

Damjan Jakšič*
Milan Cvetković

**RELIABILITY OF TESTS FOR ASSESSING
EXPLOSIVE STRENGTH IN PHYSICAL
EDUCATION STUDENTS: HOW
RELIABLE ARE THEY AND HOW CAN
THE PROPER ONE BE CHOSEN?**

**ZANESLJIVOST TESTOV ZA UGOTAVLJANJE
EKSPLOZIVNE MOČI PRI ŠTUDENTIH
ŠPORTNE VZGOJE: KAKO ZANESLJIVI
SO IN KAKO IZBRATI USTREZNE?**

ABSTRACT

A sample of 155 first-year male students at The Faculty of Sport and Physical Education in Novi Sad, Serbia underwent three motor ability tests for assessing explosive strength: standing long jump, standing high jump and standing triple jump. Based on the obtained results, it was possible to conclude that the reliability of all the tests was more than satisfactory, that the highest projections of the replication vector on the first principal component are seen in the second replication of each test and that the hypothetical latent dimension of explosive strength had the best manifestation in the standing triple jump test, followed by standing long jump. Finally, the weakest manifestation but nonetheless still very good was obtained in the standing high jump test.

Key words: composite motor test, metric characteristics, 20-year-old students

IZVLEČEK

Na vzorcu 155 študentov prvega letnika Fakultete za šport in telesno kulturo v Novem Sadu, Srbija smo primerjali tri različne teste za ugotavljanje eksplozivne moči: skok v daljino z mesta, skok v višino z mesta in troskok z mesta. Na podlagi dobljenih rezultatov je bilo mogoče sklepati, da je zanesljivost vseh testov več kot zadovoljiva, da so najvišje projekcije ponovitvenega vektorja na prvo glavno komponento v drugem poizkusu in da je hipotetična latentna razsežnost eksplozivne moči najbolj izražena v troskoku z mesta, nato v skoku v daljino z mesta, najšibkejša, a vseeno zadovoljiva, pa v skoku z mesta v višino.

Ključne besede: sestavljeni gibalni testi, merske lastnosti, 20-letni študentje

*Faculty of Sport and Physical Education, University of
Novi Sad, Novi Sad, Serbia*

***Corresponding author:**

Faculty of Sport and Physical Education,
Lovčenska str. 16, Novi Sad, Serbia
Tel.: +381 63 8 280 546
Fax: +381 21 450 199
E-mail: jaksicd@neobee.net

INTRODUCTION

Reliability testing is usually a task that a researcher needs to perform before he proceeds to the applied battery of tests. Unfortunately, in practice, this does not happen often enough. Reliability is specific; it depends on each sample of participants and it is to be controlled permanently. Thus, the high quality selection of a battery of tests is provided, its correction is made if necessary and the course of the research is monitored and guided.

Reliability is usually described as one of the two basic metric characteristics of a test (the other one being validity) (Fajgelj, 2005).

However, as far as this issue is concerned, it would be best to say that here we have predominantly dealt with test score reliability and then the reliability of measuring instruments (Braumgartner, 2006; Haase, 1998). This terminology clarification is very important, but it does not substantially or practically change the basic aims of this approach. In any case, it is worth taking into consideration.

It is very important to control reliability. “Everyone does it that way” -- according to Fajgelj (2005), that statement is frequently used in psychological testing. However, it fell out of use by serious researchers long ago. The already mentioned specificity of each sample of participants, as well as that of each applied test has long been understood. Therefore, it is necessary to fulfil minimal standards of reliability so that our further research has high quality interpretation and application.

Bearing in mind all of this, the aim of this research is to test the reliability of the applied motor measuring instruments, to discover in which test replication the best results are achieved and finally to give practical advice to future researchers regarding which test is the best representative of explosive strength.

METHODS

The research was carried out on students at the Faculty of Sport and Physical Education from Novi Sad, Serbia. The total number of participants was 155, all of them were male, their average age: 20.15 (SD=0.83) years.

Students are frequently part of the sample of participants in many professional and scientific papers. In this paper, the sample examinees might be qualified as even more appropriate, because the groups have already been formed in advance and treated as such further in the research.

In this case, we ran the tests on the students who were attributed above average motor status in some areas by some researchers (i.e. Metikoš, Prot, Horvat, Kuleš, & Hofman, 1982).

Three motor measuring instruments were applied to participants, assessing the same hypothetical latent dimension: explosive strength. All the three tests were composite and consisted of three replications. The applied motor measuring instruments were standing high jump, standing long jump and standing triple jump.

