

Gustav Bala***SOME INTERNAL MEASURING
CHARACTERISTICS OF
COMPOSITE MOTOR TESTS****NEKATERE NOTRANJE MERSKE
ZNAČILNOSTI SESTAVLJENIH
GIBALNIH TESTOV****Abstract**

The article applies some psychometrics concepts and measures of composite measuring instruments are to the kinesiometric (kinesmetric) issues in motor test theory. Measures of internal metric characteristics of seven composite motor tests were analysed on two samples of 117 children and 260 students of the Faculty of Physical Education in Novi Sad, Serbia. The analyses were performed under the classical and Guttman measurement models, and they included item representativeness, reliabilities and internal validities, as well as complete test representativeness, reliability, and homogeneity. It was found that the majority of the analysed motor tests had quite good metric characteristics, but some should be further improved to become measuring instruments of sufficient quality.

Key words: representativeness, reliabilities, internal validities, homogeneity, children, physical education students

***Corresponding author:**

Faculty of Physical Education

University of Novi Sad

Lovčenska 16

21000 Novi Sad

Tel.: +38121 450 188

Fax: +38121 450 199

E-mail: gustavbala@yahoo.com

Povzetek

Članek prenaša nekatere psihometrične koncepte in mere sestavljenega merskega instrumentarija na kineziometrično problematiko teorije gibalnih testov. Na dveh vzorcih 117 otrok in 260 študentov Fakultete za športno vzgojo v Novem Sadu, Srbija, smo analizirali mere notranjih merskih karakteristik sedmih gibalnih testov. Analize so bile izvedene na podlagi klasičnega in Guttmanovega modela merjenja in so vključevale reprezentativnost, zanesljivost in notranjo veljavnost, pa tudi celoten test reprezentativnosti, zanesljivosti in homogenosti. Izkazalo se je, da je imela večina gibalnih testov dobre merske značilnosti, nekatere pa bi bilo potrebno še nekoliko izboljšati, da bi postali dovolj kvalitetni merski instrumentarij.

Ključne besede: reprezentativnost, zanesljivost, notranja veljavnost, homogenost, otroci, študenti športne vzgoje

INTRODUCTION

The researcher or the teacher would like to collect a score for each person in one day, and he would also like to have faith in the reliability of the score. The reliability of motor test evaluation using the test-retest method is impractical: it lasts two or more days, and there are often fewer examinees at the second testing. This problem could be solved with composite motor tests. One of the more demanding problems in kinesometrics (kinesmetric)¹ is the basic construction of composite motor tests.

Motor testing involves assessment of very complex human abilities. The assessment of motor abilities is based on certain manifest indicators (certain motor tasks in motor tests), since motor abilities are, by definition, of latent character (construct) and cannot be (directly) measured (at all). These measurement procedures represent indirect measurement, which means that several indicators (motor instruments – tests) are needed for each motor ability (construct).

The motor tasks in a certain motor test are simple or more complex stimuli, with which certain neurophysiologic – motor processes are provoked in subjects. Human motor abilities cannot be measured or assessed on the basis of one or a small number of motor tasks and merely by registering the reaction to the test situation. This indicated the need to perform testing of a certain motor ability with several motor tests. Motor tests (instruments) never show the same level of measured abilities with equal validity and reliability. Therefore, a motor test should consist of a larger number of motor test items (replications of the same motor task), which usually means performing the same task in the test several times consecutively, without or with a short pause between repetitions. In this way, the so-called composite motor tests are obtained.

The assumption of a composite motor test is that a single latent trait underlies each person's performance in each test item (unidimensionality). This means that all testing items in the motor test measure a single latent trait or ability (e.g., gross-motor function in children). In practice, this assumption of unidimensionality may not be met perfectly because more than one trait, such as motivation and familiarity, or some other motor trait, may be involved in the performance of a motor task. Thus, the practice is to consider the assumption adequately met as long as a dominant factor or component determines the performance of the task. Furthermore, such thinking is based on the assumption that all test items in a motor test measure a single construct, which is equally difficult.

The age factor is reflected by the fact that very young and very old examinees tend to be less consistent in their performance than other age groups. Thus, the number of items has to be increased for older examinees, but this is not possible with children. The reason is that little examinees simply do not want to repeat the same motor task several times successively. For this, a number of items should be smaller, and there are usually three items in the motor tests with informative components (the test to estimate e.g. co-ordination, flexibility, balance, agility, etc), whereas there are only two items in the motor tests with energetic component of longer duration (different types of strength and endurance) (Bala, 1999a; 1999b; Bala, Popović and Stupar, 2002).

