Original Article

Validity and reliability of the volleyball serve accuracy-test

CICERO LUCIANO ALVES COSTA[1], NADIA CRISTINA VALENTINI[2], EMANUEL MOITA DO NASCIMENTO[3], HERBERT UGRINOWITSCH[4]

[1]*Department of Physical Education. Regional University of Cariri. Crato, Brazil.*

[2]*School of Physical Education, Physiotherapy and Dance. Federal University of Rio Grande do Sul. Porto Alegre, Brazil.*

[3]*Department of Physical Education and Sports. Federal Institute of Education, Science and Technology of Ceará. Limoeiro do Norte, Brazil.*

[4]*School of Physical Education, Physiotherapy and Occupational Therapy. Federal University of Minas Gerais. Belo Horizonte, Brazil.*

ABSTRACT

This study aimed to examine the content, face, and sensitive validity, and test-retest reliability of a volleyball overhand serve accuracy-test. Four experts – university professors evaluated the content validity; 50 volleyball coaches, teachers, and athletes were enrolled in the face validity. Thirty-three undergraduate students (18 experienced and 15 novices) participated in the study to assess the test capability to detect different performance levels. The sensitive validity was examined by comparing differences in the serving accuracy among experience levels using the t-test for independent samples. A cluster analysis (hierarchical cluster) was conducted using the between-groups linkage method and the Euclidean quadratic distance measure. The test-retest reliability was analysed using intraclass correlation coefficients and the standard error. Satisfactory results were found for experts' agreement regarding the test's accuracy and content validity. Face validity was adequate regarding clarity, pertinence, and applicability (78% to 86% of agreement). A high degree of test-retest reliability was observed for both novices (ICC = 0.81) and experienced (ICC = 0.84) participants; experienced group (p = .001) had higher means. The volleyball serve accuracy-test is a viable alternative to assess beginners' performance in a teaching-learning context.

Keywords: Performance analysis, Sport performance, Test-retest, Motor learning.

Cite this article as:
Costa, C. L. A., Valentini, N. C., Do Nascimento, E. M., & Ugrinowitsch, H. (2024). Validity and reliability of the volleyball serve accuracy-test. *Journal of Human Sport and Exercise, 19*(2), 570-578. https://doi.org/10.14198/jhse.2024.192.12

Corresponding author. *School of Physical Education, Physiotherapy and Occupational Therapy. Federal University of Minas Gerais. Belo Horizonte, Brazil. https://orcid.org/0000-0003-0317-1940*
E-mail: herbertu@ufmg.edu.br
Submitted for publication December 06, 2023.
Accepted for publication January 07, 2024.
Published April 01, 2024 *(in press* January 26, 2024).
JOURNAL OF HUMAN SPORT & EXERCISE ISSN 1988-5202.

doi: 10.14198/jhse.2024.192.12

INTRODUCTION

Researchers of motor learning have aimed to run experiments with complex motor tasks. Some studies used a daily motor task such as salon dart throwing (Coca-Ugrinowitsch et al., 2014; Romano Smith et al., 2019). Other studies used complex motor skills from different sports, such as golf (Porter et al., 2007), tennis (Buszard et al., 2017; Hebert et al., 1996), badminton (Goode & Magill, 1986), baseball (Hall et al., 1994) and volleyball (Afrouzeh et al., 2020). Regarding volleyball, some studies adopted open motor skills such as the bump and two-hand volley/set (Bortoli et al., 1992; French et al., 1990), but the volleyball serve was the most used motor skill. The volleyball serve is characterized by the environment stability while it is executed, resulting in solid generalizability and specific practical applicability.

Although some tests assess volleyball skills such as spike or bump (Gabbett & Georgieff, 2006; Palao & Valadés, 2012; Palao & Valades, 2016), most motor learning studies have used the volleyball serve (Alishah et al., 2017; Costa et al., 2018; Cruz et al., 2018; Ghorbanzadeh et al., 2017), adopting the performance accuracy as the dependent variable. However, few instruments were submitted to a validity process. Test validity is essential to improve the ecological soundness of tests for teaching contexts (Robertson et al., 2014; Silverman, 1994).