All the measuring and testing procedures were taken in their entirety from the work of Metikoš, Prot, Hofman, Pintar, & Oreb (1989).

According to Fajgelj,

It is a well-grounded belief that for the assessing of reliability, it is not enough to apply only one sign. There are too many reliability coefficients (tens of them) and each one is supported by a thoroughly-developed measuring model and a developed assumption. The necessity of using more different coefficient is shown through the fact that when calculated all of these coefficients give different reliability assessments from the same data (2005).

For the purposes of this paper, as well as of the assay, aside from validity as the most significant internal metric characteristic of a measuring instrument (Bala, Stojanović, & Stojanović, 2007), reliability was tested by means of five different reliability coefficients: Spearman, Brown et al.'s α reliability coefficient, which is usually used for calculating the reliability of measuring instruments, Momirović's λ_7 , Lord, Kaiser & Caffrey's β , Momirović's lower reliability limit β_7 and Guttman & Nicewander's coefficient ρ .

It is necessary to note that the first two coefficients (α and λ_7) are based on the classic summary model, the next two (β and β_7) represent the reliability of the first principal component, and the last one, coefficient ρ , represents a coefficient which is based on Guttman's measurement model.

For calculating the reliability of motor measuring instruments, RTT11G software was used (Momirović, 2001), which was written in Matrix programming language in order to be carried out in the standard SPSS environment. The definitions and formal mathematic presentation of all the measures implemented in that program can be found in the work of Momirović, Wolf, & Popović (1999).

Basic descriptive statistics were calculated for all the replications of individual tests (measures of central tendency, measures of variability and measures of the form of distribution), while the highest replication of each test was defined by vector projection on the first principal component, where the highest value represented the best performance within the measuring instrument.

Whether or not this is the question of the same hypothetic latent dimension was tested by factor analysis, which further in the research had a double function: a) to show that it is one hypothetic latent dimension, thus it will be a confirmatory factor analysis and b) and thus to identify the test bearing the greatest amount of information of this dimension.

A number of significant principal components were calculated based on the criterion suggested by Guttman (1953) and Kaiser (1964) (according to e.g. Gredelj, Metikoš, Hošek, Momirović, 1975; Tenjović, 2002), and which is based on the size of each characteristic root, implying that the number of these values, equal to or higher than 1, represents a significant factor, i.e. component.

RESULTS

Tables 1 to 3 show the results for each test separately. For each replication, in each test the following was calculated: the arithmetic mean (AM), standard deviation (SD), the minimal results (MIN), maximal results (MAX) and the measures of distribution, skewness (Ske) and kurtosis (Kur).

Further calculations for each test included an inter-correlation matrix of the replications (R), first eigenvalue (λ_1), common variance (V), the first principal component (H1) and communalities (h^2).

Table 1: Descriptive statistics, coefficients of item's correlation (lower triangle), statistic significance of correlations (upper triangle), first principal component (h1), communalities (h²) for the standing long jump test

Item	[cm]						R			H1	h ²
	Mean	SD	Min	Max	Ske	Kur	1	2	3		
1	234.93	19.82	160	299	-0.11	1.21		0.00	0.00	0.94	0.88
2	239.75	18.73	198	295	0.25	0.07	0.86		0.00	0.96	0.92
3	242.07	18.97	190	295	0.13	0.33	0.84	0.90		0.95	0.91
$\lambda_1=2.73$										$V=91.14\%$	

Legend: S – standard deviation, Min – minimal results, Max – maximal results, Ske – skewness, Kur – Kurtosis, R – Intercorrelation matrix, H1 – First principal component, h² – Communalities, λ_1 – First eigenvalue, V – Common variance

Analyzing the values from Table 1, it can be seen that the highest value of the arithmetic mean occurs during the third performance of the test: 242.07 cm. Skewness and kurtosis values do not deviate significantly from the normal value.

The values of the inter-correlation matrix show that the highest correlation is between the second and third replication. We can see that the highest consistence occurs within the results in items shown in the second replication, which is evident since it has the highest value of the coefficient of vector replication on the first principal component. That value is 0.96. Upon inspection of the range, it can be seen that the smallest value is also in that replication, as with standard deviation, SD=18.73. One principal component was extracted, whose eigenvalue showed, as expected, a very high value, $\lambda_1=2.73$, and that proportion was explained with 91.14% of common variability. The expected extremely high values of communalities were also confirmed here; the second of which was the highest: $h^2=0.92$.

