Faris Rašidagić[1]

Objectivity in the Evaluation of Motor Skill Performance in Sport and Physical Education

[1]*Faculty of Sport and Physical Education in Sarajevo, Bosnia and Herzegovina*

Original scientific paper
UDC 796.012-053-4

Summary
The aim of this study was directed towards presenting the model of objective evaluation in performing elements of gymnastics as well as presenting the models' appropriate positive results applicable in practice. As a part of sport and physical education teaching curriculum, sometimes it is impossible to measure learners' motor skills by measuring instruments. When this happens we take another approach in differentiating successfulness of performing certain elements based on teacher's belief. This is a complex task for any teacher considering that in a short time it is necessary to disclose a great number of data resulting from a subject's performance. Of course the greater experience of a teacher the greater quality will be in recording this information and evaluating them. A possible favourable circumstance can be defined as a request from a student to repeat certain motor task at least three times, thus allowing the evaluator to complete information on the quality of performed technique. Raising objectivity can be accomplished by introducing the student in evaluation process, the one who performs the given task at the highest level (for example additional two students from school and a teacher). Naturally it is necessary to previously develop an evaluation scale, which will be made available to all the evaluators and students undergoing evaluation. Only under these circumstances, will it be possible to determine a matter-of-fact quality level of performed technique so as to make appropriate student evaluation. The evaluation was performed by applying the acceptable statistical methods and comparing the evaluation made by three competent educators/judges and an individual independent teacher. Finally, using a sample of students aged from 15 to 16, it was determined that the adjustment of evaluation criteria made by the competent educators/judges had an appropriate-positive effect on the evaluation objectivity.

Key words: **Gymnastics, Success Scale Evaluation, Factor Evaluation Analysis, ANOVA**

Introduction

As a part of sport and physical education curriculum it is necessary to continually ad systematically evaluate the work of students throughout curriculum implementation. The evaluation is organised so as to have a constant insight into an individual progress. As a result of the evaluation a student is awarded a grade, which must be based on versatile and permanent evaluation (Hadžikadunić, 2004). One of the methods to objectivise the evaluation during the performance of motor tasks is elaborated in this paper. Unifying the evaluation must not represent an interfering factor for the teachers and should not be based on the teachers' intuition. In order to eliminate the aforementioned interfering factors (Višnjić, Ilić, Martinović, 2011), the quality of success evaluation during the technique performance of a certain element, should be established according to the highest standards. With an aim to acquire greater objectivity we could in advance determine unified forms (Federal Ministry of Education, 1999) for students (e.g. separately for each sport activity). These would be used in comparison with the performed technique when making an evaluation. Regardless of the non existence of the forms in questions, teacher should adhere to the specific principles during the evaluation. These principles can be synthesized as the following (Najšteter, 1997): systematic (planned and continuous during the whole school year), individual (there are certain anthropological differences among the students which should be taken into account) versatile (to include all the content covered by the curriculum) and concrete (to eliminate all the prejudices and student relationships from the evaluation, likewise to acknowledge the presence of diversity among students). These conditions enable recurrence of the quality which assists in the accomplishment of the most important values as a result of physical education and school sport influence. The evaluation can be descriptive or numerical, allowing for the fact that both scales have the same grading system, i.e. the number from 1 to 5 carries the appropriate descriptive value. Evaluation of the performed technique for the appropriate element is usually conducted by observing. In order to carry out the observation in as much objective way as possible (Findak, 2001), next to having a set scale for the evaluation beforehand, it is necessary to define goal and the subject of observation, observational time frame, to arrange the method of the observation, data register method and methods which would

be used to analyse the collected data. Using this approach we can accomplish objective observation, and evaluate student's success much more realistically (Jonsson, 2007). This research demonstrates one of the possible methods in how to perform evaluation success objectively, i.e. quality evaluation of the performed technique, where elements of sport gymnastics served as an example. All the participants who went through the evaluation process were familiarised with the criteria and grading method.
The aim of this study was directed towards presenting the model of objective evaluation in performing elements of gymnastics as well as presenting the models' appropriate positive results applicable in practice.

