

## Evaluation Criteria for Quality Education of Physical Education Lessons Based on Logical Analysis

<https://doi.org/10.3991/ijet.v16i21.26867>

Guirong Ning

Guizhou University of Commerce, Guiyang, China

lion4859@163.com

**Abstract**—To cultivate high-quality talents, it is important to develop a set of scientific and complete evaluation criteria for comprehensive quality. The effective evaluation could promote and guide teaching and learning and improve the overall quality of students. Considering the features of Physical Education (PE) lessons, this paper resorts to multiple methods, e.g., questionnaire survey, Delphi's method, mathematical statistics, and logical analysis, to establish a system of evaluation criteria for the quality education of PE lessons, the proposed system includes 4 primary indices, 10 secondary indices, and 24 tertiary indices. The weight of each index was determined through analytic hierarchy process (AHP), making the criteria more scientific. Taking the PE major of a college for example, the proposed criteria were investigated empirically. The results show that experts, teachers, and students made consistent evaluations. Thus, the proposed system and evaluation criteria are feasible, and capable of evaluating the overall quality of students objectively and fairly. The research results provide a tool and idea for improving the quality education of PE lessons in China, and promoting the evaluation of students' overall quality.

**Keywords**—logical analysis, physical education (PE), quality education, evaluation criteria

### 1 Introduction

Since the successful hosing of the Beijing Olympic Games in 2008, the sports industry has developed vigorously in China, at the same time, the booming economy of China has posed urgent needs on high-quality talents, in such context, the Chinese society's demand for high-quality sports talents is growing gradually. Colleges and universities are important bases for cultivating talents, however, in recent years, under the influence of the popularized higher education, the teaching quality of schools and the competencies of graduates have been questioned by the society, graduates of the sports major are facing an increasing employment pressure, therefore, developing a set of scientific and complete evaluation criteria for the quality education of PE lessons, promoting teaching and learning through evaluation, and cultivating excellent sports talents with outstanding professional ability and comprehensive quality are important works for college PE lessons.

Foreign studies prefer to carry out student evaluation starting from the learning effect of persons who receive the education [1]. British scholar Fisher is the first one to research student evaluation, and he had proposed a set of student evaluation criteria in the early 19th century [2]. Afterwards, countries all over the world have carried out research on the theory of education evaluation and obtained fruitful results. Theories such as the theory of education standardization and the theory of the diversification of education evaluation have been proposed successively [3]. The earliest research on the evaluation of higher education appeared in the United States, educationalists represented by Taylor believe that the evaluation of students should assess their overall quality, and textbooks shouldn't be the only evaluation criteria [4]. In China, scholars have achieved abundant results in terms of the research on the evaluation systems of college students' comprehensive quality, including the significance and current state of college students' comprehensive quality, as well as the research of relevant issues [5]. In terms of the understanding and improvement suggestions for such evaluation systems, Chinese scholars have researched the indicators and criteria of the evaluation of college students' comprehensive quality, the establishment of evaluation systems, and other aspects [6]. However, in terms of studies on the evaluation criteria for quality education of PE lessons, there're not much research results, let alone the construction of systems for evaluating the quality education of PE lessons [7].

After carefully reviewing relevant literatures published by domestic and foreign scholars, this paper took the features and current state of PE lessons into consideration. Drawing on theories in the fields of logic study, pedagogy, and psychology, this study proposed a set of quality education evaluation criteria for PE lessons with 4 primary indices, 10 secondary indices, and 24 tertiary indices. The weight value of each index was determined in the paper, and the feasibility and effectiveness of the proposed criterion system were verified through an empirical test.

## **2 The evaluation criterion system**

### **2.1 Methodology**

Whether the evaluation criteria are selected reasonably or not will directly affect the effectiveness and orientation of the quality education of PE lessons [8]. This study resorted to multiple methods such as questionnaire survey, Delphi's method, mathematical statistics, and logical analysis to establish the evaluation criterion system. Our research team invited 16 experts (professional teachers of sports major in colleges and experts engaged in related fields) to participate in two rounds of expert consultation questionnaire survey, 15 questionnaires were returned after each round of survey, after reliability and validity check, all returned questionnaires were verified to be valid and qualified [9].

