Süleyman Ulupınar [1,*]
Cebrail Gençoğlu [1]
Serhat Özbay [1]

TEST-RETEST RELIABILITY, SMALLEST WORTHWHILE CHANGE, AND MINIMAL DETECTABLE CHANGE SCORES FOR FREQUENCY SPEED OF KICK TEST IN JUNIOR KICKBOXING ATHLETES

TESTNO-RETESTNA ZANESLJIVOST, NAJMANJŠA VREDNA SPREMEMBA IN MINIMALNA ZAZNAVNA SPREMEMBA PRI TESTU HITROSTI UDARCA PRI MLAJŠIH ŠPORTNIKIH V KICKBOXINGU

ABSTRACT

While there exists a significant body of research dedicated to performance tests specific to combat sports, the current literature lacks valid and functional methodologies for assessing kickboxing-specific tasks. The present study endeavored to establish the test-retest reliability, standard error of measurement (SEM), minimal detectable change (MDC), smallest worthwhile change (SWC), and typical error (TE) values of the Frequency Speed of Kicks Test (FSKT). Furthermore, this study sought to compare the reliability values of the FSKT with those of the countermovement jump test (CMJ). The study cohort consisted of twenty-eight junior male kickboxers. Participants performed the CMJ and FSKT twice across test and retest sessions. Pearson correlation analysis identified a significant correlation (r=0.717) between the CMJ and FSKT. Paired t-tests revealed no significant disparities between the test and retest values for both the CMJ and FSKT. However, the difference between the test and retest in the CMJ demonstrated a small effect size, while the FSKT showed a trivial effect size. Intraclass correlation coefficient (ICC) values for the CMJ and FSKT were deemed "good" (r=0.855) and "excellent" (r=0.963) respectively, in terms of reliability. The FSKT displayed superior absolute agreement between test and retest scores due to its lower SEM values when compared to the CMJ. The TE, used to estimate trial-to-trial variation, was lower in the FSKT than in the CMJ. Moreover, the results indicated a lower MDC value in the FSKT than the CMJ, suggesting that the FSKT could be more effective at detecting smaller performance changes compared to the CMJ. In conclusion, this study posits that the FSKT could be considered a reliable method, demonstrating reproducible results in the performance evaluation of kickboxing athletes, pending the development of a functional kickboxing-specific field test.

Keywords: sport-specific, field-test, combat sport, martial arts, practical importance

[1]Faculty of Sport Sciences, Erzurum Technical University, Erzurum, Turkey

IZVLEČEK

Medtem ko obstaja precej raziskav, posvečenih testom zmogljivosti, značilnim za borilne športe, v trenutni literaturi ni veljavnih in funkcionalnih metodologij za ocenjevanje nalog, značilnih za kickboxing. Namen študije je bil ugotoviti zanesljivost testiranja, standardno napako merjenja (SEM), najmanjšo zaznavno spremembo (MDC), najmanjšo vredno spremembo (SWC) in tipično napako (TE) testa hitrosti udarcev (FSKT). Poleg tega je bil cilj te študije primerjava vrednosti zanesljivosti testa FSKT z vrednostmi zanesljivosti skoka z nasprotnim gibanjem (CMJ). Študijsko skupino je sestavljalo osemindvajset mlajših kickboxerjev. Udeleženci so dvakrat izvedli test CMJ in FSKT s tremi dnevi počitka med testiranjem in ponovnim testiranjem. Pearsonov koeficient korelacije je pokazal pomembno povezanost (r=0.717) med CMJ in FSKT. Parni t-testi niso pokazali bistvenih razlik med testnim in ponovnim testiranjem za CMJ in FSKT. Vendar je razlika med testom in ponovnim testom pri CMJ pokazala majhno velikost učinka, medtem ko je FSKT pokazala trivialno velikost učinka. Vrednosti koeficienta interklasne korelacije (ICC) za CMJ in FSKT so bile v smislu zanesljivosti ocenjene kot "dobre" (r=0.855) oziroma "odlične" (r=0.963). FSKT je pokazal boljše absolutno soglasje med rezultati testa in ponovnega testa zaradi nižjih vrednosti SEM v primerjavi s CMJ. TE, ki se uporablja za oceno variacije med posameznimi poskusi, je bila pri FSKT nižja kot pri CMJ. Poleg tega so rezultati pokazali nižjo vrednost MDC pri FSKT kot pri CMJ, kar kaže na to, da bi bil FSKT lahko učinkovitejši pri odkrivanju manjših sprememb uspešnosti v primerjavi s CMJ. Ta študija torej kaže, da bi lahko FSKT veljal za zanesljivo metodo, ki kaže ponovljive rezultate pri ocenjevanju zmogljivosti športnikov v kickboxingu, dokler ne bo razvit funkcionalni terenski test, specifičen za kickboxing.

