Research in Physical Education, Sport and Health
2025, Vol. 14, No. 2, pp.135-140
ISSN(Print):1857-8152; ISSN(Online):1857-8160

TEST-RETEST RELIABILITY OF STANDING LONG JUMP TEST IN ADOLESCENTS

DOI: https://doi.org/10.46733/PESH25142135t
(Original scientific paper)

Miodrag Todorovic
Ss. Cyril and Methodius University, Faculty of Physical Education, Skopje, Macedonia

Abstract

Test-retest reliability of the Standing long jump test has been conducted in 40 adolescents aged 13-14 years old .The participants were in good psychophysical health and regularly attended physical education classes. The test included performance of two trials and the best result was recorded for analysis The results showed a high level of test reliability, with an interclass coefficient (ICC) of 0.985, which exceeds the threshold for excellent reliability. (0.9) . The Standard measurement error (SEM) was 1.82 cm, whereas the smallest detectable difference(SDD) was 5.05 cm, which indicates high measurement precision. The paired t-test showed that there were not significant differences between the first and the second measurement, which additionally confirms the consistency of the results.. Bland-Altman analysis revealed minimal systematic differences between the measurements, with a random distribution of points around the mean line, implying a high degree of methodological agreement. The standing long jump test has proved to be a reliable and applicable tool for assessing explosive lower-limb strength. The results support findings from previous studies highlighting the practicality of the test for evaluating physical abilities in adolescents. The results provide test standardization in local and international frames. Additional research involving larger sample sizes and diverse contexts is recommended for broader validity.

Key words: *test-retest , explosive strength, adolescents, fitness, field tests*

Introduction

Muscular fitness is a key indicator of health throughout the lifespan. (Ortega et al., 2008; Ruiz et al., 2008). In children and adolescents it is particularly associated with a better cardiovascular profile and reduced risk of metabolic disorders. (Ruiz et al., 2009; Cohen et al., 2014) Additionally, it is in positive correlation with the health of the bones and the psychological wellbeing, such as higher self-confidence , which highlights the importance of the assessment of muscular strength since earliest age. (Smith et al., 2014).In reference to the assessment of muscular strength, testing of explosive lower-limb strength presents important component of the programs for physical fitness for young people(Castro-Piñero et al., 2010). Although laboratory tests provide high precision and security, their appliance is very often limited due to high expenses and the necessity for technical expertise.(Hopkins, 2000).Instead, field tests, such as long standing jump represent practical alternative with low expenses and easy implementation, which makes them suitable for broader usage in school environments. (España-Romero et al., 2010).

The standing long jump test is an integral part of several internationally recognized fitness batteries for assessing health-related physical fitness in youth. For example, it is included in the ALPHA fitness battery, which is designed to evaluate physical abilities in children and adolescents in Europe (España-Romero et al.,2010). Additionally, the standing long jump is used as a part of Eurofit tests and it is used as an indicator of explosive strength of lower limbs as well as a general index for muscular condition (Castro-Piñero et al., 2010) and many other national and international fitness batteries.

The test has been used for many years in the former Yugoslav region as well as in our country. Nevertheless, standardization and testing protocols varied from one country to another. In 2024 the study (YFIT) Delfi project was conducted, in which 169 experts participated from 50 countries and territories .This Delfi study is grounded in two key sources of evidence :the project ALPHA , financed by EU and the report from Medical Institute (now known as National medical academy) from the USA. The goal of the ALPHA project was to identify set of valid, reliable ,applicable and safe field tests for evaluation of health-related physical fitness in children and adolescents aged 6 to 18 years old, which could be used in standardized public health monitoring in European Union. The evidence for decision-making was based on four different reviews, including cross- associations between physical fitness and health outcomes, as well as validity of the tests for health outcomes, criteria validity of the tests and test-retest reliability of the tests .Additionally, methodological studies were conducted to address identified gaps in knowledge and the selected tests were examined for both feasibility and safety.

