

RELIABILITY OF V SIT-AND-REACH TEST USED FOR FLEXIBILITY SELF-ASSESSMENT IN FEMALES

Roman Cuberek*, Iva Machová, Michaela Lipenská

Faculty of Physical Culture, Palacký University, Olomouc, Czech Republic

Submitted in February, 2013

BACKGROUND: The V sit-and-reach (VSR) test seems to be an appropriate instrument of self-assessment of hamstring and low-back flexibility for its ease of execution and the need for only a small amount of materials, space, and examination skill requirements. It is assumed that the specificity of self-assessment (in general) can be the cause of other sources of measurement error.

OBJECTIVE: This study aimed to analyze reliability of VSR when used as an instrument of self-assessment of flexibility in adolescent females.

METHODS: The sample comprised 43 students (female; age 21.2 ± 0.5 years) from Palacký University in Olomouc (Czech Republic). *T*-test ($p < .05$) and Pearson correlation coefficient was used to assess systematic bias and to determine intra-individual reliability of the flexibility test; the standard error of measurement (*SEM*) and Bland and Altman's 95% limits of agreement were used to assess absolute reliability of the flexibility test.

RESULTS: The average intra-individual difference of 1.14 cm (increasing test performance) was found to be statistically significant ($t = -5.375$; $df = 42$). It was observed that high intra-individual reliability ($r = .98$); the absolute reliability (*SEM*) is equal to 0.139 cm.

CONCLUSIONS: This study provides evidence supporting the usage of VSR as a relevant instrument of self-assessment of hamstring and low-back flexibility in adolescent females.

Keywords: Health-related fitness, measurement error, testing, university student.

INTRODUCTION

The sit and reach test (SR – originally named *Canadian Trunk Forward Flexion Test*) is a field test used to measure hamstring and low-back flexibility (Baumgartner & Jackson, 1995; Wells & Dillon, 1952). It is believed that maintaining a good level of flexibility in these areas is an important part of health related fitness (HRF) (Martin, Jackson, Morrow, & Liemohn, 1998), because it prevents risk of falling, gait limitations or postural deviations and the most acute or chronic musculoskeletal injuries and lower back problems (ACSM, 2000). SR or some of its variations (for example *Unilateral seated sit-and-reach test*, *Chair sit-and-reach test*, *Modified sit-and-reach test*, *Back saver sit-and-reach test*, *Toe-touch test* and *V sit-and-reach test*) are commonly used in health related physical fitness test batteries (for example EUROFIT, FITNESSGRAM, President's Challenge, Unifittest (6–60) and so on). The choice of the test to be employed is more often based on the examiner's prefer-

ence, ease of use, professional discipline, or tradition, rather than scientific evidence (Davis, Quinn, Whiteman, Williams, & Young, 2008). The reason probably comes from the fact that there is still no convincing proof or conclusively provided evidence of which test is the most appropriate to use to assess hamstring and low-back flexibility (Ayala, Sainz de Baranda, de Ste Croix, & Santonja, 2012; Baltaci, Un, Tunay, Besler, & Gerçeker, 2002; Hui & Yuen, 2000; López-Miñarro, de Baranda, & Rodríguez-García, 2009). Taking into account this fact as well as tests' weaknesses and advantages, the choice is mostly related to conditions and situations in which the test is realized.

Although SR and its alternatives are frequently used due to ease of use, understandable procedures, and minimal skills requirements, they are usually unsuitable for self-assessment because of low availability of their material needs (e.g. "test box"). For the purpose of self-assessment the *V sit-and-reach test* (VSR) seems to be the most appropriate, because it requires only an adhesive tape and a measurement tool.

Some sources of measurement errors can be expected while HRF tests are realized as self-assessment in comparison when testing process is performed by an experienced or specially trained examiner. Therefore, if

* Address for correspondence: Roman Cuberek, Institute of Active Lifestyle, Faculty of Physical Culture, Palacký University, tř. Miru 115, 771 11 Olomouc, Czech Republic. E-mail: roman.cuberek@upol.cz

a validity of VSR (performed as self-assessment) is suggested from the content validity of the classical VSR (with an examiner) then some decrease of the validity might be considered as a result of the influence of measurement error related to the self-assessment process.

