

THE PHILOSOPHY OF PERFORMANCE-BASED EVALUATION AND GRADING STANDARDS IN LIGHT OF CONTEMPORARY EDUCATIONAL TRENDS IN SOME COURSES AT THE FACULTY OF PHYSICAL EDUCATION.

Ashraf Sobhy Younis Kabil

Assistant Professor, Department of School Sport, faculty of physical education for Boys, Alexandria University, Egypt.

Introduction

Universities seek to achieve excellence and scientific development in all fields and educational activities through the collaboration of all elements of the educational system. Evaluation and setting grades is one of the main pillars of the development process and a tool for judging the extent of achieving educational goals and controlling the future of students.

Evaluation is a major process in the work of educational institutions and an integral part of the educational system, it works mainly and thoroughly with the processes of learning and teaching, and this is reflected in the form of grades that determine students' levels and abilities (4:23).

The contemporary educational reform movement has sought great interest in the evaluation process by placing increasing demands on students learning, which required showing higher levels of learning and mastery of content, and this resulted in increasing demands for accountability at a time of increasing dissatisfaction with the traditional forms of the evaluation process (42: 652).

The grades obtained by the learners control many decisions related to graduation and academic honor and determining whether students have achieved educational goals, the grades also have a positive or negative impact in obtaining a job or a job opportunity (40: 87).

Therefore, the grading system for students is one of the main objectives of the evaluation process during university teaching, and these grades reflect the levels of learners and their abilities as well as reflect the responsibility of the educational institution towards achieving its goals (1).

The performance-based evaluation shows the skills that the learner possesses and the extent of his ability to employ them in real learning situations, and this is evident in the light of indicators of rating the grade and the level of quality of performance through process evaluation and product evaluation, that is reflected to determine to what extent the educational objectives are achieved (7: 117, 118).

The performance-based evaluation process is also run through the process or the product evaluation using a list of criteria that are predefined; the evaluation grades are recorded by experienced evaluators in the content or skills that are evaluated to ensure the understanding and consistency of evaluation criteria and grading systems (42: 657).

Therefore, the grading systems are associated to the evaluation elements and the evaluation criteria determine how to distribute them, because it gives the ability to distinguish different levels of performance, there is also a link between grades and evaluation criteria, and grading systems contribute significantly to the integrity and reliability of evaluation and have great value, especially when some students contest and appeal the grades results , grading plans are also essential when more than one evaluator participates, which is in line with university requirements, grading also helps judge the validity of the assessment and how it relates to learning outcomes (6: 20, 21).

There are many different points of view regarding grades in physical education, as many methods are used in classifying students and grading systems, whereas grading systems are based on teachers' philosophy and educational regulations, and this makes the points of views different on the appropriate procedure for grades in the field of physical education (25: 243).

Therefore, the university educational system witnesses changes in the roles of the faculty member from a knowledge expert to a knowledge facilitator; universities are also keen on democratic practice for the evaluation process in light of transparent standards in which information is exchanged and students participate in negotiation, communication, dialogue, respect, equality, shared responsibility; the democratic evaluation means the building of a shared society in which nothing is hidden from the evaluation process about the students (36).

At the beginning of the twentieth century, some criticisms emerged about university evaluation methods that were directed towards the teaching staff, which made universities strive to achieve distinction in light of many challenges such as teaching, learning and evaluation (37).

In academic communities, there is also a disconnection between students and faculty staff in anticipation and perception of the academic evaluation process, and this is evident in the different

perceptions in the systems and standards of grading, whether through the evaluator or the student who is being evaluated (46).

Although grading is important, grades are rarely included in curriculum reform or assessment discussions (28).

Problem of the study

Morrow Jr, et al (2016) illustrates that there are no specific rules for grading systems that can be clearly established in every educational situation, but grading systems differ from one country to another and from one educational institution to another due to several factors, including the educational system, course objectives, teaching methods, available tools, assessor and learners' diverse abilities (42: 609). Some view it as academic freedom and not an appropriate subject for external intervention (21).

Therefore, defining and determining a grading system is one of the biggest problems faced by the field of physical education, and it must be done according to an accurate assessment of the level of learning and student achievement (8: 381).

Performance-based assessment in the field of physical education is the basis for the current educational reform as it enhances learning and achievement; developing or using any type of performance-based assessment must be also done according to clear criteria for performance in order to ensure (consistency, validity, and fairness) (42: 696).

Lorente-Catalán, E., & Kirk, D. (2014) show that there are some difficulties that faculty staff face in higher education during the evaluation process and grading standards, so there is a need for more research and scientific study on democratic practices for the evaluation process as it is mentioned in the project report submitted to the Seventh National Conference on Formative Assessment in Higher Education held at the University of Vic (Spain) (36).

The study of Brookhart, S.M. et al(2016) indicates a review of many studies on grading systems in different educational stages, including the undergraduate level, to confirm that there is a difference in the evaluators' grades, which the future research should take into account to research the process of variation in grades and the extent to which this affects the achievement of Educational goals and future success of education (21).

The problem of evaluation and university grade-setting systems is not limited to a particular department or college, therefore it is necessary to conduct studies in the various departments and branches of the university to provide an important data base on the impact of these grades on the extent of change in students' achievement; and the need to conduct analytical studies that enable us to identify the achievement of the goals and philosophy of grading or its various purposes (1).

In light of what is confirmed by studies that dealt with evaluation, as for study no. (15), which confirms that there is a difference in the educational evaluation process between faculty staff and the lack of scientific and objective standards in evaluating students in the applied units; studies numbers (10) (11) (13) (14) indicate the necessity to adopt performance-based evaluation because of the educational benefits achieved; whereas references and studies No. (1) (25) (32) (42) (44) confirm that the grading standards represent a difficult problem and a controversial issue related to the evaluator's experience and affect students' equal opportunities.

Leirhaug, P.E (2016) (35) also confirms that educational practices and future research in the field of physical education must be taken into account that there is a better match between the formative evaluation and the summative evaluation in the grading setting process, so the researcher conducted an exploratory study on a random sample of third year students in the academic year (2018/2019) by using meta-analysis to identify the extent of disparity in grading setting between the formative and summative evaluations in some of academic and applied courses, which confirmed the existence of a clear difference in the degree setting systems for those courses. In light of what is mentioned previously, the researcher decided to conduct this study in order to identify with the philosophy of performance-based evaluation and grading setting standards in some courses at the faculty of physical education for boys in Alexandria to analyze the reality in the light of contemporary educational trends.

Table (1)Meta-analysis and stats, of year works, practical and written works (formative and summative evaluation) of the theoretical and applied courses under research. N= 22

Stats	Mean	Standard deviation	T-Value	P value	Eta ²	Effect Size Cohen
Measurement and evaluation year works	18.64	1.14	*5.42	0.00	0.57	1.19
Injuries and first aid year works	13.77	4.62				
Written measurement and evaluation	43.46	10.03	*2.73	0.01	0.26	0.54
Written injuries and first aid	48.41	7.56				
Gymnastics year works	22.32	5.55	1.75	0.09	0.12	0.47
Sports games year works	19.36	6.91				
Sports games year works	19.36	6.91	0.53	0.60	0.01	0.16
Track and field events year works	18.55	2.22				
Gymnastics year works	22.32	5.55	*3.22	0.00	0.32	0.85
Track and field events year works	18.55	2.22				
Practical Gymnastics	17.91	5.40	0.00	1.00	0.00	0.00
Practical sports games	17.91	2.11				
Practical sports games	17.91	2.11	*7.48	0.00	0.72	2.11
Practical track and field events	24.09	3.53				
Practical Gymnastics	17.91	5.40	*6.84	0.00	0.68	1.27
Practical track and field events	24.09	3.53				
Written Gymnastics	22.95	2.94	*9.27	0.00	0.80	2.02
Written sports games	13.59	5.38				
Written sports games	13.59	5.38	*8.16	0.00	0.75	1.47
Written track and field events	21.14	4.79				
Written Gymnastics	22.95	2.94	1.76	0.09	0.12	0.45
Written track and field events	21.14	4.79				

* (t) value is significant at the level of 0.05 = (2.08) the effect size Cohen is less than .50 (low) - from .50 to less than .80 (medium) - .80 and more is (high)

Table no. (1) related the meta-analysis and the stats for year works; the practical and the written (the formative and summative evaluation) for the theoretical and applied courses shows that there are statistically significant differences at the level of (0.05) between the works of the year, for the measurement and evaluation courses, injuries and first aid. The effect size of this difference reached 1.19, and this also expresses a fundamental difference in grading setting. There is also a difference between the writing tests of measurement, evaluation, injury and first aid courses. The effect size of this variance reached 0.54 which expresses a fundamental difference in grading setting and the significance of this difference is medium in grading

Also, there is a difference between the year works for gymnastics, field and track competitions courses in favor of the gymnastics course. The effect size of this difference reached 0.85, which refers to a high difference effect size when grading; and it is evident that there is a significant difference in sports games and field and track events courses in favor of the field and track events. The effect size of this difference reached 2.11, which is a high effect size for setting grades in as well as among the practical parts for the courses of gymnastics, and track and field events in favor of track and field events courses. The effect size of this difference reached 1.27, which is a high effect size, which indicates a difference in grading setting. Also, it is evident that there is a difference between the written works of the gymnastics and sports games courses in favor of the gymnastics course and the effect size of this difference reached 2.02, which is the a high effect size due to the difference in grading setting; as well as between the courses of sports games and track and field events in favor of the field and track events. The effect size of this difference reached 1.47, which represents a high difference effect size in setting grades and this indicates a clear and apparent difference in setting year works grades and the written works between two of the academic courses. Also, among three of the applied courses, the difference is also evident in grading setting when comparing the practical grades of the three applied courses for a random sample of students, which highlights the fundamental problem of grading standards during the formative and summative evaluation, which leads to the necessity of conducting this study.

Objectives of the study

- 1- Identifying the performance-based evaluation philosophy (product evaluation - process evaluation) and the grading standards according to the summative evaluation of field and track events.
- 2- Identifying the philosophy of performance-based evaluation and grading standards according to the formative and summative evaluations of the applied courses from the viewpoints of the students (second year - third year - fourth year major in school sports and sports training).
- 3- Identifying the elements of performance-based evaluation and the appropriate grades as for formative evaluation (year works) for the applied courses from the faculty staff's points of view.
- 4- Identifying the contemporary educational trends of performance-based evaluation philosophy and grading standards in the field of physical education.

Questions of the study

- 1- What is the performance-based evaluation (product evaluation - process evaluation) and the grading standards according to the summative evaluation of field and track events?
- 2- What is the philosophy of performance-based evaluation and grading standards according to the formative and summative evaluations of applied courses from the viewpoints of the students (second year – third year – fourth year major in school sports and sports training)?
- 3- Are there statistically significant differences between students of the second, third and fourth years, specializing in school sports and sports training?
- 4- What are the elements of performance-based evaluation and the appropriate grades according to the formative evaluation (year works) for the applied courses from the viewpoints of the faculty staff?
- 5- Are there statistically significant differences in the elements of the performance-based evaluation and the appropriate grading standards according to the formative evaluation (year woks) for the applied courses from the viewpoints of the faculty staff?
- 6- What are the contemporary educational trends of the performance-based evaluation philosophy and grading standards in the field of physical education?

