Pacific International Journal, Vol. 9(3); 2026
ISSN (Print) 2663-8991, ISSN (Online) 2616-4825[1]
DOI:10.55014/pij.v9i3.1026
https://rclss.com/index.php/pij

A Performance-Data-Driven Evaluation Model for Physical Education Teaching Quality in Higher Vocational Colleges

Zhiqiang Li

Linyi Vocational College, Linyi, China

Abstract: This study develops and preliminarily validates a teaching-quality evaluation model for physical education (PE) in higher vocational colleges using student sports-performance data. The model responds to two persistent limitations in conventional PE evaluation: an excessive focus on terminal test scores rather than learning processes, and insufficient attention to occupation-specific physical demands. Guided by the reform orientation of integrating job requirements, courses, competitions and certificates in vocational education, the study constructs an indicator system covering four dimensions: general physical fitness, occupation-specific physical fitness, motor skills and participation. The Delphi method, analytic hierarchy process (AHP) and Markov-chain modelling are integrated to estimate indicator weights and to analyse dynamic transitions from basic fitness attainment to occupational fitness adaptation. A 10-week pilot application was conducted in an automotive maintenance programme and a nursing programme at a higher vocational college in Shandong Province, China. Results indicate that sports performance data collected through wearable devices and professional training equipment were strongly associated with PE teaching evaluation outcomes ($r = 0.85$), with occupation-specific physical fitness contributing 42% of the explanatory weight. Student participation mediated the relationship between sports performance and teaching evaluation, with a stronger mediating effect in occupation-specific training modules than in general fitness modules. Dynamic monitoring also showed a moderating effect (effect value = 0.41), particularly in engineering-related training scenarios. The findings suggest that embedding occupational adaptability into data-driven PE evaluation can provide a dual reference framework that integrates general physical development with professional competence requirements. The model offers a practical basis for process-oriented assessment, curriculum optimisation and professionalised PE reform in higher vocational education.
Keywords: higher vocational physical education; sports performance data; occupation-specific physical fitness; evaluation model; Markov chain; vocational education reform

1. Introduction

Physical education in higher vocational colleges is part of both general education and vocational talent cultivation. It develops students' health, movement literacy, and the physical readiness needed for future work. International discussions of quality physical education emphasise lifelong participation, inclusive learning, and evidence-informed teaching [1]. Research on physical literacy also indicates that assessment should capture progression in physical, cognitive, and affective learning rather than only final performance scores [2].

In many higher vocational colleges, however, PE evaluation still follows a general university model. It usually emphasises running, jumping, and other standard fitness indicators. These indicators are important, but they do not fully reflect the physical demands of occupational posts. Automotive maintenance and logistics students may need lumbar-back strength, repeated handling endurance, and movement stability. Nursing students may need cardiopulmonary endurance, standing tolerance, and fine-motor coordination during emergency-care procedures. A teaching-quality evaluation model for vocational PE should therefore consider whether students' physical capacities support their professional training tasks.

Another limitation is the dependence on final tests and teacher judgement. Such methods provide only a partial picture of learning. Higher vocational students improve through PE classes, occupational simulations, and practicum activities. A more useful evaluation system should record the learning process and identify how students move from basic fitness attainment to occupational fitness adaptation. Formative assessment research shows that data literacy, collaboration, and pedagogical knowledge are important conditions for using assessment to improve teaching [5].

Wearable devices and professional training-monitoring equipment create new possibilities for PE evaluation. Wearable technologies can provide repeated measures of heart rate, exercise intensity, activity duration, and movement patterns [3]. Recent research also shows that wearable IoT devices can influence PE outcomes among college students, and that student acceptance or engagement can mediate this effect [4]. These findings support the use of data-driven and process-oriented assessment, but they need to be adapted to vocational settings.

