

# Determination of Trials Needed for Measurement Consistency of Standing Long Jump in Female Collegiate Volleyball Athletes: *A Brief Report*

Robert Ah Sue<sup>1</sup>, Chad Harris<sup>2</sup>, Joseph Berning<sup>3</sup>, Trish Sevene<sup>4</sup>, Kent J. Adams<sup>4</sup>, Mark DeBeliso<sup>1,\*</sup>

<sup>1</sup>Department of Physical Education and Human Performance, Southern Utah University, Cedar City, UT, USA

<sup>2</sup>College of Professional Studies, Metropolitan State University of Denver, Denver, CO, USA

<sup>3</sup>Department of Kinesiology and Dance, New Mexico State University, Las Cruces, NM, USA

<sup>4</sup>Kinesiology Department, California State University Monterey Bay, Seaside, CA, USA

**Abstract** Lower body power is an important physical attribute of competitive volleyball athletes. In order to assess such an attribute, it is important that the test reliability be established using a sport specific physical test and a participant pool consisting of competitive athletes. Furthermore, identification of the appropriate number of trials to use during assessment insures accuracy of scores and time efficiency for coaches. **PURPOSE:** The aim of this investigation was to establish the minimum number of trials needed for reliability of the standing long jump (SLJ) in female collegiate volleyball players. **METHODS:** Nine female athletes (age:  $18.3 \pm 0.5$  yrs, height:  $179.2 \pm 5.3$  cm, mass:  $68.8 \pm 7.9$  kg) completed five trials of a SLJ. Jumps were performed at 2, 6, 10, 14 and 18 minutes following a dynamic warm-up. In an attempt to simulate the active nature of sport, the athletes alternated between walking and jogging the length of a basketball court between successive trials. For the SLJ, subjects performed a standard countermovement SLJ and distance in centimeters was measured from the starting line to the rear-most heel upon landing. Pearson (PCC) and Intraclass (ICC) Correlation Coefficients were calculated between successive trials. **RESULTS:** Means for the for the successive trials were as follows: Trial 1:  $191.4 \pm 10.1$  cm; Trial 2:  $196.1 \pm 10.2$  cm; Trial 3:  $199.9 \pm 10.3$ ; Trial 4:  $199.6 \pm 11.8$  cm; Trial 5:  $200.7 \pm 9.6$  cm. Mean difference between successive trials (i.e. Trials 2 to 1, 3 to 2, 4 to 3 and 5 to 4) were;  $4.8 \pm 6.9$ ,  $3.8 \pm 5.5$ ,  $-0.3 \pm 5.6$  and  $1.1 \pm 5.4$  (cms). PCC and ICC for successive trials were: 0.77 and 0.82; 0.86 and 0.89; 0.88 and 0.91; 0.89 and 0.91. **CONCLUSIONS:** Given the parameters of this study, the SLJ is a reliable field test of horizontal muscular power output and adequate assessment of SLJ performance is achieved using three trials.

**Keywords** Broad jump, Reliability, Horizontal plane power output

## 1. Introduction

Lower body power is an important attribute of competitive athletes, and is the focus of key elements of the athlete's strength and power training program. Appropriate testing to evaluate lower body power is crucial in supplying data for development of the athlete's individualized exercise prescription. With specific data, one can better manipulate the acute program variables in an attempt to optimize training outcomes related to muscular power, and ultimately game performance.

In order to best evaluate any physical attribute, it is important that test reliability is established using a sport

specific athlete group and used in a realistic application to the particular sport. Furthermore, identification of the appropriate number of trials to use during assessment insures accuracy of scores and efficient use of time by coaches.

The standing long jump (SLJ) is a common test for the assessment of lower body power, specifically focusing on horizontal leg power [1-4]. The SLJ is a relatively easily administered test, feasible in a variety of settings, with basically no cost. Whilst literally everyone can test their athlete's SLJ anywhere with a simple yard stick and chalk, even a reliable vertical jump test requires equipment costing a few hundred dollars. To better enhance the feasibility of use of the SLJ by practitioners, the purpose of this investigation was to establish the minimum number of trials needed to establish meaningful reliability of the SLJ in female collegiate volleyball players.