¹Kinesometrics is a kinesiological discipline that deals with the measurement problems of all for kinesiological important anthropological characteristics and abilities, which are important for kinesiology. It studies the lawfulness, models and application of the measurement theory methods to measurement, evaluation and estimation of kinesiological phenomena (originators in former Yugoslavia are Konstantin Momirović and Miloš Mraković, 1971). Kinesmetrics is a very similar concept (a measurement theory applicable to the movement sciences (Zhu, 2006a)).

A possible solution to the problem of obtaining information about the motor test scores in the same day of testing and obtaining scores that are valid and reliable is to use an established valid test, with several repetitions of the same motor task. However, the number of repetitions is limited due to physiological and mental reasons. The problem is that 3 repetitions (items) is the absolute minimum that can be considered in kinesiometrics, but this is still insufficient for a realistic analysis of the metric characteristics of the motor tests in psychometrics according to Guttman's theory (domain sampling theory), since a very large number of items is needed in composite psychological measuring instruments (Momirović, Wolf, & Popović, 1999). Due to this, Momirović made the RTT11G programme for analysis of the metric characteristics of motor tests (Momirović, 2001), which computes measures of representativeness, reliability, homogeneity and internal validity of the composite measuring instruments consisting of a small number of repetitions of the same motor task. In fact, all metric characteristics refer to the test scores and measuring situation, but not to the motor measuring instruments (tests), as also indicated by Zhu (2006a).

The entire measurement procedure must of course be standardised, since only the same (or as similar as possible) measurement conditions ensure minimisation of measurement error and the influence of certain other unwanted factors on the measurement results. This is especially important with small children, whose attention for one activity is very short, and whose tempers, motives and interests are often changeable.

With such a composite motor test, there are problems concerning the definition of internal metric characteristics of the motor test as a whole, as well as its constituent items. If all these problems are solved, we can obtain information on the representativeness, reliability, homogeneity and validity of the used items, as well as on the representativeness, reliability, and homogeneity of the total result in the motor test. The validity of the total result in the motor test must be defined on the basis of external metric characteristics (factor and pragmatic validity). The reader can find more information on the comprehensive and detailed aspects of this in formal mathematical language in Momirović, Wolf and Popović (1999).

Internal metric characteristics of the items the motor test is composed of are:

- representativeness – amount of information of a test consisting of a certain number of items, in relation to the amount of information that would be obtained with a large (infinite) number of items;
- reliability – degree to which an item accurately measures what it purports to measure;
- internal validity – degree to which an item measures what it purports to measure.

Internal metric characteristics of the motor test as a whole are:

- representativeness – amount of information on the motor ability, which is the object of test measurement;
- reliability – degree to which a motor test accurately measures what it purports to measure;
- objectivity (inter-tester reliability) – the degree to which different testers can obtain the same scores for the same subjects;
- homogeneity – degree to which a motor test score depends on only one object of measurement (single motor ability);

- informativeness – amount of information emitted by a motor test in relation to the variability which a so-called zero-test would emit (with zero reliability).

External metric characteristics of the motor test are:

- validity – degree to which a motor test measures what it purports to measure; can be categorised as:
 - a) logical (face) validity – condition that is satisfied when the measure obviously involves the ability being measured;
 - b) factor validity – the degree to which scores of a test are related to some recognised standard or criterion (well-known motor factor) that is administered at about the same time;
 - c) pragmatic (predictive) validity – degree to which scores of predictor variables can accurately predict criterion scores (usually established by relating the test results to some motor behaviour).

It is quite important to emphasise that a reliable test for a certain motor ability of adults cannot be used to measure the same ability in small children. Such example is the “Standing broad jump” test, used to measure an adult’s explosive strength of legs. In the case of young children, the main problem in performing this task is connecting the simple movements of arms and legs for take-off. However, this test can be used for measuring co-ordination of young children. Another such test is “Arm tapping”, which estimates an adult’s speed of alternate hand movements (speed frequency of hand movement). In the case of small children, this test should also be used to measure co-ordination. These statements are based on the result of the author’s work with young children over a number of years.

The aim of this paper is to analyse a possibility of applying some psychometric concepts and measures of composite measuring instruments to the motor test theory issues in kinesiometrics (kinesmetrics). For this purpose, the author analysed the motor tests which are most frequently used for children, but which are also used for adults. The current study deals with the problem of assessing internal metric characteristics of motor tests used for measuring children and students of physical education and the evaluation of their motor status, as well as for other special analyses, such as finding relations of this status with some other anthropologic dimensions. As students of physical education are often used as a sample of average athletes, one needs to have valid, reliable, representative and homogenous motor tests. Such tests form a base for kinesiological research, since without valid, reliable and objective data, no data analysis, however excellent, will give results that can be scientifically explained in order to solve the set problems, aims and hypothesis of kinesiological research.

This study was done for the research project “Anthropological status and physical activity of population in Vojvodina”, which was partially supported by the Provincial Secretariat for Science and Technological Development, Autonomous Province of Vojvodina, Republic of Serbia.