Methods

The sample of subjects

Subject sample for this study was comprised of 100 male students, who attend first and second year of high school, from 15 to 16 years of age old. There were no students with somatic illnesses. Subjects had all the right conditions for regular school attendance concerning sport, physical and health education.

Variable samples

Based on the previously mentioned characteristics, to complete the evaluation of gymnastic elements success performance in sport and physical education curriculum, three competent educators/judges were chosen, as well as the teacher's evaluation of the previously conducted curriculum.

Methods of collecting data

Subject performs all the elements regardless of the performance speed (Horga, 2010), and the evaluation of acquired sport techniques will be performed by the teacher who conducted the curriculum (I - independent educator), along with three (A, B, C) competent educators/judges – physical education teachers (Black at al., 2004). Ranking of subjects' acquired knowledge in a chosen technique has been defined on a scale of 5 grades (1 to 5). The grading was conducted exclusively by „full" grades. Other grades were not allowed. In order to homogenise the factors which effect the assessment of knowledge and techniques of acquired elements, before the start of evaluation the entire procedure was explained to the evaluators, as well as the course and evaluation criteria (Jarvis, 1999). The evaluated technique did not change its sequences during the entire evaluation process.

Table 1. Description of evaluated elements

SPORTS	ELEMENT DESCRIPTION
Gymnastics	Standing position, handstand, forward roll, Standing position, backward roll, standing position

Competent evaluators/judges have approved and agreed upon the criteria before the evaluation, placing emphasis on: body posture, hand coordination, leg coordination and general coordination. The independent teacher did not adjust his own criteria with those of evaluators/judges, but based his evaluation on personal experience in teaching process and student work.

Table 2. Criteria for grade distinction

GRADE	ELEMENT DESCRIPTION
Grade 1	From a standing position attempt handstand, perform a forward roll which cannot be completed or is not performed in the direction of movement (aslant), without getting up or with some individual attempts to perform remaining elements.
Grade 2	From a standing position perform a handstand and without remaining in that position perform a forward roll in the right direction, attempt to take on the standing position, or the attempt is performed later, failed attempts to perform other moves, performed with stances and irregularities.
Grade 3	From a standing position perform a handstand and without any delay and remaining in that position perform a forward roll in line, while taking on to a standing position, attempt to perform backward roll, but it is either performed incorrectly or not in line, followed by an attempt or performance of other elements but incorrectly and not in order.
Grade 4	From a standing position perform a handstand and while remaining in that position perform a forward roll in line, taking on a standing position and then performing a backward roll in the direction of the movement, an attempt or insecure stand for the final upright position.
Grade 5	From a standing position perform a handstand and with remaining in that position perform a forward roll in line taking on a standing position, all elements performed fluently and with confidence.

Data analysis

For the data analyses the following were used: Descriptive statistical methods, Factor Evaluation Analysis and ANOVA analysis, using SPSS 22.0 IBM.

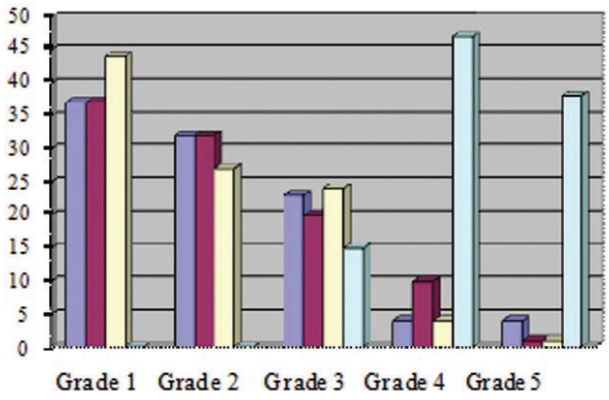
Results

During the data analysis, initially the number and height of grades awarded by the three (A, B, C) educator/judges and (I) independent teacher were analysed.