Importance of the study

- 1- Activating the use of performance-based evaluation as an important educational attitude to achieve the quality of the educational process.
- 2- Directing the educational practices of the faculty staff in the faculties of physical education towards determining objective standards for grading systems in order to ensure fair evaluation.
- 3- Building a tool for the performance-based evaluation philosophy and grading standards according to the latest contemporary educational trends in the field of physical education.
- 4- Opening new horizons towards determining objective criteria for grading standards in university courses that guarantee implementing equal opportunities for students.

Terms of the study

Performance-based evaluation

Performance-based evaluation indicates the skills that the learner possesses and the extent of his ability to employ them in real learning situations; and this is evident in the light of indicators of rating grades and the level of performance quality through process evaluation and product evaluation; and this is reflected in the extent to which the educational objectives are achieved (7: 117, 118).

Process Evaluation

Diagnosing points of strength and weakness in performance by determining the specifications for the best technical performance of the skill using rating scales (25: 230).

Product Evaluation

It is concerned with asking objective quantitative questions such as number of times, correct attempts, distance or lumped time. It does not deal with technical considerations of (technique) performance, but rather measures performance results in the form of record achievements (2: 497).

Grading Standards

Distinguishing the students' levels and determining the appropriate grades for their abilities and potentials are basically done through the use of specifications or performance criteria (procedural definition).

Methodology

Curriculum method : The researcher used the two following methods. They are curriculums of meta-analysis and the descriptive curriculum to adapt to the study nature.

Study community: it includes the faculty staff and the students of the grades two, three and four (specialized school sport and sport training) years of (2018 -2019) (2019 - 2020).

The First study Sample: It was chosen by the random method from the third year. They were (22) students (2018- 2019).

The second study Sample: It was chosen by the random method from the two, three and four years specialized school sport and sport training the total number is (30) students.

The main study Sample: : It was chosen by the following :-

-(43) of faculty staff of physical education for men in Alexandria .They were chosen as follows:

-(10) faculty staff of the department of track and field to determine the philosophy of the performance based evaluation (Product Evaluation - Process Evaluation) and the grading standards according to the summative evaluation to the track and field events.

-(33) faculty staff that teach the practical courses to determine the elements of performance based evaluation and the suitable grades according to formative evaluation (year works) .

- (95) from the second year students, and (88) from the third year students, and (38) from the fourth year of school sports students, and (41) from the fourth year of training sports students, the total of the main study (262) students.

Study tools

The researcher chooses these tools:

1- Design a form to limit the philosophy performance based evaluation (Product Evaluation - Process Evaluation) and the grading standards according to the summative evaluation to the track and field events. It was prepared according to the course skills for the grades of the study (first, second, third) in addition to the references numbers (2) (25). The form was shown to a number of experts in field of track and field training events and curriculum, teaching methods, they are (8) appendix (1). This is for recognizing:-

- How the elements of the form suit study subject.

The experts agreed to design the form and its adaptation to the study goal and the researcher chooses agreement 80% to become the final form appendix (2).

2- A questionnaire was prepared for the philosophy of performance-based evaluation and grading standards in light of contemporary educational trends for students of the second, third and fourth years, specializing in school sports and training.

- The researcher designed a questionnaire for the philosophy of performance-based evaluation and grading standards through theoretical readings and analyzing the content of references and studies (2) (5) (7) (9) (10) (23) (24) (25) (34) (35) (38).

- The performance-based evaluation philosophy questionnaire scale and the grading standards are presented in its initial form to (7) experts and jurors in the field of curricula, teaching methods and sports psychology appendix (1), with the aim of identifying:

- Appropriateness of the vocabulary for the study topic.

The clarity and correlation of each (single) statement to the aim of the questionnaire, and the researcher has been satisfied with an agreement rate of 80%, as some phrases numbers (9) (10) (21) have been modified. Phrase no. (16) has been deleted so that the scale becomes in its final form 40 phrases; appendix (3). Positive phrases come in numbers (1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 14, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40). The responses are as follows; yes = 3 to some extent = 2 No = 1 while negative phrases come in numbers (13, 15, 16, 27) so that the response is, yes = 1 to some extent = 2, no = 3, as in appendix (3).

3- Designing a form for the elements of the performance-based evaluation and determining the appropriate grades according to the formative evaluation (works of the year) for the applied courses, by referring to references and studies (5) (15) (39) (41) (45). The form has been presented to a number (6) of experts and jurors in the field of curricula and teaching methods, appendix (1) for the purpose of identifying:

- Suitability of the questionnaire elements to the topic of study.

The experts agreed on the design of the form and its suitability for the purpose of the study, and the researcher has been satisfied with an agreement rate of 80%, to become in its final form appendix (4).

- The study tools were applied in the period from the end of the first semester to the beginning of the second semester of the academic year 2019/2020.

Scientific coefficients of the scales used under research

First: Validity

A) Content Validity Ratio (CVR)

The three study tools were presented to the jurors in order to find the validity of the jurors, and the researcher has been satisfied with an 80% agreement of the jurors for the study tools, appendixs (2) (3) (4).

(B) Construct validity

The researcher presented a questionnaire (the philosophy of performance-based evaluation and grading standards) to an exploratory sample of (30) students to find the validity factor of the questionnaire as shown in the following table No. (2):

Table (2)

Construct validity between the upper quartile and lower quartile in total of the performance-based evaluation philosophy questionnaire and the grading standards.

Scale	Stats	Lower quartile N=7		Upper quartile N=7		Mean Differences	(T) Value	Validity
		Mean	± Std	Mean	± Std			
A questionnaire on the performance-based evaluation philosophy and grading standards		86.71	3.30	63.86	3.34	22.86	*12.88	0.97

* The value of (t) is significant at the level of (0.05) = (2.18)

Table (2) shows the differences between the higher quartile and the lower quartile in total of the questionnaire under research to find the validity coefficient of the performance-based evaluation philosophy questionnaire and the grading standards. There are statistically significant differences, as the calculated value of (t) reached (12.88) and this value is greater than the value of tabular (V) at the level of (0.05) = (2.18), while the value of the validity factor was (0.97), which confirms the validity of the questionnaire under research.

Second: Reliability

The researcher presented a questionnaire (the philosophy of performance-based evaluation and grading standards) for an exploratory sample of (30) students to find the reliability coefficient of the questionnaire as shown in the following table:

Table (3) shows the reliability coefficient of Cronbach's Alpha and the Spearman- Brown coefficient for the performance-based evaluation philosophy questionnaire and the grading standards. N = 30

Phrases	Cronbach's Alpha	Spearman- Brown coefficient
A questionnaire on the performance-based evaluation philosophy and grading standards	*0.797	*0.778

It is clear that table (3) of the Cronbach's Alpha and Spearman- Brown coefficient of the questionnaire for the philosophy of performance-based evaluation and grading standards is that the values of the Cronbach's alpha factor reached 0.797, and it is clear that the values of the Spearman-Brown coefficient reached 0.778, and these values are greater than the value of (0.700), which indicates the reliability of the questionnaire of performance-based evaluation philosophy and grading standards.

Statistical treatments:

Statistical treatments were conducted using the SPSS program, at a significance level (error probability) 0.05, matched by a confidence level (0.95), that are as follows: -

- Meta analysis
- Eta Square
- Effect size cohen
- T independent samples T test

- Cronbach's Alpha
- Spearman-Brown Coefficient
- percentage
- Weighted Mean
- Chi Square
- Mean
- Stander Deviation
- one - way Anova
- L.S.D (least significant dif)

Presentation and discussion of the results

First: Presentation and discussion of the results of the philosophy of the performance-based Evaluation (product Evaluation – process Evaluation) and the grading Standards according to the summative assessment of the track and field events.

Table (4) the philosophy of the performance-based assessment (product evaluation – process evaluation) and grading standards according to the summative Evaluation of the track and field events for the students of the first, second and third years at the Faculty of Physical Education (n = 10)

Student year	Events		Performance based Evaluation			Grading Standards	
			record level	holistic skill performance	phases skill performance	Criterion Reference	Norm Reference
First	100m sprinting	count	10	0	0		√
		%Event	%100.0	%0.0	%0.0		
		%Performance	%33.3	%0.0	%0.0		
	Race Walking	count	0	10	0		√
		%Event	%0.0	%100.0	%0.0		
		%Performance	%0.0	%100.0	%0.0		
	Long Jumb	count	10	0	0		√
		%Event	%100.0	%0.0	%0.0		
		%Performance	%33.3	%0.0	%0.0		
	Shot Put	count	10	0	0		√
		%Event	%100.0	%0.0	%0.0		
		%Performance	%33.3	%0.0	%0.0		
Second	110m Hurdles	count	10	0	0		√
		%Event	%100.0	%0.0	%0.0		
		%Performance	%25.0	%0.0	%0.0		
	800m Running	count	10	0	0		√
		%Event	%100.0	%0.0	%0.0		
		%Performance	%25.0	%0.0	%0.0		
	Hammer Throw	count	10	0	0		√
		%Event	%100.0	%0.0	%0.0		
		%Performance	%25.0	%0.0	%0.0		
	Triple Jump	count	10	0	0		√
		%Event	%100.0	%0.0	%0.0		
		%Performance	%25.0	%0.0	%0.0		
Third	m 400×4 Relags	count	0	10	0		√
		%Event	%0.0	%100.0	%0.0		
		%Performance	%0.0	%33.3	%0.0		
	400m Hurdles	count	10	0	0		√
		%Event	%100.0	%0.0	%0.0		
		%Performance	%33.3	%0.0	%0.0		
	Pole Vault	count	10	10	0		√
		%Event	%50.0	%50.0	%0.0		
		%Performance	%33.3	%33.3	%0.0		
	Shot Put (Rotation Technique)	count	10	10	0		√
		%Event	%50.0	%50.0	%0.0		
		%Performance	%33.3	%33.3	%0.0		

Table (5) the philosophy of the performance-based Evaluation (product evaluation – process evaluation) and grading standards according to the summative Evaluation of the track and field events for the students of the first, second and third years at the Faculty of Physical Education (n = 10)

Student year		Performance based Evaluation			Total	Grading Standards	
		record level	holistic skill performance	phases skill performance		Criterion Reference	Norm Reference
First	count	30	10	0	40		√
	% year	%75.0	%25.0	%0.0	%100.0		
	% Performance	%30.0	%25.0	%0.0	%28.6		
Second	count	40	0	0	40		√
	% year	%100.0	%0.0	%0.0	%100.0		
	% Performance	%40.0	%0.0	%0.0	%28.6		
Third	count	30	30	0	60		√
	% year	%50.0	%50.0	%0.0	%100.0		
	% Performance	%30.0	%75.0	%0.0	%42.9		
Total	count	100	40	0	140		√
	% year	%71.4	%28.6	%0.0	%100.0		
	% Performance	%100.0	%100.0	%0.0	%100.0		

Table (4) (5) indicate The final evaluation process for the track and field events mainly focuses on product assessment (record performance) in all events of the first and second year students, except the race walking event, whereas the evaluation for the third year event was based on both the process evaluation and the product evaluation for each two events. Also, the grades which are placed in all teams and events are based on the norm referenced level (overall average of students' performance).