Table 2: Descriptive statistics, coefficients of item's correlation (lower triangle), statistic significance of correlations (upper triangle), first principal component (H1), communalities (h²) for the standing high jump test

Item	[cm]						R			H1	h ²
	Mean	SD	Min	Max	Ske	Kur	1	2	3		
1	287.59	11.84	264	322	0.25	-0.41		0.00	0.00	0.97	0.98
2	288.55	11.79	264	324	0.20	-0.44	0.99		0.00	0.98	0.98
3	288.22	13.74	200	325	-1.52	9.87	0.89	0.90		0.95	0.95
$\lambda_1=2.86$										$V=95.32\%$	

The values from Table 2, on the Standing high jump motor test, show that the arithmetic mean had its the highest value in the second test replication: 288.55 cm. Skewness and kurtosis values in the third replication indicate deviation from a theoretic curve. In the above replication, we can see that the skewness value is negative (-1.52), which means that most of the results are higher than the average value of replication. The kurtosis value (9.87) also shows that the curve is leptokurtic.

The values from the part of the matrix where correlations (R) are given show that the greatest correlation existed between the first and the second replication. The fact that the highest consistency

among the items shown during the second attempt (even in this case) is perceived based on the projection of replication vectors on the first main component. The component showed the value 0.98, testing the range where the lowest value as well as the standard deviations are perceived, being also the lowest, with the value of $S=11.79$. One of the main components was extracted, whose characteristic root showed even a higher value here with respect to the values in the first test: $\lambda_1=2.86$. Proportion was accounted for by 95.32% of common variability, while very high values of communality were achieved in this case: $h_2=0.98$.

Table 3: Descriptive statistics, coefficients of item's correlation (lower triangle), statistic significance of correlations (upper triangle), first principal component (H1), communalities (h_2) for the standing triple jump test

Item	[cm]						R			H1	h^2
	Mean	SD	Min	Max	Ske	Kur	1	2	3		
1	658.04	59.02	490	838	0.12	0.71		0.00	0.00	0.95	0.90
2	669.78	56.67	520	851	0.25	0.37	0.88		0.00	0.97	0.94
3	675.61	56.23	500	850	0.09	0.66	0.87	0.93		0.96	0.94

$\lambda_1=2.80$ $V=93.27\%$

The values of the triple jump test are given in Table 3. The arithmetic mean appeared to have its highest value during the third performance of the test, amounting to 675.61 cm. The skewness/kurtosis test values do not considerably deviate from the normal value. The values from the part of the matrix where the correlations (R) are given show that the highest connection occurs between the second and the third replication. The fact that the highest consistency (among the items shown during the second attempt) is perceived based on the projection of replication vectors on the first main component (which showed the value 0.97) testing the range where the lowest value as well as the standard deviations are perceived (being also the lowest with the value of $S=11.79$). One of the main components was extracted, whose characteristic root showed a high value of $\lambda_1=2.80$. The connection was accounted for by 93.27% of common variability, while the communality value of the best replication is $h_2=0.94$. As it was stated earlier, the motor reliability test was run applying more measures that represent reliability under different measuring models. Theoretically speaking, the hypothetical limit of good reliability of measuring instruments must not be lower than 0.90 (e.g. Gredelj, Metikoš, Hošek, & Momirović, 1975; Momirović et al., 1999; Bala et al., 2007), even though the tests which have reliability quotient values higher than 0.85 are also considered to meet the criteria.

Table 4: Motor tests reliability

Variable	α	λ_7	β	β_7	ρ
Standing long jump	0.951	0.967	0.951	0.967	0.988
Standing high jump	0.975	0.983	0.975	0.983	0.999
Standing triple jump	0.963	0.975	0.963	0.976	0.994

Legend: α – Spearman-Brown-Kuder-Richardson-Guttman-Cronbach's reliability coefficient, λ_7 – Momirović's reliability coefficient, β – Lord-Kaiser-Caffrey's reliability coefficient, β_7 – Momirović's lower reliability limit, ρ – Guttman-Nicewander's reliability coefficient

Table 4 quite explicitly and evidently shows that the applied tests showed excellent reliability as the values within each criterion are above 0.90.