Ključne besede: športno specifični, terenski testi, borilni šport, borilne veščine, praktični pomen

Corresponding author:* Süleyman Ulupınar,
Faculty of Sport Sciences, Erzurum Technical University, Erzurum, Turkey
E-mail: slymnulpnr@gmail.com
https://doi.org/10.52165/kinsi.29.2.183-194

INTRODUCTION

Kickboxing is a full-contact combat sport including both an athlete's efforts of punching and kicking against an opponent and avoiding the punches and kicks of the opponent (Slimani, Chaabene, Miarka, & Chamari, 2017). However, studies investigating the performance of kickboxing athletes have widely used non-specific tests such as isokinetic strength, vertical jumps, cycle ergometer, treadmill or field running (Haugen, Breitschädel, Wiig, & Seiler, 2020; Machado et al., 2009; Ouergui et al., 2014; Salci, 2015). It is crucial to note that tests with non-specific tasks may decrease ecological validity since the nature of a kickboxing match requires intermittent or repeated actions with short rest periods (Salci, 2015; Slimani et al., 2017). However, in novel studies, a sport-specific test involving the maximum number of kicks for 10 s, which is called the frequency speed of the kick test (FSKT), has been used in taekwondo athletes (Santos & Franchini, 2016; Santos, Valenzuela, & Franchini, 2015; Santos, Lopes-Silva, Loturco, & Franchini, 2020). Although the FSKT also appears to be an appropriate test to measure kickboxing performance, no studies have used this test in kickboxing athletes.

The countermovement jump (CMJ) test has been one of the most popular and commonly reported methods to measure athletic performance such as muscular power, neuromuscular adaptation, and reactive strength in sports sciences over the last two-three decades (Barker, Harry, & Mercer, 2018; Haugen et al., 2020; Mirzaei, Norasteh, & Asadi, 2013; Özbay & Ulupınar, 2020). However, although vertical jump ability is directly related to competitive performance in some sports such as volleyball, basketball, football, and high jump, this movement pattern may not be crucial to evaluate performance in many sports. In addition, CMJ contains some difficulties in practice due to differences in instruments, testing procedures, calculation methods, and software, therefore, it may not be possible to conclude that the measurement results are free from many possible errors (Haugen et al., 2020; Street, McMillan, Board, Rasmussen, & Heneghan, 2001).

Reliability refers to the consistency of or the reproducibility of values of measurement in repeated trials on the same participants (Hopkins, 2000). Reliability analysis presents crucial information indicating the degree to which scores of a test or a method are free of measurement errors (Wilkinson et al., 2019). Thus, a reliable test allows us to reach a better precision of measurements and to detect the changes in test scores in research and practical settings (Hopkins, 2000; Zeljko, Gilic, & Sekulic, 2020). Reliability basically consists of both absolute and relative reliability (Bruton, Conway, & Holgate, 2000; Haley & Fragala-Pinkham, 2006;

Weir, 2005). Absolute reliability refers to the consistency of scores of individuals, whereas relative reliability refers to the consistency of the position or rank of individuals in the group relative to others (Weir, 2005). The main reliability values in investigations of sports sciences usually include within-subject random variation, systematic change in the mean, and retest correlation (Hopkins, 2000; Šuker, Grgantov, & Milić, 2015).

