FACTA UNIVERSITATIS
Series: **Physical Education and Sport**, Vol. 22, No 2, 2024, pp. 87 - 95
https://doi.org/10.22190/FUPES240728008T

Research article

TEST-RETEST AND INTER-RATER RELIABILITY OF VOLLEYBALL ACCURACY TESTS IN SCHOOL CHILDREN

UDC 796.325-053.5

Nebojša Trajković[1], Elzan Bibić[2], Dejan Ćeremidžić[2], Marjan Mladenović[3]

[1]Faculty of Sport and Physical Education, University of Niš, Niš, Serbia
[2]Faculty of Physical Education and Sports, University of East Sarajevo, Bosnia and Herzegovina
[3]Elementary School "Đura Jakšić", Turekovac, Serbia

ORCID iDs: Nebojša Trajković https://orcid.org/0000-0001-6276-3052
Elzan Bibić N/A
Dejan Ćeremidžić https://orcid.org/0009-0000-1806-3497
Marjan Mladenović N/A

Abstract. *The aim of this study was to examine the test-retest and inter-rater reliability of volleyball accuracy tests in a sample of school children. The study involved a sample of 46 school children aged 12-13 years (Height: 158.97 ± 5.39 cm; Mass: 52.20 ± 5.58 kg; Age = 12.4 ± 0.7 years) who had been participating in their school volleyball programs for at least one year. The sessions were organized over two separate occasions to assess test-retest reliability, with the same procedures followed each time. To evaluate inter-rater reliability, two raters independently recorded the performance of the participants, ensuring that the scoring was consistent and unbiased. Players were tested for the overhead and underhand passing, setting and serving. Perfect test-retest reliability (Interclass correlation coefficient (ICC) > 0.90; coefficient of variation (CV)% < 10%) was observed for setting and serving. Additionally, we found nearly perfect reliability for overhead (ICC = 0.84) and underhand passing (ICC = 0.72) with high CV% (> 10%). The results of the inter-rater reliability showed very high ICC's (>0.89) for all the variables. The study demonstrated that the volleyball skill testing protocol provides consistent and reliable assessments for players, with no significant differences between test and retest sessions across all the evaluated skills.*

Key words: *testing, volleyball, school children, reliability*

Received July 28, 2024 / Accepted 16 October, 2024
Corresponding author: Nebojša Trajković
Faculty of Sport and Physical Education, University of Niš, Čarnojevićeva 10A, 18000 Niš, Serbia
Phone: • E-mail nele_trajce@yahoo.com

INTRODUCTION

Volleyball is an interval sport characterized by short bursts of high-intensity activity followed by periods of rest, necessitating training programs that reflect this dynamic. Such activities rely on both anaerobic and aerobic energy systems; therefore, exercises should be designed to engage the ATP-CP and aerobic pathways while accurately simulating the kinematic patterns of competitive play (Shepard, Gabbet, & Stanganelli, 2009). The ability to accurately execute volleyball skills is a crucial aspect of the game, particularly for school-aged children who are developing their athletic abilities (Zetou, Tzetzis, Vernadakis, & Kioumourtzoglou, 2002).

The development of volleyball skills is crucial for effective performance in both competitive and educational settings. According to Zetou, Michalopoulou, Giazitzi, and Kioumourtzoglou (2007), the application of contextual interference can significantly impact the learning and retention of volleyball skills, suggesting that varied practice conditions enhance skill acquisition. Additionally, Harrison et al. (1999) highlighted the importance of instructional models, such as skill teaching and mastery learning in improving skill development, knowledge, self-efficacy, and game play among students. These studies highlight the need for well-structured physical education programs that focus on skill refinement to foster both physical and cognitive growth in volleyball players.

Reliable and valid assessment tools are necessary to accurately evaluate the skill level of young volleyball players and children in physical education and monitor their progress over time (Bartlett, Smith, Davis, & Peel, 1991; Lidor, Hershko, Bilkevitz, Arnon, & Falk, 2007; Gabbet and Georgief, 2006). Previous research has highlighted the importance of establishing the measurement properties of motor competence assessments in children and adolescents (Hulteen, et al., 2020). Specifically, studies (Gabbet and Georgief, 2006; Milić et al., 2016) have found strong support for the validity and reliability of assessments that measure skills such as overhead and underhand passing, setting, serving.