The assessment of students' performance is an important issue and it can be achieved in several ways, the assessment rules are used in the field of physical education and sport to assess both the product and the process whereas the product evaluation refers to the number of times or correct attempts whereas the process evaluation refers to diagnosing the points of strength and weakness in performance through the specification of the correct technical performance of the skill using rating scales (25: 229, 230). The grade placement process can also be a holistic process that looks at the student's performance in a holistic manner or an analytical process that divides performance into different elements, each of which is evaluated separately (3: 214).

Carroll, B. (2005) indicates that the main focus is on the assessment process in the field of physical education through product assessment as the performance assessment in sports activities such as track and field events focuses on measuring time or the final result of performance (record level), it is an objective and reliable measure that is used by the sports institutions that the time and distance criteria are based on the students' grades, which differ from the open skills as in team games (23:37).

The researcher observes that the focus on product evaluation (the level of record performance) is due to the ease of this type of assessment. It helps to save time by a large percentage compared to the process of criterion referenced evaluation which depends on the details of technical performance using assessment measures and this requires greater effort, especially because of the large numbers of students. Therefore, the performance-based evaluation philosophy for all the first and second year events focuses on product evaluation, with the exception of race walking event, which was evaluated using overall skill performance (and not by using rating scales). The assessment process was carried out according to a limited distance of performance, which is difficult to legally and completely conduct as a event that has specific legal specifications. It is evident that there is a difference in the evaluation of the third year events, especially the pole vault and shot put rotational technique events using the overall skill performance evaluation. These two events are characterized by difficulty in movement performance (technique) as complex skills that are taught during a short period of the first semester of the academic year. The duration of learning doesn't suit the difficulty of movement performance, which results in the presence of difficulties for some students due to their varied abilities. Regardless of the ease or difficulty of performance, the assessment process and grade setting must take into account the evaluation of technical performance according to previously announced criteria for students (using rating scales) due to the contrast of individual differences (physical-skillful - practiced activity) among students. Also, the general objectives of studying these courses are to teach students how to perform the technical performance of these events and contribute mainly to

preparing them for specializations such as teaching and training in track and field events to be able to perform those kinetic models for these events.

Morrow Jr, et al (2016) confirms that the process evaluation refers to the fundamental elements to reach the best form of kinetic performance, as it examines the quality of performance, while the outcome evaluation refers to the learning outcomes that represent the outcome of students' performance. He adds that there is a basic rule means that raters shouldn't determine the product criteria until they determine the optimal form of performance, noting that the correct performance is reflected in the evaluation of the output (record level), also, the recording of performance-based evaluation scores is done through a check list, which includes the features or characteristics that represent the quality of performance, so the feature is present or not, all the features and characteristics also have the same value, which the evaluators use in process evaluation, such as arm movement ... etc. The checklist is also simple to prepare and is useful as a product for documenting the performance observation accurately (42: 670, 671).

The researcher also indicates that the track and field events differ from each other in the characteristics that represent the technical dimension (technique). They also differ according to the importance of the technical stages from one event to another, which is reflected in the quality and efficiency of the movement performance. For example, the landing phase plays a large role in the level and efficiency of technical performance in the events of the long jump and triple jump, while the flying stage plays the largest role in the high jump event, which mainly affects the criteria for obtaining a higher score for that stage compared to other technical stages.

Miller, D. (2010) confirms that the classification and grading systems in the field of physical education are done through the criterion referenced evaluation or the norm referenced evaluation (40: 91).

The evaluation process provides information for evaluators to make judgments about students' performance based on certain criteria that are divided into two categories: the norm referenced evaluation that requires the use of evaluation information in order to evaluate the student compared to the peers' performance, it also focuses on the results of performance instead of results processes, as for the criterion referenced evaluation; it includes a comparison of the student's performance according to a set of predetermined criteria, also, the evaluation of physical education is typically done according to the criterion referenced evaluation, according to this criterion, grading is connected to specific and clear specifications that include the correct criteria for performance, consequently, this type of evaluation provides clear information about students' achievement to a certain level (25: 240, 241).

Second: Presentation and discussion of the results of the questionnaire of the philosophy of the performance-based assessment and the grading criteria according to the contemporary educational trends for the students of the second, third and fourth years, school sport and sport training specializations.

Table (6) Frequencies, percentages, Chi square, and approval percentage of the phrases of the questionnaire of the philosophy of the performance-based evaluation and grading standards for the students of the second year (N = 95)

Items	yes		Sometimes		No		Chi square	Mean	%
	count	%	count	%	count	%			
1	7	7.37%	44	46.32%	44	46.32%	*28.82	1.61	30.53
2	5	5.26%	23	24.21%	67	70.53%	*64.25	1.35	17.37
3	24	25.53%	42	44.68%	28	29.79%	5.70	1.96	47.87
4	50	53.19%	38	40.43%	6	6.38%	*33.02	2.47	73.40
5	20	21.05%	34	35.79%	41	43.16%	*7.22	1.78	38.95
6	27	28.42%	33	34.74%	35	36.84%	1.10	1.92	45.79
7	1	1.06%	24	25.53%	69	73.40%	*76.36	1.28	13.83
8	2	2.11%	3	3.16%	90	94.74%	*161.20	1.07	3.68
9	2	2.11%	27	28.42%	66	69.47%	*65.71	1.33	16.32
10	5	5.26%	10	10.53%	80	84.21%	*111.05	1.21	10.53
11	4	4.21%	66	69.47%	25	26.32%	*62.80	1.78	38.95
12	14	14.89%	27	28.72%	53	56.38%	*25.17	1.59	29.26
13	74	77.89%	12	12.63%	9	9.47%	*85.03	2.68	84.21
14	78	82.11%	13	13.68%	4	4.21%	*102.97	2.78	88.95

Items	yes		Sometimes		No		Chi square	Mean	%
	count	%	count	%	count	%			
15	70	73.68%	16	16.84%	9	9.47%	*70.38	2.64	82.11
16	41	43.16%	23	24.21%	31	32.63%	5.14	2.11	55.26
17	21	22.11%	37	38.95%	37	38.95%	5.39	1.83	41.58
18	34	35.79%	31	32.63%	30	31.58%	0.27	2.04	52.11
19	6	6.38%	37	39.36%	51	54.26%	*33.85	1.52	26.06
20	18	19.15%	46	48.94%	30	31.91%	*12.60	1.87	43.62
21	41	43.16%	44	46.32%	10	10.53%	*22.38	2.33	66.32
22	6	6.32%	23	24.21%	66	69.47%	*60.40	1.37	18.42
23	4	4.26%	12	12.77%	78	82.98%	*105.28	1.21	10.64
24	24	25.26%	19	20.00%	52	54.74%	*19.98	1.71	35.26
25	18	18.95%	24	25.26%	53	55.79%	*22.13	1.63	31.58
26	3	3.16%	8	8.42%	84	88.42%	*130.13	1.15	7.37
27	77	81.05%	15	15.79%	3	3.16%	*99.62	2.78	88.95
28	5	5.26%	5	5.26%	85	89.47%	*134.74	1.16	7.89
29	7	7.37%	34	35.79%	54	56.84%	*35.14	1.51	25.26
30	2	2.11%	15	15.79%	78	82.11%	*104.36	1.20	10.00
31	23	24.21%	56	58.95%	16	16.84%	*28.82	2.07	53.68
32	85	89.47%	10	10.53%	0	0.00%	*59.21	2.89	94.74
33	9	9.47%	25	26.32%	61	64.21%	*44.80	1.45	22.63
34	10	10.64%	58	61.70%	26	27.66%	*38.13	1.83	41.49
35	2	2.13%	21	22.34%	71	75.53%	*81.09	1.27	13.30
36	6	6.32%	47	49.47%	42	44.21%	*31.60	1.62	31.05
37	11	11.58%	35	36.84%	49	51.58%	*23.33	1.60	30.00
38	9	9.47%	39	41.05%	47	49.47%	*25.35	1.60	30.00
39	2	2.11%	10	10.53%	83	87.37%	*125.83	1.15	7.37
40	3	3.16%	5	5.26%	87	91.58%	*145.10	1.12	5.79

* Chi square significant at the level of 0.05 where the significance level ≤ 0.05 at freedom degree 2 = 5.99, freedom degree 1 = 3.84

Likert triple scale: Weighted arithmetic mean: 1.00-1.66 (no), 1.67-2.33 (somehow), 2.34-3.00 (yes)
 Table (6) which illustrates the frequencies, percentages, Chi square, and approval percentage of the questionnaire phrases for the second year students shows: significant differences in all of the phrases of the questionnaire of the philosophy of the performance-based evaluation and grading standards, where the calculated Chi square value was greater than the tabular Chi square value at the level of 0.05 = (5.99) except the phrases (3, 6, 16, 17, 18), the values of the approval percentages of the questionnaire phrases ranged between (3.68% to 94.74%), and Phrase No. (32) had the highest approval rates, while Phrase No. (8) had the lowest approval percentage.

Table (7) Frequencies, percentages, Chi square, and approval percentage of the phrases of the questionnaire of the philosophy of the performance-based evaluation and grading standards for the students of the third academic year (N = 88)

Items	yes		Sometimes		No		Chi square	Mean	%
	count	%	count	%	count	%			
1	4	4.55%	36	40.91%	48	54.55%	*35.27	1.50	25.00
2	1	1.14%	14	15.91%	73	82.95%	*36.27	1.18	9.09
3	14	15.91%	37	42.05%	37	42.05%	*37.27	1.74	36.93
4	26	29.55%	44	50.00%	18	20.45%	*38.27	2.09	54.55
5	1	1.16%	34	39.53%	51	59.30%	*39.27	1.42	20.93
6	7	8.05%	35	40.23%	45	51.72%	*40.27	1.56	28.16
7	3	3.41%	7	7.95%	78	88.64%	*41.27	1.15	7.39
8	0	0.00%	3	3.45%	84	96.55%	*42.27	1.03	1.72