The purpose of the study was to examine reliability of the V sit and reach test used for the self-assessment of hamstring and low-back flexibility in females as one of the components of health related fitness.

METHODS

Participants

Forty three female students (age 21.2 ± 0.5 year, BMI 22.61 ± 3.10 kg·m⁻²) from the Faculty of Education of Palacký University in Olomouc volunteered to participate in this study, which is in accordance with recommended number of a minimum of forty persons for reliability studies (Atkinson & Nevill, 1998). All subjects were informed about the purpose of the study and all procedures in this study before the testing. No female was obese, or a professional athlete and none had a physical impairment.

The project in which this study was realized, was approved by the Ethics Committee of the Faculty of Physical Culture of Palacký University in Olomouc (document number 65/2011). Participants could leave the study at any time during the practical component. All participants signed a statement of informed consent. All data of the study was used only for the purpose of the study.

Data procedure

Testing took place in the athletic sport hall in Olomouc, Czech Republic. All participants performed an eight minute warm up and static stretching routine once they had received the test card with the procedure of modified VSR test and a scoring sheet. On the scoring sheet was the place for measured results (in cm) and subjects had to enter basic information such as date of birth and their height and weight (for calculation of BMI index). Subjects were instructed to perform testing alone without any additional information except that given on the test card. All testing tools were free to be used and available. Description and information about the VSR test was identical with the original VSR modification as presented on the internet – INDARES.com. The test card has a verbal description, two pictures with the starting and reaching position of VSR together with the description of scoring the result. Subjects were instructed to perform two trials with a rest of 10–15 minutes between.

The instructions on the test card (originally written in Czech language) – the subject removes their shoes and sits on the floor with the measuring line between their legs, the soles of their feet are placed immediate-

ly behind the baseline, heels about 20 cm apart. The thumbs are clasped so that hands are together, palms facing down and placed on the measuring line. With the legs held flat, the subject slowly reaches forward as far as possible, keeping the fingers on the baseline and feet flexed. Movement should be smooth and without bouncing. Here, the subject stays for 2 s. The test is done twice with a short break in between as stated above.

Scoring: Zero point is at the level of feet. (We note negative values towards our body and positive values outward from our body.) The best trial is recorded in centimeters.

Statistical analysis

A level of participants' flexibility was assessed by basic statistics (mean, median, standard deviation, 95% confidence interval) and compared to population norms implemented in INDARES.com software. After homoscedasticity (using the Pearson product-moment correlation coefficient – the correlation between the absolute differences of two repeated measurements and the mean of two repeated measurements) and normality (*Shapiro-Wilks W test*, $p < 0.05$) were verified, two tailed *t*-test of the repeated measurements ($p < 0.05$) and Cohen's effect size *d* were used to detect and assess systematic bias. Subsequently the systematic bias was considered as significant only in case when null hypothesis (neither difference between repeated measurements; *t*-test) was rejected and at the same large effect size (classification according to Cohen (1988)) was detected by Cohen's *d*. The Pearson product-moment correlation coefficient (test – re-test) as parameter of intra-individual reliability was determined to assess the intra-individual reliability. The standard error of measurement (*SEM*) and Bland and Altman's 95% limits of agreement for two repeated measurements were used to express and assess the reliability in the original measurement unit, e.g. absolute reliability (Atkinson & Nevill, 1998). To calculate the *SEM*, the following formula was used (Thomas, Nelson, & Silverman, 2005): $SEM = SD \cdot \sqrt{1-r}$, where *SD* is the standard deviation of the intra-individual differences and *r* is the determined correlation coefficient. The statistical software package SPSS 17.0 (SPSS Inc., Chicago, IL) was used to compute all of the statistical characteristics.