Establishing the reliability of volleyball accuracy tests is crucial, as these assessments can provide valuable information to coaches, physical education teachers, and researchers regarding the skill level and progress of young athletes. This study will contribute to the existing literature by providing evidence on the test-retest and inter-rater reliability of volleyball accuracy tests, which can inform the use of these assessments in school-based settings. The aim of this study was to examine the test-retest and inter-rater reliability of volleyball accuracy tests in a sample of school children.

METHODS

Participants

The study involved a sample of 46 school children aged 12-13 years (body height: 158.97 ± 5.39 cm; body mass: 52.20 ± 5.58 kg; Age = 12.4 ± 0.7 years) who had been participating in their school volleyball programs for at least one year. Participants were recruited from a local school, ensuring a diverse representation of both male and female students. The inclusion criteria for the study required the participants to be within the specified age range and to be currently enrolled in their school's volleyball program. Additionally, the participants had to have parental consent and be willing to commit to both the test and retest sessions. Exclusion criteria included any children who had sustained injuries or had medical conditions that could affect their ability to perform the

volleyball accuracy tests safely. Students who were involved in external volleyball training that could interfere with their school volleyball participation or introduce confounding variables into the study were also excluded. The study protocol was carried out in accordance with the Declaration of Helsinki and was approved by the ethics committee of the Faculty of Sport and Physical Education, University of Nis.

Procedure

The testing procedures were conducted in a sports hall to ensure a controlled environment with standardized conditions for all the participants. Each participant was given a thorough explanation of the testing protocols and a brief demonstration of the volleyball accuracy tests. The tests involved targeting specific areas on the volleyball court, with each participant completing a series of trials. The sessions were organized over two separate occasions to assess test-retest reliability, with the same procedures followed each time. To evaluate inter-rater reliability, two raters independently recorded the performance of the participants, ensuring that the scoring was consistent and unbiased. Raters were physical education teachers who are engaged with volleyball courses in schools. All of the equipment that was used, including volleyballs and target markers, was standardized to maintain uniformity across tests.

Overhead and underhand passing

Target hitting with the overhead and underhand pass from zone VI to position III, with ball tossing from the assistant from zone VI from the other side of the court (Fig. 1). The test the participants' service reception will be evaluated by determining their ability to pass the ball with their fingers to a target positioned on the net, 3 meters from the right sideline. The target dimensions are 1.5 meters in length and 2 meters in width. The coach is positioned at the serving position, about 1 meter above the ground and 10 meters from the receiving player, tosses the ball overhead and passes it with their fingers to the receiving player. The participants were required to pass the ball to another player who stands with their hands above their head in the target zone. Players who successfully pass the ball to the player in the target area receive 2 points. The second target area is for balls that did not reach the main target area but would likely reach the player in a match situation. The second target area extends from the right sideline and measures 3 meters in length and 4 meters in width. Players who successfully pass the ball to the second target area receive 1 point. Finally, passes that do not reach the target areas receive 0 points. The final score was the sum of points from 6 attempts. This test showed good reliability in this age group (Gabbett, et al., 2007).