This study also contributes to the standardization of tests and protocols used worldwide, which will enable data pooling, international monitoring and surveillance, the generation of international and continental reference standards specific to sex and age, health-related cut-off points, and ultimately the development of a Global Youth Fitness Map. As part of this study, the standing long jump test is included.

Examining the test–retest reliability of standardized tests is of crucial importance for the accurate interpretation of results and for ensuring consistency in performance evaluation over time. Based on these considerations, and on the recognition that this issue has not yet been addressed in the Republic of North Macedonia, this study was conducted with the primary aim of determining the test–retest reliability of the standing long jump test in Macedonian adolescents.

Methods

Sample of Participants

The study was conducted on a sample of 40 participants of both sexes, aged 13–14 years. The sample included all students whose parents provided consent for participation in the study, who were psychophysically healthy, and who regularly attended physical and health education classes. All procedures involving the participants were carried out in accordance with the Declaration of Helsinki.

Measurement Procedure

All measurements were conducted indoors in a sports hall between 8:00 and 12:00 a.m.

Equipment: A non-slip surface, a stick for identifying the landing point, a measuring tape for measuring jump distance, adhesive tape for marking the starting line, and cones for marking the testing area.

Instructions for the ExaminerЧ The standing long jump is performed from a standing position using both legs simultaneously. A horizontal line is drawn or marked with tape to identify the starting line. A measuring tape is placed perpendicular to and alongside the starting line. The jump distance is measured from the take-off line to the point where the back of the heel closest to the starting line lands on the ground. The child is asked to place the toes just behind the starting line. While bending the knees and swinging the arms, the child is instructed to jump as far forward as possible, maintaining balance on both feet. If the child falls backward or touches the floor with the hands or any part of the body behind the feet, the attempt must be repeated. If the child falls forward but maintains the rear foot on the ground after landing, the test is considered valid.

Instructions for the Participant: “Stand with your feet shoulder-width apart, with your toes slightly behind the line. Bend your knees, keeping your arms in front of you, parallel to the ground. While swinging both arms, push off forcefully and jump as far forward as possible, taking off and landing on both feet simultaneously. Try to land with your feet together and remain upright. If you fall forward after landing, try to keep the rear foot on the ground so the jump can be measured; otherwise, the jump must be repeated. If you fall backward after landing, the jump must be repeated.”

Demonstration: Two trials are performed, and the best result is recorded. The result is recorded to the nearest whole centimeter. For example, a jump of 1 m 56 cm is recorded as 156 cm.

Statistical Analysis

All values are presented as mean ± SD unless otherwise stated. Data normality was assessed using the Kolmogorov–Smirnov test.

Relative and absolute reliability were calculated (Baumgartner, 1989). Relative reliability was assessed using the intraclass correlation coefficient (ICC) based on a two-way random-effects model (absolute agreement, ICC2.1) (Weir, 2005). Test–retest reliability was considered good when ICC values ranged from 0.61 to 0.80, and excellent when values ranged from 0.81to 1.00 (Shrout, 1988) .

Before calculating absolute reliability, heteroscedasticity (non-uniform variance) was assessed by examining Pearson correlation coefficients between the absolute difference of test and retest scores and the mean of the test and retest scores (Bland & Altman, 1986; Bland & Altman, 1999). If the correlation coefficient (r) ranged between 0 and 0.1, the data were considered homoscedastic. In such cases, absolute reliability was recommended to be assessed using the standard error of measurement (SEM) (Atkinson & Nevill, 1998).

If r exceeded 0.1, the data were considered heteroscedastic and, consequently, the logarithmically transformed coefficient of variation (CV) was used to assess absolute reliability (Bland & Altman, 1999).

The SEM was used to estimate the smallest detectable difference (SDD), also referred to as the minimal detectable change (MDC) or smallest detectable change (SDC). To enable comparison of the reliability of the handgrip dynamometry test with most studies conducted in pediatric populations, the percentage SEM (SEM%) and the normalized smallest detectable difference (nSDD) were also calculated.