Paired t-test and Bland-Altman plot methods only refer to the absolute agreement while the Pearson correlation coefficient only refers to the measure of correlation (Koo & Li, 2016; Özbay, Ulupınar, Çınar, & Akbulut, 2019; Wilkinson et al., 2019). Therefore, these analyzes are considered less-ideal methods for deciding the precision of a test (Bruton et al., 2000; Koo & Li, 2016; Portney & Watkins, 2009). On the contrary, intra-class correlation coefficients (ICC), which is an index reflecting a more reasonable measure of reliability, refers to both degrees of correlation and agreement between measurements (Haley & Fragala-Pinkham, 2006; Hopkins, 2000; Koo & Li, 2016). ICC is the common metric in test-retest, intra-rater, and inter-rater reliability analyses (Koo & Li, 2016). ICC reflects not only the degree of correlation but also agreement between measurements unlike the Pearson product-moment correlation coefficient (Bruton et al., 2000). However, ICC may not be enough to interpret that measurement scores are reliable due to hypersensitive to the heterogeneity of the research sample, therefore, this value needs to be supported by other reliability analyses (Hopkins, 2000; Weir, 2005).

Absolute reliability provides information for differentiating a real change in performance from a change due to individual variation and measurement error and is mostly quantified as standard error of measurement (SEM) (Segura-Ortí & Martínez-Olmos, 2011). SEM, which is used as an indicator of agreement between measurements, is calculated from the ICC and standard deviation values (Weir, 2005). SEM is also used to calculate minimal detectable change (MDC), which is a more clinically useful value for interpreting reliability (Haley & Fragala-Pinkham, 2006). MDC refers to the smallest amount of reliable change in the scores to conclude that the difference is not attributable to measurement errors (Segura-Ortí & Martínez-Olmos, 2011). Thus a change amount exceeding the MDC is used to define a real (or true) change that is free from measurement errors (Haley & Fragala-Pinkham, 2006).

A measurement error indicates the observed value of a measure differs from its real value (Hopkins, 2000). In sports sciences, experimental research designs using test-retest analyses to determine the reliability of a method generally consist of the measurements that the same

investigator performed using the same equipment. The effects of learning, motivation, and fatigue are considered the possible reasons for systematic changes in the mean of a measure between successive trials, and their effects need to be eliminated from estimates of within-subject variation (Hopkins, 2000). Typical error refers to the standard deviation of a participant's repeated measurements and it is another practical indicator used to estimate trial-to-trial variation (Weir, 2005). It is stated that studies designed to determine the reliability of a test or method should be supported by calculations such as SEM, MDC, and TE together with ICC. The aim of the present study was to determine the test–retest reliability, estimated SEM, MDC, SWC, and TE of FSKT, and compare the reliability values of FSKT with CMJ.

METHODS

Participants

Twenty-eight junior male kickboxing athletes (mean and standard deviation; age: 16.4 ± 1.2 years; body mass: 67.7 ± 7.1 kg; height: 175.6 ± 7. cm; training experience: 5.0 ± 1.2 years) participated in this study. Athletes were competing at national level and training at least three sessions weekly. All athletes were free from any lower-body injury or neuromuscular disorder. Participants were instructed not to exercise strenuously or not to use any supplement during the research period. All participants provided written consent after being informed about the purpose, procedures, and possible risks. Parental written consent has been obtained for participants under 18 years. The research was approved by the Ethics Committee of the Institute of Winter Sports, Atatürk University, Erzurum, Turkey (SBFEK190066441-42).

Procedures

Participants performed the CMJ and FSKT with three days of rest between the test and retest sessions. In addition, participants performed these tests twice times with five min-breaks during the same session, however, the best score was considered valid for that session. In other words, the best scores collected from the test and retest sessions were used to evaluate test-retest analyses. Actually, the common method for determining the valid score of a participant in research designs measuring athletic performance is to choose the highest score in several attempts in a session and assume that this score is the best for the participant. However, we aimed to evaluate the reliability of the tests using the variation between two trials with a few days of rest, not within-session variation.

The purpose of FSKT was to perform the maximal quantity of kicks during the 10-second test. FSKT tests were recorded using an ultra-slow motion video camera (Sony PXW-FX9 6K Full-Frame, Japan). A kicking was considered a valid score when a participant hit the target area on the sandbag with the appropriate kicking technique. CMJ height was measured with the My Jump 2 mobile application and its validity and reliability were tested by previous studies (Balsalobre-Fernández, Glaister, & Lockey, 2015; Haynes, Bishop, Antrobus, & Brazier, 2019). To measure the CMJ height via a smartphone (iPhone 7) using this app, the same investigator manually selected the initial contact frame, the take-off frame from the floor, and the final landing frame and results were automatically presented by the app. Athletes were asked to rest their hands on the sides of their waist, keep their knees in full extension, and then flex their knees and jump vertically at the fastest possible speed during the CMJ test. Both tests were applied in the same period of the day to minimize circadian variation influence on performance.