## 2.2 System establishment

Based on the actual features of PE lessons and existing research results of world field scholars, at first, this paper established an initial evaluation criterion system for the target matter with 4 primary indices, 10 secondary indices, and 24 tertiary indices, as shown in Table 1 [10].

**Table 1.** Evaluation criterion system for quality education of PE lessons (First-round)

Primary index	Secondary index	Tertiary index
Ideological and moral quality U <sub>1</sub>	Political quality U <sub>11</sub>	Law and politics class resultsU <sub>111</sub>
		Political attitudeU <sub>112</sub>
		Political performanceU <sub>113</sub>
	Moral QualityU <sub>12</sub>	Style of study performanceU <sub>121</sub>
		Spirit of serviceU <sub>122</sub>
		Collective mindsetU <sub>123</sub>
Physical and mental qualities U <sub>2</sub>	Physical fitness U <sub>21</sub>	Physical fitness test resultsU <sub>211</sub>
		Exercise fitness habits U <sub>212</sub>
	Mental quality U <sub>22</sub>	Interpersonal harmonyU <sub>221</sub>
		Emotion regulationU <sub>222</sub>
		Objective self-evaluation U <sub>223</sub>
		Mental Health Scale for College StudentsU <sub>224</sub>
	Knowledge and skill level U <sub>3</sub>	Knowledge level U <sub>31</sub>
Public knowledgeU <sub>312</sub>		
Humanistic literacy U <sub>313</sub>		
Skill levels U <sub>32</sub>		Specific skill levelU <sub>321</sub>
		Skill level of other projectsU <sub>322</sub>
Ability U <sub>4</sub>	Competence for clubs and social activities U <sub>41</sub>	Association participationU <sub>411</sub>
		Social work and part-time workU <sub>412</sub>
	Quality of scientific research abilityU <sub>42</sub>	Subject competitionU <sub>421</sub>
		Academic papersU <sub>422</sub>
	Culture, arts, and sports ability U <sub>43</sub>	cultural activityU <sub>431</sub>
		Sports competitionU <sub>432</sub>
	Employability U <sub>44</sub>	Professional qualification certificateU <sub>441</sub>
		Non-professional qualification certificateU <sub>442</sub>
		Graduation internship U <sub>443</sub>
		Physical education abilityU <sub>444</sub>

**Screening results of first-round expert consultation questionnaire survey.** 5-point Likert scale was adopted in the survey and experts were asked to evaluate the degree of reasonableness of the initial evaluation criteria, the evaluation standards are given in Table 2 [11]. Then, two indicators, the arithmetic mean (Formula 1) and the coefficient of variation (Formula 2) were used to analyze the reasonableness of initial

evaluation criteria and the coordination of expert opinions. The more unified the opinions of experts, the higher the importance of a criterion, then the larger the arithmetic mean and the smaller the coefficient of variation [12].

**Table 2.** Evaluation standards for judging the reasonableness of evaluation criteria

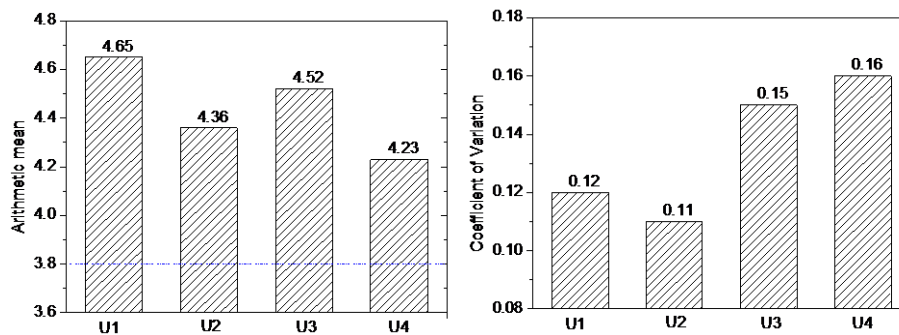
Degree of reasonableness	Very unreasonable	Unreasonable	General	Reasonable	Very reasonable
Points	1	2	3	4	5

$$\bar{X}_k = \frac{1}{k} \sum_{n=1}^k C_k \tag{1}$$

$$V_k = \frac{S_k}{\bar{X}_k} \tag{2}$$

where,  $k$  is the number of experts who returned valid questionnaires,  $C_{zk}$  is the score given by the  $z$ -th expert to the importance degree of index  $k$ ,  $S_k$  is the standard deviation. After calculation, indices that satisfy  $\bar{X}_k > 3.8$  and  $V_k < 0.2$  were included in the evaluation criterion system [13].