Items	yes		Sometimes		No		Chi square	Mean	%
	count	%	count	%	count	%			
9	5	5.68%	35	39.77%	48	54.55%	*43.27	1.51	25.57
10	2	2.27%	3	3.41%	83	94.32%	*44.27	1.08	3.98
11	1	1.14%	73	82.95%	14	15.91%	*45.27	1.85	42.61
12	3	3.45%	19	21.84%	65	74.71%	*46.27	1.29	14.37
13	80	90.91%	3	3.41%	5	5.68%	*47.27	2.85	92.61
14	83	94.32%	3	3.41%	2	2.27%	*48.27	2.92	96.02
15	75	85.23%	11	12.50%	2	2.27%	*49.27	2.83	91.48
16	31	35.23%	38	43.18%	19	21.59%	*50.27	2.14	56.82
17	10	11.49%	42	48.28%	35	40.23%	*51.27	1.71	35.63
18	11	12.64%	37	42.53%	39	44.83%	*52.27	1.68	33.91
19	10	11.36%	43	48.86%	35	39.77%	*53.27	1.72	35.80
20	17	19.32%	42	47.73%	29	32.95%	*54.27	1.86	43.18
21	24	27.27%	55	62.50%	9	10.23%	*55.27	2.17	58.52
22	6	6.98%	10	11.63%	70	81.40%	*56.27	1.26	12.79
23	0	0.00%	14	15.91%	74	84.09%	*57.27	1.16	7.95
24	75	85.23%	4	4.55%	9	10.23%	*58.27	2.75	87.50
25	6	6.90%	37	42.53%	44	50.57%	*59.27	1.56	28.16
26	1	1.14%	9	10.23%	78	88.64%	*60.27	1.13	6.25
27	64	72.73%	15	17.05%	9	10.23%	*61.27	2.63	81.25
28	46	52.27%	19	21.59%	23	26.14%	*62.27	2.26	63.07
29	4	4.60%	39	44.83%	44	50.57%	*63.27	1.54	27.01
30	4	4.60%	10	11.49%	73	83.91%	*64.27	1.21	10.34
31	23	26.14%	53	60.23%	12	13.64%	*65.27	2.13	56.25
32	68	77.27%	13	14.77%	7	7.95%	*66.27	2.69	84.66
33	4	4.55%	39	44.32%	45	51.14%	*67.27	1.53	26.70
34	6	6.82%	56	63.64%	26	29.55%	*68.27	1.77	38.64
35	5	5.68%	19	21.59%	64	72.73%	*69.27	1.33	16.48
36	5	5.75%	49	56.32%	33	37.93%	*70.27	1.68	33.91
37	6	6.82%	15	17.05%	67	76.14%	*71.27	1.31	15.34
38	7	7.95%	52	59.09%	29	32.95%	*72.27	1.75	37.50
39	0	0.00%	8	9.09%	80	90.91%	*73.27	1.09	4.55
40	0	0.00%	3	3.41%	85	96.59%	*74.27	1.03	1.70

* Chi square significant at the level of 0.05 where the significance level ≤ 0.05 at freedom degree 2 = 5.99, freedom degree 1 = 3.84

Likert triple scale: Weighted arithmetic mean: 1.00-1.66 (no), 1.67-2.33 (somehow), 2.34-3.00 (yes)
Table (7) which illustrates the frequencies, percentages, Chi square, and approval percentage of the questionnaire phrases for the for the third year students shows: significant differences in all of the questionnaire phrases where the calculated Chi square value was greater than the tabular Chi square value at the level of 0.05 = (5.99), the values of the approval percentages for the questionnaire phrases ranged between (1.70% to 96.02%), Phrase No. (14) had the highest approval percentage, while Phrase No. (40) had the lowest approval percentage.

Table (8) Frequencies, percentages, Chi square, and approval percentage of the phrases of the questionnaire of the philosophy of the performance-based evaluation and grading standards for the students of the fourth academic year specialized in school sport (N = 38)

Items	yes		Sometimes		No		Chi square	Mean	%
	count	%	count	%	count	%			
1	12	31.58%	18	47.37%	8	21.05%	4.00	2.11	55.26
2	3	7.89%	11	28.95%	24	63.16%	*17.74	1.45	22.37
3	8	22.22%	24	66.67%	4	11.11%	*18.67	2.11	55.56
4	6	16.22%	25	67.57%	6	16.22%	*19.51	2.00	50.00
5	3	8.33%	16	44.44%	17	47.22%	*10.17	1.61	30.56
6	14	38.89%	17	47.22%	5	13.89%	*6.50	2.25	62.50
7	1	2.70%	11	29.73%	25	67.57%	*23.57	1.35	17.57
8	1	2.70%	3	8.11%	33	89.19%	*52.11	1.14	6.76
9	1	2.63%	13	34.21%	24	63.16%	*20.90	1.39	19.74
10	3	7.89%	8	21.05%	27	71.05%	*25.32	1.37	18.42
11	7	18.42%	22	57.89%	9	23.68%	*10.47	1.95	47.37
12	8	21.05%	21	55.26%	9	23.68%	*8.26	1.97	48.68
13	30	81.08%	6	16.22%	1	2.70%	*38.97	2.78	89.19
14	30	81.08%	4	10.81%	3	8.11%	*38.00	2.73	86.49
15	24	63.16%	12	31.58%	2	5.26%	*19.16	2.58	78.95
16	12	31.58%	17	44.74%	9	23.68%	2.58	2.08	53.95
17	21	55.26%	10	26.32%	7	18.42%	*8.58	2.37	68.42
18	13	34.21%	13	34.21%	12	31.58%	0.05	2.03	51.32
19	7	18.92%	17	45.95%	13	35.14%	4.11	1.84	41.89
20	13	35.14%	19	51.35%	5	13.51%	*8.00	2.22	60.81
21	16	42.11%	16	42.11%	6	15.79%	5.26	2.26	63.16
22	4	10.53%	20	52.63%	14	36.84%	*10.32	1.74	36.84
23	6	15.79%	14	36.84%	18	47.37%	5.90	1.68	34.21
24	18	47.37%	10	26.32%	10	26.32%	3.37	2.21	60.53
25	10	27.78%	18	50.00%	8	22.22%	4.67	2.06	52.78
26	1	2.63%	10	26.32%	27	71.05%	*27.53	1.32	15.79
27	31	81.58%	6	15.79%	1	2.63%	*40.79	2.79	89.47
28	6	16.67%	8	22.22%	22	61.11%	*12.67	1.56	27.78
29	22	57.89%	15	39.47%	1	2.63%	*18.05	2.55	77.63
30	9	23.68%	8	21.05%	21	55.26%	*8.26	1.68	34.21
31	18	47.37%	13	34.21%	7	18.42%	4.79	2.29	64.47
32	26	68.42%	11	28.95%	1	2.63%	*25.00	2.66	82.89
33	8	22.22%	21	58.33%	7	19.44%	*10.17	2.03	51.39
34	25	65.79%	13	34.21%	0	0.00%	3.79	2.66	82.89
35	15	40.54%	11	29.73%	11	29.73%	0.87	2.11	55.41
36	3	7.89%	22	57.89%	13	34.21%	*14.26	1.74	36.84
37	14	36.84%	21	55.26%	3	7.89%	*13.00	2.29	64.47
38	20	52.63%	13	34.21%	5	13.16%	*8.90	2.39	69.74
39	5	13.51%	17	45.95%	15	40.54%	*6.70	1.73	36.49
40	2	5.26%	10	26.32%	26	68.42%	*23.58	1.37	18.42

* Chi square significant at the level of 0.05 where the significance level ≤ 0.05 at freedom degree 2 = 5.99, freedom degree 1 = 3.84

Likert triple scale: Weighted arithmetic mean: 1.00-1.66 (no), 1.67-2.33 (somehow), 2.34-3.00 (yes)
 Table (8) which illustrates the frequencies, percentages, Chi square, and approval percentage of the questionnaire phrases for the fourth year students specializing in school sport shows: significant differences in most of the questionnaire phrases, where the calculated Chi square value was greater than the tabular Chi square value at the level of 0.05 = (5.99), while there are no statistically significant differences in phrases (1, 16, 18, 19, 21, 23, 24, 25, 31, 34, 35), the values of the approval

percentages of the questionnaire phrases ranged between (6.76% to 89.47), Phrase No. (27) had the highest approval percentage, while Phrase No. (8) had the lowest approval percentage.

Table (9) Frequencies, percentages, Chi square, and approval percentage of the phrases of the questionnaire of the philosophy of the performance-based evaluation and grading standards for the students of the fourth academic year specialized in sport training (N = 41)

Items	yes		Sometimes		No		Chi square	Mean	%
	count	%	count	%	count	%			
1	26	%63.41	7	%17.07	8	%19.51	*16.73	2.44	71.95
2	2	%5.00	13	%32.50	25	%62.50	*19.85	1.43	21.25
3	13	%31.71	24	%58.54	4	%9.76	*14.68	2.22	60.98
4	14	%34.15	18	%43.90	9	%21.95	2.98	2.12	56.10
5	9	%22.50	17	%42.50	14	%35.00	2.45	1.88	43.75
6	9	%21.95	22	%53.66	10	%24.39	*7.66	1.98	48.78
7	11	%27.50	14	%35.00	15	%37.50	0.65	1.90	45.00
8	3	%7.32	15	%36.59	23	%56.10	*14.83	1.51	25.61
9	16	%40.00	17	%42.50	7	%17.50	4.55	2.23	61.25
10	5	%12.20	10	%24.39	26	%63.41	*17.61	1.49	24.39
11	18	%45.00	15	%37.50	7	%17.50	4.85	2.28	63.75
12	13	%32.50	20	%50.00	7	%17.50	*6.35	2.15	57.50
13	23	%56.10	9	%21.95	9	%21.95	*9.56	2.34	67.07
14	30	%76.92	7	%17.95	2	%5.13	*34.31	2.72	85.90
15	23	%56.10	11	%26.83	7	%17.07	*10.15	2.39	69.51
16	15	%36.59	14	%34.15	12	%29.27	0.34	2.07	53.66
17	14	%35.00	16	%40.00	10	%25.00	1.40	2.10	55.00
18	16	%39.02	12	%29.27	13	%31.71	0.63	2.07	53.66
19	10	%25.00	20	%50.00	10	%25.00	5.00	2.00	50.00
20	14	%34.15	20	%48.78	7	%17.07	*6.20	2.17	58.54
21	16	%40.00	16	%40.00	8	%20.00	3.20	2.20	60.00
22	15	%36.59	13	%31.71	13	%31.71	0.20	2.05	52.44
23	6	%14.63	17	%41.46	18	%43.90	*6.49	1.71	35.37
24	15	%37.50	16	%40.00	9	%22.50	2.15	2.15	57.50
25	16	%39.02	15	%36.59	10	%24.39	1.51	2.15	57.32
26	11	%27.50	13	%32.50	16	%40.00	0.95	1.88	43.75
27	31	%77.50	6	%15.00	3	%7.50	*35.45	2.70	85.00
28	11	%28.21	15	%38.46	13	%33.33	0.62	1.95	47.44
29	16	%40.00	16	%40.00	8	%20.00	3.20	2.20	60.00
30	10	%24.39	14	%34.15	17	%41.46	1.81	1.83	41.46
31	12	%29.27	15	%36.59	14	%34.15	0.34	1.95	47.56
32	29	%72.50	5	%12.50	6	%15.00	*27.65	2.58	78.75
33	16	%39.02	17	%41.46	8	%19.51	3.56	2.20	59.76
34	17	%41.46	15	%36.59	9	%21.95	2.54	2.20	59.76
35	11	%28.21	15	%38.46	13	%33.33	0.62	1.95	47.44
36	14	%35.00	21	%52.50	5	%12.50	*9.65	2.23	61.25
37	9	%22.50	17	%42.50	14	%35.00	2.45	1.88	43.75
38	12	%30.00	18	%45.00	10	%25.00	2.60	2.05	52.50
39	12	%29.27	11	%26.83	18	%43.90	2.10	1.85	42.68
40	6	%14.63	15	%36.59	20	%48.78	*7.37	1.66	32.93

* Chi square significant at the level of 0.05 where the significance level ≤ 0.05 at freedom degree 2 = 5.99, freedom degree 1 = 3.84

Likert triple scale: Weighted arithmetic mean: 1.00-1.66 (no), 1.67-2.33 (somehow), 2.34-3.00 (yes)

Table (9) which illustrates the frequencies, percentages, Chi square, and approval percentage of the questionnaire phrases for the fourth year students specializing in sport training shows:

significant differences in the phrases (1, 2, 3, 6, 8, 10, 12, 13, 14, 15, 20, 23, 27, 32, 36, 40) where the calculated Chi square value was greater than the tabular Chi square value at the 0.05 level = (5.99), while there are no statistically significant differences in the rest of the phrases, the values of the approval percentages for the questionnaire phrases ranged between (21.25% to 85.90%), and Phrase No. (14) had the highest approval rates, while Phrase No. (2) had the lowest approval percentage.