Statistical Analysis

The SPSS package version 21.0 (IBM SPSS Statistics for Windows, Armonk, NY) was used for data management and analysis. The level of significance was predetermined to be $p \leq 0.05$ for all statistical analyses. Data are reported as mean ± standard deviation. Paired t-test was used to compare the mean values between the test and retest. To evaluate the magnitude of difference, Cohen´s d effect size was calculated and d values were classified to Hopkins scale. Pearson product–moment correlation coefficients were calculated to determine the strength of the relationship between the best scores of CMJ and FSKT.

Intra-class correlation coefficient ($ICC_{3,k}$), based on a two-way mixed model (the rater was considered a fixed effect and the participants were considered a random effect) with a consistency definition, was used to measure the test−retest relative reliability. An ICC above 0.750 was considered to demonstrate good reliability while ICC above 0.900 was considered to indicate excellent reliability (Portney & Watkins, 2009). Typical error of measurement was computed to estimate trial-to-trial variation (do Nascimento et al., 2017; Weir, 2005). Minimal detectable change (MDC) basically defines the smallest change that is not within the expected range of error, therefore, it was used to identify the required amount of change in a measurement to ensure that the difference was not attributable to error (Charter, 1996; Wilkinson et al., 2019). The smallest worthwhile change (SWC) was used to define the smallest meaningful change or smallest clinically important difference (Duthie, Pyne, Ross, Livingstone, & Hooper, 2006;

Özbay & Ulupınar, 2019; Özbay et al., 2019). The formulas used to support ICC in the reliability analysis are presented below.

Formula 1: $TE = SD_{differences} / \sqrt{2}$

Formula 2: $MDC = SEM \times 1.96 \times \sqrt{2}$

Formula 3: $SEM = SD_{mean} \times \sqrt{1 - r}$

Formula 4: $CV_{SEM} = SEM / Mean \times 100$

Formula 5: $SWC_{small} = 0.2 \times$ between-subject SD_{best}

Formula 6: $SWC_{moderate} = 0.6 \times$ between-subject SD_{best}

where $SD_{differences}$ = the standard deviation of the difference between successive two trials; SD_{mean} = the average of the standard deviations for the participants in two trial; SD_{best} = the standard deviation of best scores; r = the intra-class correlation coefficient; CV_{SEM} = the SEM as the coefficient of variation. The 1.96 value represents the z-score at the 95% confidence interval. SEM is the standard error of measurement.

RESULTS

Paired t-test demonstrated that there was no significant difference between test and retest mean values for both CMJ and FSKT (Table 1). However, difference between test and retest in CMJ had a small effect size. Reliability analysis results of the test-retest for ICC, TE, SWC, CV_{SEM}, and MDC are presented in Table 2. Results of the test-retest for ICC, TE, SWC, CV_{SEM}, and MDC are presented in Table 2. ICC values for CMJ and FSKT were considered to have good and excellent reliability, respectively. SEM was classified as “good” when lower than SWC. Figure 1 indicated that CMJ and FSKT had a significant Pearson product–moment correlation coefficient (r = 0.717).

Table 1. Performance results during the CMJ and FSKT (n = 28).

	Trial 1 (mean ± SD)	Trial 2 (mean ± SD)	Paired t-test value	p-value	Cohen's d (effect size)
CMJ (cm)	33.57 ± 3.83	34.51 ± 4.97	1.617	0.118	0.21, small effect
FSKT (kicks)	20.64 ± 2.56	20.82 ± 2.64	0.961	0.345	0.07, trivial effect