In Tables 1 to 3, the highest values of the first principal component are in bold. These are the replications of each test which have the highest results consistency. The values of these replications consisted of a new matrix of secondary data with two goals: to show that it is one hypothetic latent dimension, thus it will be a confirmatory factor analysis and to identify the test bearing the greatest amount of information of this dimension.

In the analysis of the principal components of the inter-correlation matrix (which represented the second replication of every motor test), and on the basis of Kaiser-Guttman's criterion, only one principal component was extracted (Table 5). That had been expected while observing the inter-correlation matrix (R). High communality values further confirm this, as well as the high percentage of the already-explained variance of the first and only main component. The values of the other two characteristic roots are below 1 and do not represent new factors on their own.

Table 5: Coefficients of items correlation (lower triangle), statistic significance of correlations (upper triangle), first principal component (H1), communalities (h^2), eigenvalues (λ), percent of common variance (V)

Variable	R			H ¹	h ²
	1	2	3		
Standing long jump		0.000	0.000	0.888	0.788
Standing high jump	0.621		0.000	0.865	0.748
Standing triple jump	0.749	0.699		0.919	0.845
		$\lambda_1=2.381$		V ₁ =	79.37%
		$\lambda_2=0.384$		V ₂ =	12.80%
		$\lambda_3=0.235$		V ₃ =	7.83%

Legend: R – Inter correlation matrix, H¹ – First principal component, h² – Communalities, λ_1 – First eigenvalue, V – Common variance

DISCUSSION

Generally speaking, when discussing the analyzed motor tasks, it can be said that there are simple or complex stimuli which provoke neurophysiologic, i.e. motor processes in the examinees (Bala, & Krneta, 2006). In this paper, it was confirmed that motor abilities cannot be measured or assessed based on only one motor task or only one replication. The necessity for applying tests comprised of more than one replication has been understood for quite some time as human bodies, the basic subject of kinesiology, are very complex systems and not easy to define.

It is always necessary to provoke a person's maximum motor ability. This represents his/her maximal manifestation in testing and a reduction of all the accompanying residual factors to a minimum, which overlaps and correlates with the analyzed ability in many different ways. This most often implies performing repetition of the same task, without a break or shorter breaks between the repetitions. As has already been said, in practice it is impossible to obtain the isolation of a motor ability, as measuring always depends on other factors, such as motivation, body constitution, earlier experience in the tests, but also the fatigue that occurs. The factor last mentioned

is specifically significant in the case of composite tests in kinesiological research; this requires particular attention

Tables 1, 2 and 3 show that the tests applied to this sample in this research need to be performed at least twice in a row. This is due to the fact that first performance of a test is usually an attempt for replication when the examinees are learning or trying to remember the moving structures which are necessary for performing all the movements included in those tests and whose final product is a jump (long jump, high jump or triple jump). The last, third attempt shows a great dispersion of results, which is probably the consequence of the fact that, in the physically less fit examinees, muscle fatigue occurs in the extensors of the lower extremities: m. quadriceps femoris, especially in its basic and m. rectus femoris the most important fibre for jump. However, it must not be overlooked that the muscles of the rear shinbone preceded by m. triceps surae significantly contribute to the manifestation of explosive strength in lower extremities. More physically fit examinees obtained better values even during the third attempt.

Because this is about explosive strength in lower extremities, the dimension structure of explosive strength was explicitly confirmed. This coincides with the whole series of earlier research projects (e.g. Zaciorski, 1975; Kurelić, Momirović, Stojanović, Šturm, Radojević, & Viskić-Štalec, 1975; Metikoš et al., 1982; Metikoš et al., 1989; Bala, 1999; Madić, 2000, and many others), which identified the structure of the latent motor area in the examinees of various characteristics many times before. What makes these findings different from previous ones is the fact that by far the largest part of the latent dimension of explosive strength in lower extremities is seen in the triple jump test. This was somewhat expected since the triple jump itself contains three jumps and thus it can reveal the greatest amount of information on explosive strength. The test is somewhat more demanding than the other two as it requires more coordination, but it is absolutely applicable to the examinees of this age, abilities and acquired skills and knowledge of more complex motor tasks. The results dispersion is also quite satisfactory within the range of hypothetical Gaussian distribution, and therefore, this piece of information completes the afore-mentioned claim.

The next test that provides most information is the standing long jump test, which again finds its practical application among somewhat younger examinees (e.g. Bala, Popović, & Sabo, 2006), while the standing high jump test would probably be carried out on the examinees from the sports where movements of that type are predominant, i.e. basketball and volleyball.