METHOD

Participants

The entire sample of participants in this study consisted of two subsamples. The subjects in the first sub sample were 117 children (85 boys and 32 girls) at the ages between 4 and 8 years old (average 5.83 years). The children attended a special training program for the development of

motor behaviour in “Munchkin Sport School” in Novi Sad (Serbia). The sample included all 117 children, because additional analyses of the subgroups of boys and girls showed that there were no significant differences in the reliability of the motor tests, in terms of age or gender, for which reason all the children were analysed together.

The second sub sample of subjects consisted of 260 first and second year male students of the Faculty of Physical Education in Novi Sad (Serbia), 18-22 years of age, who were selected on admission to the Faculty according to their biological, health, motor and psychological development.

Instruments

The motor behaviour of small children has a certain general characteristic, but we are familiar with the names of motor abilities (constructs) used for older children and adults. For this reason, the sample of motor tests for this study is selected based on the model of motor abilities of youth (Kurelić, Momirović, Stojanović, Šturm, Radojević and Viskić-Štaleb, 1975; Gredelj, Metikoš, Hošek and Momirović, 1975), even though this model is not suitable for pre-school children.

A) MECHANISM FOR MOVEMENT STRUCTURING

I Functional co-ordination of primary motor abilities

1) Test “Backwards obstacle course”

B) MECHANISM FOR TONUS AND SYNERGETIC REGULATION

II Frequency of simple movements

2) Test “Arm tapping”

III Flexibility

3) Test “Forward bend”

C) MECHANISM FOR REGULATION OF EXCITATION INTENSITY

IV Explosive strength

4) Test “Standing broad jump”

5) Test “20 m dash”

D) MECHANISM FOR REGULATION OF EXCITATION DURATION

V Repetitive strength of the trunk

6) Test “Crossed-arm sit-ups”

VI Static strength of arms

7) Test “Bent arm hang”.

The battery of seven motor tests was administered with the next test administration procedures (a short description of the motor tests follows):

1. *Backwards obstacle course.* The examinee has to walk backwards on all fours and cover the distance of 10m, climb the top of Swedish bench and go through the frame of the bench. The task is measured in tenths of a second and is repeated three times, with an appropriate rest time in between. The “Backwards obstacle course” test was administered only to the sub sample of children.

2. *Arm tapping*. For fifteen seconds the examinee has to tap alternately two plates on the tapping board with his dominant hand, while holding the other hand in between the two plates. The result is the number of alternate double hits. The task is performed three times, with an appropriate rest in between.
3. *Forward bend*. The examinee stands on a bench and bows as deep as possible. A straight-angle ruler which points down with the 40 cm mark at the child's feet, and 40 cm below it, is next to him/her. The result is the depth of the reach measured in cm. The task is performed three times without any rest.
4. *Standing broad jump*. The examinee jumps with both feet from the reversed side of Reuter bounce board onto a carpet which is marked in cm. The result is the length of the jump in cm. The task is performed three times without rest.
5. *Crossed-arm sit-ups*. The examinee lies on his/her back with his/knees bent and arms crossed on opposite shoulders and raises into a seated position and returns into the starting position. The instructor's assistant holds the child's feet. The result is the number of correctly executed raises to the seated position (no longer than 60 seconds). The children repeated the task twice and the students three times, with a rest in between.
6. *Bent arm hang*. The examinee under-grips the bar and holds the pull-up as long as he/she can (chin above the bar). The result is the time of the hold measured in tens of a second. The children repeated the task twice and the students three times, with a rest in between.
7. *20 m dash*. A pair of examinees run 20 meters from standing position. The task is measured in tens of a second and is repeated three times, with an appropriate rest time in between.

These motor tests are routinely performed on older children and youth, but they need to be modified before they can be applied to young children. The modifications occurred in the following tests:

1. *Arm tapping*. The height of the table and the chair is adjusted for comfortable seating of young children (if the seat is not made for children, this is accomplished by putting an additional layer on the chair as well as under the child's feet). The inside distance between the spaces that hands can touch is 50 cm (61 cm for adults), although this distance should be even smaller.
2. *Forward bend*. The children perform a deep bow beside a straight-angle ruler, so that it does not hurt them.
3. *Bent arm hang*. The bar has a smaller radius than the regular one, so the child can grip the bar firmly when holding a pull-up.