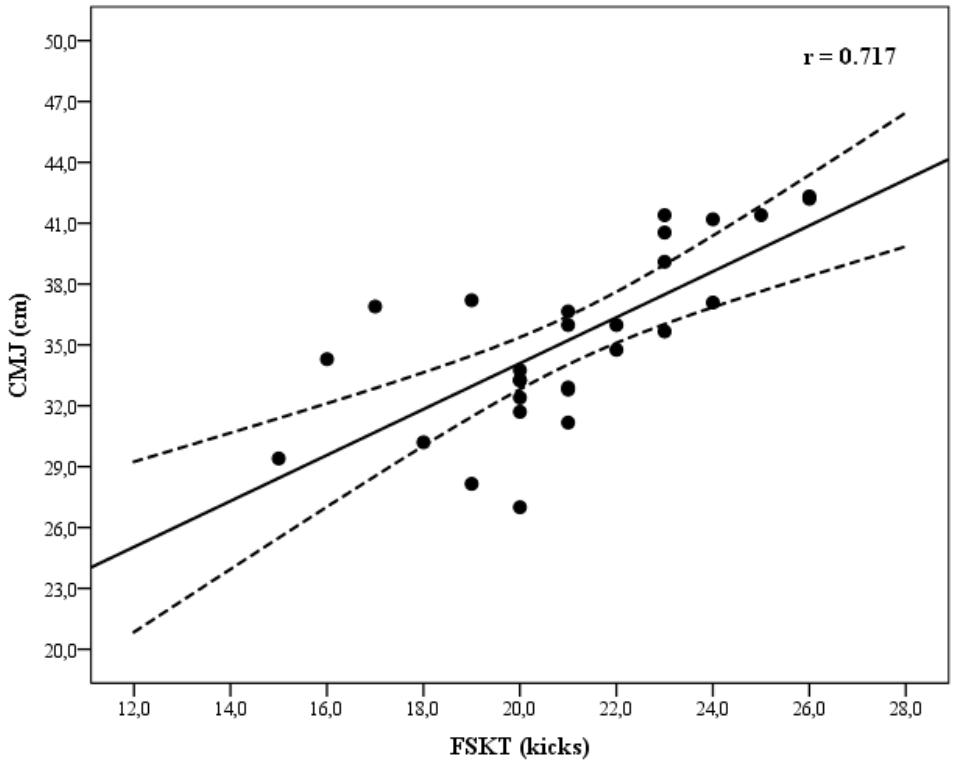
CMJ= countermovement jump; FSKT= frequency speed of kick test

Table 2. Test–retest reliability analyses of the CMJ and FSKT (n = 28).

	ICC (95% CI)	TE	CV_{TE} (%)	SWC_{small}	$SWC_{moderate}$	SEM	CV_{SEM} (%)	MDC
CMJ	0.855 (0.69–0.93)	1,14	3,40	0,76	2,28	0,62	1,82	1,73
FSKT	0.963 (0.92–0.98)	0,51	2,47	0,50	1,50	0,10	0,48	0,27

CMJ= countermovement jump; FSKT= frequency speed of kick test; ICC= intra-class correlation coefficient; CI= confidence interval; TE= typical error; CV_{TE}= the TE as coefficient of variation; SWC_{small}= smallest worthwhile change for small effect size; $SWC_{moderate}$= smallest worthwhile change for moderate effect size; SEM= standard error of measurement; the SEM as coefficient of variation; MDC= minimal detectable change.

Figure 1. Relationship between the best scores of CMJ and FSKT (n=28).



DISCUSSION AND CONCLUSION

To our knowledge, this is the first study to investigate the reliability of FSKT in kickboxing athletes. In this study, the reliability of FSKT for junior kickboxers was evaluated using complementary statistical procedures consisting of relative and absolute reliability analyses. The ICC criterion adopted in this study resulted in good (r = 0.855) and excellent (0.963) reliability classification for CMJ and FSKT, respectively (Portney & Watkins, 2009). However, if the lower limits of the confidence interval were considered, it could be made inference that the relative reliability of FSKT (95% CI = 0.92–0.98) is higher than that of CMJ (95% CI = 0.69–0.93). Current studies reported that ICC values of FSKT were 0.95 for taekwondo athletes with different competition levels (Santos et al., 2020), 1.00 for black-belt taekwondo athletes (Silva Santos & Franchini, 2016), and 0.95 for women taekwondo athletes (Santos & Franchini, 2018). Thus, the present study in addition to various studies proven that the relative reliability of FSKT is excellent.