SEM was calculated using the following equation:

SEM = SD × (1 − ICC)^0.5,

where SD represents the pooled standard deviation of the test and retest scores, and ICC represents the calculated intraclass correlation coefficient. SEM was divided by the mean of the measurements from test 1 and test 2 and multiplied by 100 to obtain SEM%.

CV was calculated as the standard deviation of √d divided by the mean and multiplied by 100. SDD is a linear transformation of SEM, calculated as 1.96 × √2 × SEM (Weir, 2005; Schreuders et al., 2003). nSDD represents SDD expressed as a percentage of the average maximal voluntary contraction (Molenaar et al., 2008).

ICC2.1 was calculated using MedCalc for Windows, version 15.2.2 (MedCalc Software Inc., Mariakerke, Belgium) (Schoonjans et al., 1995).

Results

The results of the comparison between measurements T1 and T2 show a high level of agreement, with average values of 167.17 and 168.28, respectively, and a minimal average difference of −1.11. The standard deviations (15.20 for T1 and 14.43 for T2) indicate similar variability around the mean values, further supporting the consistency of the two measurements. The paired t-test yielded a t-statistic of −1.81 with a significance value of 0.089, indicating that the difference between the measurements was not statistically significant ($p > 0.05$).

These findings demonstrate that the measurements are methodologically consistent, and any observed differences can be attributed to random variation rather than systematic bias. This confirms the validity and reliability of the measurement instrument.

Table 1. Differences between the first and second measurements in the standing long jump test among adolescents.

Metric	Value
Mean (T1)	167,17
Mean (T2)	168,28
SD (T1)	15,20
SD (T2)	14,43
Mean Difference	-1,11
T-statistic	-1,81
P-value (Sig.)	0,09

Table 2. Statistical reliability coefficients for the standing long jump test in adolescents

Metric	Value
Intraclass Correlation Coefficient (ICC)	0,985
Standard Error of Measurement (SEM)	1,820
Smallest Detectable Change (SDD)	5,045
Coefficient of Variation (CV %)	8,717

The results of the analysis indicate that the measurement method demonstrates high reliability, as evidenced by an Intraclass Correlation Coefficient (ICC) value of 0.985. This value, which substantially exceeds the 0.9 threshold, indicates excellent agreement between measurements, demonstrating minimal error in the repeatability of the instrument.

Additionally, the Standard Error of Measurement (SEM) was 1.82, representing very low variability and allowing for high precision in individual measurements. The Smallest Detectable Change (SDD), calculated at 5.05, represents the minimal meaningful change that can be attributed to a true change rather than measurement error or random variation, which is particularly important for clinical and research interventions.

The Coefficient of Variation (CV) of 8.72% further confirms the quality of the instrument, indicating low relative variability in the data. Overall, the obtained results provide strong empirical support for the applicability of this method in both research and practical applications that require high reliability and measurement precision.

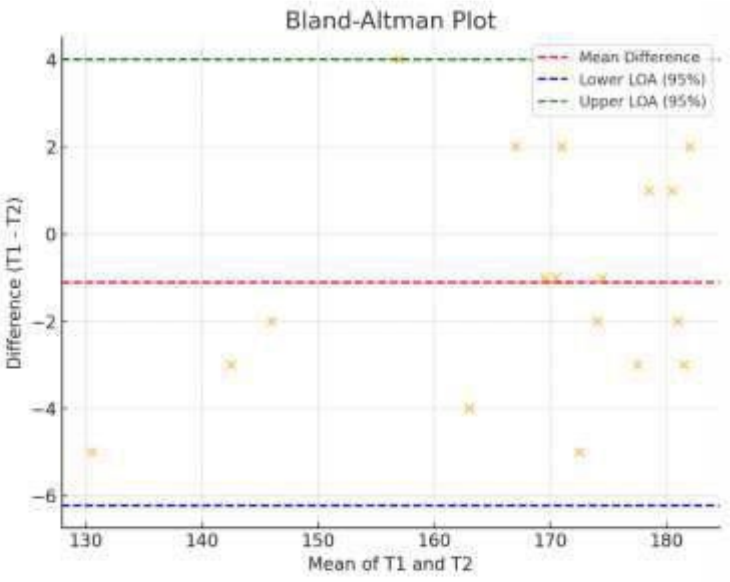

Figure 1. Bland-Altman анализа (LOA = inter-trials mean difference ± 1.96 SD the inter-trials difference)