Figure 1 gives the arithmetic mean and variation coefficient of primary indices, as can be seen from the figures, all 4 primary indices had an arithmetic mean greater than 3.8 and a variation coefficient less than 0.2, indicating that the expert opinion coordination and index importance of the four indices were all at a high level, and they could enter the second-round expert evaluation.



**Fig. 1.** Arithmetic mean and variation coefficient of primary indices

Figure 2 shows the arithmetic mean and variation coefficient of secondary indices, as can be seen from the figure, all 10 secondary indices had an arithmetic mean greater than 3.8 and a variation coefficient less than 0.2, indicating that the expert opinion coordination and index importance of the ten indices were both at a high level, and they could enter the second-round expert evaluation.

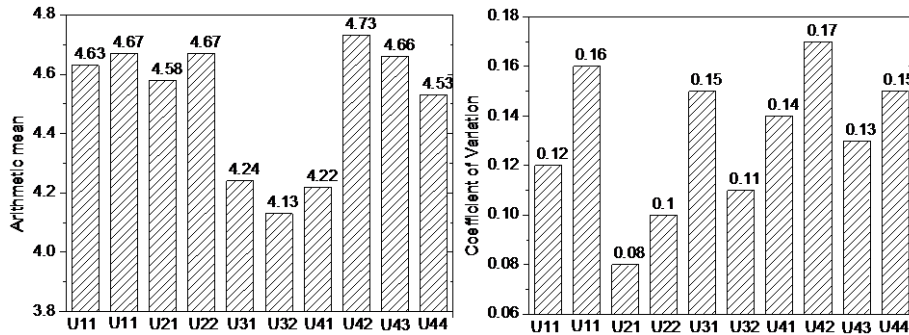


Fig. 2. Arithmetic mean and variation coefficient of secondary indices

Figure 3 shows the arithmetic mean and variation coefficient of tertiary indices, in the figure, three tertiary indices  $U_{223}$ ,  $U_{313}$ , and  $U_{443}$  had an arithmetic mean less than 3.8 and a variation coefficient greater than 0.2, indicating that the expert opinion coordination and index importance of the three indices were at a relatively low level, the rest indices met the requirements and could enter to the second-round expert evaluation.