Procedures

In this study, the RTT11G programme for analysis of the metric characteristics of motor tests (Momirović, 2001) was used, which computes measures of representativeness, reliability, homogeneity and internal validity of the composite measuring instruments consisting of a small number of replications of the same motor task. The definitions and formal mathematical presentation of the measures implemented in this programme can be found in Momirović, Wolf and Popović (1999). From the measures available in the programme, the following estimators of metric characteristics are used and showed in the tables:

A) For the item variables of the composite motor test:

- MSA – Kaiser-Rice measure of item representativeness:

$$MSA_J = 1 - \left(\sum_{k=1}^m a_{jk}^2 - u_j^2 \right) / \left(\sum_{k=1}^m r_{jk}^2 - 1 \right)$$

where a_{jk} are elements of the matrix $\mathbf{A} = \mathbf{U}^2 \mathbf{R}^{-1} \mathbf{U}^2$,

r_{jk} – elements of the item correlation matrix \mathbf{R} ,

u_j^2 – elements of the unique variance matrix of items, $\mathbf{U}^2 = (\text{diag} \mathbf{R}^{-1})^{-1}$.

- SMC – squared multiple correlation of the item with the other items in the test;
- H – internal validity of items based on Hotelling's principal component method:

$$H = \mathbf{x} \lambda^{1/2}$$

where \mathbf{x} – first eigenvector of item correlation matrix \mathbf{R} ,

λ – first eigenvalue of \mathbf{R} .

- B – internal validity of items based on Burt's simple summation method:

$$B = \mathbf{R} \mathbf{e} \sigma^{-1}$$

$$\sqrt{\sigma^2} = \sqrt{\sum_{j=1}^m \sum_{k=1}^m r_{jk} / m}$$

where σ is standard deviation of scores of items =

\mathbf{R} – item correlation matrix,

\mathbf{e} – identity vector (summation vector of order (m,1)),

m – number of items.

B) For the complete composite motor test:

- COMVAR – common variance (sum of squared multiple correlations – SMC – each item with all others of the motor test):
- ψ_2 – Kaiser-Rice measure of representativeness:

$$\psi_2 = 1 - (\mathbf{e}^t ((\mathbf{B} - \mathbf{I}) \circ (\mathbf{B} - \mathbf{I})) \mathbf{e}) (\mathbf{e}^t ((\mathbf{R} - \mathbf{I}) \circ (\mathbf{R} - \mathbf{I})) \mathbf{e})^{-1}$$

where \mathbf{e} – identity vector (summation vector of order (m,1)),

$\mathbf{B} = \mathbf{U} \mathbf{R}^{-1} \mathbf{U}$ – correlation matrix of antiimage items of the motor test,

\mathbf{I} – identity matrix,

\circ – symbol for Hadamard product of vectors,

\mathbf{R} – matrix of item intercorrelations.

- α – Spearman-Brown-Kuder-Richardson-Guttman-Cronbach coefficient of reliability under the classical summation model:

$$\alpha = (m / (m - 1)) (1 - \sum_{j=1}^m \sigma_j^2 / \sigma^2)$$

where m – number of items,

$\sigma_j^2, j = 1, \dots, m$ – variance of items,

σ^2 – variance of total score.

- λ_7 – assessed Momirović’s coefficient of reliability under the classical summation model:

$$\lambda_7 = 1 - \sigma^{-2} + m^{-1}$$

where σ^2 – variance for the sum across all items, and
 m – number of items.

- β – Lord-Kaiser-Caffrey measure of reliability of the first principal component;

$$\beta = m / (m - 1)(1 - \lambda^{-1})$$

where m – number of items,
 λ – first eigenvalue of item correlation matrix (\mathbf{R}).

- β_7 – assessed Momirović’s lower bound of reliability of the first principal component:

$$\beta_7 = 1 - \lambda^{-1} + m^{-1}$$

where m is a number of items, and
 λ – the largest eigenvalue of item correlation matrix \mathbf{R} .

- ρ – Guttman-Nicewander coefficient of reliability under the Guttman’s model of measuring based on the first Harris component:

$$\rho = \mathbf{x}^t (\mathbf{R} - \mathbf{U}^2) \mathbf{x} / \mathbf{x}^t \mathbf{R} \mathbf{x},$$

where \mathbf{x} is $(m,1)$ vector chosen under the condition $\mathbf{x}^t \mathbf{R} \mathbf{x} = 1$ maximize the reliability coefficient ρ ,
 \mathbf{R} – item correlation matrix,
 $\mathbf{U}^2 = (\text{diag } \mathbf{R}^{-1})^{-1}$ – diagonal matrix which contents error variances of measurement of the test items.

- H_1 – average correlation of the variables, as a measure of test homogeneity:

$$h_j = \left(\left(\sum_{k=1}^m r_{jk} \right) - 1 \right) / (m - 1).$$

for $j=1, \dots, m$, where r_{jk} are the elements of matrix \mathbf{R} .

