**Construction and verification of performance appraisal index for physical education teachers under multi-dimensional evaluation system**

Yanhong lv 1,a；Haixin Li1,b；Zhenguo Shi1 ,c\*

1. Shandong University, Jinan, Shandong,250061,China;

aEmail: 202315403@mail.sdu.edu.cn

bEmail: 202415405@mail.sdu.edu.cn

cEmail: shizg@sdu.edu.cn

Abstract: This study focuses on developing and validating performance evaluation metrics for university physical education faculty within a multi-dimensional assessment framework. It examines common issues in current evaluations, such as oversimplification and utilitarian tendencies. Guided by principles of comprehensiveness and specificity, the research establishes an evaluation system across teaching, research, and competition coaching dimensions. Through empirical testing and reliability/validity analysis, the framework undergoes refinement based on validation results. The findings are then applied to optimize the evaluation system, breaking free from traditional limitations while enhancing physical education quality and promoting holistic teacher development.

Key words: multiple evaluation; physical education teachers; assessment indicators; construction and verification

**foreword**

Physical education plays a pivotal role in talent development within higher education institutions. The quality of physical education instructors directly impacts students' physical fitness and overall development. However, current performance evaluations for PE teachers predominantly rely on quantitative metrics while neglecting qualitative dimensions, coupled with a single evaluation framework. These limitations hinder the advancement of physical education. This paper proposes a multi-dimensional evaluation system to establish scientific performance assessment criteria, addressing existing challenges. By introducing innovative evaluation approaches, we aim to drive high-quality development in sports education through systematic improvement of teacher assessment mechanisms.

**I. Definition and importance of multiple evaluation system**

The diversified evaluation system refers to a comprehensive assessment model that transcends traditional single-dimensional performance metrics or quantitative indicators. It evaluates subjects through multiple dimensions including knowledge mastery, competency development, emotional attitudes, and practical innovation. This approach utilizes diverse evaluators such as teachers, students, parents, and social organizations, along with varied methods like observation records, project outcomes, peer reviews, and growth portfolios, combined with standardized criteria to holistically reflect the subject's actual proficiency and developmental potential[1]. Its significance lies in avoiding the limitations of "score-centric evaluation," respecting individual differences and personalized development needs, stimulating initiative and creativity, while providing educators and administrators with comprehensive feedback to optimize training programs. Additionally, it enables subjects to recognize their strengths and weaknesses, clarify development directions, ultimately driving their holistic and sustainable growth.

**2. The present situation and problems of performance appraisal of college physical education teachers**

The current performance evaluation system for physical education teachers in universities exhibits notable tendencies toward oversimplification and utilitarianism. Most institutions still prioritize quantitative metrics like class hours, competition achievements, and publication counts as core assessment criteria, while neglecting qualitative dimensions such as teaching quality, student feedback, curriculum innovation, and professional ethics. The evaluation process remains predominantly administered by school administrative departments, lacking diverse participation from students, peers, and society, resulting in a one-sided assessment perspective[2]. Some evaluation standards are disconnected from practical sports teaching realities, emphasizing research over instruction and focusing on outcomes rather than processes. This forces teachers to concentrate their efforts on short-term quantifiable achievements, neglecting fundamental tasks like curriculum reform and sports injury prevention guidance. Such practices not only dampen teaching enthusiasm but also fail to comprehensively reflect teachers' overall competencies, ultimately hindering the improvement of physical education quality and the holistic development of students.

**3. Construction of performance assessment indicators for college physical education teachers under the multi-dimensional evaluation system**

**(1) Principles of index system construction**

The development of performance evaluation metrics for college physical education faculty should adhere to four fundamental principles. The comprehensiveness principle requires covering multiple dimensions including teaching, research, community service, and professional ethics, with emphasis on measurable outcomes while equally valuing qualitative aspects such as pedagogical innovation and student fitness improvement. The specificity principle emphasizes discipline-specific characteristics by highlighting professional indicators like skill instruction, ideological integration in curriculum design, and competition guidance, preventing homogenization with other academic assessment standards[3]. The developmental principle balances short-term achievements with long-term potential by incorporating growth-oriented metrics such as teacher training engagement and innovative teaching reforms, encouraging continuous professional development. The operability principle mandates clear, quantifiable criteria that are easy to measure or document, avoiding ambiguous language while balancing evaluation costs and effectiveness to ensure efficient assessment processes. Additionally, the dynamic principle requires regular optimization of metrics based on educational policy adjustments and evolving needs in physical education.

**(2) The proposal of the preliminary index system**

The preliminary established indicator system proposes specific metrics across five dimensions. In the teaching dimension, it includes the quality of class hour completion, student physical health compliance rates, effectiveness of integrating ideological and political education into courses, and the practicality of sports injury prevention guidance. The research dimension evaluates the quality of sports teaching research papers, achievements in textbook compilation, and participation levels in teaching reform projects. The competition guidance dimension covers students' award rankings in competitions, daily training attendance rates, and outcomes of athlete psychological counseling. The social service dimension assesses contributions to campus sports activities, community fitness guidance duration, and effects of sports science popularization[4]. The teacher ethics dimension encompasses satisfaction with teacher-student relationships, adherence to teaching discipline, and emergency response performance. Additionally, it incorporates process records from self-evaluations by teachers, student evaluations, and peer reviews, forming a preliminary indicator pool supported by multi-source data to facilitate subsequent screening processes.