The Bland–Altman analysis reveals that the mean difference between measurements T1 and T2 is minimal, indicating a high level of methodological agreement. The limits of agreement (95% LOA), defined as the mean difference ± 1.96 × SD, show that the differences between measurements fall within a narrow range, further confirming the reliability of the test.

The random distribution of data points around the mean line, with no visible systematic trend, implies the absence of significant bias between the two measurement occasions. The uniform dispersion of points within the limits of agreement suggests that the differences between measurements are not statistically significant and arise from random variation rather than measurement error or systematic differences.

These findings support the applicability of the measurement method in contexts where reliability and precision are of key importance.

Discussion

The obtained results confirm the high test–retest reliability of the standing long jump test in adolescents. According to the analysis, the intraclass correlation coefficient (ICC) was 0.985, indicating excellent reliability of the measurements. This value substantially exceeds the 0.9 threshold, highlighting the reliability and consistency of this measurement instrument (Shrout, 1998). Furthermore, the smallest detectable difference (SDD) of 5.045 cm indicates high precision in identifying meaningful changes, making the test suitable for monitoring performance over time.

The standard error of measurement (SEM) of 1.820 cm is relatively low, further supporting the accuracy of individual measurements (Atkinson & Nevill, 1998). In addition, the coefficient of variation (CV) of 8.717% indicates low relative variability, which is consistent with previous research confirming the robustness of the test across different contexts (Bland & Altman, 1999).

The results of this study are consistent with previous research that has examined the reliability of the standing long jump test. Findings such as the high ICC value of 0.985 indicate excellent reliability across repeated measurements. This is in agreement with the findings of Castro-Piñero et al. (2010), who also reported high reliability of this test in children and adolescents. Such consistency highlights the applicability of the test across different populations and settings.

In the study by España-Romero et al. (2010), which analyzed 600 participants across different age groups, ICC values for the standing long jump test ranged from 0.80 to 0.95, depending on age and sex. Although these values are slightly lower than those obtained in the present study, the difference may be attributed to the smaller sample size or to stricter protocol control in the current research. Additionally, the study by Markovic et al. (2004) emphasized that this test is particularly suitable for assessing explosive strength of the lower extremities, as it provides accurate information on adolescents' motor abilities.

An interesting aspect is that the family of tests used to assess explosive strength, including the standing long jump, often demonstrates higher reliability when applied to youth with similar levels of physical fitness (Hopkins, 2000). This aligns with the characteristics of our sample, in which all participants regularly attended physical education classes, which may have contributed to performance homogeneity.

On the other hand, research by Bobbert et al. (1996) indicates that smaller measurement errors, such as the low SEM observed in the present study (1.820 cm), are crucial for accurately monitoring individual progress. These characteristics make the test suitable not only for research purposes but also for practical applications in sports medicine and education.

This test is of great importance for teachers and coaches, as it is easy to implement, does not require expensive equipment, and provides rapid results. The standardization of protocols, as proposed in this study, facilitates data comparison across different institutions. For example, the results of this research are aligned with the ALPHA project, which recognizes field-based tests as fundamental tools for assessing physical fitness in youth (España-Romero et al., 2010).

Although this study confirms a high level of reliability of the test, questions remain regarding its validity in different contexts, such as comparisons between adolescents from urban and rural environments. Further studies with larger samples would allow for a more precise examination of these potential differences.

