) evaluation levels \( \{C_1, C_2, C_3, \ldots, C_t\} \) can be excellent, good, qualified, unqualified; and so on.

The evaluation model in gray clustering technology has the following three features for each element in the evaluation level set \( C \): certainty, which means that each element either belongs to a particular set or does not belong to the set; inequality, which means that the elements within the same set are distinct from one another; and disorder, which means that the order of the elements in the set can be changed without affecting their membership in the set (Guerra et al., 2016). Although each evaluation level in the set \( C \) meets the disorder and the position of each element in the set can be arbitrary, comprehensively, in terms of the whole set, each element in the set is in a certain order. Because the set \( C \) is a collection composed of multiple evaluation levels, each element is the evaluation obtained according to the division of different grades, and each evaluation’s grade range has no intersection; therefore, the elements in the set that are classified within one range have no intersection or a continuous subrange.

Each element of the evaluation object set \( A \) in the model needs to satisfy the similarity principle, namely, each element must be homogeneous. For example, when we undertake learning-outcome assessment of students, we cannot classify sports students and students who were admitted through national unified recruitment in the same sample, because there is an essential difference between them. The evaluation model aims to produce a unique evaluation result for each evaluation object, meaning that each evaluation object can correspond to only one evaluation result (Zhang, 2017). The model can be considered as a mapping function that satisfies the many-to-one mapping principle (Mâsse et al., 2017). For instance, although multiple students’ performance in a learning-outcome assessment can be classified as good, a student cannot be evaluated as both excellent and good.
For the \( n \) single-class evaluation indexes, the index values are obtained by test; by constructing the white function, clustering analysis is conducted for \( m \) clustering objects (students) to determine the gray class that the clustering objects belong to among gray classes. \( m \) clustering objects are taken as \( A_1, A_2, A_3, ..., A_m \), the clustering index is \( B_1, B_2, B_3, ..., B_n \), and the gray class is \( C_1, C_2, C_3, ..., C_t \).

Now, assuming that the sample matrix of all object clusters in all indexes is \( d \):

\[
d = \begin{bmatrix}
d_{11} & d_{12} & d_{1m} \\
d_{21} & d_{22} & d_{2m} \\
& \vdots & \vdots \\
d_{m1} & d_{m2} & d_{mn}
\end{bmatrix}
\tag{1}
\]

Among them \( d_{ij} \) is the sample value of the \( i \) cluster object \( A_i \) on the \( j \) cluster of the index \( B_j \). When \( 1 < i < m, 1 < j < n \), construct a white function \( f^k_j \). When \( 1 < j < n, 1 < k < t \), \( f^k_j \) is the white function that the \( j \) index \( B_j \) belongs to the \( k \) gray class.

Calculate the grayscale vector as shown in (2):

\[
\sigma_i = (\sigma^1_i, \sigma^2_i, ..., \sigma^t_i), 1 \leq i \leq m
\tag{2}
\]

The mapping \( f^k_j \rightarrow \sigma^k_i \) is called gray clustering. Among them, there are (3) and (4):

\[
\sigma^k_i = \sum_{j=1}^{n} f^k_j (d_{ij}) \cdot \eta^k_j
\tag{3}
\]

\[
\eta^k_j = \frac{\lambda^k_j}{\sum_{j=1}^{n} \lambda^k_j}
\tag{4}
\]

The steps for using the comprehensive evaluation model of learning effect based on gray clustering are as follows:

1) Select evaluation samples, determine the evaluation objects, expressed as \( A_1, A_2, A_3, ..., A_m \); determine the evaluation indexes, expressed as \( B_1, B_2, B_3, ..., B_n \); determine evaluation of the gray class, expressed as \( C_1, C_2, C_3, ..., C_t \).
2) According to the actual sample value, determine the sample matrix \( d = (d_{ij})_{mn} \); \( d_i \) is the sample value of the \( i \)-clustering object \( A_i \) on the \( j \)-clustering index \( B_j \).
3) Determine the evaluation gray class’s white function as the white function of the \( j \) index \( B_j \) belonging to the \( k \) gray class.
4) Figure out the calibration clustering; \( \lambda^k_j \) is the threshold of \( f^k_j \), \( \eta^k_j \) is the weight of the \( j \) index belong to the \( k \)-clustering class.
5) Figure out the gray clustering; \( f^k_j (d_{ij}) \) is the white weight of the sample value \( d_{ij} \) belonging to the \( k \) gray class, comprehensively reflecting the weight of the \( i \)-evaluation object belonging to the \( k \) gray class.
6) Figure out the gray vector \( \sigma_k \), \( \sigma_k = (\sigma^1_k, \sigma^2_k, ..., \sigma^t_k) \).
7) \( \sigma^*_i = \max \{ \sigma^1_i, \sigma^2_i, ..., \sigma^t_i \} \); the \( i \)-clustering object \( A_i \) belongs to the \( k^* \) gray class \( C_k \).
EXPERIMENTAL ANALYSIS