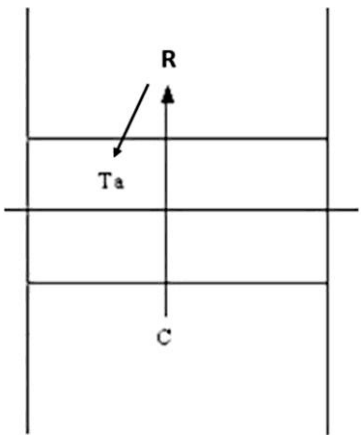

Fig. 1 Overhead and underhand passing; R - passer; C - coach; Ta - target

Setting

The testing procedures for assessing the accuracy of volleyball setting were conducted in a sports hall to ensure a controlled environment. The test involved participants aiming to hit a horizontal target with their fingers positioned in front of their heads. The objective was to measure the accuracy of setting the ball to a specific target area in position IV from zone III, with ball feeds from zone VI on the same side of the court. The equipment included an 80 cm diameter hoop placed vertically at a height of 2.70 meters in position IV, a cart with 10 balls, and a scoring sheet. The participants' setting ability was evaluated based on their skill in hitting the target positioned near the net at a height of 2.7 meters and a distance of 5.5 meters from the player performing the set. This target was chosen to mimic the position of a hitter preparing to spike the ball during a match. The coach, positioned 5 meters from the player, tossed the ball overhead to the center player, who then attempted to set the ball through the 80 cm hoop. Successful sets through the target earned 3 points, balls hitting the outer part of the hoop and not passing through earned 2 points, and sets landing 2.3 meters from the net (1.5 meters from the target) earned 1 point, marked by adhesive tape. Balls not landing in any target zone earned 0 points. The final score was the sum of points from 6 attempts. The participants were allowed one practice attempt, and the next ball feed began after the previous set was completed and either hit or missed the target. This test was a modification of a previously conducted test by Milić et al. (2016).

Serving

The test aimed to measure the precision of serves using a volleyball court, ten volleyballs, a measuring tape, and boundary tape. During the procedure, each player performed ten consecutive serves underhand using their dominant hand, aiming to direct the ball towards target zones of higher value. They stood in the serving zone, positioned 9 meters from the volleyball net (which is 2.24 meters high). The participants were instructed to aim for designated areas marked on the floor of the opposite court. Instructions included assigning points for each serve based on the hit target areas, with zero points awarded if the ball landed outside the court. They were instructed to serve with maximum force to ensure the ball crossed the net, while also aiming to accurately hit the far corners of the opposite court. The opposite court was divided into four areas. Serving the ball to the far corners earned 5 points, landing in the middle zone earned 3 points, and hitting the area nearest the net earned 1 point. Missing the court resulted in a score of 0. Higher points were given if the ball landed between two zones. Players were allowed to choose their desired position behind the service line and had 8 seconds for each serve, matching the preparation time permitted in actual matches. This test was developed and explained and its reliability was determined for the mentioned age groups (Lidor, & Mayan, 2005; Lidor, et al., 2007).

Statistical analysis

The statistical analysis for inter-rater reliability and test-retest reliability was conducted using the IBM SPSS Version 25 software (IBM SPSS Inc., Armonk, MY). To assess inter-rater reliability, Intraclass Correlation Coefficients (ICCs) were calculated, considering a two-way random effects model to account for the consistency and absolute

agreement among raters. The ICC values were interpreted based on established guidelines, with values interpreted as: low = < 0.10, moderate = 0.11 – 0.30, high = 0.31 – 0.50, very high = 0.51 – 0.70, nearly perfect = 0.71 – 0.90 and perfect = 0.91 – 1.0 (Hopkins, 2000). Test-retest reliability was determined using the Intraclass Correlation Coefficient (ICC). The within-subject variation for each of the tests was determined by calculating the coefficient of variation (CV). Additionally, a paired sample t-test was conducted to compare the mean scores from the two testing occasions, ensuring that there were no significant differences that could indicate variability in the test conditions or participants' performance over time. The level of significance was set at $p < 0.05$ for all statistical tests.

RESULTS

Test – retest reliability

Similar overhead passing (test = 6.6 ± 2.2; retest = 7.2 ± 1.8), underhand passing (test = 5.3 ± 1.3; retest = 5.7 ± 1.2) setting (test = 7.4 ± 2.9; retest = 8.3 ± 2.9), and serving (test = 29.2 ± 8.2; retest = 30.4 ± 8.4) values were observed between two volleyball skill testing (Table 1). Non-significant differences ($p > 0.05$) were observed between testing sessions for all volleyball skills. High test-retest reliability (ICC > 0.90; CV% < 10%) was observed for setting and serving. However, we found good reliability for overhead passing (ICC = 0.84) and underhand passing (ICC = 0.72) with a high CV% (> 10%).