Tables (6), (7), (8) and (9) which illustrate the data of the second, third and fourth years (specialization of school sport and sport training) show that Phrase No. (8) had the lowest approval percentage, followed by the second year 3.68%, the third year 1.72%, specialization of school sport 6.76%, and specialization of sport training 25.61%, which means that the grade is based on participatory assessment criteria (between the student and the faculty member).

Giving the opportunity of participation in the assessment process to the students is a fair method to the assessment process (30).

In this regard, the study of Lorente-Catalán, E., & Kirk, D. (2016) shows that the university aims at changing its practice from the traditional model that depends on delivering information to promote the direct and active participation of the students in the learning and assessment process. The university has ten guides to help the faculty staff improve their skills and their practice of the evaluation process. Two of them (the first and third guides) were specifically designed to assist in the assessment and learning process. The first guide includes an introduction to teaching and assessment and it shows that whoever performs the assessment process must have sufficient scientific information in a way that suits the students, while the third guide (assessment for learning AFL) specifically targets the newcomers in teaching and assessment to acquire the scientific information with its contents. The details of the most effective assessment practices are also highlighted in order to apply and include them in all academic courses within the university, and emphasizing the importance of involving the students in setting standards using participatory assessment strategies such as self-assessment and peer-assessment (37).

There is also a growing interest in higher education towards the use of assessment strategies that focus on the students' participation in the assessment process. Students' participation is crucial and democratic manifestation of bearing responsibility for their own learning and developing lifelong learning skills in a rapid changing world (36).

Therefore, grading systems must be an integration of student and teacher assessments during the evaluation process (25: 234).

The researcher illustrates that embedding a culture of participatory assessment requires a full understanding by the faculty staff through specialized training courses for this type of assessment and forms of practice in order to be able to engage and educate students about the nature of participatory roles in the assessment and grading process. The results of that study are consistent with the results of the two studies (26) and (36). However, there is an obvious lack of students' participation in the assessment process in the physical education field. Therefore, we need more discussions about the effectiveness of assessment with the participation of teachers and students. The study of Redelius, K., & Hay, P (2009) confirms that there are contradictions about the students' perceptions of the grading standards, systems and the assessment process. The study confirms the contradiction between the official assessment standards and the students' observations on the contributing factors of achieving high levels of the achievement process (44).

Phrase No. (7) had low percentages in the following order, the second year 13.83%, the third year 7.39%, and specialization of school sport 17.57%. It confirms that the grade is assessed through a declared standard card specifying the physical or technical performance specifications for each test. The assessment process in the physical education field includes a standard card that focuses on observing performance according to criteria to set the grade (9:144). It is also necessary to provide the students with the assessment rubrics and criteria before the assessment process so that they have a sufficient understanding of their physical, technical, or other assessment criteria (25: 230).

The study of Haerens, L., et al (2018) confirms that the students' knowledge of the standards and specifications of the tests before their performance is an important component of the quality of the assessment process (29).

The results of this study conform to the study of Maen Al-Shaalan and Mahmoud Al-Wedian (2018) which confirms that applying practical tests is easy, but there is a difficulty in giving meaning to the grades because the assessment process depends on the grades estimated by the evaluator and

not using a specific and unified standard for all students. Also, this may cause disparities in giving the grades, and this may affect the reliability and fairness of the assessment (12).

Phrase No. (2) had low percentages in the next order, the second year 17.37%, the third year 9.09%, specialization of school sport 22.37%, and specialization of training 21.25%. This gives the sufficient grades according to the level of your physical or technical assessment (formative – summative assessment).

Morrow Jr, et al (2016) point to the importance of providing the students with the sufficient information of assessment methods before starting teaching through all types and procedures of assessment and what the assessment specifications and criteria are, whether formative or summative (42: 614).

The researcher points out that the Faculty of Physical Education embraces various (individual and team) sports, which require different kinetic and physical performances. Therefore, when previously information is built on performance specifications; we get a clearer perception of the grade obtained by the student. Also Miller, D. (2010) confirms that when the students get informed with the assessment methods, standards, and grading systems in the physical education field, they usually do not get surprised by the grades they have. Therefore, most students like to be informed with their level of achievement and this represents a challenge for some students to get higher grades (40: 87).

Phrase No. (9) indicates that the grade that the student obtains in the applied tests is done according to sufficient time to demonstrate his abilities. The phrase had low approval percentages according to the following order, the second year 16.32%, specialization of school sport 19.74%, while the third year had 25.57%. The researcher attributes that the assessment of large numbers in a limited period of time within the time of the lecture is not enough to give the student sufficient time to perform the assessment process, which actually requires that the assessment of each student takes enough time not only related to the time of the lecture (formative evaluation) in order for the assessment process to be done fairly and this percentage appears in the large numbers of the second and third grades students. However, the specialization of school sport had low percentage between the last two, although the number of students of specialization is less. The researcher explains that all the applied courses for the students of specialization focus on evaluating each student's performance in teaching skills (planning, implementation and evaluation). They require a long time to evaluate the teaching performance of each student, unlike the students of the fourth grade, specialization of sports training with 61.25% . The number of students in this specialization decreases (various training branches in the faculty for individual and team sports), which may provide an opportunity to perform the assessment process in sufficient time. Therefore, it is necessary for the faculty staff to consider giving sufficient time for each student to complete the assessment process fairly and accurately.

Carroll, B. (2005) confirms that the difficulties faced by the faculty staff when assessing the skills during the physical activities do not lie in applying performance standards or criteria, but the difficulty lies in applying them to large numbers of learners in a short time. When the time is not enough for large numbers, it becomes difficult to assess various performances, and in turn it becomes difficult to set grading fairly and reliably (23:38).

Phrase No. (10), had low approval percentages according to the following order, the second year by 10.53%, the third year 3.98%, specialization of school sport 18.42% and specialization of sports training 24.39%. This shows that the grades are placed for colleagues' assessment systems for their level of performance as one of the assessment aspects. During the development of peer-assessment tasks, the teacher gets relieved from being the only one to decide the students' success.

Peer-assessment also increases communication and observation skills, develops learners' understanding of the effective technical performance and also increases the full use of learning time (25: 233).

The study of Lorente-Catalán, E., & Kirk, D. (2016) emphasize that the assessment process practices should include peer-assessment as one of the main assessment strategies within the university (36).

Morrow Jr, et al (2016) points out that when using peer-assessment, it is better for the evaluator to conduct the assessment only because when the performer records the assessment results at the time of performance (passing the ball to the person who performs the assessment), it becomes difficult to pass and accurately observe the performance at the same time, which means, passing and observing equally efficiently becomes more difficult (42: 660).

The researcher indicates that there are many practices for the peer-assessment process in some courses, as in practicing reciprocal learning style, especially for the students of the school sport specialization, or by applying this method in external field training for the third and fourth year students in the period of practical education, but it does not exceed in this case being an education activity and not one of the elements of setting the grade.

Phrase No. (13) had high approval percentages according to the following order: the second year 84.21%, the third year 92.61%, specialization of school sport 89.19%, and specialization of sports training 67.07%, and this indicates a feeling of anxiety about not being informed with the tasks of the tests and the details of the grade. When the students are notified of the assessment criteria in advance, they no longer need to assess the details of the test.

When the students get informed with the grading procedures and criteria that reduces the anxiety when studying courses (42: 624, 670).

Phrase No. (14) had high approval percentages according to the following order, the second year 88.95%, the third year 96.02%, specialization of school sport 86.49%, and specialization of training 85.90%, and this indicates a feeling of more confidence resulting from prior information to the assessment details and appropriate grades.

The study of Krijgsman, C et al (2017) shows that when setting grades, the evaluator must take into account the clarity of the assessment criteria as well as the feelings of the learners by reducing the feeling of pressure and avoiding feelings of failure or frustration by providing a positive environment that stimulates learners and increases motivation towards better performance (34).

These results conform to the study of Barkoukis, V., et al (2014) which showed that developing students' motivation helps acquire information, improve performance and reduce internal pressures that affect the students' achievement and obtain better grades in the physical education field (19). It also conform to the results of the study of Haerens, L., et al (2018) which showed that the information of test standards has an effect on the students' motivations and anxiety while taking tests in the physical education field, as the results confirmed that the students who have prior information of the test standards are more effective in achieving the objectives (29).

Phrase No. (15), had high approval percentages, according to the following order, the second year 82.11%, the third year 91.48%, specialization of school sport 78.95%, and specialization of training 69.51%, which states that the lack of grades clarity in the details of the tests makes the student feel the absence of (objectivity) fairness. Therefore, among the procedures that must be taken into account when setting grades fairly, reliably and properly, students should be informed with the procedures, systems, details of settings the grading and details of the various tests and assessments (42: 621).

Phrase No. (22), had low approval percentages, according to the following order, the second year by 18.42% and the third year 12.79%, which shows that the assessment system depends on placing a grade for each test according to rating scales. The rating scales are of the assessment tools in the physical education field, in which a report of the students' level and abilities is prepared (2: 498).

It is important to consider this through observation using the rating scales when assessing the techniques of team or individual sports (42: 621). The study of Dinan Thompson, M., et al (2015) also confirms that assessment is done through clear criteria of technical performance, such as the rating scales that include items that clarify details of kinetic performance (26).

Phrase No. (23) had low approval percentages according to the following order, the second year 10.64% and the third year 7.95%, which shows that the assessment systems and the execution of applied tests provide opportunities (retesting) to improve the grade.

One of the difficulties for the evaluators is a one-time process of examining the practical performance only, in addition to other difficulties, depending on the purpose of the assessment process, whether it is to identify the level of the learner's accuracy such as the level of technical performance for one time or frequently, through good or bad technical performance, or the ability to perform in more difficult or complex situations and the extent of the effectiveness of the teammates or the opponent players regarding the performance and other factors such as the level of effort and personal characteristics (23: 60).