Example Profile

A class in the Wuhan Institute of Physical Education was taken as the sample to select objects. There were 46 students, one class adviser, and 10 teachers in this class. Six students were randomly selected as the evaluation objects from this class. The questionnaire, called *The Survey on the Comprehensive Learning Evaluation of Sports Students in the Teaching Mode of Physical Education*, was used to obtain the learning outcomes of the six evaluation objects. The evaluation process involved distributing questionnaires to three groups: the class adviser, all the teachers, and the 6 students. Each group received 22 questionnaires, which were used to gather information about an evaluation object. This process was repeated six times for a total of six evaluation objects.

The content involved in the questionnaire was divided into four types, effectively collecting 20 questionnaires, and the results of these 20 questionnaires were taken as the basis of data acquisition. The evaluation of grade, learning attitude, and quality development was conducted for the six students, who were numbered 1, 2, 3, 4, 5, and 6. The results of the questionnaires on these four aspects were processed to obtain the sample data shown in Table 1.

The evaluation of the six students’ learning was made according to four grades, namely, excellent, good, qualified, and unqualified. The following comprehensive analysis of the students’ learning effect was performed.

1) According to the evaluation samples, determine the evaluation objects as A₁, A₂, A₃, A₄, A₅, and A₆ with the corresponding number of 1, 2, 3, 4, 5, and 6. Determine evaluation indexes as B₁, B₂, B₃, or B₄, corresponding to the student’s grade, learning attitude, and quality development. Determine evaluation gray class as C₁, C₂, C₃, or C₄, corresponding to the student’s learning effect as excellent, good, qualified, or unqualified.

2) According to the actual sample values in Table 1, the sample matrix \( d \) can be obtained, as shown in (5):

\[
d = \begin{bmatrix}
90.23 & 80.26 & 86.36 & 77.41 \\
86.21 & 65.23 & 85.29 & 78.54 \\
79.23 & 78.25 & 84.19 & 76.36 \\
91.25 & 59.69 & 52.36 & 85.26 \\
60.28 & 72.84 & 62.36 & 96.03 \\
74.23 & 83.26 & 82.69 & 65.23
\end{bmatrix}
\]

(5)

3) Determine the white function of evaluation gray class \( f_j^k (x) \), based on the experience of experts and relevant standards.

<table>
<thead>
<tr>
<th>Table 1. Single-class evaluation results</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Indexes</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Student achievement</td>
</tr>
<tr>
<td>Learning attitude</td>
</tr>
<tr>
<td>Quality developing</td>
</tr>
</tbody>
</table>
Results Analysis

The evaluation results of the test in this study show that a student having excellent performance in an evaluation index of the learning-outcome assessment system does not necessarily mean the student’s learning effect is excellent. Student Number 2’s comprehensive ability is excellent, but his learning effect is just good. Therefore, considering comprehensive factors can more objectively reflect the overall learning effect of the evaluation objects. The 20 students that were randomly selected from the class in which the six evaluation students were, all 10 teachers of this class, and the class adviser constitute the comprehensive learning-effect evaluation group. The 31 evaluation team members conducted comprehensive evaluations on 6 students. The comprehensive evaluation of student learning outcomes includes four outcomes: excellent, good, qualified, or unqualified. When an evaluation result accounts for more than 80% of the total evaluation, it serves as a reference result for student evaluation. The statistical results are shown in Table 2.