Table 1 Test-retest reliability of volleyball accuracy tests; data are mean ± SD

	Test	Retest	ICC	CV%
Overhead passing	6.6 ± 2.2	7.2 ± 1.8	0.84 (0.68-0.92)	15.4 (12 – 21.6)
Underhand passing	5.3 ± 1.3	5.7 ± 1.2	0.72 (0.48-0.86)	14.3 (11.2 – 20.1)
Setting	7.4 ± 2.9	8.3 ± 2.9	0.99 (0.98-1.00)	6.0 (4.7 – 8.3)
Serving	29.2 ± 8.2	30.4 ± 8.4	0.98 (0.96-0.99)	4.7 (3.7 – 6.5)

Inter-rater reliability

Very high ICC's were found > 0.89 for all variables. The results of the inter-rater reliability of the two raters are shown in Table 2.

Table 2 Inter-rater reliability; data are mean ± SD

	Rater 1	Rater 2	ICC
Overhead pass	6.6 ± 2.2	7.2 ± 2.1	0.93 (0.85 – 0.97)
Underhand pass	5.3 ± 1.5	5.8 ± 1.6	0.89 (0.76 – 0.95)
Setting	7.4 ± 2.9	8.2 ± 2.8	0.97 (0.95 – 0.99)
Serving	29.2 ± 8.2	29.7 ± 8.2	0.99 (0.99 – 1.00)

DISCUSSION

This study aimed to determine the test-retest and inter-rater reliability of volleyball accuracy tests in a sample of school children. The results of the volleyball skill testing

reveal consistent performance between the initial test and the retest sessions across all evaluated skills, with no significant differences observed. This indicates that the participants' performance was stable over time. Perfect test-retest reliability was particularly evident for setting and serving, suggesting that these skills are more consistently executed by players, possibly due to the repetitive and practiced nature of these actions. Overhead passing and underhand passing, while also showing nearly perfect reliability, exhibited more variability. This could be due to the inherent complexity and variability in these skills, which may be influenced by factors such as ball trajectory, player positioning, and coordination. The greater variability in these skills suggests that they may be more susceptible to fluctuations in player performance or external conditions, highlighting potential areas for targeted training and assessment refinement to improve consistency. Additionally, the results of the inter-rater reliability demonstrated very high intraclass correlation coefficients (ICCs) above 0.89 for all items, indicating that the two raters were highly consistent in their evaluations. Overall, the findings highlight the robustness of the testing protocol for evaluating volleyball skills, while also pointing to specific areas where variability may need to be addressed.

Reliability and inter-rater reliability are critical factors in assessing volleyball skills, ensuring that the evaluations are consistent and objective. Gabbett, Georgieff, and Domrow (2007) highlight the importance of using a combination of physiological, anthropometric, and skill data to predict player selection in junior volleyball squads. This multi-faceted approach can enhance the reliability of the assessment by considering various aspects of an athlete's performance. Gabbett and Georgieff (2006) further emphasize the development of a standardized skill assessment for junior players, which can significantly improve the reliability of skill evaluations by providing a consistent framework for assessment. Lidor et al. (2007) demonstrate that the measurement of talent in volleyball requires longitudinal tracking to ensure that the assessment tools remain reliable over time. Their study on elite adolescent players over a 15-month period underscores the need for continuous validation of skill tests to maintain their reliability. Bartlett et al. (1991) contribute to this discussion by developing a valid volleyball skills test battery, which serves as a benchmark for other skill assessments and reinforces the importance of standardized testing procedures.

Inter-rater reliability, which ensures that different evaluators provide consistent ratings, is another crucial aspect. Standardized protocols, as discussed by Gabbett and Georgieff (2006), help mitigate subjective biases and improve inter-rater reliability. The comprehensive approach taken by Gabbett, Georgieff, and Domrow (2007) also supports this by incorporating objective measures alongside skill assessments, thereby enhancing the overall reliability of the evaluations. Thus, the integration of standardized assessments, objective measures, and longitudinal tracking collectively contribute to the reliability and inter-rater reliability of volleyball skill evaluations.