Conclusion

This study confirmed the high test–retest reliability of the standing long jump test in adolescents, demonstrating that this method is a reliable and practical tool for assessing explosive strength of the lower extremities. The results showed excellent consistency, with an intraclass correlation coefficient (ICC) of 0.985, exceeding the threshold for high reliability. These values confirm that the instrument is capable of accurately monitoring performance over time, with minimal variation attributable to measurement error.

Reference

Atkinson, G., & Nevill, A. (1998). Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Medicine*, 26(4), 217–238.

Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*, 1(8476), 307–310.

Bland, J. M., & Altman, D. G. (1999). Measuring agreement in method comparison studies. *Statistical Methods in Medical Research*, 8(2), 135–160.

Bobbert, M. F., et al. (1996). Why is countermovement jump height greater than squat jump height? *Medicine and Science in Sports and Exercise*, 28(11), 1402–1412.

Castro-Piñero, J., Artero, E. G., España-Romero, V., Ortega, F. B., Sjöström, M., Suni, J., & Ruiz, J. R. (2010). Criterion-related validity of field-based fitness tests in youth: A systematic review. *British Journal of Sports Medicine*, 44(13), 934–943. https://doi.org/10.1136/bjsm.2009.058321

Castro-Piñero, J., et al. (2010). Assessing muscular strength in youth: Usefulness of standing long jump as a general index of muscular fitness. *Journal of Strength and Conditioning Research*, 24(7), 1810–1817.

Cohen, D. D., Gómez-Arbeláez, D., Camacho, P. A., Pinzón, S., Hormiga, C., Trejos-Suárez, J., ... & López-Jaramillo, P. (2014). Low muscle strength is associated with metabolic risk factors in Colombian children: The ACFIES study. *PLoS One*, 9(4), e93150.

España-Romero, V., Artero, E. G., Jiménez-Pavón, D., Cuenca-García, M., Ortega, F. B., Castro-Piñero, J., & Ruiz, J. R. (2010). Assessing health-related fitness tests in the school setting: Reliability, feasibility, and safety; the ALPHA study. *International Journal of Sports Medicine*, 31(7), 490–497. https://doi.org/10.1055/s-0030-1251990

España-Romero, V., et al. (2010). ALPHA fitness test battery: A practical approach for health-related fitness assessment in children and adolescents. *International Journal of Obesity*, 34(8), 1050–1055.

Hopkins, W. G. (2000). Measures of reliability in sports medicine and science. *Sports Medicine*, 30(1), 1–15.

Markovic, G., et al. (2004). Reliability and factorial validity of squat and countermovement jump tests. *Journal of Strength and Conditioning Research*, 18(3), 551–555.

Ortega, F. B., Ruiz, J. R., Castillo, M. J., & Sjöström, M. (2008). Physical fitness in childhood and adolescence: A powerful marker of health. *International Journal of Obesity (London)*, 32(1), 1–11. https://doi.org/10.1038/sj.ijo.0803774

Ortega, F. B., Zhang, K., Cadenas-Sanchez, C., Tremblay, M. S., Jurak, G., Tomkinson, G. R., ... & Delphi Fitness Expert Group. (2024). The Youth Fitness International Test (YFIT) battery for monitoring and surveillance among children and adolescents: A modified Delphi consensus project with 169 experts from 50 countries and territories. *Journal of Sport and Health Science*, 101012.

Ruiz, J. R., Castro-Piñero, J., Artero, E. G., Ortega, F. B., Sjöström, M., Suni, J., & Castillo, M. J. (2009). Predictive validity of health-related fitness in youth: A systematic review. *British Journal of Sports Medicine*, 43(12), 909–923. https://doi.org/10.1136/bjsm.2008.056499

Smith, J. J., Eather, N., Morgan, P. J., Plotnikoff, R. C., Faigenbaum, A. D., & Lubans, D. R. (2014). The health benefits of muscular fitness for children and adolescents: A systematic review and meta-analysis. *Sports Medicine*, 44(9), 1209–1223. https://doi.org/10.1007/s40279-014-0196-4

Weir, J. P. (2005). Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *Journal of Strength and Conditioning Research*, 19(1), 231–240.