The researcher shows that there are sports such as track and field events that give the students the performance opportunities in each contest that include three attempts during the assessment process and the grade is calculated according to the best attempt, which is not available in many sports as the objective of the evaluation process and the possibility of fair assessment by giving all

students the opportunities to re-test. This process needs good organization and sufficient space of time, especially in light of the large numbers of students.

Phrase No. (26) had low approval percentages, according to the following order, the second year 7.37%, the third year 6.25%, and specialization of school sport 15.79%, which is “the student can communicate and discuss with the faculty staff about the level of grade that he has got”. Tippin, G. K, et al. (2012) shows that communication between the students and faculty staff about the purpose and process of academic evaluation should be improved, and this is obvious in the various perceptions in the grading systems, whether through the evaluator or the evaluated student (46).

Therefore, it is necessary that the discussion takes place among the evaluators within the same department about the performance criteria and the grading criteria and how to use them in the process of assessing the learners (23:59).

Phrase No. (27), had high approval percentages according to the following order, the second year 88.95%, the third year 81.25%, specialization of school sport 89.47%, and specialization of sports training 85.00%, and this shows the difference in assessment systems and the status of the grade from one evaluator to another.

The study of Annerstedt, C., & Larsson, S. (2010) confirms that there is a clear difference in the grading systems from one evaluator to another within the same school, and the teachers need in-service training about the assessment methods and the criteria of setting the grading as well as thinking about achieving the objectives and standards through developing a grading system that is applicable and compatible with achieving the target objectives and standards (16).

Therefore, when a teacher’s observation is subjective, this affects the grade decision as it lacks objectivity and fairness, and it is also difficult to achieve consistency in grades due to the differences between the teachers and the students, which is a natural phenomenon in any educational environment, and the lack of consistency in grades is one of the weakness points of many grading schemes, so any evaluator must give the same grade for the same level of performance (42: 616, 622).

The researcher shows that the main reason for the difference in the evaluation and the grade setting systems from one evaluator to another is due to the lack of clear performance criteria that define the standards and details of the grades, as many faculty staff depend on subjective observation of performance and this is consistent with the results of the study of Maen Al-Shaalan and Mahmoud Al-Wedian (2018) the assessment process is based on the grades estimated by the evaluator and not on a specific and unified standard for all students, which may cause a variation in setting the grades, which may expose the assessment process to a lack of reliability or fairness of assessment (12).

Phrase No. (28) had low approval percentages according to the following order: The second year 7.89%, while the fourth year , specialization of school sport 27.78%, which means that the grading systems in the applied tests (open skills) depend on the separate technical performance test (not in a real play situation).

Many official assessments focus on evaluating the sports skills in different games separately (isolated as a stand-alone skill) and do not evaluate actual performance in the game situations as the actual performance in the game situations is called a real assessment which determines the student’s success in applying the skills (25: 217).

Also, controlling the assessment of team sports is not an easy process to obtain difficulty, variety, complexity and frequency according to the same criteria for all learners at the same time due to the diversity of abilities and performances, for example the performance of the free shot in basketball outside the real playing situations lacks real playing conditions, also the evaluator needs to have detailed information of the elements and criteria of evaluation and how to apply them, such as the information of correcting the technique of a skill, such as finger positions to control the tool, feet positions, and various tactics such as defense and attack positions, and these elements differ according to the activities and the nature of skills, whether closed or open (23: 55, 56, 58).

Morrow Jr, et al (2016) confirms that assessing techniques and grading systems in team sports is one of the difficult tasks, where the separate (isolated) sports skills tests are often used, as they lack relevance and validity, also the assessment and grading setting according to the team rank regarding providing equal opportunities for the learners to prove ability is a difficult process, as one of the students may be weak and have been selected in a good team and obtain a high grade and vice versa if his performance is good and he is selected in a weak team, he may get a low grade. This strategy is not fair for all students because the grading standards depend on the performance of the others. (42: 616)

Therefore, judging the effectiveness of performance in open skills requires assessment of techniques and other processes such as decision-making, tactical awareness, and application of skills in playing situations that represent reactions in competition situations as an integral part of the performance-based evaluation (23: 37, 38).

Phrase No. (30) had low approval percentages according to the following order: the second year 10.00%, the third year 10.34%, specialization of school sport 34.21%, and specialization of training 41.46%, which shows that the tests provide the self-assessment task as part of the grading system. The researcher illustrates that it is necessary to carry out the self-assessment task. The student must have awareness of his capabilities with an understanding of the details and tasks of performance, which may be the product of learning, practice and experience (and this is obvious somehow in the students of specialization) so that the self-assessment can be used as part of the grading systems. Although a large number of students allegedly participate in the self-assessment for some time, 41% of the faculty members deny that the students participate in the process, which indicates a lack of common meaning between the students and the faculty members about their perceptions and understanding of the practices of the assessment process (38).

Teaching the students how to evaluate their performance helps them to develop independence and the ability to develop their strategies that help define and achieve their goals and in order to get the self-assessment correct and objective, students should have learned through the opportunities of various practices of self-assessment process (25: 233, 234).

Phrase No. (32) had high approval percentages according to the following order, the second year 94.74% the third year 84.66%, specialization of school sport 82.89%, and specialization of sports training 78.75%, which shows that your performance tasks for the applied tests and grading systems are affected by the environmental factors (weather conditions - availability and validity of the tools – the nature of the ground).

Carroll, B. (2005) points that the objective measures such as time and distance have no difficulty in the assessment process, as in track and field events, but changes such as weather conditions and the type of the tracks may have an impact on the times and therefore the standards and rubrics for performance-based evaluation are affected (23:53).

This is obvious among the faculty students, as they sometimes face some assessments in the open playgrounds under variable weather conditions (cold during the first semester) where the open playground may get affected, which may have some effect on their performance level and grades setting.

Phrase No. (39), had low approval percentages according to the following order, the second year 7.37%, the third year 4.55%, specialization of school sport 36.49%, and specialization of sports training, 42.68%, and this show that the assessment and grading systems depend on taking advantage of the technological applications (video recording). Koekoek, J et al. (2018) shows that the professionals in the physical education field have become clearly interested in using technological tools in their educational practices, but many educational environments are not ready to use such tools (33: 1).

The traditional form of assessment in physical education is conducted through the teacher's visual vision as a basis for judging the technical performance and it is considered a basic form of the assessment process (25: 23)

Also, in the field of kinetic performance, students must have the ability to understand the kinetic meaning of performance and may include video clips of the students' performance and feedback, whether spoken or written, providing the students with clear standards of performance, and providing reference points for their own reflections and appropriate learning (30).

The study of Haerens, L., et al. (2018) confirms that in order to achieve the quality of the assessment, teachers must communicate with students more clearly through the assessment criteria using video forms (29).

The researcher believes that using technology effectively requires great capabilities that must be available, which causes difficulty to a large extent in light of the large numbers, limited capabilities and the lack of availability of these technologies significantly in the faculty stadiums, but despite of the fact that there is a benefit somehow through some limited technical capabilities to enrich the learning process, it is also obvious in the small numbers of the students of the sports training specialization.

Capel, S. et al (2013) point that using the digital video and the kinetic performance capturing tools by the learners is an excellent way to record learning and the level of progress, as it is useful in assessing the level of learners (22: 139).

Video recording proves the student's ability to practice a skill well and helps to monitor the learner's progress and evaluate this progress through the summative assessment of the product in addition to the process evaluation that indicates the type and quality of performance (3: 213, 214).

The results of Phrase No. (40) showed low approval percentages according to the following order, the second year 5.79%, the third year 1.70%, specialization of school sport 18.42%, and specialization of sports training, 32.93%, which shows the presentation of a clear record of the assessment systems and transcripts of the students. The evaluation process depends on a record that documents the student's activity and contains several elements such as the results of physical and technical tests, written reports, pictures, illustrations, and visual or audio recordings (42: 663). Also, the assessment portfolios are an effective tool for evaluating the level of students' performance and are considered as a guide or record of what the learner has achieved regarding the objectives and what has been achieved. It also shows the extent of progress made by the learner (25: 238).

Differences between the study years

Table (10)

The statistical description of the questionnaire of the philosophy of the performance-based evaluation and the grading standards for the sample under research. N = 262

Variables	Academic Level	N	Mean	Std.	Min	Max	Rank
A questionnaire on the performance-based evaluation philosophy and grading standards	Second year	95	64.86	7.00	49.00	86.00	3
	Third year	88	63.99	5.79	53.00	80.00	4
	Fourth year (school sports)	38	74.92	8.30	56.00	88.00	2
	Fourth year (training)	41	78.63	11.89	49.00	100.00	1
	Total	262	68.18	9.70	49.00	100.00	

Table (10) shows the statistical description of the questionnaire of the philosophy of the performance-based evaluation and the grading standards, where the fourth year students specializing in sport training came in the first rank in the questionnaire total with an average of 78.63 ± 11.89 , followed by the fourth year students specializing in school sport in the second rank in the questionnaire total with an average of 74.92 ± 8.30 , the second year students in the third rank in the questionnaire total with an average of 64.86 ± 7.00 , and finally the third year students in the fourth rank in the questionnaire total with an average of 63.99 ± 5.79 .

Table (11) shows the significance of the differences between the study years in the questionnaire of the philosophy of the performance-based evaluation and the grading standards under research. N = 262

Stats / Variable	Variance	Sum of Squares	df	Mean Square	F	P Value
A questionnaire on the performance-based evaluation philosophy and grading standards	Between Groups	8798.72	3	2932.91	*48.09	0.00
	Within Groups	15734.49	258	60.99		
	Total	24533.21	261			

* Tabular (P) value is significant at the level of $0.05 = (2.65)$

Table (11) shows that there are statistically significant differences at the level of (0.05) between the study years in the questionnaire of the philosophy of the performance-based evaluation and the grading standards under research, as the calculated P value was (48.09) and this value is greater than the tabular P value at the level of $(0.05) = 2.65$ with significance level less than 0.05.

Table (12) L.S.D test at the level of 0.05 to determine the significance and direction of the differences in the performance-based evaluation philosophy questionnaire and the grading standards

Variables		Mean Difference	P Value	
A questionnaire on the performance-based evaluation philosophy and grading standards	Second year	Third year	0.87	0.45
		Fourth year (school sports)	*10.06-	0.00
		Fourth year (training)	*13.77-	0.00
	Third year	Fourth year (school sports)	*10.93-	0.00
		Fourth year (training)	*14.65-	0.00
	Fourth year (school sports)	Fourth year (training)	*3.71-	0.04

Table (12) which shows the data of the L.S.D test at the level of 0.05 to determine the significance and direction of the differences in the performance-based evaluation philosophy questionnaire and the grading standards under research shows:- There are statistically significant differences between the fourth year (school sport – sport training) and the second year, in favor of the fourth year (school sport – sport training).- There are statistically significant differences between the fourth year (school sport – sport training) and the third year, in favor of the fourth year (school sport – sport training).- There are no statistically significant differences between the second and third years.- There are no statistically significant differences between the fourth year teaching methods and training specializations.