\* Corresponding author:

markdebeliso@suu.edu (Mark DeBeliso)

Published online at <http://journal.sapub.org/sports>

Copyright © 2017 Scientific & Academic Publishing. All Rights Reserved

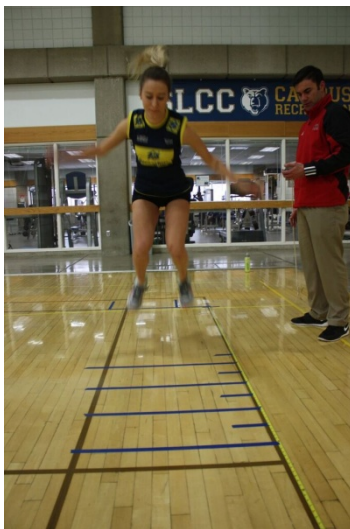
## 2. Methods

### 2.1. Participants

Participants were a sample of female volleyball athletes at a Community College in the Intermountain West of the USA. Permission from the school athletic administration and team coach was granted prior to the study initiation. The athletes were verbally solicited to volunteer for the study. Permission from the Institutional Review Board for use human subjects in research was obtained prior to conducting any training or assessments of the participants. Participants provided written consent prior to study involvement. Further, all of the student athletes were 18 years or older.

### 2.2. Instruments and Apparatus

Data collection sessions were held in the Activities Center at the Community College. The equipment necessary to conduct this study included a measuring tape, blue painter masking tape to set up jump markers (see Figure 1), and a measuring stick.



**Figure 1.** Athlete setting up and attempting the standing long jump

### 2.3. Procedures

The initial meeting with the participants consisted of an orientation period where age, height, and weight were recorded. The investigators then reviewed the proper technique for the execution of the SLJ with the participants.

Following the orientation the participants reassembled for the collection of SLJ trials. The data collection period started with a dynamic warm-up (WU). The dynamic WU protocol was identical to the WU protocol implemented in the athlete's regular season physical condition program; intended to maintain consistency/specificity. The WU consisted of (in specific order): 20-meter high knees, 20-meter butt kickers, 20-meter lunges, 20-meter karaoke, 20-meter A-skips, 20-meter side shuffle, 30 second line hops (both feet), 15 second line hops (one set each leg), 10-meter broad jumps, 10-squat hops, leg sweeps (hamstring warm-up), 20-meter reverse lunges, 20-meter walking quad/hamstring stretch, 20-meter straight leg bounding, 10-leg swings, 10-medium arm circles and 10-large arm circles.

Five trials of SLJs were then collected at 2, 6, 10, 14, and 18 minutes. Athletes started in standstill position with toes just behind the starting line [1]. The athlete then performed a countermovement and jumped forward as far as possible [1]. The athlete was required to land on their feet for the jump to be scored; otherwise the trial was repeated [1]. In the event of a fault the athlete was allowed only one immediate re-jump to avoid fatigue or hindrance of another athletes timing. A mark was placed at the back edge of the athlete's rear most heel, and the tape measure determined the distance between the starting line and the mark [1].

In order to mimic an "active" athletic setting, instead of seated rest between jumps, athletes alternated between walking and jogging the length of the basketball court. Thirty seconds prior to their jump, athletes were notified to prepare for their SLJ. SLJs were strictly regulated to the 5 established time markers of 2, 6, 10, 14, and 18 minutes.

### 2.4. Statistical Analysis

A combination of statistical approaches were used to determine the needed number to trials to achieve reliable measurements of SLJ scores. We examined all the trials until we found the minimal number of trials to reach a meaningful level of reliability between two consecutive trials of SLJs. Initially, interclass ( $r$ ) and intraclass reliability coefficients (ICC) were calculated between sequential pairs of SLJ trial scores until a plateau was achieved (i.e. Trials 2 to 1, 3 to 2, 4 to 3 and 5 to 4). Once two consecutive trials were identified the following analysis/calculations were conducted: mean difference between trials and the standard error of the measure ( $SE_m$ ) was calculated, a scatter diagram was plotted to visually inspect linearity between trails, Bland-Altman plots were constructed in order to explore uniformity of error, and trial data was log-transformed in order to assess typical error as expressed as a coefficient of variation percent (CV%). Further, 90% confidence limits were calculated for the aforementioned reliability statistics. Statistics were calculated in Excel with a spreadsheet prepared by Hopkins [5]. This particular reliability analysis is near identical to that reported by White et al. [6] and Tano et al. [7].