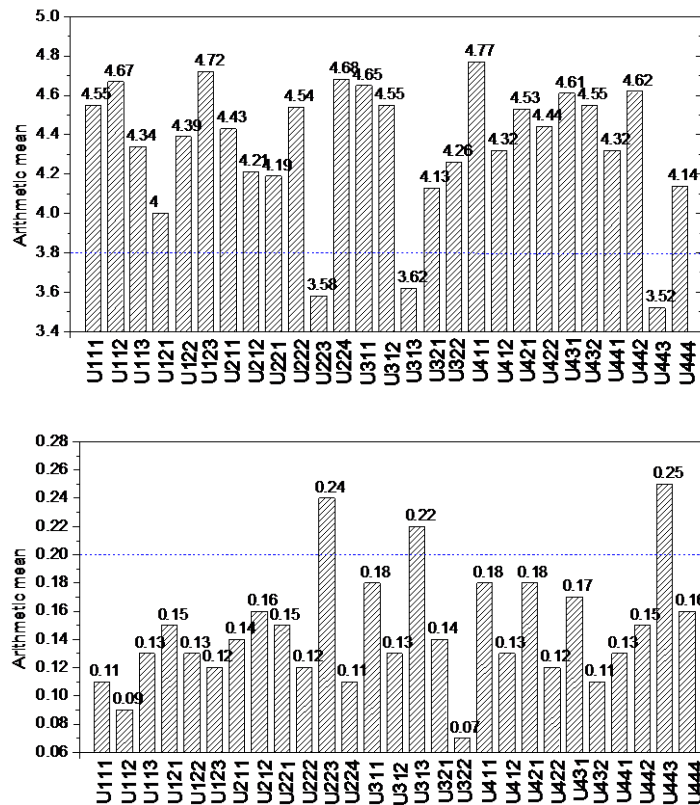


Fig. 3. Arithmetic mean and variation coefficient of tertiary indices

**Screening results of second-round expert consultation questionnaire survey.** According to the screening results of the first-round expert consultation questionnaire survey, three tertiary indices with an arithmetic mean less than 3.8 and a variation coefficient greater than 0.2 were deleted. 4 primary indicators, 10 secondary indicators and 24 tertiary indicators entered the second-round survey. Then, questionnaires of the two rounds were summarized and analyzed, experts basically held a satisfied attitude towards the modified indices, and all indices had an arithmetic mean greater than 3.8 and a variation coefficient less than 0.2. Finally, a PE lesson quality-education evaluation criterion system with 4 primary indicators, 10 secondary indicators and 24 tertiary indicators was established, as shown in Table 3.

**Table 3.** The proposed criterion system, indices, and weights

Primary index (weight)	Secondary index (weight)	Tertiary index (weight)
Ideological and moral quality U <sub>1</sub> (0.133)	Political quality U <sub>11</sub> (0.325)	Law and politics class resultsU <sub>111</sub> (0.123)
		Political attitudeU <sub>112</sub> (0.389)
		Political performanceU <sub>113</sub> (0.488)
	Moral Quality U <sub>12</sub> (0.675)	Style of study performanceU <sub>121</sub> (0.253)
		Spirit of serviceU <sub>122</sub> (0.342)
		Collective mindsetU <sub>123</sub> (0.405)
Physical and mental qualities U <sub>2</sub> (0.294)	Physical fitness U <sub>21</sub> (0.5)	Physical fitness test resultsU <sub>211</sub> (0.466)
		Exercise fitness habits U <sub>212</sub> (0.534)
	Mental quality U <sub>22</sub> (0.5)	Interpersonal harmonyU <sub>221</sub> (0.325)
		Emotion regulationU <sub>222</sub> (0.311)
		Mental Health Scale for College StudentsU <sub>223</sub> (0.364)
Knowledge and skill level U <sub>3</sub> (0.479)	Knowledge level U <sub>31</sub> (0.496)	Professional knowledgeU <sub>311</sub> (0.503)
		Public knowledgeU <sub>312</sub> (0.497)
	Skill levels U <sub>32</sub> (0.504)	Specific skill levelU <sub>321</sub> (0.634)
		Skill level of other projectsU <sub>322</sub> (0.366)
Ability U <sub>4</sub> (0.009)	Competence for clubs and social activities U <sub>41</sub> (0.173)	Association participationU <sub>411</sub> (0.363)
		Social work and part-time workU <sub>412</sub> (0.637)
	Quality of scientific research ability U <sub>42</sub> (0.231)	Subject competitionU <sub>421</sub> (0.643)
		Academic papersU <sub>422</sub> (0.357)
	Culture, arts and sports ability U <sub>43</sub> (0.452)	cultural activityU <sub>431</sub> (0.489)
		Sports competitionU <sub>432</sub> (0.511)
	Employability U <sub>44</sub> (0.144)	Professional qualification certificateU <sub>441</sub> (0.375)
		Non-professional qualification certificateU <sub>442</sub> (0.202)
Physical education abilityU <sub>443</sub> (0.423)		

### 3 Determination of evaluation standards

#### 3.1 Determination of index weight

**Method and process.** AHP was employed to determine the weight values of indices, first, experts were issued with the index importance consultation questionnaire and asked to score the importance degree of each index using 1-9 points, wherein 1, 3, 5, 7, and 9 points stand for that, when comparing two indices, the former index is equally important, slightly more important, more important, much more important, and extremely more important than the latter index; and 2, 4, 6, and 8 points stand for that, when comparing two indices, the importance degree of the two indexes is between the two adjacent degrees [14].