Experiments show that by using gray clustering to conduct the comprehensive evaluation of physical-education teaching, the evaluation index system is reasonable, the evaluation model is feasible and practical, and the evaluation results are basically in line with the objective conditions, which can play a positive role in promoting students’ learning. Twenty-two questionnaires were sent to each of these six evaluation objects in this study, and 20 pieces were recycled to obtain the required information; the 32 questions in the questionnaire reflect all the content of the comprehensive evaluation index system of the learning effect. A comprehensive analysis of the existing research on the learning-effect evaluation indicates that the number of questionnaires is generally above 40, and can be up to 200, and the number of questions is generally 40 to 80. By comparing the evaluation results of the model in this study, it can be concluded that even in the case of small sample data, the evaluation results of the model are basically in line with the actual situation. Therefore, the overall evaluation scheme that this research puts forward is feasible, effective, and can be used in practice. It should be noted that although this study was conducted on a relatively small sample database, in actual evaluations, it is important to obtain more representative and diverse sample data to improve the reliability and accuracy of the evaluation results.

In practical applications, this method can be applied to students’ physical-education teaching, guiding course improvement and optimization by evaluating the comprehensive effectiveness of physical-education courses. At the same time, this method can also be used to evaluate the learning effectiveness of different students, identify their weaknesses and advantages in sports learning, and develop personalized learning plans and educational plans.

CONCLUSION

This paper uses a gray clustering model to evaluate the learning outcomes of six students based on factors such as student achievement, learning attitude, and quality development. The results

Table 2. Learning effect comprehensive performance results

<table>
<thead>
<tr>
<th>Student Number</th>
<th>Excellent (%)</th>
<th>Good (%)</th>
<th>Qualified (%)</th>
<th>Unqualified (%)</th>
<th>Evaluation of the Results (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>83.25</td>
<td>12.6</td>
<td>3.22</td>
<td>0</td>
<td>Excellent</td>
</tr>
<tr>
<td>2</td>
<td>9.36</td>
<td>87</td>
<td>3.25</td>
<td>0</td>
<td>Good</td>
</tr>
<tr>
<td>3</td>
<td>3.24</td>
<td>90.23</td>
<td>6.54</td>
<td>0</td>
<td>Good</td>
</tr>
<tr>
<td>4</td>
<td>83.57</td>
<td>9.36</td>
<td>3.45</td>
<td>3.25</td>
<td>Excellent</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>9.36</td>
<td>80.26</td>
<td>9.45</td>
<td>Qualified</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>87.36</td>
<td>12.36</td>
<td>Qualified</td>
</tr>
</tbody>
</table>
demonstrated that the gray clustering model was effective in evaluating students’ learning progress and had high applicability in this context.

Many experts and scholars are trying to explore scientific theories of physical education teaching evaluation. The physical-education evaluation model proposed in this paper can be well used in practice and proves its practicability. Although science is a long historical process, as long as sports continue to develop, then the scientific evaluation of physical education will exist. In the future, modern technological means can be utilized to collect, analyze, and mine a large amount of student-behavior data and teaching data, construct a more refined and intelligent physical-education teaching-evaluation model, and optimize teaching effectiveness through prediction and recommendation.

AUTHORS NOTE

The data used to support the findings of this study are included in the article. The authors declare that there are no conflicts of interest. The authors acknowledge the 2021 Bozhou University Sports Virtual Simulation Teaching and Experiment Center Project (No. 2021XJXM015), Research on the Innovation and Development of Social Science in Anhui Province in 2021 (2021CX507), the Anhui University Humanities and Social Science Research Project (SK2020A0574), the 2023 Anhui Province Action Plan for the Cultivation of Young and Middle aged Teachers in Universities: Key Research Projects for the Cultivation of Excellent Young Teachers(YQZD2023085), and the 2023 Anhui Provincial Quality Engineering Teaching Research Project for Higher Education Institutions: “Construction and Practice Research of Evaluation Index System for University Physical Education Curriculum Teaching under OBE Concept” and “Construction Research of Competency Evaluation Index System for University Physical Education Teachers in the New Era.”
REFERENCES


Jinxin Jiang is a professor at Bozhou University.

Yoo Sang Keon is currently an assistant professor in the Department of Sport Information, Communications and Technology Convergence in the graduate school at Sangmyung University in Korea.