- H_2 – Momirović measure of test homogeneity:

$$h_2 = \delta^2 / c$$

where $\delta^2 = \mathbf{y}^t \mathbf{G} \mathbf{y}$; \mathbf{y} is eigenvector refers to the largest eigenvalue of the characteristic equation of the matrix δ^2 , and

$$\mathbf{G} = \mathbf{R} + \mathbf{U}^2 \mathbf{R}^{-1} \mathbf{U}^2 - 2\mathbf{U}^2$$

$$c = \text{trace } G = \text{trace} (\mathbf{I} - \mathbf{U}^2) = m - \sum_{j=1}^m u_j^2;$$

where $u_j^2, j=1, \dots, m$ unique variance of items, and $\mathbf{U}^2 = (\text{diag } \mathbf{R}^{-1})^{-1}$.

All explanations, comments and formal mathematical language of these, and many more concepts and measures can be found in the textbook “Introduction to the theory of measurement and internal metric characteristics of composite measuring instruments” (Momirović, Wolf, Popović, 1999), but, unfortunately, only in the Serbian language.

RESULTS

In this study, it was assumed that the chosen motor tests have good validity, which has already been proved in several studies (ex. Kurelić et al., 1975; Gredelj et al., 1975; Metikoš et al., 1989; Bala, 1981, 1999a; Madić, 2000 and others). Therefore attention was paid to:

- a) representativeness, reliability and internal validity of the items of motor tests (Table 1 and Table 2);
- b) common variance, representativeness, reliability, and homogeneity of the complete motor test (Table 3 and Table 4).

Table 1: Representativeness, reliability and internal validity of each item in a motor test in children

ITEM VARIABLE	Representativeness			Reliability			Internal validity					
	MSA			SMC			Hotelling (H)			Burt (B)		
	1	2	3	1	2	3	1	2	3	1	2	3
Obs. cour. back.	.997	.996	.997	.897	.927	.854	.971	.983	.963	.971	.983	.963
Arm tapping	.995	.993	.994	.831	.874	.855	.960	.971	.965	.960	.971	.965
Forward bend	.998	.997	.997	.826	.929	.921	.957	.981	.977	.958	.981	.971
St. broad jump	.992	.988	.993	.812	.867	.799	.952	.970	.949	.952	.970	.949
Sit-ups	.961	.961		.802	.802		.973	.973		.973	.973	
Bent arm hang	.901	.901		.686	.686		.956	.956		.956	.956	
20m dash	.997	.996	.996	.879	.902	.899	.972	.977	.976	.972	.977	.976

Table 2: Representativeness, reliability and internal validity of each item in a motor test in students

ITEM VARIABLE	Representativeness			Reliability			Internal validity					
	MSA			SMC			Hotelling (H)			Burt (B)		
	1	2	3	1	2	3	1	2	3	1	2	3
Arm tapping	.919	.910	.938	.559	.586	.460	.887	.900	.851	.885	.897	.856
Forward bend	.999	.999	.999	.972	.986	.977	.993	.997	.994	.993	.997	.994
St. broad jump.	.986	.989	.988	.810	.774	.792	.956	.946	.951	.956	.946	.951
Sit-ups	.996	.994	.992	.742	.873	.888	.935	.964	.972	.937	.963	.971
Bent arm hang	.989	.976	.973	.168	.827	.830	.636	.938	.945	.709	.906	.916
20m dash	.920	.892	.915	.466	.563	.491	.854	.896	.863	.857	.893	.864

Table 3: Common variance, representativeness, reliability and homogeneity of the motor tests in children

TEST VARIABLE	Comvar	Repres		Reliability			Homogeneity		
	%	ψ^2	α	λ^7	β	β^7	ρ	H1	H2
Obs. cour. back.	89.30	.997	.971	.981	.971	.981	.996	.919	.986
Arm tapping	85.38	.994	.963	.975	.963	.975	.992	.899	.982
Forward bend	89.15	.997	.970	.980	.970	.980	.997	.917	.988
St. broad jump.	82.65	.991	.954	.969	.954	.969	.990	.875	.978
Sit-ups	80.27	.961	.945	.972	.945	.972	.979	.895	.948
Bent arm hang	68.62	.901	.906	.953	.906	.953	.946	.828	.914
20m dash	89.37	.997	.974	.983	.974	.983	.996	.927	.987

Table 4: Common variance, representativeness, reliability and homogeneity of the motor tests in students

TEST VARIABLE	Comvar	Repres		Reliability			Homogeneity		
	%	ψ^2	α	λ^7	β	β^7	ρ	H1	H2
Arm tapping	53.50	.921	.853	.902	.853	.902	.911	.659	.931
Forward bend	97.81	.999	.994	.996	.994	.996	.999	.983	.997
St. broad jump.	79.19	.987	.947	.965	.947	.965	.984	.856	.974
Sit-ups	83.45	.993	.954	.969	.954	.969	.993	.873	.981
Bent arm hang	60.85	.976	.798	.865	.811	.874	.985	.568	.959
20m dash	50.67	.908	.841	.894	.841	.894	.896	.638	.925

DISCUSSION

Representativeness of items, computed by the Kaiser-Rice procedure on the basis of Guttman's theory of representativeness of variable sampling, is an important metric characteristic of a composite motor test. The magnitudes of coefficients of representativeness (MSA = Measure of Sampling Adequacy) in Table 1 and Table 2 show extremely high values for all measured tests. This shows that each item is representative, i.e. carries a sufficient amount of information about the motor ability assessed by the relevant motor test. It is interesting to note that, practically, the first item proves to be the most representative in all tests, which is usually not the case when assessing reliability, where the second item is usually the best. The coefficient in the test "20m dash – second item" in students is somewhat lower but still fully satisfies the usual kinesiological criteria. It is therefore possible to conclude that all the items have adequate representatives of the relevant motor tests.