The findings of this study align with previous research on volleyball skill assessment instruments, highlighting their effectiveness and reliability. Zonifa (2020) emphasized the importance of developing robust skill test instruments for advanced-level students, which is corroborated by our results showing high test-retest reliability, particularly for setting and serving. This suggests that well-designed assessments can effectively gauge advanced players' abilities and track their development over time. Yudiana, Hidayat, Hambali, & Slamat (2017) discussed the significance of content validity in assessment instruments, particularly through volleyball information systems, ensuring that the tests measure the intended skills accurately. Our study’s high inter-rater reliability further supports the validity of the testing protocol, as consistent evaluations across different raters reinforce the

accuracy of skill measurement. Additionally, Sopa and Pomohaci (2018) highlight the evaluation of motor development and skills in younger volleyball players, emphasizing the need for reliable assessments to monitor growth. While our study focused on a broader age range, the observed consistency and reliability of skill assessments can similarly be applied to younger athletes, aiding in their early development and long-term improvement. The current results along with the above mentioned references highlight the critical role of reliable and valid skill assessment tools in enhancing volleyball training and player development across different skill levels and age groups.

Studies on volleyball skill assessment have several limitations and strengths. One limitation is that they often have limited sample sizes or focus on specific groups, such as junior players, which may affect the generalizability of the findings to broader populations. Additionally, many studies provide short-term assessments, potentially missing long-term skill development and changes in reliability over time. Even with standardized protocols, subjectivity can persist in evaluating complex skills and gameplay situations, impacting inter-rater reliability. Variability in testing conditions, such as differences in environments, equipment, and procedures, can also introduce inconsistencies, affecting the validity of the assessments. Furthermore, some assessments focus on specific skills, potentially overlooking other important aspects of a player's performance. The reliance on specific technologies or testing equipment can limit the applicability of assessments in different settings or regions lacking such resources.

Despite these limitations, there are notable strengths in these studies. The development of standardized testing protocols enhances the reliability and comparability of skill evaluations across different contexts. Combining physiological, anthropometric, and skill data provides a comprehensive evaluation of players, offering a holistic understanding of their abilities and potential. Longitudinal studies allow for the assessment of skill development over time, providing valuable insights into the progression and consistency of skill acquisition. Including objective metrics alongside subjective evaluations helps mitigate biases and improve the overall reliability of the assessments. Creating validated skill test batteries sets benchmarks for other studies, ensuring assessments are grounded in reliable and valid measurement principles. Lastly, focusing on talent identification contributes to optimizing selection processes in competitive sports, leading to better development pathways for young athletes.

CONCLUSION

The study demonstrated that the volleyball skill testing protocol provides consistent and reliable assessments for players, with no significant differences between test and retest sessions across all evaluated skills. Perfect test-retest reliability was observed for setting and serving, indicating that these skills are more consistently executed by players. Overhead passing and underhand passing also showed nearly perfect reliability, though with greater variability, suggesting these skills are more susceptible to fluctuations in performance. The very high inter-rater reliability, with ICCs above 0.89 for all items, confirms that the evaluation method is also consistent across different raters. This reliability indicates that the testing protocol can be effectively used by various coaches and evaluators, ensuring uniformity in skill assessment. Coaches can use this reliable testing protocol to identify specific areas where players need improvement, particularly focusing on the more variable overhead and underhand passing. This targeted approach can help in developing

more consistent skill execution. Moreover, the consistent and reliable nature of the testing protocol makes it a valuable tool for talent identification, helping coaches and scouts make informed decisions about player selection. Overall, the reliable assessment of volleyball skills through this protocol can enhance training effectiveness, player development, and talent identification, contributing to the overall improvement of volleyball performance.

REFERENCES

Bartlett, J., Smith, L., Davis, K., & Peel, J. (1991). Development of a valid volleyball skills test battery. *Journal of Physical Education, Recreation & Dance*, 62(2), 19-21.

Gabbett, T. J., & Georgieff, B. (2006). The development of a standardized skill assessment for junior volleyball players. *International journal of sports physiology and performance*, 1(2), 95-107.

Gabbett, T., Georgieff, B., & Domrow, N. (2007). The use of physiological, anthropometric, and skill data to predict selection in a talent-identified junior volleyball squad. *Journal of Sports Science*, 25, 1337 – 1344.

Harrison, J. M., Preece, L. A., Blakemore, C. L., Richards, R. P., Wilkinson, C., & Fellingham, G. W. (1999). Effects of two instructional models—skill teaching and mastery learning—on skill development, knowledge, self-efficacy, and game play in volleyball. *Journal of teaching in Physical Education*, 19(1), 34-57.

Hopkins, W. (2000). Reliability from consecutive pairs of trials. *A new view of statistics*.