RESULTS

The data of the first and the second measurement are normal according to the normality of *Shapiro-Wilkins W test*, $p < .05$ ($W = 0.961$, $p = .150$ for first, and $W = 0.995$, $p = .038$ for second measurement). The low value of Pearson product-moment correlation coefficient ($r = -.135$) shows at the homoscedasticity of data

and this signifies no risk of increased measurement error along with increasing test results.

The flexibility performances of female students are listed in Table 1 together with the basic statistical differences describing repetition of the test (intra-individual variability of performance).

It was found, that a difference -1.14 cm between the first and the second measurement (shows an increase of performance) is statistically significant on $p < 0.05$ ($t = -5.375$; $df = 42$). This is in contradiction to the low effect size (Cohen, 1988) found by $d = 0.16$. High value of Pearson product-moment correlation coefficient ($r = .980$) suggests a high level of intra-individual reliability. Estimated standard error of measurement (SEM) is 0.139 cm, where for two repeated measurements are Bland and Altman's 95% limits of agreement among -1.3 cm to 4.1 cm. The SEM represents only 7.19% of the unit of measurement in the test which can be considered as practically insignificant.

DISCUSSION

Reliability is a theoretical concept used for a description of the quality of the measuring instruments and procedures. Researchers and practitioners need to know the level of reliability (as same as validity) to justify their choice of an adequate measurement procedure employed in data collection during specific activities. Knowledge of different sources of errors and level of their influence help researchers and practitioners to better understand, to interpret and to study particular phenomenon.

The purpose of this study was to evaluate errors of the field test (*V sit-and-reach*), which was realized as a self-assessment test. This meant that an individual was the examiner and subject, as the subjects had to measure and record on their own. Logically, this process can cause increasing influence of various sources of measurement error. During the self-assessment, the typical sources of measurement error can be the process of learning. It can be understood as a process to learn the move (that particular practical task) or to learn how to operate the test. This can cause an in-

crease of the test performance. The next natural source of measurement error is a sequence of tenses of tasks during the test. There can be some moment during testing when the subject (and examiner in one) cannot process all necessary tasks at the same time even as is required by the test itself. This results in a latency of subtasks and overall increase of the test duration. This disturbing effect is more important in those tests where the score is related to the time (time as a result or served as an interval of the test duration). Another source of measurement error is an effect of a wrong score reading from the measurement tool, which is used during the testing. Mostly this is determined by conflict between the final position of the subject and the location of a measurement tool (set by the test procedure) which enables the subject to read the test score properly. The true score can be also influenced by a subjective attitude of the subject to the recorded result – subject intentionally shifts the score into his/her performance-benefits (rounds off the score into the performance-positive values or simply shifts the score). A partial source of measurement error can be caused by an experience with a process of testing. Some previous experience with testing (in general) of individual can lead to a form of anticipation which reduces potential of measurement error.

The sample character allow us to consider our results as representative for a population of non-obese females in age of 20–30 years with good or good-to-moderate level of flexibility of lower back and hamstrings who have a normal amount of daily physical activities.

The findings based on participants' test performances in a range of 0–20 cm show that the magnitude of measurement error in VSR does not depend on the magnitude of measured value. This indicates that the final subject's position in the test should not influence the reading of the values (reaching position) and therefore it does not have to be considered as a source of measurement error. Although our results do not indicate this influence, it can be recommended to eliminate potential influence by the flat rectangular object (i.e. a book) which is moved forward by the subject during the forward bending.

Table 1
Females' performance in the test *V sit-and-reach* ($N = 43$; the values are presented in cm)

Characteristics	<i>M</i>	95% <i>CI</i>	<i>Mdn</i>	<i>SD</i>
The first measurement	10.1	(8.0, 12.2)	12	6.91
The second measurement	11.3	(9.1, 13.4)	13	7.10
Intra-individual differences*	-1.14	$(-1.6, -0.7)$	1	1.39

Note. *CI* = confidence interval, *Mdn* = median.

* the values are computed as $X_1 - X_2$, where X_1 is the first and X_2 is the second measurement (testing process).