The reliability of the items of the used motor test significantly differs from the reliability of the complete motor test, i.e. the values are usually significantly lower. They are Squared Multiple Correlations (SMC) of one item with the others, where we can obtain complete test reliability in different ways and this is considered adequate if the coefficient is at least 0.90. It is therefore difficult and not very wise to try to explain the coefficients of item reliability of a certain motor test, without having information on complete test reliability. However, the presented coefficients

of reliability show very low item reliability of the following motor tests: “Bent arm hang” with the children, and “Arm tapping”, as well as “20m dash” with students. These tests merit greater attention in further analyses since their item reliability is much lower than that obtained for the other motor tests. It seems that the lower reliability is mostly the characteristic of the tests which are energetically demanding, even if their duration is not very long. This is an already known characteristic of “endurance tests”, especially in children, but it is rather surprising for tests such as the 20m dash, especially due to the fact that the subjects (students of PE) are supposed to be physically fit and enough time for recuperation was given between the repetitions.

The reliability coefficients of the items show that it is mostly the second repetition that is the most reliable. Therefore, if one result per test is planned, which is of course not the best strategy since combinations of items (sum or projection on the first principal component) give more reliable data, it is best to measure each test twice and take the second result as the valid one.

The information on the internal validity of the items in the used motor tests, based on Hotelling’s method of principal components (H) and Burt’s method of simple summation (B), is presented in Table 1 and Table 2. The obtained coefficients are very high for this kind of measurement instruments (motor tests). According to Burt’s method, the main object of measurement of a motor test and validity of the items of that test is defined under the assumption that all items are oriented in the same direction and that the total result in this test is the sum of the results in the individual items; therefore, the correlations between the so-defined items and the main object of measurement represent coefficients of validity of these items. According to Hotelling’s method, these coefficients are a correlation between the item results and the first principal component of standardised results in these items. It can be seen that the least valid items (which does not automatically mean bad) are included in the “20m dash” and “Arm tapping” tests in students, which was also the case when analyzing their item representativeness. It was almost always the first and then the third item with the lowest validity; therefore, here too the second item seems best. However, even if there were differences between the items, it can be said that practically all items were valid for the relevant motor test. This holds true for classical measurement theory, since the analysis was performed with these assumptions.

Analysis of metric characteristics of the complete motor tests (Table 3 and 4) includes explanation of the obtained information on the real (usable) values of each test for practical applications. By assessing the common variance of each test on the basis of its constituents (items), it can be seen that special care should be taken when using the following motor tests: “Bent arm hang” in children, but “20m dash”, “Arm tapping”, and “Bent arm hang” in students. These tests, being again more energy demanding ones, can be suspected of having a tendency of having the most frequent and largest measurement errors as well as specificities, in comparison with the other analysed tests.

The Kaiser-Rice measure of representativeness (ψ^2 in Table 3 and 4) is extremely high for all the analysed motor tests; this shows that they are good representatives for assessing the relevant hypothetical motor abilities.

The reliability of the motor tests in assessing the relevant motor ability was analysed on the basis of the classical model of measurement (Spearman-Brown-Kuder-Richardson-Guttman-Cronbach coefficient of reliability, better known as Cronbach α and λ_7 – assessed Guttman’s coefficient for determining the lower bound of reliability), as well as on the basis of Guttman’s model of measurement (β – Lord-Kaiser-Caffrey measure of reliability of the first principal component and

β_7 – assessed Momirović's lower bound of reliability and ρ – Guttman-Nicewander's coefficient of reliability). The first group are coefficients of reliability, which assume that the result in the relevant test also contains error variance; therefore, it can be concluded that the following tests have low reliability: "Bent arm hang", "20m dash", "Arm tapping" in students.

In case of the students, the same motor tests also show lower validity when analysed with the Lord-Kaiser-Caffrey coefficient of reliability (β), but not in the other coefficients, in which the result does not contain error variance: Guttman's model of measurement. It can be seen that only the "20m dash" has really low reliability according to all the used methods of assessing validity, while the others have quite good reliability according to the coefficients β_7 (assessment of Momirović's lower bound of reliability of the first principal component) and ρ (Guttman-Nicewander's coefficient of reliability of the first Harris's component).