Regarding validated tests for the volleyball serve, few instruments are available, such as the AAHPERD Serve Test (AAHPERD, 1969) and the Serve Test of the Volleyball Skills Test Battery (Bartlett et al., 1991); the score in these tests is related to the most challenging areas for reception in the actual game situation. Although the court area is a relevant feature considering the context of the game, it brings limitations in measuring changes in the accuracy characteristics of the learning process. For instance, the highest score when the ball reaches the back edge of the court (Bartlett et al., 1991) may impair the reproducibility and validity of the test because experienced individuals may attempt to reach the back edge of the court and miss by one centimetre, resulting in a score of zero. On the other hand, novices can aim for areas with lower scores, ensuring a minimum score (Morrow Jr et al., 2015). Thus, in the initial learning phase characterized by inaccuracy and lack of performance consistency (Fitts & Posner, 1967), low-sensitivity tests can make it difficult to assess the learning process of the serve.

Developing new tests and examining their validity and reliability are necessary to assess learning processes and verify the differences in performance related to the novice's accuracy are critical to practitioners. This investigation is also relevant to motor learning researchers since few validated instruments to assess accuracy in complex skills with adequate sensitivity to identify small changes in performance are available. The present study examines the content, face, and sensitive validity, and test-retest reliability of a volleyball overhand serve accuracy-test.

METHODS

Participants

Four university professors (three experts in Motor Behaviour and one in Sports Pedagogy with emphasis on Volleyball) assessed the validity of the test regarding clarity, pertinence, and applicability; a Likert scale (4 points) was used. Further, 50 professionals, volleyball coaches' youth and adults' teams (N = 21), volleyball teachers working in schools (N = 13), and volleyball athletes (N = 16) independently participated in the face validity procedure using the same Likert scale to assess the face validity of the volleyball overhand serve accuracy-test. All coaches and teachers had large professional experience (5 to 10 years = 43%; 10 to 20 years = 12%; 20 to 30 years = 30%; 30 to 40 years = 15%), they also had previous experiences as athletes

(14% professionals, 86% amateurs). To analyse sensitive validity, 33 volunteers, undergraduate students, self-declared right-handers, aged between 18 and 35 years, participated in the study. Two groups were formed, characterized by two-level of experience. The experienced group had 18 volunteers who participated in regular volleyball training (minimum of two weekly training sessions during the last two years) enrolled in a local university. The novice group had 15 volunteers with no volleyball training experience. The Ethics Committee in Research of the Federal University of Minas Gerais approved the project (Code: 1.939.735), and all procedures followed the APA ethical guidelines 7th Edition.

Measures

The volleyball serve was performed with participants positioned five meters away from the "*A*" side of the court, with both feet resting on the ground and facing the target laid on the floor on the "*B*" side (Figure 1). The task required to perform the overhand volleyball serve and hit the target bull's eye (located four meters from the net on the "*B*" side of the court).

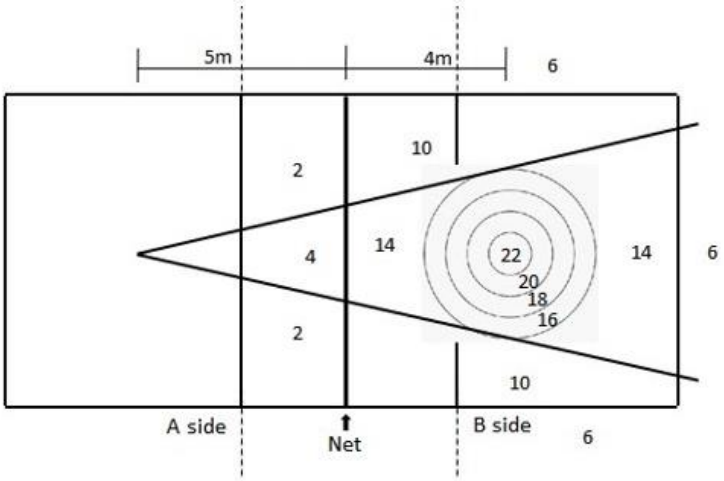

Figure 1. Representation of the area for execution of task.