Table 3. Distribution of ratings A, B, C, educators/judge compared with I- Independent Educator

Grade	1	2	3	4	5	Summ. A
A - educator statistics	37	32	23	4	4	100
B - educator statistics	37	32	20	10	1	100
C - educator statistics	44	27	24	4	1	100
I - educator statistics	**0**	**0**	15	47	38	100

A – Educator/Judge, B – Educator/Judge, C – Educator/Judge, I – Independent Educator



A – Educator/Judge, B – Educator/Judge, C – Educator/Judge, I – Independent Educator

Figure 1. Graphic relationship score Distribution of ratings A, B, C, educators/judge compared with I- Independent Educator

Analysing grade distribution one can state that the A, B, C evaluators/judges were acquainted with the grading scale from 1 to 5, considering that there is not even one badly formulated grade (e.g. above 5, or a grade 4.5 or something similar). However the independent teacher who did the evaluation individually obviously failed to apply the grades from the scale. Among the evaluators the highest marks for technique performance were awarded by the educator/judge listed under the letter A, taking into account the number of performances which were ranked under 5, while the lowest performance grades (the most rigorous) were awarded by the educator/judge C (the highest number of 1's and the lowest number of 5's). Analysing the grades awarded by the independent teacher we can state that 1 or 2 were never awarded to any of the students for the performance of the given element, i.e. he/she did not pay close attention to coordination, performance fluency, precision technique, control, efficiency and effectiveness of the movements (Honeyburne, 2006). Further on, factor analyses was applied so as to examine the statistical reality of grading performed by A, B, C , educator/judges and the independent educator. Within this context, Bartlett's Test was applied, which is primarily used to check whether or not the use of factor analysis was justified.

Table 4. KMO and Bartlett's Test

Bartlett's Test	Approx. Chi-Square	,677
df		6
Sig.		**,000**

After Bartlett's Test pointed out towards the justified application at a level ,000 the next in line was the analysis and component extraction which will represent an imaginary educator/judge structured by grades. Total variance displayed only one significant function, which was to be expected.

Table 5. Total Variance Explained

Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2,384	59,591	59,591	2,384	59,591	59,591

The next table (Table 6) indicates the correlation between grades for all those who participated in grading the student performance (A, B, C and I educators). The correlation was taken into account regarding the independent imaginary judge who was defined through the first main component and according to all the grades. It is clearly visible that teacher, who participated in adjusting the criteria, has diverged from the grading system of other evaluators.

Table 6. Communalities and Component Matrix(a)

	Communalities	**Extraction**	**Component Matrix**
A-educator	1,000	,724	,851
B- educator	1,000	,875	,935
C- educator	1,000	,759	,871
I- educator	1,000	,026	-,161

A- Educator/Judge, B- Educator/Judge, C- Educator/Judge, I- Independent Educator

Unlike correlation which proved to be at an appropriate level, cross-correlation defined by the data displayed in Table 7 has not fully exhibited the expected level of similarity between the grades appointed by the A, B, C educators/judges, considering that not even one correlation coefficient was above ,765. Anticipated coefficient which was expected to demonstrate maximum similarity evaluation made by the judges was ,800 and higher.

Table 7. Correlation ratings for Educators/Judges

Correlation	**A- educator**	**B- educator**	**C- educator**
A-educator	**1**		
B-educator	,715	**1**	
C-educator	,567	**,765**	**1**
I-educator	-,087	-,118	-,044

A- Educator/Judge, B- Educator/Judge, C- Educator/Judge, I- Independent Educator

In order to confirm whether evaluators shared appropriate similar understandings regarding the characteristics under evaluation, resulting from confusion caused by relatively low correlation coefficient within the Correlation matrix, the ANOVA analysis was applied for all the judges, the grades appointed by the judges, and for a connection between the grades appointed by each educators/judges individually and grades awarded by the independent educator who was not acquainted with the evaluation criteria.