On the basis of all types of reliability coefficients, it is possible to conclude that most of the analysed composite motor tests have quite good reliability; however, for the motor tests with lower values in reliability coefficients in the analysed sample it would make sense to increase the number of items, or use for further analysis the results from which error variance has been removed.

Since homogeneity of a motor test is dependent on the result in this test on a single object of measurement, it is a very important metric characteristic. With a homogenous test, it is very easy to assess the relevant relatively "clean" or unique motor ability. However, with such a test it is difficult, usually impossible, to assess some complex motor ability, which is often the case in kinesiological research. Two of the existing measures of homogeneity of a complete composite motor test were used in the current study: 1) average correlation between all items (H1) and 2) on the basis of relative variance of the first principal component of the items transformed into image form (Momirović's measure of homogeneity – H2) (Table 3 and 4).

On the basis of average correlation between the motor test items, it is possible to notice that the lowest coefficients belong to the tests which also had the lowest measures of reliability and representativeness. The obtained coefficients are relatively low, since they depend on error variance, and are therefore not a very good basis for assessing homogeneity of motor tests. Momirović's measure of test homogeneity (H2) is independent of its reliability, since it is defined only by the proportion of total variance of the real test result that is explained by the test's main object of measurement. So-defined coefficients of homogeneity show a different picture. It is significant that they are all above 0.91. Therefore, when measurement error is removed from the variance of the total test result, very high values of homogeneity coefficients are obtained, but still the lowest values tend to be obtained for those tests that also have the lowest average correlations, measures of representativeness and reliability, both in terms of items as well as of the complete tests.

On the basis of the obtained results, it is possible to conclude that further work is necessary on test construction as well as on their application to various subject samples for the following motor tests:

1. Arm tapping, to evaluate frequency of simple movements;
2. 20m dash, to evaluate explosive strength at the students, as well as co-ordination in the children, and
3. Bent arm hang, to evaluate general strength.

A relatively easy way to improve the internal metric characteristics would be by increasing the number of items in the composite motor tests. However, in this case certain other problems can emerge, which are not within the scope of this work (ex. transfer of motor learning on succeeding test items, physiological and mental fatigue, decrease of motivation for repeating the same test, etc). It is, therefore, much easier to improve the data quality by removing error variance, either by using factor scores (projection on the first principal component of the items) or by transforming the data to some other metric (image, universal metric). It is also very important to fully observe the standardised performance of each test in order to decrease as much as possible the possible sources of error in the execution and measurement phases of each item (testing conditions, requisites, measuring instruments, subject, measurer, entry of the measured data on the data sheet, etc).

Other motor tests have sufficiently good internal metric characteristics and can be recommended as measuring instruments for assessing those motor abilities in kinesiological research with subjects of similar characteristics. The most important thing is that the subjects completely observe the requirements set by the standardization of the tests, which was the case in the current study.

Finally, a word of warning: this analysis of metric characteristics of motor tests was performed on data obtained in children and students of physical education. These students are as a rule well co-ordinated and in good physical condition; therefore, motor learning effects and degradation due to fatigue should be minimal. However, it is obvious from the results that it is precisely these tests that have the worst metric characteristics. If this is the case for such subjects, these effects can be expected to be much stronger for the general (adult) population or children and especially for those with below-average motor efficiency.

External validity was not analyzed in this study, but care must be taken in this regard. Namely, it is possible, for example, that a test assumed to measure explosive power of legs (Standing broad jump) really does measure that motor ability on well co-ordinated subjects, who have no problem co-ordinating the action of the arms and legs at take-off and judging how much they can push legs forward without falling back on landing. However, with less co-ordinated subjects or small children, it might just as well measure co-ordination more than explosive power. It is, therefore, quite possible that the same test measures different motor abilities – or at least the same ability to a different degree – on different subjects.

This means at least two things: first, metric characteristics of the same tests should be established for each different type of a subject sample, and second, it is quite possible that a certain motor test can be an adequate measuring instrument for some subjects and completely inadequate for others. Care must therefore be taken not to use motor tests indiscriminately.

In this paper, the author wanted to show the possibility of applying some psychometric concepts and measures of composite measuring instruments to the kinesiometric (kinesmetric) issues in motor test theory. This theory is based mostly on the classical test theory, which is the most commonly used test theory in kinesiology.

It is obvious that all coefficients of practically all metric characteristics are too high, i.e. higher than we can expect in psychometrics analysis of some psychological tests. This situation could point to a different nature of the items in motor and in psychological test, which means that the composite measuring instruments in motor space should be treated in a different way, and that

the psychometric methods for such composite motor tests are “soft” for the kinesiometric issues. The author recommends further research to be based mostly on the Guttman’s test theory, as well as on the methods explained in the papers written by Zhu (2006a; 2006b), Zhu and Cole (1996), Zhu, Timm, and Ainsworth (2001), which deal with the problems based on a relatively new test theory called item response theory (IRT) and Rasch’s model in educational measurement practice and kinesiology.