The target had four circular areas of 1, 2, 3, and 4 meters in diameter. Scores related to accuracy concerning the target's bull's eye ranged from 2 to 22 points. Adhesive tape delimited a triangular area whose vertex originated at the position of the serve and passed through the side of the target. The width of all lines was five centimetres, and if the ball landed on a line, a higher score was obtained. Along with the trials, the serves that reached the target's central area scored 22 points, and as the graduations moved away from the central region, the score decreased by 2 points. The serves that reached the triangular area outside the target and within the court scored 14 points; if the serve touched the upper band of the net and landed toward the target; the score was 12 points. The serves that reached the areas in the court and outside of the triangle scored 10 points – if the serve touched the net and went over the net but did not land in the target's direction, the score was 8 points. Off-court serves were equivalent to 6 points. The triangular area on the "*A*" side of the court scored 4 points. The areas on the "*A*" side outside the triangle were 2 points. The experiment was run on an official volleyball court with the net set at the height of 2.24 m, using adhesive tape, a target, and official volleyball balls.

Procedure

Experts were invited to be a part of the content validity. After the final version was obtained, the target population (volleyball coaches, teachers, and athletes) were invited to assess face validity. Then, volunteers

were invited to participate in the study. Information regarding the study objectives, procedures, and risks was provided; all participants signed the informed consent. Before trials started, participants were instructed about the task goal and with a clear description of the kind of serve to be performed - overhand. After this initial instruction, each participant was directed individually to the serve position; each participant's performance was private. One researcher stranded on the left side of the participant and provided a ball, controlled the inter-trials interval (six to eight seconds), and prompted the action with the commends "*prepare*" and "*go*" before each trial. Another researcher stood on the court at side "*B*", close to the net and the triangle line, and scored each serve immediately after the trials. The experimenter did not provide feedback regarding the trials, but the location where the ball hit the ground was available throughout the trials. After seven days, the retest was conducted following the same protocol. Three participants from the experienced group and one participant from the novice group did not attend the retest; no information was provided regarding the reasons; their data were not included in the data analysis.

Analysis

Content and face validity were conducted using a Likert scale, ranging from 1 (totally disagree) to 4 (totally agree), with experts and the target public (coaches, teachers, and athletes) to assess test pertinence, clarity, and applicability. For content validity coefficient (CVC) was calculated (Hernández-Nieto, 2002), adopting above .80 as a satisfactory coefficient (Landis & Koch, 1977). Percent agreement was used to analyse the face validity of the professionals' responses.

Descriptive statistics of mean and standard deviation were provided. The normality of the data was verified with the Shapiro-Wilk test; parametric inferential statistics were used. The discriminant validity of the test was examined by comparing differences in the serve accuracy between experienced and novice groups using independent t-tests. A hierarchical cluster analysis was conducted using the between-groups linkage method and the Euclidean quadratic distance measure, with 2 clusters through the single solution method. A Chi-square test was used to verify associations between groups and clusters. The test-retest reliability was analysed using intraclass correlation coefficients (ICC) and the standard error of measurement (Weir, 2005). The statistical package SPSS for Windows (version 22.0) was used for all analyses, and the level of significance was set at $p < .050$.

RESULTS

Content and face validity

Initially, experts assess the test content validity. According to the experts' recommendations, the original test underwent two evaluations, undergoing modifications in its description and punctuation. After the two rounds, the final version was obtained. Experts agree that the test measures the accuracy of the volleyball serve; satisfactory levels of the agreement were obtained for clarity (CVC = .93), pertinence (CVC= 1.0), and applicability (CVC = .93). The face validity evaluated by the 50 professionals resulted in 78% concordance among them for clarity, 78% for pertinence, and 86% for applicability.

Sensitive validity

For sensitive validity, the independent variable - levels of experience - was analysed and the results showed the accuracy of the serve test. The t-test showed higher scores for the experienced group than the novice group at the test and the retest, with a large effect size(J, 1988). Cohen's index d showed a high magnitude effect of experience level on volleyball serve accuracy for both the test (d = 2.20) and retest (d = 1.59).

Table 1. Groups comparisons: Means, standard deviations, p values and Cohen's effect size (d).

	Groups	N	Mean	Standard deviation	*p*	Cohen's d
Test	Novice	15	15.92	1.47	.001*	2.20#
	Experienced	18	18.84	1.17		
Retest	Novice	14	16.54	1.93	.001*	1.59#
	Experienced	15	19.10	1.21		

*Note: * Student's t-test p-values; # Cohen's d large effect size.*

Table 2 shows that the cluster analysis grouped most participants in a similar way to the original division of the groups by level of experience. In the cluster analysis by the level of experience groups, a high number of the cases were correctly classified (90.9%). Only one participant from the experienced group was grouped in cluster 1, whereas among the novices, two participants were grouped in cluster 2. A significant association between clusters and groups was found (Chi2 p = .001).