Table 8. ANOVA – Correlations between educators/judges who are familiar with criteria

	Sum of Squares	df	Mean Square	F	Sig.
Correlation A/B	,325	1	,325	,296	**,587**
Correlation A/C	,135	1	,135	,133	**,716**
Correlation B/C	,135	1	,135	,135	**,714**

Correlation A/B: Between A and B educators/judges, Correlation A/C: Between A and C educators/judges, Correlation C/B: Between C and B educators/judges

After conducting ANOVA analysis, if we look at the significance which in this case indicates whether or not the grades of independent competent judges are different in regards to statistical significance we can state that by analysing the coefficient of significance for all the judges there were no statistically significant differences between the grades awarded for performance, therefore we can acknowledge that the criteria was adjusted at an appropriate level. In order to once again appropriately evaluate quality of adjusted criteria evaluation, the difference in grading performed by every judge personally and the independent teacher was evaluated, likewise using the ANOVA analysis (Table 9).

Table 9. ANOVA – Correlation between educators/judges who are familiar with the criteria and independent educator

	Sum of Squares	df	Mean Square	F	Sig.
Correlation A/I	235,445	1	235,445	292,552	,000
Correlation B/I	235,445	1	235,445	303,998	,000
Correlation C/I	269,120	1	269,120	380,885	,000

Correlation A/I: Between A and I educators, Correlation B/I: Between B and I educators, Correlation C/I: Between C and I educators

Considering that in this case the significant difference between grades is statistically significant and at the highest level, 000 we can state that the evaluation performed by the all three educators/judges differed from that of the independent educator who was not acquainted with the evaluation criteria, and the grades awarded by the "I" educator varied from the previously confirmed criteria.

Discussion

Considering that the process of motor skill evaluation has always been exposed to a certain criticism and sometimes even followed by dissatisfaction of those who were evaluated (Magill, 2007), this paper is represented as a model indicating a method in how to make it more objective. This entire process was organised according to the criteria adjustments made by the three (A, B, C) independent judges for the evaluation of specific gymnastic elements (Lund, 1992). The teacher (I) who is directly involved in a teaching process and was not familiar with these evaluation criteria, also performed the evaluation independently. The students who were evaluated were familiar with the evaluation and grading criteria. Variations in student evaluation have been noticed immediately with the grade distribution (Table 3 and Chart 1). To be more exact the independent educator who was evaluating the students did not have any evaluation or a grade for performance and achievement of elements with grade 1 or 2, therefore the lowest grades started with 3. In further analysis, all the grades were analysed through factor analysis (Wolf, Rađo, 1998), after which the presented KMO and Bartlett Test was at a level of significance .000, a good introduction for further Factor Evaluation Analysis application. Further on, the first main component was isolated with a characteristic root of 2.384 and variance explanation of 59,59%. Considering the specificity of the analysis, that is factorisation of grades for only one technique, it was expected that only one significant function would be isolated, but it was not expected that the level of variance explanation would be in range of 60%, considering that the A, B, C, educators/judges adjusted their criteria before the evaluation. However, analysing the relations between the grades awarded by individual judges even if compared with the first main component, we can still see the positive result after the adjustment because A, B, C, educators/judges are connected by the isolated main component through high grades projections (Table 6), so we can conclude that each of them knew what they were evaluating, what characteristic student movements (posture, hand coordination, leg coordination and general coordination) should be observed. If the A, B, C, educators/judges were not familiar with the criteria and grading system, the isolated coefficients of significance would be significantly below the level. We must emphasise here that the educator/judge B was connected with the first main component by the highest coefficient, so it can be concluded that he was clear with the procedure of assigning any type of grade in accordance with the specific level of performance, which indicated that he was amongst the first unbiased ones. If we pay closer attention to the correlation coefficient of the independent educator's (I) evaluation who was not acquainted with the evaluation criteria, but performed the evaluation based on personal experience and in the usual way, it becomes noticeable that the stated coefficients are at a relatively low correlation level with the first and main component, indicating that there is no correlation (Noonan and Duncan, 2005). In order to assess the adjustments of A, B, C educators/judges' evaluation, mutual assessment of grade correlation was performed. Unfortunately, coefficients calculated in this way were not higher than ,765 so the authors opinion was to perform an additional assessment of the grade modification made by the A, B, C judges and the I educator.