It was found that during repeated measurement the test performance increased in average by 1.14 ± 1.39 cm. This observed difference was recognized as a statistically significant, but interpretation of value of effect size (low effect) is not in agreement with this finding. Based on the requirement of the conformity of statistical significance and the effect size (see methods) we have to classify observed test – re-test difference as insignificant. This conclusion corresponds to high level of observed intra-individual reliability (an interpretation of the value of the test – re-test correlation) and also it identified indicators of absolute reliability (the *SEM* and the Bland and Altman's 95% limits of agreement). Both, high level of absolute relative reliability do not presuppose the existence of significant sources of measurement error. These findings indicate low probability of sources of measurement error which originate from the self-assessment process. This means that all changes in modification of VSR for self-assessment are not (do not include) significant sources of measurement error. This means that modified VSR can be considered as an eligible instrument to measure a reach distance in sitting position while performed by an individual, even better when VSR is used for the purpose of self-assessment of physical fitness. Regarding, self-assessment of physical fitness it is more important to carry out individual assessment than the outcome achieved in the test. This can be an indication that the individual is aware of the need for change and it can be considered as a first step leading to an active life style.

It must be kept in mind that this study was carried on under the specific conditions and their change(s) can result in different findings. (Especially) For example, the implementation and realization of the warm-up before the test can be considered (generally) as a significant sources of measurement error. This simple fact of warming or previous stretching of muscle fibers can result increased flexibility in particular joints. This effect of warm-up on the level of performance in flexibility tests, in our case of hamstring and low-back flexibility tests, was verified by numerous studies (Arazi, Asadi, & Hoseini, 2012; Golden, Hoffman, Pavol, & Wallace, 2005; O'Sullivan, Murray, & Sainsbury, 2009; Zakaz, Grammatikopoulou, Zakas, Zahariadis, & Vamvakoudis, 2006). This influence of a warm-up on reliability of VSR should be studied more and it's correction of realization of the form of warm-up can lead to higher standardization of VSR or similar tests.

CONCLUSIONS

This study provides the evidence of a high level of reliability, resp. appropriate reliability, of the *V sit-and-reach* test used as an instrument of the self-assessment of hamstrings and low-back flexibility in adolescent fe-

males. According to these findings it is not important to consider the double role of an examiner as well as a subject in the test as a source of measurement error.

ACKNOWLEDGMENTS

The authors of this study would like to thank the students for their participation in the study. The study has been supported by the student's research grant from the Palacký University (IGA No. FTK_2012:021) "Self-assessment of health-related fitness: Reliability assessment of four fitness tests".

REFERENCES

- ACSM. (2000). *Guidelines for exercise testing and prescription* (6th ed.). Baltimore, MD: Lippincott, Williams & Wilkins.
- Arazi, H., Asadi, A., & Hoseini, K. (2012). Comparison of two different warm-ups (static-stretching and massage): Effects on flexibility and explosive power. *Acta Kinesiologica*, 6(1), 55–59.
- Atkinson, G., & Nevill, A. M. (1998). Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Medicine*, 26(4), 217–238.
- Ayala, F., Sainz de Baranda, P., de Ste Croix, M. S., & Santonja, F. (2012). Fiabilidad y validez de las pruebas sit-and-reach: Revisión sistemática [Reliability and validity of sit-and-reach tests: Systematic review]. *Revista Andaluza de Medicina del Deporte*, 5(2), 57–66.
- Baltaci, G., Un, N., Tunay, V., Besler, A., & Gerçeker, S. (2002). Comparison of three different sit and reach tests for measurement of hamstring flexibility in female university students. *British Journal of Sports Medicine*, 37(1), 59–61.
- Baumgartner, T. A., & Jackson, A. S. (1995). *Measurement for evaluation in physical education and exercise science*. Dubuque, IA: Brown & Benchmark.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Davis, D. S., Quinn, R. O., Whiteman, C. T., Williams, J. D., & Young, C. R. (2008). Concurrent validity of four clinical tests used to measure hamstring flexibility. *Journal of Strength and Conditioning Research*, 22(2), 583–588.
- Golden, G. M., Hoffman, M. A., Pavol, M. J., & Wallace, D. A. (2005). The effect of warm-up routine on sit and reach, muscle onset, and vertical jump performance. *Journal of Athletic Training*, 40(Suppl. 2), S93.
- Hui, S. S., & Yuen, P. Y. (2000). Validity of the modified back-saver sit-and-reach test: A comparison

with other protocols. *Medicine and Science in Sports and Exercise*, 32(9), 1655–1659.