### 3. Results

Nine college female volleyball players participated in the study. The average age, height, and body mass of the participants is presented in Table 1. All of the participants completed all SLJ trials without incident.

**Table 1.** Participant Descriptive Information

N	Age (years)	Height (cms)	Mass (kgs)
9	18.3 ± 0.5	179.2 ± 5.3	68.8 ± 7.9

College female volleyball players, N=9 (mean±sd)

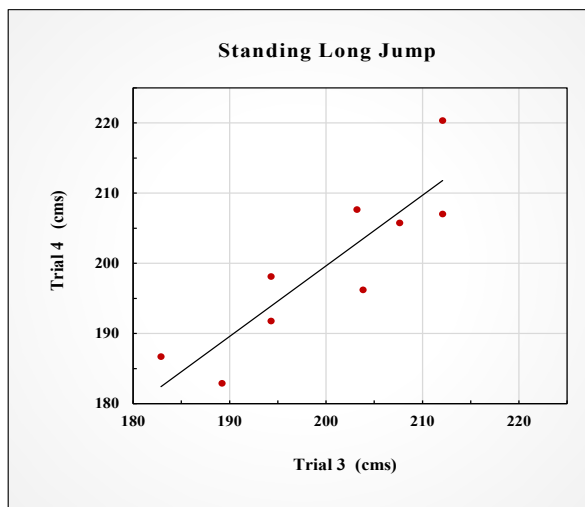
The SLJ trial scores are presented in Table 2. Visual inspection of the trial mean scores showed a close agreement between trials 3 and 4 that were recorded at 10 and 14 minutes following the dynamic WU (and confirmed by intra and inter class reliability coefficients reported in the abstract). As such, further reliability analyses were applied to the SLJ pairing of trials 3 and 4.

**Table 2.** Standing Long Jump (SLJ) Trial Scores

2 Min	6 Min	10 Min	14 Min	18 Min
191.4 ± 10.1	196.1 ± 10.2	199.9 ± 10.3	199.6 ± 11.8	200.7 ± 9.6

SLJ mean trial scores in centimeters (mean±sd)

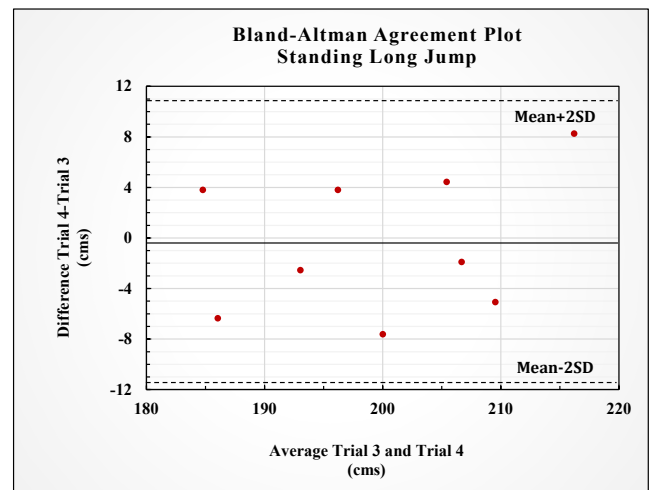
The reliability analysis is based one SLJ trials 3 and 4 that were recorded at 10 and 14 minutes following the dynamic WU (n=9). The mean of trial 3 and trial 4 scores were  $199.9 \pm 10.3$  and  $199.6 \pm 11.8$  centimeters respectively, with a mean difference between trials of  $-0.3 \pm 5.6$  cms.



**Figure 2.** Scatter Plot SLJ Trial 3 and 4 Scores

Figure 2 is a scatter plot of trial 3 and 4 of the SLJ scores while Figure 3 is a Bland-Altman plot comparing the trial average scores versus the difference scores. The scatter plot suggests a linear relationship between the trial scores. The Bland-Altman plot demonstrates that 0 of 9 data pairs exceeded the 95% limits of agreement [8]. Neither the scatter plot nor the Bland-Altman plot suggested bias or heteroscedasticity (non-uniform error). Of the 9 pairs of SLJ trials scores, 5 were lower during trial 4, and 4 were greater

during trial 4.