CONCLUSION

The results of this research have proved that all the applied motor tests to have assessed their own object of measurement. It was confirmed that by using more than one reliability criterion based on different determination type, it is possible to assert their reliability is at a very high level. The tests are absolutely applicable and we recommend them for further use in the student population at the Faculty of Sport and Physical Education.

REFERENCES

Bala, G. (1999). *Motoričke dimenzije studenata fizičke kulture* [Motor Dimensions of P. E. Students], Technical Report. Novi Sad: Faculty of Physical Culture.

- Bala, G., & Krneta, Ž. (2006). O metrijskim karakteristikama motoričkih testova za decu [About Metric Characteristics of Some Motor Tests for Children]. In G. Bala (ed.) *Proceedings of Interdisciplinary Scientific Conference with International Participation "Anthropological Status and Physical Activity of Children and Youth"*, 13-20. Novi Sad: Faculty of Sport and Physical Education.
- Bala, G., Popović, B., & Sabo, E. (2006). Istraživanja na predškolskoj deci u Novom Sadu [Examinations of Preschool Children in Novi Sad]. In G. Bala (Ed.) *Physical Activity of Preschool Girls and Boys* (pp. 75-102). Novi Sad: Faculty of Physical Culture.
- Bala, G., Stojanović, M., & Stojanović, M. (2007). *Merenje i definisanje motoričkih sposobnosti dece* [The Measurement and Defining Motor Abilities of Children]. Novi Sad: Fakultet sporta i fizičkog vaspitanja.
- Braungartner, T. A. (2006). Reliability and Error of Measurement, 27-52. In T. Wood, & W. Zhu (Eds.) *Measurement Theory and Practice in Kinesiology*. Champaign, IL: Human Kinetics.
- Fajgelj, S. (2005). *Psihometrija* [Psychometrics]. Beograd: Centar za primenjenu psihologiju.
- Gredelj, M., Metikoš, D., Hošek, A., & Momirović, K. (1975). Model hijerarhijske strukture motoričkih sposobnosti [Model of Hierarchical Structure of Motor Abilities]. *Kineziologija*, 5 (1-2), 7-81.
- Haase, V. (1998). Reliability Generalization: Exploring Variance in Measurement Error Affecting Score. *Educational and Psychological Measurement*, 58, 6-20.
- Kurelić, N., Momirović K., Stojanović M., Šturm J., Radojević, D., & Viskiće-Štalec, N. (1975). *Struktura i razvoj morfoloških i motoričkih dimenzija omladine* [Structure and Development of Morphological and Motor Dimensions of Youth]. Belgrade: Institute for Science Researches of Faculty of Physical Education.
- Madić, D. (2000). *Povezanost antropoloških dimenzija studenata fizičke kulture sa njihovom uspešnošću vežbanja na spravama* [The Relationship Between Anthropological Dimensions and His Successfulness in Exercise on Utilities at PE Students]. Unpublished doctoral dissertation, Novi Sad: Faculty of Physical Culture.
- Metikoš, D., Prot, F., Horvat, V., Kuleš, B., & Hofman, E. (1982). Bazične motoričke sposobnosti ispitanika natprosečnog motoričkog statusa. [Basic Motoric Capacities of Individuals with Above Average Motoric Status]. *Kineziologija*, 14(5), 21-62.
- Metikoš, D., Prot, F., Hofman, E., Pintar, Ž. & Oreb, G. (1989). *Mjerenje bazičnih motoričkih dimenzija sportaša* [Measurement of Basic Motor Abilities of Sportsmen]. Zagreb: Faculty of Kinesiology.
- Momirović, K. (2001). *RTT11G: Programme for Analysing Metric Characteristics of Composite Measurement Instruments Consisting of a Small Number of Replications of a Same Task*. Technical Note. Belgrade: Institute for Criminological and Sociological Researches.
- Momirović, K., Wolf, B., & Popović, D. A. (1999). Uvod u teoriju merenja I. *Interne metrijske karakteristike kompozitnih mernih instrumenata* [Introduction in Theory of Measurement: Internal Metric Characteristics of Composite Measure Instruments]. Priština: Faculty for Physical Culture.
- Tenjović, L. (2002). *Statistika u psihologiji* [Statistics in Psychology]. Belgrade: Faculty of Philosophy.
- Zaciorski, V. M. (1975). *Fizička svojstva sportiste* [Physical characteristics of sportsmen]. Beograd: NIP Partizan.