López-Miñarro, P. A., de Baranda, A. P., & Rodríguez-García, P. L. (2009). A comparison of the sit-and-reach test and the back-saver sit-and-reach test in university students. *Journal of Sports Science and Medicine*, 8(1), 116–122.

Martin, S. B., Jackson, A. W., Morrow, J. R., & Liemohn, W. P. (1998). The rationale for the sit and reach test revisited. *Measurement in Physical Education and Exercise Science*, 2(2), 85–92.

O'Sullivan, K., Murray, E., & Sainsbury, D. (2009). The effect of warm-up, static stretching and dynamic stretching on hamstring flexibility in previously injured subjects. *BMC Musculoskeletal Disorders*, 10, 37. doi: 10.1186/1471-2474-10-37

Thomas, J. R., Nelson, J. K., & Silverman, S. J. (2005). *Research methods in physical activity* (5th ed.). Champaign, IL: Human Kinetics.

Wells, K. F., & Dillon, E. K. (1952). The sit and reach. A test of back and leg flexibility. *Research Quarterly*, 23, 115–118.

Zakaz, A., Grammatikopoulou, M. G., Zakas, N., Zahariadis, P., & Vamvakoudis, E. (2006). The effect of active warm-up and stretching on the flexibility of adolescent soccer players. *Journal of Sport Medicine and Physical Fitness*, 46(1), 57–61.

RELIABILITA V SIT-AND-REACH TESTU VYUŽITÉHO K SEBEHODNOCENÍ FLEXIBILITY U ŽEN (Souhrn anglického textu)

ÚVOD: Ukazuje se, že V sit-and-reach test (VSR) je vhodným nástrojem k sebehodnocení flexibility hamstringů a oblasti bederní páteře. Důvodem je především fakt, že test je poměrně snadno proveditelný, je nenáročný na materiální vybavení a nevyžaduje příliš znalostí a zkušeností z oblasti užívání motorických testů. Vzhledem ke specifickým formy realizace testu, tj. při sebehodnocení, lze u tohoto testu očekávat existenci několika zdrojů chyb měření. Důsledkem může být snížení reliability testu.

CÍL: Záměrem studie je posoudit reliabilitu VSR při jeho použití formou sebehodnocení při posuzování flexibility u adolescentů (ženy).

METODY: Výzkumný soubor tvořilo 43 studentek (věk $21,2 \pm 0,5$ let) Univerzity Palackého v Olomouci. Ke stanovení systematické chyby a vymezení intra-individuální reliability testu byl použit *t*-test pro opakovaná měření ($p < 0,05$), resp. Pearsonův korelační koeficient. K posouzení absolutní reliability byly použity směrodatná chyba měření (*SEM*) a Bland-Altmanovy 95% limity shody.

VÝSLEDKY: Průměrný intra-individuální rozdíl mezi opakovanými měřeními (1,14 cm) byl shledán jako statisticky významný ($t = -5,375$; $df = 42$). Intra-individuální reliabilita byla klasifikována jako vysoká ($r = 0,98$); absolutní reliabilita odpovídá hodnotě 0,139 cm.

ZÁVĚRY: Studie poskytuje důkazy, že VSR je vhodným nástrojem při sebehodnocení flexibility hamstringů a oblasti bederní páteře u adolescentních dívek.

Klíčová slova: zdravotně orientovaná zdatnost, chyba měření, testování, vysokoškolský student.