**Calculation of index weight.** This paper took the primary indices in the proposed system as examples to show the calculation process of the index weight values. Table 4 lists the weight assignment and consistency check results of the 4 primary indices [5].

**Table 4.** Weight assignment and consistency check results of 4 primary indices

	Ideological and moral quality U <sub>1</sub>	Physical and mental qualities U <sub>2</sub>	Knowledge and skill level U <sub>3</sub>	Ability U <sub>4</sub>	Weights w <sub>i</sub>
Ideological and moral quality U <sub>1</sub>	1	1/3	1/4	2	0.133
Physical and mental qualities U <sub>2</sub>	3	1	1/2	3	0.294
Knowledge and skill level U <sub>3</sub>	4	2	1	4	0.479
Ability U <sub>4</sub>	1/2	1/3		1	0.009

Consistency check: n=4, λ<sub>max</sub>=4.082, CI=0.027, RI=0.9, CR=CI/RI=0.03<0.1,

$$\text{where, } \lambda_{\max} = \frac{1}{n} \sum_{i=1}^n \left[ \frac{\sum_{j=1}^n a_{ij} w_j}{w_i} \right], \lambda_{\max} = \frac{1}{n} \sum_{j=1}^n \left[ \frac{\sum_{i=1}^n a_{ij} w_j}{w_j} \right], CI = \frac{\lambda_{\max} - n}{n - 1} \quad CR = \frac{CI}{RI}$$

In the same way, the weight values of secondary and tertiary indices could be calculated. Table 3 gives the determined indices and their weight values.

#### 3.2 Determination of evaluation standards

**Evaluation standards.** This paper selected 5 grades: excellent (90-100), good (80-89), average (70-79), pass (60-69), and fail (<60) as the evaluation standards for judging the quality education of PE lessons. The standard value took 95, 85, 75, and 65 points, respectively [15].

**Evaluation process.** Subjects in the evaluation included two parts: teachers and students. Since counsellor teachers and course teachers usually have more chances to con-

fact with students, the teacher subjects in this study were mainly composed of counselor teachers and course teachers [16]. The student subjects in this study were student representatives democratically elected by class students. All teacher and student subjects gave evaluations according to the given evaluation table for quality education of PE lessons, and they were asked to put the scores under the corresponding evaluation results, as shown in Table 5. After that, the numbers of student or teacher subjects who chose the corresponding grades of each index were counted, then the specific score of a student for a certain evaluation index could be obtained, after that, according to the weight coefficients, the final quality score of each student being evaluated could be obtained [17]. For example, Ideal and moral quality score=Political quality score×32.5%+Moral quality score×67.5%.

**Table 5.** Evaluation form for quality education of PE lessons

Primary index (weight)	Secondary index (weight)	Tertiary index (weight)	Evaluation grade				
			Excellent	Good	Average	Pass	Fail
Ideological and moral quality U <sub>1</sub> (0.133)	Political quality U <sub>11</sub> (0.325)	U <sub>111</sub> (0.123)					
		U <sub>112</sub> (0.389)					
		U <sub>113</sub> (0.488)					
	Moral Quality U <sub>12</sub> (0.675)	U <sub>121</sub> (0.253)					
		U <sub>122</sub> (0.342)					
		U <sub>123</sub> (0.405)					
Physical and mental qualities U <sub>2</sub> (0.294)	physical fitness U <sub>21</sub> (0.5)	U <sub>211</sub> (0.466)					
		U <sub>212</sub> (0.534)					
	Mental quality U <sub>22</sub> (0.5)	U <sub>221</sub> (0.325)					
		U <sub>222</sub> (0.311)					
		U <sub>223</sub> (0.364)					
Knowledge and skill level U <sub>3</sub> (0.479)	knowledge level U <sub>31</sub> (0.496)	U <sub>311</sub> (0.503)					
		U <sub>312</sub> (0.497)					
	Skill levels U <sub>32</sub> (0.504)	U <sub>321</sub> (0.634)					
		U <sub>322</sub> (0.366)					
Ability U <sub>4</sub> (0.009)	Competence for clubs and social activities U <sub>41</sub> (0.173)	U <sub>411</sub> (0.363)					
		U <sub>412</sub> (0.637)					
	Quality of scientific research ability U <sub>42</sub> (0.231)	U <sub>421</sub> (0.643)					
		U <sub>422</sub> (0.357)					
	Culture, arts, and sports ability U <sub>43</sub> (0.452)	U <sub>431</sub> (0.489)					
		U <sub>432</sub> (0.511)					
	Employability U <sub>44</sub> (0.144)	U <sub>441</sub> (0.375)					
		U <sub>442</sub> (0.202)					
		U <sub>443</sub> (0.423)					