This study develops and preliminarily validates a dual-dimensional evaluation model for PE teaching quality in higher vocational colleges. The model combines general physical fitness with occupation-specific physical fitness. It uses expert consultation and AHP to determine indicator weights, Markov-chain analysis to describe student state transitions, and structural equation modelling to test mediation and moderation mechanisms. The study focuses on three questions: (1) How can vocationally relevant PE evaluation indicators be selected and weighted? (2) How can repeated performance

[1] [Received 15 April 2026; Accepted 16 May 2026; Published (online) 20, June, 2026]

data be used to track students' transition toward occupational fitness adaptation? (3) How do participation and dynamic monitoring influence the relationship between sports-performance data and teaching evaluation outcomes?

2. Methods and Evaluation Framework

2.1 Research design and participants

The study used a quasi-experimental pilot design. The pilot was conducted in one higher vocational college in Shandong Province, China. Two programmes were selected because they represent different occupational physical demands: automotive maintenance and nursing. Each programme included three intact classes with 90 students. Within each programme, one set of intact classes was assigned to the experimental condition and one set to the control condition. Each group contained 45 students, yielding a total sample of 180 students.

The intervention lasted 10 weeks, including 8 weeks of PE classroom training and 2 weeks of practicum-related monitoring. The experimental groups used the proposed evaluation model and added occupation-specific physical training modules. The control groups used routine PE training and conventional evaluation. For automotive maintenance, the occupation-specific module focused on lumbar-back strength, upper-limb pushing force, handling endurance, and safe movement during parts handling. For nursing, it focused on cardiopulmonary endurance, standing tolerance, movement stability, and emergency-care simulation.

2.2 Expert consultation and indicator selection

Three rounds of Delphi consultation were used to select and refine the indicators. To improve methodological transparency, the expert panel is now described more fully. The panel consisted of 15 experts: six PE teachers from higher vocational colleges, three professional programme leaders from automotive or nursing programmes, three enterprise or hospital training specialists familiar with occupational physical demands, and three researchers in educational evaluation or sports science. Most experts had more than 10 years of professional experience, and eight held senior professional titles. The experts reviewed the relevance, feasibility, and measurability of each proposed indicator.

Indicators were retained when they met three criteria: direct relevance to PE teaching quality, observable or measurable data sources, and clear linkage to either general fitness development or occupation-specific physical tasks. Indicators with low consensus were revised or removed. This process produced four first-level dimensions: general physical fitness, occupation-specific physical fitness, motor skills, and participation. The selection logic was kept simple so that colleges with different resources can adapt the model without excessive technical burden.

2.3 Evaluation indicators and data sources

Table 1 presents the revised indicator framework. Compared with the original version, the table has been shortened and reorganised. It explains why each dimension is included, how it can be measured, and how it can be adapted to different programme types.

Dimension	Rationale	Core measures	Typical data source and adaptation
General physical fitness	Provides the baseline physical foundation required by all students.	Heart-rate response; exercise intensity; running, rope-skipping or similar basic fitness tasks.	Smart wristbands or PE monitoring systems; applicable to all programmes.
Occupation-specific physical fitness	Connects PE evaluation with the physical demands of occupational tasks.	Lumbar-back strength, handling endurance, cardiopulmonary endurance, standing tolerance, fine-motor coordination.	Dynamometers, step-test instruments, motion capture, practicum logs; measures are selected by programme.
Motor skills	Assesses whether students can perform movements safely, efficiently and consistently.	Movement standardisation; technical mastery; integrated motor performance in sport and practicum contexts.	Video analysis, skill tests and teacher rubrics; criteria are aligned with classroom and occupational scenarios.
Participation	Reflects engagement in process-oriented learning and predicts whether data feedback is used.	Attendance, active exercise duration, special-module participation, teacher-student and peer interaction.	Behavioural records, device logs and classroom/practicum observation; applicable to all programmes.

Table 1. Revised indicator framework for vocational PE teaching-quality evaluation

2.4 AHP weighting procedure

AHP was used to determine indicator weights because it supports structured comparison among multiple criteria [6]. Experts completed pairwise comparisons using a nine-point scale. The first-level dimensions and second-level indicators were compared separately for the automotive maintenance and nursing programmes. This allowed the model to maintain a shared framework while adapting the weights to different occupational demands.