**(3) Selection and optimization of index system**

The initial indicator system underwent multiple rounds of screening and refinement. Through expert consultation, we engaged sports education specialists, frontline teachers, and student representatives to validate the scientific validity of indicators. We eliminated redundant or impractical metrics while removing research-related indicators loosely connected to physical education. Utilizing empirical research methods, pilot evaluations were conducted to collect data and assess the differentiation and practicality of indicators. The "student classroom satisfaction" and "post-class feedback response speed" were consolidated into a comprehensive "teaching interaction quality" metric. Finally, through a feedback adjustment mechanism, we optimized indicator formulations based on teachers 'actual performance and evaluation results. Ambiguous terms like "curriculum innovation" were refined into specific items such as "frequency of adopting innovative teaching methods" and "number of developed distinctive courses," ensuring the indicators comprehensively cover assessment content while accurately reflecting teachers' authentic work status.

**(4) Weight allocation of the index system**

The allocation of indicator weights should holistically consider the core objectives of physical education and teachers' priorities, determined through a combination of Analytic Hierarchy Process (AHP) and Delphi method. The teaching dimension holds the highest weight as the core responsibility, with key indicators like "student fitness improvement outcomes" and "innovative teaching methods" carrying greater weight than basic class hours. The research dimension maintains moderate weight, emphasizing the transformation of teaching research achievements while reducing emphasis on pure theoretical papers and increasing weight for teaching reform projects. The competition guidance and social service dimensions each account for approximately 15%, highlighting the educational value of competitions and practical social services. Teacher ethics and conduct are established as mandatory baseline requirements with a "one-vote veto" system. The weight distribution undergoes multiple rounds of expert voting calibration to ensure both teaching centrality and balanced consideration of research and social services, forming a scientifically sound weight structure[5].

**4. Verification strategies of performance appraisal index system for physical education teachers under the multi-dimensional evaluation system**

**(1) Selection and design of verification methods**

Scientifically designed validation methods are crucial for ensuring the effectiveness of indicator systems. A "multi-complementary, quantitative and qualitative integrated" validation framework should be established, with core methodologies including empirical testing, reliability and validity analysis, expert review, and comparative experiments. The empirical testing phase requires selecting representative samples covering universities at different educational levels—such as Double First-Class universities, local undergraduate institutions, and sports-specialized colleges (e.g., those focusing on ball games, track and field, martial arts, etc.). Pilot programs must involve no fewer than five institutions and 300 physical education teachers. This is achieved through standardized assessment scales, teaching video collection, and student fitness monitoring data to verify the indicators 'applicability across scenarios. For reliability and validity analysis, the Cronbach's α coefficient and retest reliability are used to assess internal consistency (α ≥0.7) and stability, respectively. Validity assessment relies on content validity by inviting over 10 physical education experts to evaluate indicator coverage and structural validity. Exploratory and confirmatory factor analyses ensure factor loadings ≥0.5, while criterion-related validity is validated through correlation analysis with annual teacher excellence evaluation results for multidimensional verification. The expert review method requires forming a panel comprising university administrators, frontline teachers, and student representatives to evaluate the new system through criteria such as indicator rationality and operational feasibility. The comparative experimental method involves conducting a significance test between the new system and traditional assessment results, while utilizing questionnaires with a minimum sample size of 500 to collect teachers' subjective evaluations on indicator clarity and fairness, thereby establishing a multi-dimensional verification evidence chain.

**(2) Verification process and data analysis**

The validation process should be implemented in a phased manner to ensure data quality and analytical appropriateness, as shown in Table 1. The first phase involves multi-source data collection, spanning two teaching semesters. This stage utilizes a digital assessment platform to collect real-time teaching information (e.g., class hours, course types) and process-related metrics (e.g., student classroom satisfaction, sports skill attainment rates), along with outcome-based indicators (e.g., competition awards, research achievements). Through semi-structured interviews with 10-15 teachers per school and open-ended questionnaires, we gather teacher feedback on these metrics to ensure comprehensive coverage of the entire teaching process. The second phase focuses on data preprocessing. Using SPSS 26.0 software for data cleaning, we apply multiple imputation techniques to minimize missing values below 5%. Extreme data is excluded using Z-score method (|Z|>3 considered outliers) and K-S test for normality verification. Non-normal data undergoes logarithmic transformation or non-parametric testing. The third phase conducts quantitative analysis: calculating discriminant power of indicators through independent samples t-test (p<0.05) and Pearson correlation analysis (highly correlated indicators>0.8 warrant re-evaluation). Qualitative analysis employs NVivo software to code interview transcripts, identifying core issues such as "inappropriate indicator weighting" and "evaluation cycles deviating from pedagogical principles." Finally, based on the cross-analysis of the assessment results of teachers with different teaching years and different specialties, a visual report containing data distribution characteristics and index performance differences is formed.