FSKT had a greater absolute agreement between test and retest scores due to having lower SEM and CV_{SEM} values compared with CMJ. In addition, the effect size calculated from paired t-test analysis supported SEM and CV_{SEM} results since the differences between test and retest showed that CMJ had a small effect size while FSKT had a trivial effect size. Similarly, TE and CV_{TE}, which were used to estimate trial-to-trial variation, were lower in the FSKT than CMJ. A study reported that SEM, CV_{SEM}, and MDC values were 0.60, 2.85, and 1.67, respectively (Santos et al., 2020). These results are quite similar to the findings of our study. Hence, the various findings demonstrated that FSKT could present highly precise results and error-free. The MDC is the smallest change that falls outside the expected range of error and a difference amount exceeding the MDC threshold can be considered a “true” change (Haley & Fragala-Pinkham, 2006; Segura-Ortí & Martínez-Olmos, 2011). The current findings indicated that MDC value was lower in the FSKT than CMJ thus FSKT could allow to determine lower performance changes compared with CMJ.

Pearson product–moment correlation coefficient revealed a significant relationship between CMJ and FSK. Although this result suggested to be a relationship between the skills measured by these two tests, other analyzes indicated that FSKT can be a more reasonable test for kickboxing athletes. For example, SWC refers to the smallest value considered important in practice and the findings showed that FSKT represented a higher sensitivity for detecting an important change in practice, as SWC was lower in the FSKT than CMJ (Bernards, Sato, Haff,

& Bazyler, 2017; Duthie et al., 2006; Santos et al., 2020). TE and SEM values should be lower than SWC values for proper use of SWC (Bernards et al., 2017; do Nascimento et al., 2017; Santos et al., 2020). SEM values were lower than SWC_{small} (threshold value emphasizing the importance of a small effect in practice) for both tests. But TE value for CMJ was greater than SWC while TE value for FSKT was similar to SWC. Thus, $SWC_{moderate}$ (threshold value emphasizing the importance of a moderate effect in practice) values may be more appropriate to conclude that the results were completely outside the limits of error.

CMJ has been commonly used to assess athletic performances such as muscular power (İnce & Şentürk, 2019; İnce & Ulupınar, 2020), neuromuscular adaptation (Claudino et al., 2017; Mirzaei et al., 2013), and reactive strength (Barker et al., 2018) in a wide variety population for decades. However, using the biomechanical movement pattern required for CMJ as a test method may not be a valid approach to assess the performance of kickboxing athletes (Haugen et al., 2020; Slimani et al., 2017). Moreover, considering the difficulties in the CMJ measurements that can be caused by instruments, testing procedures, calculation methods, and software, using a sport-specific test seems a more appropriate method (Street et al., 2001). This study showed that relative and absolute reliability results were lower in the CMJ than in FSKT and the indicators of measurement error such as SEM indicating the absolute agreement and TE indicating the variation of trial-to-trial were higher in the CMJ. In addition, the CMJ results less allow monitoring of smaller changes because the results indicating that a change in performance can be considered error-free and practically important are higher in the CMJ than FSKT.

A comprehensive study investigating sport-specific performance tests in Olympic sports such as amateur boxing, fencing, judo, karate, taekwondo, and wrestling revealed to be more than 40 different tests in total (Chaabene et al., 2018). However, to our knowledge, unlike other combat sports, kickboxers do not have a valid and functional choice in terms of kickboxing-specific tests. Since FSKT only measures kicking ability, its use as a performance test for kickboxers cause a limitation, but it appears to be the most appropriate for kickboxers among valid and reliable tests described in the literature. Based on the many similarities of Kickboxing with combat sports such as karate, taekwondo, boxing, and muay thai (Salci, 2015; Slimani et al., 2017; Taskin & Akkoyunlu, 2020; Vertonghen & Theeboom, 2008), we aimed to investigate the extent to the reliability of a test developed for these sports can be when used in kickboxers. For example, FSKT was originally developed to measure a skill based on kicking as fast as possible in 10 seconds for taekwondo athletes (Santos & Franchini, 2016, 2018; Santos et al.,

2020). Therefore, our hypothesis was that FSKT could present reliable results in kickboxers since it appears to be an appropriate test to measure kickboxing performance compared with CMJ. Our findings confirmed the initial hypothesis and presented evidence that FSKT can provide reliable results until a specific test for kickboxers is designed. In conclusion, this study suggested that FSKT could be considered a reliable method and presented reproducible results in evaluating the performance of kickboxing athletes until a functional kickboxing-specific field test is developed.