The pairwise comparison matrices were tested for consistency. Matrices with a consistency ratio above 0.10 were returned to the experts for revision. After revision, the consistency ratios of the first-level matrices were below 0.10 in both programmes, indicating acceptable consistency. All indicators were normalised before aggregation. The composite evaluation score was calculated as $Q = \Sigma(w_k x_k)$, where x_k is the normalised value of indicator k and w_k is the AHP weight.

First-level dimension	Automotive weight	Nursing weight	Reason for programme difference
General physical	0.30	0.28	Both programmes require a basic fitness foundation; the

First-level dimension	Automotive weight	Nursing weight	Reason for programme difference
fitness			weight is slightly lower in nursing because cardiopulmonary occupational indicators are emphasised.
Occupation-specific physical fitness	0.35	0.38	This dimension receives the highest weight because it directly reflects job-related physical readiness. Nursing gives more weight to cardiopulmonary endurance and standing tolerance.
Motor skills	0.25	0.24	Movement quality matters in both programmes, including safe handling in automotive maintenance and standardised movement in care procedures.
Participation	0.10	0.10	Participation is treated as a process factor and a mediator rather than as the main outcome.

Table 2. Simplified AHP first-level weights by programme

2.5 Markov-chain state transition modelling

Markov-chain modelling was used to describe changes in students' physical readiness across repeated observations. A Markov chain is suitable when the next state can be estimated from the current state and observed transition probabilities [7]. In this study, states were defined by combining general physical fitness and occupation-specific physical fitness. The purpose was not to predict long-term labour-market performance. It was to provide a process indicator showing whether PE teaching helped students move toward occupational fitness adaptation.

Let $S(t)$ denote the physical readiness state of a student at observation time t. The transition probability from state i to state j was estimated as $p_{ij} = n_{ij} / \Sigma_j n_{ij}$, where n_{ij} is the number of observed transitions from state i to state j. The transition matrix $P = [p_{ij}]$ therefore represents the probability that students remain in the same level, improve to a higher level, or decline to a lower level after a period of instruction or practicum-related training.

State	Meaning	Operational interpretation
State 1: Below basic standard	General fitness is below the required baseline and occupation-specific fitness is insufficient.	Student needs basic conditioning before occupationally oriented training can be effective.
State 2: Basic standard attained	General fitness meets the baseline, but occupational fitness is unstable.	Student can complete simple practicum tasks but fatigue or movement instability appears.
State 3: Occupationally adapted	General and occupation-specific fitness support stable task completion.	Student can complete repeated practicum tasks with acceptable safety and efficiency.
State 4: Occupationally excellent	Physical readiness exceeds minimum job-related requirements.	Student can sustain higher-intensity practicum or competition-related tasks.

Table 3. Markov-chain state definitions

2.6 Statistical analysis

The revised statistical strategy is reported explicitly in response to the reviewers' comments. First, descriptive statistics were calculated for all indicators. Second, Pearson correlation analysis examined the relationship between sports-performance data and PE teaching evaluation scores. Third, structural equation modelling (SEM) tested the pathway from sports-performance data to teaching evaluation outcomes. Student participation was tested as a mediator, and dynamic monitoring was tested as a moderator. Bootstrap estimation with 5,000 resamples was used to estimate confidence intervals for indirect effects. Fourth, cross-validation and programme-consistency analysis were used to examine whether the model performed similarly in automotive maintenance and nursing.

Model fit was evaluated using common SEM indices, including chi-square divided by degrees of freedom (χ^2/df), comparative fit index (CFI), Tucker-Lewis index (TLI), root mean square error of approximation (RMSEA), and standardised root mean square residual (SRMR). The analysis was exploratory because the sample came from one college and two programmes. Therefore, the results are interpreted as preliminary evidence rather than as proof of universal applicability.