The same was performed with the ANOVA analysis, and data produced pointed out towards the indicator that the A, B and C educators' grades are not statistically different, so we can acknowledge the fact that the judges knew and applied already adjusted criteria. The assessment was conducted to check whether or not evaluation corresponds to the grading system (using the same ANOVA analysis) comparing each judge personally with the independent educator. This has indicated that there is statistically significant difference between the performed evaluations, and therefore we can state that the independent educator not knowing the criteria did not award the same or similar grades as the A, B, C educators/judges.

In a conclusion we can state that as a result of previously provided consultations, and the training on the quality levels of performed technique where each was carrying a certain grade, had a positive impact on criteria adjustments, later evaluation and grading in which case the A, B, C educators/judge's grade corresponded with the students' performance (Harrison, Marshall and Wiliam,

2004). In this case we must stress that the process of motor learning most commonly indicates learning based on a great number of repetitions and that some sources state that in order to acquire one motor skill it is necessary to repeat it between 10000 and 15000 times (Schmidt and Wrisberg, 2000), so that the presence of uniqueness does not even come as a surprise, or the presence of a certain percentage of unexplained variance which in this study goes up to 40 %. Not so great satisfaction comes from the presence of this unexplained variance percentage, which can be clarified by the disturbing factors or requests for movement modification (Gentile, 2000), because these are very difficult to explain in any other way apart from certain technique adjustments for the one who performs exercises, but also subjectivity to some extent, which is present with the judges regardless of the previously set criteria. Also, (Van Wersch, Trew, Turner, 1992) stated that the students' motivation in accomplishing activities from the teaching curriculum is inversely proportional to their growth, so that partly this can be perceived as a part of unexplained variance. Summing up all previously stated we can conclude that the aim of this paper was achieved, and we presented the model of objective evaluation in performing elements of gymnastics and how to achieve this model's appropriate results in practice during the performance of certain motor tasks.

Conclusions

As a part of sport and physical education curriculum, organised in schools, its aim is to launch and continually work on the maintenance of a systematic observance of students work during the implementation of these curricula.Observance is organised with intent to develop constant overview and control of individual improvement, and also awarding a concrete grade for the performance. Concerning this study, the primary role of awarding a grade was to offer feedback, acquired on the basis of performed task. This greatly facilitates the development of motor skills or it helps in the quality of the performance (Feltz and Landers, 1983), (Rushall and Lippman, 1997), (Rogers, 2006).

This study is for all intense and purposes a case-study, with an aim to present the model of objective evaluation in performing elements of gymnastics and how to achieve this model's appropriate results in practice during the performance of certain motor tasks.

Even though in this paper we have used elements of sport gymnastics, it is possible to establish the same model for other subjects as well (if not for everyone than at least for the content mostly used in teaching).

It is important to note that the A, B, and C educators/judges, as well as students were familiar with the grading criteria, because this immediately affected the quality of their evaluation, but also their performance, even though we must take into account the growth and motivation of learners who should perform an assigned content (Xiang et al. 2003).

The same has been confirmed by the descriptive analysis of the grades, as well as using the example of Factor Evaluation Analysis where the significance of KMO and Bartlett test was at the level of .000, indicating that its use was justified, and the later relation of isolated main component with the 59,59% explained variance, indicated that the correlation between the grades awarded by the A, B, and C evaluators/judges in comparison to grades of the independent teacher – I who did the evaluation based on personal knowledge and teaching experience.

Lesser dilemma on the quality of the adjusted criteria was indicated by the **correlation** matrix of the grades awarded by A, B, and C educators/judges, considering that not even one of the correlation coefficients of the evaluation was above, 800. In order to eliminate the misconceptions regarding this matter we applied ANOVA analysis for judges' grading – intergroup, as well as to evaluate relation between the grades of the A, B, C educators/judges individually and the grades of the independent educator (I) who was not acquainted with the criteria but performed the grading on basis of personal teaching experience.

The collected indicators confirmed the correlation between the A, B, and C judges' grades, as well as the difference once compared with the I- independent educator's grades. Therefore we can conclude that the previously provided consultations had a positive impact on the adjustment and criteria application.