### 4 Empirical test

A PE major student A from a college was taken as the evaluation object, and 12 students and 10 teachers were elected from A’s class to evaluate A’s comprehensive quality. The student and teacher evaluators were asked to fill in the evaluation form shown in Table 4, and then the scores of the comprehensive quality of evaluation object student A given by student evaluators and teacher evaluators were calculated respectively. With the primary index “Ideological and moral quality” as an example, this paper brief analyzed the evaluation process of student evaluators. Table 6 is a statistic of the scores of student A under each tertiary index. According to the data, the number of student evaluators who think student A’s performance in the law and politics course is excellent, good, average, pass, and fail were respectively 6, 3, 1, 1, and 0.

**Table 6.** Statistics of student A’s performance in ideological and moral quality

Primary index (weight)	Secondary index (weight)	Tertiary index (weight)	Evaluation grade (Standard value)					Score
			Excellent 95	Good 85	Average 75	Pass 65	Fail 55	
Ideological and moral quality U <sub>1</sub> (0.133)	Political quality U <sub>11</sub> (0.325)	U <sub>111</sub> (0.123)	6	3	1	1	0	85
		U <sub>112</sub> (0.389)	7	2	3	0	0	88.33
		U <sub>113</sub> (0.488)	8	2	1	1	0	89.17
	Moral Quality U <sub>12</sub> (0.675)	U <sub>121</sub> (0.253)	7	3	1	0	1	87.5
		U <sub>122</sub> (0.342)	6	5	1	0	0	89.17
		U <sub>123</sub> (0.405)	5	4	2	1	0	85.83

Then, student A’s score in the law and politics course was [18]:

$$S_{(U_{111})} = 6/12 \times 95 + 3/12 \times 85 + 1/12 \times 75 + 1/12 \times 65 + 0/12 \times 55 = 85$$

In the same way, the scores of other tertiary indices could also be obtained. After that, according to the weight values of tertiary indices, the score of the secondary index could be obtained, for example, the score of students A’ political quality was [19]:

$$\begin{aligned} S_{(U_{11})} &= S_{(U_{111})} \times w_{111} + S_{(U_{112})} \times w_{112} + S_{(U_{113})} \times w_{113} \\ &= 85 \times 12.3\% + 88.33 \times 0.389 + 89.17 \times 0.488 \\ &= 88.33 \end{aligned}$$

Similarly, we can get  $S_{(U_{12})}=87.39$ , then the final score of the primary index “Ideological and moral quality” was:

$$\begin{aligned} S_{(U_1)} &= S_{(U_{11})} \times w_{11} + S_{(U_{12})} \times w_{12} \\ &= 88.33 \times 0.325 + 87.39 \times 0.675 \\ &= 88.90 \end{aligned}$$

Judging from the definition of each evaluation grade, the score of student A in terms of ideological and moral quality given by student evaluators was 88.97 points, which was between 80 points and 89 points, indicating that student A's performance in terms of ideological and moral quality was good.

Using the same method, we can get  $S_{(U2)}=88.68$ ,  $S_{(U3)}=87.03$ , and  $S_{(U4)}=91.15$ . Then, student A's comprehensive quality score was:

$$\begin{aligned} S &= S_{(U1)} \times w_1 + S_{(U2)} \times w_2 + S_{(U3)} \times w_3 + S_{(U4)} \times w_4 \\ &= 88.90 \times 0.133 + 88.68 \times 0.294 + 87.03 \times 0.479 + 91.15 \times 0.009 \\ &= 80.30 \end{aligned}$$

The student evaluation result was 80.30 points, between 80-89 points, indicating that student evaluators think that student A's comprehensive quality was good.