Acknowledgments

This work was orally presented at International Khazar Conference on Scientific Research (September 18-20, 2020 Khazar University, Azerbaijan).

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

REFERENCES

Balsalobre-Fernández, C., Glaister, M., & Lockey, R. A. (2015). The validity and reliability of an iPhone app for measuring vertical jump performance. *Journal of sports sciences, 33*(15), 1574-1579.

Barker, L. A., Harry, J. R., & Mercer, J. A. (2018). Relationships between countermovement jump ground reaction forces and jump height, reactive strength index, and jump time. *The Journal of Strength & Conditioning Research, 32*(1), 248-254.

Bernards, J. R., Sato, K., Haff, G. G., & Bazyler, C. D. (2017). Current research and statistical practices in sport science and a need for change. *Sports, 5*(4), 87.

Bruton, A., Conway, J. H., & Holgate, S. T. (2000). Reliability: what is it, and how is it measured? *Physiotherapy, 86*(2), 94-99.

Chaabene, H., Negra, Y., Bouguezzi, R., Capranica, L., Franchini, E., Prieske, O., . . . Granacher, U. (2018). Tests for the assessment of sport-specific performance in Olympic combat sports: a systematic review with practical recommendations. *Frontiers in physiology, 9*, 386.

Charter, R. A. (1996). Revisiting the standard errors of measurement, estimate, and prediction and their application to test scores. *Perceptual and Motor Skills, 82*(3_suppl), 1139-1144.

Claudino, J. G., Cronin, J., Mezêncio, B., McMaster, D. T., McGuigan, M., Tricoli, V., . . . Serrão, J. C. (2017). The countermovement jump to monitor neuromuscular status: A meta-analysis. *Journal of science and medicine in sport, 20*(4), 397-402.

Duthie, G. M., Pyne, D. B., Ross, A. A., Livingstone, S. G., & Hooper, S. L. (2006). The reliability of ten-meter sprint time using different starting techniques. *The Journal of Strength & Conditioning Research, 20*(2), 251.

Haley, S. M., & Fragala-Pinkham, M. A. (2006). Interpreting change scores of tests and measures used in physical therapy. *Physical Therapy, 86*(5), 735-743.

Haugen, T., Breitschädel, F., Wiig, H., & Seiler, S. (2020). Countermovement jump height in national team athletes of various sports: a framework for practitioners and scientists. *Int J Sports Physiol Perform*.

Haynes, T., Bishop, C., Antrobus, M., & Brazier, J. (2019). The validity and reliability of the my jump 2 app for measuring the reactive strength index and drop jump performance. *The Journal of sports medicine and physical fitness*.

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

Ince, İ., & Şentürk, A. (2019). Effects of plyometric and pull training on performance and selected strength characteristics of junior male weightlifter. *Physical education of students, 23*(3), 120-128.

Ince, I., & Ulupinar, S. (2020). Prediction of competition performance via selected strength-power tests in junior weightlifters. *The Journal of sports medicine and physical fitness, 60*(2), 236-243.

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of chiropractic medicine, 15*(2), 155-163.

Machado, S., Souza, R., Simão, A., Jerônimo, D., Silva, N., Osorio, R., & Magini, M. (2009). Comparative study of isokinetic variables of the knee in taekwondo and kickboxing athletes. *Fitness & Performance Journal, 8*(6), 407-411.

Mirzaei, B., Norasteh, A. A., & Asadi, A. (2013). Neuromuscular adaptations to plyometric training: depth jump vs. countermovement jump on sand. *Sport Sciences for Health, 9*(3), 145-149.

Nascimento, M. A., Ribeiro, A. S., de Souza Padilha, C., da Silva, D. R. P., Mayhew, J. L., do Amaral Campos Filho, M. G., & Cyrino, E. S. (2017). Reliability and smallest worthwhile difference in 1RM tests according to previous resistance training experience in young women. *Biology of Sport, 34*(3), 279-285.