REFERENCES

- Bala, G. (1981). *Struktura i razvoj morfoloških i motoričkih dimenzija dece SAP Vojvodine* [The structure and development of morphological and motor dimensions of children in SAP Vojvodina]. Novi Sad: Faculty of Physical Education.
- Bala, G. (1999a). Motor behavior evaluation of pre-school children on the basis of different result registration procedures of motor test performance. In V. Strojnik, & A. Ušaj (Eds.), *Proceedings of the 6th Sport Kinetics Conference '99. Theories of Human Motor Performance and their Reflections in Practice* (pp. 62-65). Ljubljana: University of Ljubljana, Faculty of Sport.
- Bala, G. (1999b). Some problems and suggestions in measuring motor behavior of pre-school children. *Kinesiologia Slovenica*, 5(1-2), 5-10.
- Bala, G. (2002). *Sportska školica* [Sport School for Children]. Novi Sad: Kinesis.
- Bala, G., Popović, B., & Stupar, D. (2002). Pouzdanost nekih kompozitnih testova za procenu motoričkog ponašanja predškolske dece. [Reliability of some composite tests for evaluation of pre-school children’s motor behaviour]. *Zbornik sažetaka Deseti međunarodni interdisciplinarni simpozijum “Sport, fizička aktivnost i zdravlje mladih”* (pp. 85-86). Novi Sad: Novosadski maraton.
- Gredelj M., Metikoš D., Hošek A., & Momirović K. (1975). Model hijerarhijske strukture motoričkih sposobnosti. 1. Rezultati dobijeni primjenom jednog neoklasičnog postupka za procenu latentnih dimenzija. [Model of a hierarchic structure of motor abilities. 1. The results obtained using a neo-classical method for estimating latent dimensions]. *Kineziologija*, 5(1-2), 7-81.
- Kurelić N., Momirović K., Stojanović M., Šturm J., Radojević Đ., & Viskić-Štalec, N. (1975). *Struktura i razvoj morfoloških i motoričkih dimenzija omladine* [Structure and development of morphologic and motor dimensions of youth]. Belgrade: Institute for Research of the Faculty of Physical Education.
- Madić, D. (2000). *Povezanost antropoloških dimenzija studenata fizičke kulture sa njihovom uspešnošću vežbanja na spravama* [The relationship between anthropological dimensions of physical education students and successful exercising on gymnastic apparatuses]. Doctoral dissertation, Novi Sad (Yugoslavia): Faculty of Physical Education.
- Metikoš, D., Prot, F., Hofman, E., Pintar, Ž., & Oreb, G. (1989). *Mjerenje bazičnih motoričkih dimenzija sportaša* [Measurement of basic motor abilities of athletes]. Zagreb: Fakultet za fizičku kulturu.
- Momirović, K. (2001). *RTT11G: Program za analizu metrijskih karakteristika kompozitnih mernih instrumenata koji se sastoje od malog broja replikacija istog zadatka* [RTT11G: Programme for analysing metric characteristics of composite measurement instruments consisting of a small number of replications of the same task]. Technical note. Belgrade: Institute for criminological and sociological research.
- Momirović, K., Wolf, B., & Popović, D. (1999). *Uvod u teoriju merenja i interne metrijske karakteristike kompozitnih mernih instrumenata* [Introduction to the theory of measurement and internal metric characteristics of composite measuring instruments]. Priština: Fakultet za fizičku kulturu.

Mraković, M. (1971). Kineziologija [Kinesiology]. *Kineziologija*, 1(1), 1-5.

Mraković, M. (1992). *Uvod u sistematsku kineziologiju*. [Introduction to systematic kinesiology]. Zagreb: Faculty of Physical Education.

Zhu, W. (2006a). Constructing Tests using Item Response Theory. In T. Woods & W. Zhu (Eds.), *Measurement theory and practice in kinesiology* (pp. 53-76). Champaign, IL.: Human Kinetics.

Zhu, W. (2006b). Scaling, equating, and linking to make measures interpretable. In T. Woods & W. Zhu (Eds.), *Measurement theory and practice in kinesiology*, 93-111. Champaign, IL.: Human Kinetics.

Zhu, W., & Cole, L.E. (1996). Many faceted Rasch calibration of a gross motor instrument. *Research Quarterly for Exercise and Sport*, 67(1), 24-34.

Zhu, W., Timm, G., & Ainsworth, B. (2001). Rasch calibration and optimal categorization of an instrument measuring women's exercise perseverance and barriers. *Research Quarterly for Exercise and Sport*, 72(2), 104-116.