Table 2. Test-retest reliability.

Groups	N	ICC	CI (95%)	SEM	%SEM
Novice	14	.81	.41 – .94	.74	4.60
Experienced	15	.84	.53 – .95	.49	2.58
Total	29	.91	.81 – .96	.59	3.38

Note: N = participants; ICC = Intraclass Correlation Coefficient; CI = Confidence Interval; SEM = Standard Error Measure; % SEM = percentage of SEM in relation to the mean.

Reliability

Intra-class correlation coefficients, their respective confidence intervals, and the standard error of measure are presented in Table 3. A high degree of test-retest reliability was observed for both novice and experienced groups. For the temporal stability, the confidence interval was smaller for the total sample and experienced groups than for the novice groups; consequently, although adequate for novices, the test-retest was more robust for experiences and the total sample.

Table 3. Frequency distribution of clusters according to the level of experience.

	Cluster 1	Cluster 2	Total	*p*
Inexperienced	13 (86.7%)	2 (13.3%)	15	*.001
Experienced	1 (5.6%)	17 (94.4%)	18	

*Note. * Pearson's Qui-square = 22.037. Ninety-one percent of the originally grouped cases were correctly classified.*

DISCUSSION

This study examined the content, face, and sensitive validity, and test-retest reliability of a volleyball overhand serve accuracy-test. The results demonstrate that the target accuracy-test is a valid and reliable alternative for motor learning research and monitoring players' progress. The experts agreed that the test could assess the serve accuracy, and the CVCs results for clarity, pertinence, and applicability were satisfactory - all results above recognized cutoffs (Hernández-Nieto, 2002; Landis & Koch, 1977). The group of professionals confirmed face validity; the test was considered to have pertinent content to assess serve accuracy with clear and understandable instruction.

The results also demonstrated the sensitivity of the volleyball serve accuracy-test to differentiate the experience levels. This form of validity is as essential as other psychometric criteria (Milne & Reiser, 2011).

As observed in the present study, concerning to accuracy tasks such as volleyball serve and other sports skills, it is expected that experienced practitioners perform better than the novice. A plausible explanation is related to a greater capacity of the experienced individuals to identify, remember, and handle the relevant information of the environment (Arroyo et al., 2016). In addition, cluster analysis demonstrated that almost all participants were grouped consistently with their original group, and the serve test for accuracy was able to differentiate the participants' performance levels.

Volleyball coaches have used skill tests in different stages to assess the performance level of young athletes on the basic skills (Lidor et al., 2007). Regarding serve skill, accuracy tests have focused on the game demands, with higher scores in areas that impose greater difficulty for reception (AAHPERD, 1969; Bartlett et al., 1991). These kinds of assessments are consistent with the specific demands of the sports and are highly relevant for monitoring the athletes' performance. Improvements in the parameterization of aspects such as strength, speed, and direction of the ball are relevant to the motor learning process. Instruments used to assess motor skills acquisition within sport contextual goals need to identify improvements in performance accuracy, and the present test achieves this objective. It is vital to note that although Bartlett et al. (1991) test differentiate athletes' performance accuracy during the volleyball game, the one-centimetre difference between the maximum and minimum score (as mentioned early) is a problem in distinguishing performance during early learning. Thus, the relationship between the level of experience and the test sensitivity was examined in the present study might be considered in future studies. For example, the no difference in performance accuracy of elite and near-elite players (Lidor et al., 2007) may occur due to the lower sensitivity of the instrument. The accuracy test proposed in the present study overcomes these limitations, besides presenting suitable validity and reliability indexes.

The findings related to test-retest reliability were considered suitable for both groups and excellent for the total sample (Koo & Li, 2016). Good and excellent levels of reliability, like those obtained in the present study, have been restricted in several tests that assess sports motor skills performance, around only 41% of 22 studies analysed in a systematic review achieve similar indexes (Robertson et al., 2014). The indices of the present study are also superior to those obtained by the NCSU test (Bartlett et al., 1991) and resemble those found by Gabbett and Georgieff (2006) in a proposal for assessing volleyball skills for junior players. However, the latter study did not use a specific target to assess the accuracy of the serve, only whether the ball hit the opposite side of the court or not. Therefore, as well as the NCSU Volleyball Skills Test Battery (Bartlett et al., 1991), the test proposed by Gabbett and Georgieff (2006) also runs the risk of underestimating the performance of players with a better technical level and overestimating the performance of beginners.