With this the main aim of this study, which was designed to present the model of objective evaluation in performing elements of gymnastics and how to achieve this model's appropriate results in practice during the performance of certain motor tasks, has been achieved.

The recommendation would be to adjust the evaluation of successful technique performance of elements and set it as a standard for other teaching content planned and programmed to be implemented as a part of sport and physical education curriculum.

References

Black, P., Harrison, C., Lee, C., Marshall, B. and Wiliam, D. (2004). Working inside the black box.: Assessment for learning in the classroom. Phi Delta Kappan. 86 (1), 13-22

Federal Ministry of Education (1999). The curriculum for secondary schools. Bosnia and Herzegovina, the Federation of Bosnia and Herzegovina, the Federal Ministry of Education, Science Culture and Sport. Sarajevo.

Feltz, D. L. i Landers, D. M. (1983). The effects of mental practice on motor skill learning and performance: A meta-analysis. Journal of Sport Psychology, 5, 25-57.

Findak, V. (2001). The methodology of physical education. School books. Zagreb

Gentile, A.M. (2000). Skill acquisition: Action, movement, and neuromotor processes. U J.H. Carr i R.B. Sheperd (Ur.) Movement science: Foundations for physical therapy (2nd Ed), (str. 111-187) Rockville, MD: Aspen.

Hadžikadunić M. Mađarević, M. (2004). Methods of teaching physical education with the basics of exercise physiology. Pedagogical Faculty in Zenica. Zenica.

Honeybourne, J. (2006). Acquiring skill in sport. An introduction. New York: Routhledge.

Horga, S. (2010). In psychology of sport, 2nd Edition. Zagreb: Faculty of Kinesiology.

Jarvis, M. (1999). Sport Psychology. London & New York: Routhledge.

Jonsson, A. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. Educational Research review, Vol. 2., Issue 2., Pages 130 – 144.

Lund, J. (1992). Assessment and Accountability in Secondary Physical Education. QUEST - National Association for Kinesiology in Higher Education (NAKHE), Vol. 44, Issue 3., Pages 352-360.

Magill, R. A. (2007). Motor learning and control. Concepts and Applications. New York; McGraw-Hill.

Najšteter, Đ. (1997). Kinesiology didactics. Bosnia and Herzegovina, the Federation of Bosnia and Herzegovina. Federal Ministry of Education, Science, Culture and Sport. Sarajevo.

Noonan, B., Duncan, C. R. (2005). Peer and Self-Assessment in High Schools, A peer-reviewed electronic journal, Vol. 10 No. 17, (1-8).

Rushall, B. S. i Lippman, L. G. (1997). The role of imagery in physical performance. International Journal for Sport Psychology, 29, 57-72.

Rogers, R. (2006). Mental Practice and Acquisition of Motor Skills: Examples from Sports Training and Surgical Education.Obstetrics and Gynecology Clinics of North America, 33(2), 297-304.

Schmidt, R.A. i Wrisberg, C.A. (2000). Motor learning and control, 2nd edition. USA: Human Kinetics.

Van Wersch, A., Trew, K., & Turner, I. (1992). Post-primary school pupils' interest in physicaleducation: Age and gender differences. British Journal of Educational Psychology, 62, 56-72.

Visnjic, D. Ilic, J., Martinovic, D. (2011). The motivation for engaging, achievements in physical education and the age of students. 6 Th FIEP European Congress, Str. 504-511

Wolf, B., Raja, I. (1998). A cluster analysis of manifest variables. Faculty of Sport in Sarajevo. Sarajevo.

Xiang, P., McBride, R.E., Guan, J.M., & Solmon, M.A. (2003). Children's motivation in elementary physical education: An expectancy-value model of achievement choice. Research Quarterly for Exercise and Sport, 74, 25-35.

Correspondence to:
Faris Rašidagić
Patriotske lige 41, 71 000 Sarajevo
Bosnia and Herzegovina,
e-mail: faris.rasidagic@gmail.com