3. Results

3.1 Performance data and teaching evaluation

Sports-performance data were strongly associated with PE teaching evaluation outcomes. The overall correlation between the composite performance-data score and the teaching-evaluation score was $r = 0.85$. The correlation for occupation-specific physical fitness was $r = 0.72$ ($p < 0.01$), which was higher than the correlation for general physical fitness ($r = 0.65$). This result indicates that the vocationally relevant indicators added explanatory value beyond ordinary fitness indicators.

The practical meaning of this result is clear. In the automotive maintenance experimental group, the addition of lumbar-back strength and load-handling training was associated with an 18.5-point increase in the teaching-evaluation score compared with the control group. Parts-handling efficiency improved by 22%. In the nursing experimental group, participation in cardiopulmonary endurance training was 35% higher than participation in conventional running training. These findings suggest that students may be more engaged when PE tasks are visibly connected to future occupational situations.

3.2 Markov-chain transition results

The Markov-chain results showed that the experimental groups had higher probabilities of moving from State 2 to State 3 than the control groups. This transition is important because it represents movement from basic fitness attainment to occupational adaptation. In the automotive maintenance programme, the State 2-to-State 3 transition probability was 0.45. In the nursing programme, the corresponding probability was 0.50. The nursing value was slightly higher, probably because cardiopulmonary indicators respond more quickly to short-term targeted training and can be monitored with more standardised equipment.

The State 3-to-State 4 transition probability was 0.25 in automotive maintenance and 0.27 in nursing. These values should be interpreted cautiously. They show improvement within the pilot period, but they do not prove long-term occupational readiness. Longer interventions and follow-up observations are needed to test whether state improvement is retained.

3.3 Mediation, moderation and model fit

Table 4 reports the revised SEM results. The table adds the analytic information requested by the reviewers, including path coefficients, confidence intervals and model fit indices. The estimates should be understood as exploratory because the sample size and institutional scope were limited.

Model component	Estimate	95% CI	p value	Interpretation
Sports-performance data -> teaching evaluation	β = 0.62	0.49 to 0.74	< 0.001	Performance data had a strong positive association with evaluation outcomes.
Sports-performance data -> participation	β = 0.55	0.42 to 0.67	< 0.001	Better performance feedback was associated with stronger participation.
Participation -> teaching evaluation	β = 0.43	0.28 to 0.58	< 0.001	Participation partly explained why performance data related to evaluation outcomes.
Indirect effect through participation	0.43	0.27 to 0.56	< 0.001	The mediation effect was stronger in occupation-specific modules than in general fitness modules.
Performance data × dynamic monitoring	β = 0.41	0.23 to 0.58	< 0.001	Dynamic monitoring strengthened the performance-evaluation relationship.
Model fit	χ^2/df = 1.83; CFI = 0.951; TLI = 0.936; RMSEA = 0.061; SRMR = 0.048	—	—	Fit indices suggested acceptable exploratory model fit.

Table 4. Structural-equation and moderation results

The mediation result suggests that data feedback did not improve evaluation outcomes only by recording physical performance. It also increased participation, especially when students perceived the training module as relevant to future occupational tasks. The moderation result indicates that real-time or repeated monitoring made the performance-evaluation relationship stronger. This effect was stronger in practicum monitoring (0.47) than in ordinary classroom monitoring (0.36).

3.4 Validity test

Cross-validation and programme-consistency analysis showed that the proposed model performed better in the experimental groups than in the control groups. The results do not establish broad generalisability, but they support the internal feasibility of the model in the two pilot programmes.

Validation item	Automotive experimental	Automotive control	Nursing experimental	Nursing control	Inter-programme consistency
State-prediction accuracy	88.9%	71.1%	91.1%	73.3%	0.92
Transition-probability stability	0.94	0.82	0.95	0.83	0.93
Evaluation-result reliability	0.91	0.78	0.92	0.79	0.90

Table 5. Validity-test results of the evaluation model

The nursing experimental group showed slightly higher state-prediction accuracy than the automotive maintenance experimental group. This may be because nursing indicators such as heart-rate stability and step-test performance are easier to measure repeatedly with standardised instruments. The automotive maintenance indicators were more affected by task difficulty, posture, and equipment location. This difference reinforces the need for programme-specific calibration.