The present study's strength relates to the test applicability in teaching-learning context. The instrument is a viable alternative to evaluate an extensive range of performance, from novices to experienced players. One limitation of the present study was the small sample size regarding the sensitive validity, restricting our generalization capability.

CONCLUSIONS

The volleyball serve accuracy-test presents adequate validity and test-retest reliability and is a feasible alternative for motor learning research on complex motor skills; the results from this line of research may support sports intervention more than lab tasks (Christina, 1987). We recommended that in early learning, due to the difficulty in reaching the opposite side of the court because of structural and functional constraints, a shorter distance from the net can be used to facilitate the testing process - this alternative needs to be investigated regarding its psychometrics capabilities. We also suggested that other studies continue

examining the psychometrics of the serve accuracy-test regarding criterion validity and the strength of the relationship between test accuracy and movement patterns quality.

AUTHOR CONTRIBUTIONS

Costa and Ugrinowitsch designed the study. Costa and Valentini performed the data collection and data analysis. All the authors wrote the paper and approve the final submission.

SUPPORTING AGENCIES

No funding agencies were reported by the authors.

DISCLOSURE STATEMENT

No potential conflict of interest were reported by the authors.

REFERENCES

AAHPERD. (1969). Skills test manual: volleyball for boys and girls. (p. 36). American Alliance for Health, Physical Education, Recreation and Dance.

Afrouzeh, M., Konukman, F., Lotfinejad, M., & Afroozeh, M. S. (2020). Effects of knowledge of results feedback on more accurate versus less accurate trials on intrinsic motivation, self-confidence and anxiety in volleyball serve. Physical Culture and Sport, Studies and Research, 87(1), 24-33. https://doi.org/10.2478/pcssr-2020-0016

Alishah, E. R., Ates, O., & Ahmadi, M. (2017). The effects of attentional focus on the performance of volleyball jump serve in elite players. European Journal of Physical Education and Sport Science, 3(12), 49. https://doi.org/10.5281/zenodo.1066314

Arroyo, M. P. M., Domínguez, A. M., Arias, A. G., García-González, L., & Del Villar Álvarez, F. (2016). Influence of experience in procedural knowledge in volleyball: Applications for training. International Journal of Sports Science and Coaching, 11(2), 191-199. https://doi.org/10.1177/1747954116637152

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. https://doi.org/10.1080/07303084.1991.10606554

Bortoli, L., Robazza, C., Durigon, V., & Carra, C. (1992). Effects of contextual interference on learning technical sports skills. Perceptual and Motor Skills, 75(2), 555-562. https://doi.org/10.2466/pms.1992.75.2.555

Buszard, T., Reid, M., Krause, L., Kovalchik, S., & Farrow, D. (2017). Quantifying contextual interference and its effect on skill transfer in skilled youth tennis players. Frontiers in Psychology, 8, 1931. https://doi.org/10.3389/fpsyg.2017.01931

Christina, R. W. (1987). Motor Learning: Future lines of research. In The academy papers: The cutting edge in physical education and exercise science research (pp. 26-41). Human Kinetics.

Coca-Ugrinowitsch, A. A., Benda, R. N., Aburachid, L. M., De Andrade, A. G. P., Greco, P. J., Menzel, H. J. K., & Ugrinowitsch, H. (2014). Bandwidth knowledge of results on the learning of the saloon dart throwing task. Perceptual and Motor Skills, 118(2), 462-474. https://doi.org/10.2466/25.23.PMS.118k17w9

Cohen, J. (1988). Statistical power analysis for the behavioral sciences. In Pschological Bulletin (2nd ed., p. 559). Lawrence Erlbaum Associates.

Costa, C. L. A., Bandeira, P. F. R., De Oliveira Matos, C., Cruz, M. P., & Ugrinowitsch, H. (2018). Validade de construto e fidedignidade de uma lista de checagem para análise do saque do voleibol. Revista Brasileira de Cineantropometria e Desempenho Humano, 20(1), 95-101. https://doi.org/10.5007/1980-0037.2018v20n1p95

Cruz, M. P., Lage, G. M., Ribeiro-Silva, P. C., Neves, T. F., Matos, C. de O., & Ugrinowitsch, H. (2018). Constant and random practice on learning of volleyball serve. Revista Brasileira de Cineantropometria e Desempenho Humano, 20(6), 598-606. https://doi.org/10.5007/1980-0037.2018v20n6p598

Fitts, P. M., & Posner, M. I. (1967). Human Performance (2nd ed.). Brooks/Cole Publishing Company.