Table 7 is a statistic of student A's comprehensive quality result given by teacher evaluators.

**Table 7.** Student A's comprehensive quality result given by teacher evaluators

Comprehensive quality score	Primary index (weight)	Score
81.03	Ideological and moral quality U1(0.133)	89.12
	Physical and mental qualities U2 (0.294)	88.73
	Knowledge and skill level U3(0.479)	88.27
	Ability U4(0.009)	90.43

Figure 4 compares the evaluation results of the comprehensive quality of student A given by student and teacher evaluators. According to the figure, there's little difference between the evaluation results of each index given by student evaluators and teacher evaluators, and there's not much difference in the final evaluation results of student A's comprehensive quality, indicating that the proposed system designed in this paper is feasible and effective.

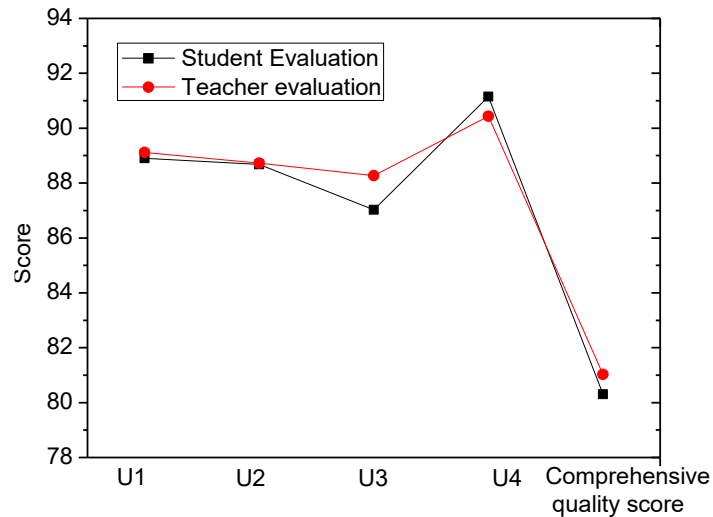


Fig. 4. Comparison of evaluation results of the comprehensive quality of student A given by student and teacher evaluators

## 5 Conclusions

How to cultivate high-quality PE major talents who can meet the requirements of social development is a key issue for higher educational schools and a key topic for the research of quality education. This paper researched the evaluation criteria of the quality education of PE lessons based on logical analysis and obtained following conclusions:

1. With the help of questionnaire survey, Delphi's method, mathematical statistics, and logical analysis, this paper established an evaluation criterion system with 4 primary indices, 10 secondary indices, and 24 tertiary indices for the said matter.
2. This paper used AHP to determine the weight of each index, then it designed a form for evaluating the quality education of PE lessons, and introduced the specific process of the evaluation.
3. This paper conducted an empirical test on the proposed system and proved its feasibility and effectiveness. The designed system can be used for the quantitative evaluation of students' comprehensive quality and the analysis of the quality of a specific index of a student, and it could facilitate teachers and students to make targeted improvements.

## 6 References

- [1] Liu, Y.R. (2021). An artificial intelligence and machine vision based evaluation of physical education teaching. *Journal of Intelligent and Fuzzy Systems*, 40(2): 3559-3569. <https://doi.org/10.3233/jifs-189392>