4. Discussion

4.1 Main findings and practical meaning

This study shows that PE teaching-quality evaluation in higher vocational colleges can be improved by combining general fitness indicators with occupation-specific physical indicators. The results support the idea that vocational PE should not be evaluated only by terminal running or jumping tests. It should also ask whether PE teaching helps students develop the physical capacities needed in professional learning and practicum tasks.

The model also provides a practical explanation of why data-driven evaluation may work. Performance data had a strong relationship with evaluation outcomes, but student participation was an important pathway. When students saw a direct connection between PE training and future work, they were more willing to participate. This finding is consistent with international research showing that acceptance and engagement can mediate the impact of wearable technologies on PE outcomes [4].

Dynamic monitoring also had practical value. In the automotive maintenance module, teachers could adjust training intensity when monitoring data indicated fatigue during handling tasks. In the nursing module, heart-rate response and step-test data helped teachers identify students who needed additional endurance support. Therefore, the model is not only an evaluation tool. It can also be used as a feedback tool for teaching adjustment.

4.2 Application in different colleges and programmes

For implementation in other colleges, the model can be applied in four steps. First, each programme should identify the physical demands of typical occupational tasks. This can be done through consultation with professional teachers, enterprise mentors, hospitals, training bases, and skill-competition standards. Second, the college should select a small number of measurable indicators. The indicators do not need to be identical across programmes, but they should be valid, observable, and feasible.

Third, colleges should choose data-collection methods according to their resources. Well-equipped colleges can use smart wristbands, electronic dynamometers, motion capture, and professional monitoring platforms. Colleges with fewer resources can use simpler tools such as step-test instruments, mechanical dynamometers, PE logs, video rubrics, and structured observation forms. The key requirement is that the data should correspond to meaningful movement tasks.

Fourth, evaluation results should be fed back into teaching. Students in State 1 need basic conditioning. Students in State 2 need support to build occupational adaptation. Students in State 3 need stable task practice. Students in State 4 may receive advanced or competition-oriented training. In this way, the model supports layered teaching rather than a single standard for all students.

Different vocational fields can adapt the model. Engineering and logistics programmes can emphasise load-handling endurance, lumbar-back strength, grip strength, and movement safety. Medical and health programmes can emphasise cardiopulmonary endurance, standing tolerance, emergency-operation rhythm, and fine-motor stability. Service programmes can emphasise posture control, prolonged standing, fatigue management, and interpersonal movement coordination. Digital and office-oriented programmes can include sedentary-risk prevention, postural health, shoulder-neck mobility, and exercise participation.

4.3 Implications for teaching reform

The model has three implications for teaching reform. First, higher vocational colleges should establish a closed loop linking programme requirements, PE curriculum design, process evaluation, and teaching feedback. PE teachers should not work separately from professional teachers. The interpretation of occupation-specific indicators requires cross-disciplinary cooperation.

Second, evaluation should become formative. Teachers can use data to identify students who improve slowly, fatigue quickly, or participate passively. The purpose is not to rank students by device data. The purpose is to provide timely feedback and adjust teaching. This is consistent with formative assessment research, which emphasises assessment literacy and teacher collaboration [5].

Third, the model can support professionalised PE reform without making PE narrowly utilitarian. General fitness remains necessary. Occupationally oriented indicators are added to strengthen relevance, not to replace the broader educational function of PE. A balanced model can therefore support both health development and vocational readiness.

4.4 Limitations and future research

The study has important limitations. First, the pilot was conducted in one college and involved only two programmes. The findings cannot be generalised directly to all higher vocational institutions. Multi-site studies with larger samples and more professional fields are needed. Second, the study used intact classes rather than full random assignment. Although this design was feasible in a teaching setting, future research should use stronger designs where possible and report baseline equivalence, missing-data handling, and sensitivity analyses.

Third, some indicators depend on device-based data collection. Colleges with limited equipment may face implementation barriers. Future studies should compare high-resource and low-resource versions of the model. Fourth, the intervention lasted only 10 weeks. Longer follow-up is needed to test whether state transitions are stable and whether improved occupational physical readiness is retained during internships or early employment.