Hulteen, R. M., Barnett, L. M., True, L., Lander, N. J., del Pozo Cruz, B., & Lonsdale, C. (2020). Validity and reliability evidence for motor competence assessments in children and adolescents: A systematic review. *Journal of Sports Sciences*, 38(15), 1717-1798.

Lidor, R., & Mayan, Z. (2005). Can beginning learners benefit from preperformance routines when serving in volleyball?. *The Sport Psychologist*, 19(4), 343-363.

Lidor, R., Hershko, Y., Bilkevitz, A., Arnon, M., & Falk, B. (2007). Measurement of talent in volleyball: 15-month follow-up of elite adolescent players. *Journal of sports medicine and physical fitness*, 47(2), 159.

Milić, V., Nešić, G., Murić, B., Kahrović, I., Radenković, O., Mavrić, F., & Memić, M. (2016). The relation between setting precision and spiking and effectiveness in the game of volleyball. *Fizička kultura,* 70(1), 70-79.

Shepard, J.M., Gabbet, T.J., & Stanganelli, LC.R. (2009). An analysis of playing positions in elite men's volleyball: considerations for competition demands and physiologic characteristics. *Journal of Strength and Conditioning Research,* 23(6).1858- 1866.

Sopa, I. S., & Pomohaci, M. (2018). Evaluation of Motor Development and Skills in Mini-Volleyball Game (10-12 Years Old). Bulletin of the Transilvania University of Braşov. Series IX: Sciences of Human Kinetics, 95-104.

Yudiana, Y., Hidayat, Y., Hambali, B., & Slamat, S. (2017, March). Content validity estimation of assessment instrument based on volleyball information system of volleyball learning: Field research. In IOP conference series: materials science and engineering (Vol. 180, No. 1, p. 012230). IOP Publishing.

Zetou, E., Michalopoulou, M., Giazitzi, K., & Kioumourtzoglou, E. (2007). Contextual interference effects in learning volleyball skills. *Perceptual and motor skills*, 104(3), 995-1004.

Zetou, E., Tzetzis, G., Vernadakis, N., & Kioumourtzoglou, E. (2002). Modeling in learning two volleyball skills. *Perceptual and motor skills*, 94(3_suppl), 1131-1142.

Zonifa, G. (2020). A volleyball skills test instrument for advanced-level students. *Journal of Physical Education & Sport*, 20.

TEST-RETEST I INTER-RATER POUZDANOST TESTOVA PRECIZNOSTI U ODBOJCI KOD ŠKOLSKE DECE

Cilj ovog istraživanja bio je ispitati test-retest i inter-rater pouzadnost testova preciznosti u odbojci na uzorku dece školskog uzrasta. Studija je obuhvatila uzorak od 46 ispitanika uzrasta 12-13 godina (telesna vsina: 158.97 ± 5.39 cm; telesna masa: 52.20 ± 5.58 kg; godine starosti: 12.4 ± 0.7 godina) koja su najmanje godinu dana učestvovala u školskim programima odbojke. Testiranje je sprovedeno u dva odvojena navrata radi procene test-retest pouzdanosti, pri čemu je svaki put praćena ista procedura. Za procenu inter-rater pouzdanosti, dva merioca su nezavisno beležila učinak ispitanika, osiguravajući konzistentno i nepristrasno bodovanje. Testirane varijable podrazumevale su gornji i donji prijem, dizanje lopte i servis. Perfektna test-retest pouzdanost (Intraklasni koeficijent korelacije (ICC) > 0.90; koeficijent varijacije (CV)% < 10%) uočena je za dizanje i servis. Pored toga, pronađena je skoro perfektna pouzdanost za gornji prijem (ICC = 0.84) i donji prijem (ICC = 0.72) sa visokim CV% (> 10%). Rezultati inter-rater pouzdanosti pokazali su veoma visoke ICC vrednosti (> 0.89) za sve varijable. Studija je pokazala da protokol testiranja veština u odbojci pruža konzistentne i pouzdane procene za igrače, bez značajnih razlika između test i retest sesija u svim procenjivanim veštinama.

Ključne reči: *testiranje, odbojka, školska deca, pouzdanost*