- [2] Simpson, D., Fincher, R., HAFler, J.P., Irby, D.M., Viggiano, T.R. (2007). Advancing educators and education by defining the components and evidence associated with educational scholarship. *Medical Education*, 41(10): 1002-1009. <https://doi.org/10.1111/j.1365-2923.2007.02844.x>
- [3] Pletinckx, J., Segers, M. (2001). Programme evaluation as an instrument for quality-assurance in a student-oriented educational system. *Studies in Educational Evaluation*, 27(4): 355-372. [https://doi.org/10.1016/s0191-491x\(01\)00035-9](https://doi.org/10.1016/s0191-491x(01)00035-9)
- [4] Gangstead, S.K., Beveridge, S.K. (1984). The implementation and evaluation of a methodological approach to qualitative sport skill analysis instruction. *Journal of Teaching in Physical Education*, 3(2): 60-70. <https://doi.org/10.1123/jtpe.3.2.60>
- [5] Wang, X.L., Wang, J.C., Li, Z.X. (2020). A Questionnaire Survey on the Sense of Educational and Social Equities among College Students in China, *International Journal of Emerging Technologies in Learning*, 15(14), 122-136. <https://doi.org/10.3991/ijet.v15i14.15357>
- [6] Guo, Q. (2020). Detection of head raising rate of students in classroom based on head posture recognition. *Traitement du Signal*, 37(5): 823-830. <https://doi.org/10.18280/ts.370515>
- [7] Bailey, R. (2010). Physical education and sport in schools: a review of benefits and outcomes. *Journal of School Health*, 76(8): 397-401. <https://doi.org/10.1111/j.1746-1561.2006.00132.x>
- [8] Cope, A.B. (1976). Areawide educational evaluation. *Journal of Continuing Education in Nursing*, 7(3): 38-41. <https://doi.org/10.3928/0022-0124-19760501-12>
- [9] Wolf, R.M. (1987). Educational evaluation: the state of the field. *International Journal of Educational Research*, 11(1), 1-143. [https://doi.org/10.1016/0883-0355\(87\)90033-4](https://doi.org/10.1016/0883-0355(87)90033-4)
- [10] Silverman, S., Subramaniam, P.R. (1999). Student attitude toward physical education and physical activity: A review of measurement issues and outcomes. *Journal of Teaching in Physical Education*, 19(1): 97-125. <https://doi.org/10.1123/jtpe.19.1.97>
- [11] Francis, N.C. (1977). Quality education and equality of educational opportunity. *NASSP Bulletin*, 61(409): 16-25. <https://doi.org/10.1177/019263657706140903>
- [12] Dello-Iacovo, B. (2009). Curriculum reform and 'quality education' in China: an overview. *International Journal of Educational Development*, 29(3): 241-249. <https://doi.org/10.1016/j.ijedudev.2008.02.008>
- [13] Weade, G., Evertson, C.M. (1991). On what can be learned by observing teaching. *Theory Into Practice*, 30(1): 37-45. <https://doi.org/10.1080/00405849109543474>
- [14] Silverman, S., Skonie, R. (1997). Research on teaching in physical education: an analysis of published research. *Journal of Teaching in Physical Education*, 16(3): 300-311. <https://doi.org/10.1123/jtpe.16.3.300>
- [15] Willgoose, C.E. (2010). Evaluation in health education and physical education. *Journal of School Health*, 31(8): 278-278. <https://doi.org/10.1111/j.1746-1561.1961.tb02806.x>
- [16] Dudley, D., Goodyear, V., Baxter, D. (2016). Quality and health-optimizing physical education: Using assessment at the health and education nexus. *Journal of Teaching in Physical Education*, 35(4): 324-336. <https://doi.org/10.1123/jtpe.2016-0075>
- [17] Tomas, L., Ritchie, S.M. (2015). The challenge of evaluating students' scientific literacy in a writing-to-learn context. *Research in Science Education*, 45(1): 41-58. <https://doi.org/10.1007/s11165-014-9412-3>
- [18] Román, M.L., Muoz, S.L., Castuera, R.J. (2019). The importance of assigning responsibility during evaluation in order to increase student satisfaction from physical education classes: A structural equation model. *PLOS ONE*, 14(9), e0209398. <https://doi.org/10.1371/journal.pone.0209398>

- [19] Pierre, R.B., Wierenga, A., Barton, M., Branday, J.M., Christie, C.D. (2004). Student evaluation of an osce in paediatrics at the university of the West Indies, Jamaica. *BMC Medical Education*, 4(1), 22. <https://doi.org/10.1186/1472-6920-4-22>

## 7 Author

**Guirong Ning** graduated from the Physical Education Department of Guizhou Normal University. Now, he works in the Physical Education Department of Guizhou business college. His main research direction is physical education and training.

Article submitted 2021-09-05. Resubmitted 2021-10-10. Final acceptance 2021-10-12. Final version published as submitted by the author.