Tables (10), (11) and (12) which illustrate the differences between the study years in the questionnaire of the philosophy of the performance-based evaluation and the grading standards show that the differences are in favor of the fourth year specializing in sport training, followed by the school sport specialization, and the researcher refers those differences to factors such as practice and experience since the students of specialization have more experience as a result of the number of years of studying as well as the conditions for admission to the specialization such as sport training through the specific tests and the level of experience in the sports of specialization, all of these factors contribute mainly to the students' understanding of the assessment process and grading criteria, also teaching in many branches of the training specialization is applied on relatively small numbers which does not exceed ten students in each specialization, which gives a greater opportunity for the learning and assessment process, and is reflected in the degree setting, and this percentage is close to the school sport specialization, as it is distributed into two groups, each group studies the applied courses separately, and the teaching practices of the students of specialization focus mainly on the assessment process, which is reflected in the grading criteria, and therefore there are clear differences between the students of specialization and the third and second years students with their large numbers and short experience. The results of the study conform to the results of the study of (Svennberg, L., et al, 2014) that the students have different views on the grading criteria in the physical education field (45).

Third: Presentation and discussion of the results of the components of the philosophy of the performance-based evaluation and determining the appropriate grades according to the formative evaluation (year activities) in the applied courses.

Table (13) shows the percentage and average grades of the components of the performance-based evaluation used in the courses of the study years under research

Evaluation items		1	2	3	4	5	Cognitive tasks (Research papers)		Cognitive tests				Social aspects	Students' assessment tasks		Other tasks	Other standards you would like to add	
		Regular attendance	Behavior	Effort	Commitment of Uniform	Tests to measure skill performance	individual	team	Objective	Essay	Objective and essay	Oral	Cooperative tasks	Self-evaluation	Peers' evaluation			
First	count	6	1	0	0	9	5	0	0	0	0	1	0	0	0	0	2	0
	%	25.00	4.17	0.00	0.00	37.50	20.83	0.00	0.00	0.00	0.00	4.17	0.00	0.00	0.00	0.00	8.33	0.00
	Grading mean	6.67	1.00	---	---	8.78	10.00	---	---	---	---	2.00	---	---	---	---	4.00	---
Second	count	4	0	1	0	4	4	0	1	0	0	2	0	0	0	0	0	0
	%	25.00	0.00	6.25	0.00	25.00	25.00	0.00	6.25	0.00	0.00	12.50	0.00	0.00	0.00	0.00	0.00	0.00
	Grading mean	5.75	---	2.00	---	8.25	5.00	---	10.00	---	---	1.00	---	---	---	---	---	---
Third	count	6	0	0	0	6	2	0	0	0	0	0	0	0	0	0	0	0
	%	42.86	0.00	0.00	0.00	42.86	14.29	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Grading mean	5.83	---	---	---	12.92	3.75	---	---	---	---	---	---	---	---	---	---	---
Fourth	count	12	5	4	5	9	12	2	0	3	0	0	1	2	1	2	2	1
	%	20.34	8.47	6.78	8.47	15.25	20.34	3.39	0.00	5.08	0.00	0.00	1.69	3.39	1.69	3.39	3.39	1.69
	Grading mean	7.83	3.00	3.00	1.60	9.33	8.58	3.50	---	4.33	---	---	2.00	3.00	2.00	6.50	1.00	

Table (13) indicates the most used elements in the assessment and grading process for the first year students are tests to measure skill performance or movement performance with a score of 37.50% with an average score of 8.78. It is followed by regularity of attendance of 25% with an average of 6.67. It is also followed by individual research papers by 20.83% with an average score of 10 grades, while the rest of the elements are in low percentage and varying grades or not included in the assessment elements and the criteria for grading.

According to the second grade, the most used elements are (regularity of attendance - tests to measure skill performance or movement performance - individual research papers), all at an equal rate of 25% with an average of grading respectively (5.75 - 8.25 - 5) which clearly indicates the variance of grading to the mostly used elements. Even though the percentages of their use are equal, the rest of the elements are in low rates, varying grades or not included in the assessment elements and the grading criteria.

As for the third year, the mostly used elements are attendance regularity - tests to measure the skill performance or movement performance by an equal rate of 42.86% respectively, with an average of grade (5.83 - 12.92), which indicates the existence of difference in setting grades, while the rest of the elements are in low rates and varying grades or not included in elements of assessment or grade criteria

As for the fourth year, it becomes clear that many assessment elements are used more than the other three grades with varying grading rates. The mostly used elements are attendance regularity - individual research papers by 20.34% with an average score (7.83-8.88), respectively, followed by tests to measure skill or kinetic performance by 15.25. % With an average score of 9.33, followed by behavior - adherence to uniform equally 8.47%, with an average of score (3 - 1.60), respectively, followed by effort at an average rate of 6.78% with an average score of 3 grade whereas other elements of varying grades or not included in the assessment elements or grading criteria. Consequently, this indicates a clear difference in the use of assessment elements and grading criteria for all academic years.

It is clear that Table (14) appendix (5) which is related to the elements of the performance-based evaluation and the details of the grading of each course, there is a clear difference within the same course as in the sports course (football - volleyball) for the first year and the contest courses for the first year (judo - boxing - Karate) and the second year (fencing - boxing - wrestling), whether in terms of the various types of the assessment elements and the appropriate grade for each element within the same course, or determining one assessment element, but at a different rate of grades. It becomes clear that the teams that mostly use the assessment elements in the fourth year courses are those who specialize in sports training. This also highlights the diverse experiences of faculty members and their various degrees (lecturer - assistant professor - professor) in addition to teaching to relatively small numbers of students that may not exceed ten students in one specialization such as training exercise training and kickboxing.

The researcher believes that there is a difference in the assessment elements between the academic year. The interest of attendance comes first for all years except for the first year due to the nature of the special organizational procedures for a number of first year students because they take a period of time during the first semester as new students in the college. Mahdi , Ezzedine (2014) confirms that the regularity of attendance and

skillful tests are one of the most important criteria that faculty members focus on when evaluating students' performance according to evaluation elements only, the researcher also shows that the difference in the use of specific types of assessment elements with a different rate of grades affects equal opportunities for students because the student's attendance in a course may affect obtaining a high score in that course, while this regularity may not affect attendance with the same amount or grade in another course and also with the rest of the assessment elements and grading criteria obtained by students.

The grading system in the field of physical education depends significantly on elements such as attendance, uniform and participation. Although attendance is an important element when teaching physical education, it should not be a reason for reducing students' grades, students may attend the class as in math, but they aren't evaluated properly, elements such as behavior and participation are important aims in physical education programs (42: 615, 616).

Miller, D. (2010) shows that attendance and participation are among the main elements to obtain grades in physical education as students who do not attend regularly, have high skills, and have the ability to achieve the objectives of the course at a good level without regular attendance, consequently, it is important to set grades on the basis of achieving the objectives of the course and not attendance, the grades should not also be used as a threat to students but rather to create an atmosphere of challenge and enjoyment during the learning process (40: 88).

It is clear that there are many elements of assessment that have not been used extensively, nor are they represented in grading, such as social aspects, self-assessment, peers, knowledge tests, and effort. Big Bailey (2014) believes that the goal of formative assessment is to guide teaching and not to give grades to students (4: 24).

Lorente-Catalán, E., & Kirk, D. (2014) point out that it is important to consider the use of faculty members in universities to cooperation skills and social skills as evaluative practices, whether for formative or final purposes, the results of this study also differ from what is confirmed by the study of Barkoukis, V. et al (2014) they suggest that the evaluators set more concentration on effort and participation, not just on skill performance (36) (19).

The report submitted to the seventh national conference to the University of Vic in Spain (2012) confirms that teachers generally use self-assessment by self-rating, and few of them use grades according to peer assessment for many reasons; students' lack of confidence, fear of assessing their peers, difficulty in accepting criticism from their peers and difficulty in converting peer assessment into grades (36).

Miller, D. (2010) confirms that the focus is on the psychomotor side more than the cognitive aspect when setting out grades in the applied courses as in the tennis unit through major factors such as (Skill test, Game performance, Tournament standing, Technique, Knowledge test).

Skill test - game performance - technique - tournament standing - knowledge test and it may include assessment procedures such as self-assessment, peer assessment, teacher assessment, or a combination of the three as criteria for grading (40: 95).

Baghurst, T. (2014) indicates that there is a diversity of views on how to assess in the field of physical education. Some notice that measuring skills are through competence and others argue that features such as effort and participation ((Attitude, Effort, Participation) are equal or have the same importance, or even more important. One of the studies that included 617 schools including aspects such as participation, effort, uniform is more than 50% of the grading standards and this means a great contradiction with educational and professional assessments based on norms (18).

It is also difficult to assess effort objectively, an individual with weak skills may seem to make more effort while a highly skilled individual appears to be making less effort, and therefore, it is possible for an individual with weak skills to get the same grade as a highly skilled person, is it fair for these two students to have the same grade? Therefore, when putting effort exerted as a condition or requirement for the assessment and grading process, objective and realistic specifications must be defined and determined for each student (40: 88).

Tippin, G. K, et l. (2012) add that students judge faculty staff as being unfair when the effort invested in completing the task isn't compensated as for poor performance (46).

Evaluators in the field of physical education usually focus on specific criteria or certain elements in the teaching unit, the evaluator may give importance to certain elements rather than other elements and focus on them when grading. Also, the importance of these elements is reflected in the relative weight (percentage of the total score) assigned to each one of them, if assessment grading average is calculated for all elements without assigning relative weights, all scores will be of equal value (40: 95), which is confirmed by the results of this study.

The researcher emphasizes that the assessment elements differ according to the different types of courses, activities and skills (closed and open) that will be evaluated. The difference in the norms for grading is due to the various philosophies and experiences of the faculty staff about the concepts and elements of performance-based evaluation, systems and standards of grading. The norms shouldn't be determined previously before teaching and evaluation when grading. The results of the study of Ahmed Odah and Mufid Al-Hawamdah confirm that there

are differences in grades and these differences are difficult to explain, consequently, the views and philosophy of faculty staff that are related to assessment and grading of university courses should come close, the faculty staff also has to adopt unified standards (1).

In order to increase confidence in the evaluation process, it must be based on clear norms and levels. This is evident in the assessment rubrics which the (Rubric) represents a group of criteria that are used in evaluating students and describe the levels of performance that students are expected to achieve so these levels direct the evaluator to the specifications he is seeking in students' performance according to a graduated scale that is previously specified (160: 5).

Fourth: Contemporary educational trends for grading standards.