Future research should also test automatic recognition of occupational movement standardisation through artificial intelligence and video analysis. However, data privacy and ethical governance must be considered. Student data should be anonymised, used only for educational purposes, and interpreted with professional judgement rather than treated as a complete representation of student ability.

5. Conclusion

This study proposed and preliminarily tested a performance-data-driven evaluation model for PE teaching quality in higher vocational colleges. The model integrates general physical fitness, occupation-specific physical fitness, motor skills, and participation. It uses AHP to determine weights, Markov-chain analysis to track state transitions, and SEM to examine the role of participation and dynamic monitoring. The pilot results suggest that occupation-specific performance data can strengthen PE teaching evaluation and provide useful feedback for curriculum improvement. The main contribution is the construction of a process-oriented and vocationally adaptive evaluation framework. Because the evidence comes from a single-college pilot, future research should test the model across more institutions, disciplines, and teaching conditions.

References

[1] UNESCO, Quality Physical Education (QPE): Guidelines for Policy-Makers. Paris: UNESCO Publishing, 2015.

[2] D. A. Dudley, “A conceptual model of observed physical literacy,” The Physical Educator, vol. 72, no. 5, pp. 236–260, 2015, doi: 10.18666/TPE-2015-V72-I5-6020.

[3] A. C. Sousa, S. N. Ferrinho, and B. Travassos, “The use of wearable technologies in the assessment of physical activity in preschool- and school-age youth: Systematic review,” International Journal of Environmental Research and Public Health, vol. 20, no. 4, Art. no. 3402, 2023, doi: 10.3390/ijerph20043402.

[4] Y. Xu, J. Peng, F. Jing, and H. Ren, “From wearables to performance: How acceptance of IoT devices influences physical education results in college students,” Scientific Reports, vol. 14, Art. no. 23776, 2024, doi: 10.1038/s41598-024-75071-3.

[5] K. Schildkamp, F. M. van der Kleij, M. C. Heitink, W. B. Kippers, and B. P. Veldkamp, “Formative assessment: A systematic review of critical teacher prerequisites for classroom practice,” International Journal of Educational Research, vol. 103, Art. no. 101602, 2020, doi: 10.1016/j.ijer.2020.101602.

[6] T. L. Saaty, The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation. New York: McGraw-Hill, 1980.

[7] J. R. Norris, Markov Chains. Cambridge: Cambridge University Press, 1997.

[8] J. Wang, “Reform paths of physical education curriculum under the background of job-course-competition-certificate integration in higher vocational education,” Vocational and Technical Education, vol. 44, no. 26, pp. 68–72, 2023, in Chinese.

[9] G. Li, “Construction of a higher vocational physical education evaluation system based on occupational physical fitness requirements,” Sport, no. 15, pp. 89–91, 2022, in Chinese.

[10] L. Zhang, “Empirical study on the connection between physical education courses and occupational competence in higher vocational nursing programmes,” Health Vocational Education, vol. 42, no. 3, pp. 112–114, 2024, in Chinese.

[11] M. Chen, “Application of Markov chains in the evaluation of occupational physical fitness improvement among higher vocational students,” Mathematics in Practice and Theory, vol. 53, no. 12, pp. 289–296, 2023, in Chinese.

[12] B. Xie, “Evaluation and optimisation of school sports-meeting quality based on importance-performance analysis,” Journal of Kashgar University, vol. 41, no. 6, pp. 82–88, 2020, in Chinese.

[13] Y. Jiang, H. Wang, and Z. Pan, “A model for evaluating primary and secondary school students’ sports ability based on core literacy,” Journal of Shenyang Sport University, vol. 38, no. 6, pp. 105–114, 2019, in Chinese.

[14] Z. Yi, X. Wang, and X. Li, “Prediction model of sports injury based on dynamic sampling and transfer learning,” Microprocessors and Microsystems, vol. 80, Art. no. 103583, 2021.