French, K. E., Rink, J. E., & Werner, P. H. (1990). Effects of contextual interference on retention of three volleyball skills. Perceptual and Motor Skills, 71(1), 179-186. https://doi.org/10.2466/pms.1990.71.1.179

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. https://doi.org/10.1123/ijspp.1.2.95

Ghorbanzadeh, B., Bayar, P., & Koruç, Z. (2017). The effect of feedback on serve and bump skills training in volleyball. Journal of Physical Education and Sport, 17(3), 995-1001. https://doi.org/10.7752/jpes.2017.s3153

Goode, S., & Magill, R. A. (1986). Contextual interference effects in learning three badminton serves. Research Quarterly for Exercise and Sport, 57(4), 308-314. https://doi.org/10.1080/02701367.1986.10608091

Hall, K. G., Domingues, D. A., & Cavazos, R. (1994). Contextual interference effects with skilled baseball players. Percept Mot Skills, 78(3), 835-841. https://doi.org/10.1177/003151259407800331

Hebert, E. P., Landin, D., & Solmon, M. A. (1996). Practice schedule effects on the performance and learning of low- and high-skilled students: An applied study. Research Quarterly for Exercise and Sport, 67(1), 52-58. https://doi.org/10.1080/02701367.1996.10607925

Hernández-Nieto, R. A. (2002). Contributions to statistical analysis. Universidad de los Andes.

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. https://doi.org/10.1016/j.jcm.2016.02.012

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. Biometrics, 33(1), 159. https://doi.org/10.2307/2529310

Lidor, R., Arnon, M., Hershko, Y., Maayan, G., & Falk, B. (2007). Accuracy in a volleyball service test in rested and physical exertion conditions in elite and near-elite adolescent players. Journal of Strength and Conditioning Research, 21(3), 937-942. https://doi.org/10.1519/R-19455.1

Milne, D. L., & Reiser, R. P. (2011). Observing competence in CBT supervision: a systematic review of the available instruments. Cognitive Behaviour Therapist, 4(3), 89-100. https://doi.org/10.1017/S1754470X11000067

Morrow Jr, J. R., Mood, D. P., Disch, J. G., & Kang, M. (2015). Measurement and Evaluation in Human Performance (5th ed.). Human Kinetics.

Palao, J. M., & Valadés, D. (2012). Validity of the standing spike test as a monitoring protocol for female volleyball players. Biology of Sport, 29(4), 281-284. https://doi.org/10.5604/20831862.1019666

Palao, J. M., & Valades, D. (2016). Validity of the standing and jump spike tests for monitoring female volleyball players of different levels of competition. International Journal of Performance Analysis in Sport, 16(3), 1102-1108. https://doi.org/10.1080/24748668.2016.11868951

Porter, J. M., Landin, D., Hebert, E. P., & Baum, B. (2007). The effects of three levels of contextual interference on performance outcomes and movement patterns in golf skills. International Journal of Sports Science & Coaching, 2(3), 243-255. https://doi.org/10.1260/174795407782233100

Robertson, S. J., Burnett, A. F., & Cochrane, J. (2014). Tests examining skill outcomes in sport: a systematic review of measurement properties and feasibility. Sports Medicine, 44(4), 501-518. https://doi.org/10.1007/s40279-013-0131-0

Romano Smith, S., Wood, G., Coyles, G., Roberts, J. W., & Wakefield, C. J. (2019). The effect of action observation and motor imagery combinations on upper limb kinematics and EMG during dart-throwing. Scandinavian Journal of Medicine and Science in Sports, 29(12), 1917-1929. https://doi.org/10.1111/sms.13534

Silverman, S. (1994). Communication and motor skill learning: What we learn from research in the gymnasium. Quest, 46(3), 345-355. https://doi.org/10.1080/00336297.1994.10484131

Weir, J. P. (2005). Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. J Strength Cond Res, 19(1), 231-240. https://doi.org/10.1519/00124278-200502000-00038