- 1- There should be guidebooks as guides for faculty staff directed towards improving their practices for evaluation processes and grading for academic courses inside the university (37).
- 2- The necessity of evaluating the student's performance level and applying their knowledge and skills in a real playing environment (real playing situations such as competitive situations not man made test environment (25: 230)
- 3- Continuous discussions about assessment methods and grading norms should be done through the participation of students and faculty staff.
- 4- The grading norms must take into account the differences of the levels of students' acquisition, application and quality of skills in the field of physical education.
- 5- There should be no separation between process assessment and product assessment as a basis for evaluation and grading systems (13: 23).
- 6- It is difficult for a single assessment tool to evaluate all aspects of students' performance appropriately so the real challenge lies in selecting or developing performance-based assessments to complete each other so that students' evaluation can be fair in the field of physical education (42: 659).
- 7- Using the model (GPAI) (Game Performance Assessment Instrument) as a tool for assessment game performance, which includes aspects such as decision-making (DM) (17).
- 8- There are two types of (Rubrics) first, analytical rubrics for assessing individual traits or performance characteristics separately; secondly, comprehensive rubrics are concerned with evaluating sports with sequential skills. Also, the consistency of form affects the quality of performance (42: 674).

Conclusions

- 1- A variation in the philosophy of performance-based evaluation (product evaluation - process evaluation) and norms for grading according to summative evaluation of track and field events.
- 2- There are differences between the opinions of students of the study teams about the philosophy of performance-based evaluation and the grading systems.
- 3- The difference in the elements of performance-based evaluation and grading systems for faculty staff during the semester works (formative evaluation) for the academic courses of all academic years.
- 4- Reaching the latest contemporary educational attitudes for the philosophy and grading standards in the field of physical education and sports as a major step for educational reform.

Recommendations

- 1- Awareness of the philosophy, applications, and tools of performance-based assessment in the field of physical education and sports.
- 2- Determine declared norms for grading before the start of teaching and evaluation.
- 3- Providing the faculty staff with guides to clarify the norms and details of grading for the various academic courses.
- 4- Holding courses for students on the culture of participatory evaluation and grading norms in university courses.
- 5- Conducting studies on norms and grading systems between each of the courses (theoretical - applied theory - applied) periodically to identify the availability of norms of justice and reliability and achieve equal opportunities for students.
- 6- Conducting studies on grading standards and their relationship to some variables such as (teaching skills - those with learning difficulties - sports achievement - psychological characteristics - assessment tools) in the field of physical education and sports in schools and universities.

First: Arabic references

- 1: Ahmed Odah & Moufid Hawamdah (1996) .Characteristics of the grades distribution in university courses in light of the grading and grading system approved at Yarmouk University.
- 2: Amin Anwar Al-Khouly & Jamal Al-Din Al-Shafei (2000).Contemporary Physical Education Curricula, Dar Al-Fikr Al-Arabi, Cairo.
- 3: Aiken Lewis R (2007) .Tests and exams: measuring abilities and performance, (Farah Al-Sarraj translation) Al-Obikan, Saudi Arabia.
- 4: Big Kelly (2014) .Formative assessment in science 75 practical strategies for linking assessment, teaching and learning (translated by Jabr bin Muhammad Al-Jabr) King Saud University house.
- 5: Jaber Abdel Hameed Jaber (2006) .Contemporary trends and experiences in evaluating the performance of the student and the teacher, Dar Al-Fikr Al-Arabi, Cairo.
- 6- University of Wales (2015): Handbook of awarding degrees within the courses system.
- 7: Salah El-Din Mahmoud Allaam (2009) .Alternative educational assessment, Dar Al-Fikr Al-Arabi, Cairo.
- 8: Afaf Abdel-Karim (1990) .Teaching to learn in physical education and sports, Dar Al Ma'arif, Alexandria.
- 9: Afaf Abdel-Karim (2008) . Development of Curriculum Design Elements in Physical Education, Knowledge Institute, Alexandria.
- 10: Fahmi Al-Balawnah (2010): The effect of the performance-based assessment strategy on developing the sports thinking and problem-solving ability among the secondary school students, (Journal of Al-Najah University for Research of Humanities, Jordan).
- 11: Mohy Al-Din Abdullah Hassan (2011) .Developing methods of assessment of the students in the faculties of education according to the performance assessment model: A future vision, Faculty of Education, University of Sirte.
- 12: Maen Al-Shaalan & Mahmoud Al-Wedyan(2018) .Finding standard grades for some handball skills for the students of the Faculty of Sports Sciences at Mu'tah University, Journal of Al-Najah University for Research of Humanities, Jordan, Vol. (10) 32, Faculty of Sports Sciences, Mu'tah University, Jordan.
- 13: Manal Abdel-Aal Mubariz (2014) .The difference in the type of performance-based assessment in the learning strategy in the web-based projects and its impact on developing problem-solving skills and cognitive control power in the computer and information technology course for the preparatory school students, Egyptian Association for Educational Technology, Cairo.
- 14: Mona Abdou Abdel Dayem (2018) .Development of the assessment tools for teaching Arabic in the preparatory stage in light of the performance-based assessment strategy, Journal of the Faculty of Education, Port Said University.
- 15: Mahdi Ezz-Eldeen (2014) .Using test batteries to evaluate the applied units of the students of physical education and sports institutes in Algeria, Journal of Sports Creativity, University of Messila, Algeria.

Second : Foreign references

- 16 - Annerstedt, C., & Larsson, S. (2010). 'I have my own picture of what the demands are...': Grading in Swedish PEH—problems of validity, comparability and fairness. *European Physical Education Review*, 16(2), 97-115.
- 17 - Arias-Estero, J., & Castejón, F. (2014). Using instruments for tactical assessment in physical education and extra-curricular sports. *European Physical Education Review*, 20(4), 525-535.
- 18 - Baghurst, T. (2014). Assessment of effort and participation in physical education. *Physical Educator*, 71(3), 505.
- 19 - Barkoukis, V., Taylor, I., Chanal, J., & Ntoumanis, N. (2014). The relation between student motivation and student grades in physical education: A 3-year investigation. *Scandinavian journal of medicine & science in sports*, 24(5), e406-e414.
- 20 - Barton, H. M. (1949). A Grading Plan for Physical Education. *Journal of the American Association for Health, Physical Education, and Recreation*, 20(8), 512-544.
- 21 - Brookhart, S. M., Guskey, T. R., Bowers, A. J., McMillan, J. H., Smith, J. K., Smith, L. F., ... & Welsh, M. E. (2016). A century of grading research: Meaning and value in the most common educational measure. *Review of Educational Research*, 86(4), 803-848.
- 22 - Capel, S., Breckon, P., & O'Neill, J. (Eds.). (2013). *A Practical Guide to Teaching Physical Education in the Secondary School*. Routledge.
- 23 - Carroll, B. (2005). *Assessment in Physical Education: a teacher's guide to the issues*. Routledge.

- 24 - Chng, L. S., & Lund, J. (2018). Assessment for learning in physical education: The what, why and how. *Journal of Physical Education, Recreation & Dance*, 89(8), 29-34, <https://doi.org/10.1080/07303084.2018.1503119>.
- 25 - Darst, P. W., Pangrazi, R. P., Brusseau Jr, T., & Erwin, H. (2014). *Dynamic physical education for secondary school students*. Pearson.
- 26 - DinanThompson, M., & Penney, D. (2015). Assessment literacy in primary physical education. *European Physical Education Review*, 21(4), 485-503.
- 27 - Duchane, K. A. (1996). *Attitude and grading practices of secondary physical educators in inclusionary educational settings..* (Doctoral dissertation, Texas Woman's University).
- 28- Guskey, T. R. (2000). Grading policies that work against standards... and how to fix them. *Nassp Bulletin*, 84(620), 20-29.
- 29 - Haerens, L., Krijgsman, C., Mouratidis, A., Borghouts, L., Cardon, G., & Aelterman, N. (2018). How does knowledge about the criteria for an upcoming test relate to adolescents' situational motivation in physical education? A self-determination theory approach. *European Physical Education Review*, 25(4), 983-1001.
- 30 - Hay, P., & Penney, D. (2009). Proposing conditions for assessment efficacy in physical education. *European Physical Education Review*, 15(3), 389-405.
- 31 - Hill, K. L. (1997). *Grading practices of ninth grade physical education teachers* (Doctoral dissertation, University of South Dakota).
- 32- James, R. A. (2018). Grading in Physical Education, *Journal of Physical Education, Recreation & Dance*, 89:5, 5-7, <https://www.tandfonline.com/doi/full/10.1080/07303084.2018.1442063>.
- 33- Koekoek, J., & Van Hilvoorde, I. (Eds.). (2018). *Digital technology in physical education: Global perspectives*. Routledge.
- 34- Krijgsman, C., Vansteenkiste, M., van Tartwijk, J., Maes, J., Borghouts, L., Cardon, G., ... & Haerens, L. (2017). Performance grading and motivational functioning and fear in physical education: A self-determination theory perspective. *Learning and Individual Differences*, 55, 202-211.
- 35 - Leirhaug, P. E. (2016). Exploring the relationship between student grades and assessment for learning in Norwegian physical education. *European Physical Education Review*, 22(3), 298-314.
- 36 - Lorente-Catalán, E., & Kirk, D. (2014). Making the case for democratic assessment practices within a critical pedagogy of physical education teacher education. *European Physical Education Review*, 20(1), 104-119.
- 37 - Lorente-Catalán, E., & Kirk, D. (2016). Student teachers' understanding and application of assessment for learning during a physical education teacher education course. *European Physical Education Review*, 22(1), 65-81.
- 38 - Maclellan, E. (2001). Assessment for learning: the differing perceptions of tutors and students. *Assessment & Evaluation in Higher Education*, 26(4), 307-318.
- 39 - Mao, Y. (1991). *The effects of three grading systems on student outcome measures in table tennis* (Doctoral dissertation, The Ohio State University).
- 40 - Miller, D. (2010). *Measurement by the physical educator why and how*. McGraw-Hill Higher Education.
- 41 - Mintah, J. K. (2003). *Authentic assessment in physical education: Prevalence of use and perceived impact on students' self-concept, motivation, and skill achievement*. *Measurement in physical education and exercise science*, 7(3), 161-174.
- 42 - Morrow Jr, J. R., Mood, D., Disch, J., & Kang, M. (2016). *Measurement and Evaluation in Human Performance, 5E*. Human kinetics.
- 43 - Muñoz, M. A., & Guskey, T. R. (2015). Standards-based grading and reporting will improve education. *Phi Delta Kappan*, 96(7), 64-68.
- 44 - Redelius, K., & Hay, P. (2009). Defining, acquiring and transacting cultural capital through assessment in physical education. *European Physical Education Review*, 15(3), 275-294.
- 45- Svennberg, L., Meckbach, J., & Redelius, K. (2014). Exploring PE teachers' 'gut feelings' An attempt to verbalise and discuss teachers' internalised grading criteria. *European Physical Education Review*, 20(2), 199-214.
- 46 - Tippin, G. K., Lafreniere, K. D., & Page, S. (2012). Student perception of academic grading: Personality, academic orientation, and effort. *Active Learning in Higher Education*, 13(1), 51-61.
- 47- Zhu, X. (2015). Student perspectives of grading in physical education. *European Physical Education Review*, 21(4), 409-420.