|  |  |  |  |
| --- | --- | --- | --- |
| stage | primary coverage | Key methods/tools | Data requirements/standards |
| data collection | Teaching data, student performance, teacher feedback | Assessment platform, interview and questionnaire | Cover the whole process |
| data preprocessing | Data cleaning, inspection and conversion | SPSS 26.0 | The missing rate is less than 5%, and the distribution is compliant |
| DA | Differences, correlations, text mining | SPSS/Nvivo | P <0.05, avoid redundancy |
| Results presented | Visual analysis report | Charting tools | Make the conclusion clear |

Table 1. Validation process and data analysis

**(3) Result analysis and discussion**

The outcome analysis should focus on three core dimensions: scientific validity, feasibility, and practical effectiveness. Scientific validity analysis emphasizes reliability and validity metrics. If a dimension's α coefficient is below 0.7 or factor loading is less than 0.5, expert consultation should be sought to identify causes. For instance, the "effectiveness of ideological education in courses" indicator might lack sufficient observation points leading to inadequate reliability. By comparing new systems with traditional assessment methods—such as using paired-sample t-tests—we can verify whether multiple indicators better reflect teachers' comprehensive capabilities. For example, analyzing the positive correlation between teaching innovation scores and student physical fitness improvement requires a correlation coefficient (r) ≥0.3. Feasibility discussions must quantify evaluation costs and operational complexity through metrics like data collection time and labor input. Comparing pilot universities 'feedback and evaluating the consistency of qualitative indicators like "sports injury prevention guidance" using Kendall's harmony coefficient (W ≥0.6) helps address potential subjective biases caused by vague evaluation criteria. Practical effectiveness analysis relies on tracking teacher behavior changes, such as comparing participation rates in curriculum reforms and tutoring hours. Evaluating how assessment indicators guide teaching practices, combined with teacher interviews, helps determine if outcomes effectively support career development. For example, the "research commercialization" indicator could prompt teachers to prioritize applying teaching research. Any issues identified during validation should be addressed with improvement strategies aligned with physical education principles, providing theoretical support for system optimization.

**(4) Adjustment and improvement of the index system**

When adjusting the indicator system based on validation results, precise measures should be implemented to establish a "dynamic optimization" mechanism. For indicators with insufficient reliability, such as "social service contribution," they can be refined into quantifiable sub-indicators like "guidance hours for campus sports clubs" and "community fitness sessions." Clear evaluation criteria must also be established. Regarding indicators with low validity, such as "research achievements," evaluation dimensions need restructuring by adding practical items like "proportion of teaching reform papers" and "patent commercialization of sports training," while reducing the weight of purely theoretical papers. Process optimization should incorporate feasibility analysis findings. If the issue of "short evaluation cycles" is prominent, long-term indicators like "annual improvement rate in student physical fitness" could be adjusted to a "semester-stage assessment + annual comprehensive evaluation" model. Mobile data collection tools should be developed to simplify operations. Indicators with insufficient differentiation should be consolidated, such as merging "competition coaching frequency" and "athlete comprehensive quality improvement" into a unified "educational effectiveness through competitions" indicator. A dynamic adjustment mechanism should be established: conduct annual implementation effectiveness evaluations, organize national expert seminars every three years to revise indicators, and add specialized indicators reflecting changes in sports education policies. This will ultimately form an assessment system that combines scientific rigor, operational feasibility, and adaptability.

**epilogue ：**

This study has established and validated a performance evaluation system for university physical education teachers under a multi-dimensional assessment framework. Breaking away from traditional single-dimensional quantification models, the system incorporates multiple evaluation metrics that have been proven scientifically sound and practically feasible. Its implementation facilitates comprehensive assessments of teaching competencies, stimulates pedagogical enthusiasm, and guides educators to focus on both instructional quality and student development. Going forward, continuous optimization through practical application will provide actionable references for evaluating physical education faculty, thereby advancing the sustainable development of sports education.

**reference documentation ：**

[1] Dong Yameng. Analysis of Performance Evaluation of Middle School Physical Education Teachers under the Perspective of Energy Level Theory [J]. Contemporary Sports Science and Technology, 2022,12(25):194-198.

[2]Liu Zihan, Wang Rongrong, and Zeng Jing. Research on Influencing Factors of Scientific Research Performance of College Physical Education Teachers. ——Grounded Theory [C]. Chinese Society of Sports Science. Abstract Compilation of the 12th National Sports Science Conference. ——Wallpaper Presentation (School Physical Education Branch). Beijing Sport University; 2022:800-802.

[3] Zheng Xiaofeng. Research on the Professional Development of Physical Education Teachers in Universities in Western China from a Institutional Perspective [D]. Shanghai University of Sport, 2021.

[4] Xu Bin. Research on the Construction of Multi-dimensional Performance Evaluation Index System for Physical Education Teachers in Chinese Universities [J]. Journal of Shenyang Sport University, 2020,39 (05):66-73.

[5] Sun Hao. Research on the relationship between work-family promotion, job satisfaction and job performance of college physical education teachers [D]. Shandong Sport University, 2020.