Ouergui, I., Hssin, N., Haddad, M., Padulo, J., Franchini, E., Gmada, N., & Bouhlel, E. (2014). The effects of five weeks of kickboxing training on physical fitness. *Muscles, ligaments and tendons journal, 4*(2), 106.

Özbay, S., & Ulupınar, S. (2019). Reliability of 1RM, 5RM and 10RM Tests in Upper Body Resistance Exercises. *The Journal of Turkish Sport Sciences for Health, 2*(1), 1-7.

Özbay, S., & Ulupınar, S. (2020). Strength-Power Tests are More Effective When Performed After Exhaustive Exercise in Discrimination Between Top-Elite and Elite Wrestlers. *Journal of strength and conditioning research*.

Özbay, S., Ulupınar, S., Çınar, V., & Akbulut, T. (2019). Reliability of Easily Applicable Non-Laboratory Methods Used for Determination of the Upper Body Strength. *Turkiye Klinikleri Journal of Sports Sciences, 11*(2).

Portney, L. G., & Watkins, M. P. (2009). *Foundations of clinical research: applications to practice* (Vol. 892): Pearson/Prentice Hall Upper Saddle River, NJ.

Salci, Y. (2015). The metabolic demands and ability to sustain work outputs during kickboxing competitions. *International Journal of Performance Analysis in Sport, 15*(1), 39-52.

Santos, S. J. F., & Franchini, E. (2016). Is frequency speed of kick test responsive to training? A study with taekwondo athletes. *Sport Sciences for Health, 12*(3), 377-382.

Santos, S. J. F., & Franchini, E. (2018). Frequency speed of kick test performance comparison between female taekwondo athletes of different competitive levels. *The Journal of Strength & Conditioning Research, 32*(10), 2934-2938.

Santos, S. J. F., Valenzuela, T. H., & Franchini, E. (2015). Can different conditioning activities and rest intervals affect the acute performance of taekwondo turning kick? *The Journal of Strength & Conditioning Research, 29*(6), 1640-1647.

Santos, S. F. J., Lopes-Silva, J. P., Loturco, I., & Franchini, E. (2020). Test-retest reliability, sensibility and construct validity of the frequency speed of kick test in male black-belt taekwondo athletes. *Ido Movement for Culture. Journal of Martial Arts Anthropology, 20*(3), 38-46.

Segura-Ortí, E., & Martínez-Olmos, F. J. (2011). Test-retest reliability and minimal detectable change scores for sit-to-stand-to-sit tests, the six-minute walk test, the one-leg heel-rise test, and handgrip strength in people undergoing hemodialysis. *Physical Therapy, 91*(8), 1244-1252.

Slimani, M., Chaabene, H., Miarka, B., & Chamari, K. (2017). The activity profile of elite low-kick kickboxing competition. *International Journal of Sports Physiology and Performance, 12*(2), 182-189.

Street, G., McMillan, S., Board, W., Rasmussen, M., & Heneghan, J. M. (2001). Sources of error in determining countermovement jump height with the impulse method. *Journal of Applied Biomechanics, 17*(1), 43-54.

Šuker, D., Grgantov, Z., & Milić, M. (2015). Intrarater and interrater reliability of the spiking efficiency assessment in top men's volleyball. *Kinesiologia Slovenica, 21*(2), 22.

Taskin, M., & Akkoyunlu, Y. (2020). Effect of anaerobic power on agility and quickness in male national taekwondo athletes. *Kinesiologia Slovenica, 26*(2), 49-57.

Vertonghen, J., & Theeboom, M. (2008). Analysis of experiences of flemish children in martial arts: an exploratory study. *Kinesiologia Slovenica, 14*(2).

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

Wilkinson, T. J., Xenophontos, S., Gould, D. W., Vogt, B. P., Viana, J. L., Smith, A. C., & Watson, E. L. (2019). Test–retest reliability, validation, and "minimal detectable change" scores for frequently reported tests of objective physical function in patients with non-dialysis chronic kidney disease. *Physiotherapy theory and practice, 35*(6), 565-576.

Zeljko, I., Gilic, B., & Sekulic, D. (2020). Validity, reliability and correlates of futsal-specific pre-planned and non-planned agility testing protocols. *Kinesiologia Slovenica, 26*(2), 25-34.