**Figure 3.** Bland-Altman plot comparing the trial average scores versus the difference scores

**Table 3.** SLJ Reliability Statistics

Statistic		Upper Limit	Lower Limit
$\Delta$ Means (cms)	$-0.4 \pm 5.6$	3.1	-3.8
r	0.88	0.97	0.61
ICC	0.91	0.97	0.71
Typical Error (CV%)*	2.0	3.4	1.4
SE <sub>m</sub>	3.95	6.76	2.84

90% Confidence limits for selected reliability statistics UL-upper limit, LL-lower limit. \*Typical error expressed as a CV% based on Log-transformed data. SEM-standard error of the measure. r- Pearson correlation coefficient. ICC- Intraclass correlation coefficient. CMS-centimeters

Table 3 provides the reliability statistics with 90% confidence limits (UL, LL). The intraclass reliability coefficient was ICC=0.91 (0.97, 0.71), which is considered “above average acceptable” [9]. The interclass reliability coefficient was  $r=0.88$  (0.97, 0.61), which is considered very high [10]. The standard error of the measure was  $SE_m=3.9$  (6.8, 2.8) cms.

While the trial data did not exhibit non-uniform error, we still wanted to quantify typical error; hence the data was Log-transformed as suggested by Hopkins [11]. The typical error as expressed as a coefficient of variation was CV%=2.0 (3.4, 1.4) percent.

### 4. Discussion

There was a twofold purpose to the current study. First, determine the reliability of the SLJ in female collegiate volleyball players. Second, assuming a meaningful reliability existed for the SLJ test, what was the appropriate number of trials to conduct in order to achieve a consistent SLJ score? The SLJ data collected in this study indicated that the SLJ was a highly reliable test and that a consistent score could be achieved at 3 trials. The SLJ scores collected in the current study (trials 3 and 4) were comparable to 20<sup>th</sup> % ile

SLJ scores of elite female athletes and would have been considered “excellent” when compared to 15-16 year old female athletes as reported by Hoffman [12].

The interclass reliability coefficient in the current study ( $r=0.88$ ) is considerably higher than test-retest reliability coefficients reported for other commonly used physical performance tests [10]. Further, it should be noted that the lower limit of the 90% confidence interval for the interclass reliability coefficient in the current study is ( $L_L=0.61$ ), which is considered to be moderate-high [10].

The intraclass reliability coefficient in the current study is  $ICC=0.91$ , which is also higher than the  $ICC$ 's reported for other commonly used physical performance tests [9]. Again, it should be noted that the lower limit of the 90% confidence interval for the intraclass reliability coefficient in the current study is ( $ICC L_L=0.71$ ), which is considered below-average acceptable [9]. Of note is agreement between the current findings and those of Markovic *et al.* [13]. Markovic *et al.* [13] reported an  $ICC=0.93$  for the SLJ in a sample of physically active college aged men ( $n=91$ ).

The standard error of measure ( $SE_m$ ) is a measure of absolute reliability [14]. The  $SE_m$  in the current study suggests that when measuring the SLJ as described in this study (and in this population) that the true score is within  $\pm 3.9$  cms of the measured score (68% confidence). The implication here is that if a coach is attempting to use the SLJ to measure of change horizontal plane power output then the score change must exceed 3.9 cms or it could likely be that the change in scores recorded was simply measurement error. It should also be noted that when scores reflect non-uniform error (not the case in the current study), the  $SE_m$  is biased and likely over estimates error in the lower scores and underestimates the error in the higher scores. Although we did not encounter evidence in the data suggesting bias or non-uniformity of error, we choose to attempt to make a correction to the data to ensure that our reliability examination was extensive.

In order to correct for non-uniformity of error Hopkins [11] suggests Log-transforming the scores and expressing the error as ‘typical error’ or a CV%. The CV% with 90% confidence limits represents ( $CV\%=2.0$ : 3.4, 1.4) a non-bias measure of error that should be used as a bench mark when attempting to measure change in the SLJ scores in this population. It should be noted here that the  $CV\%=2.0$  would be  $\approx 3.99$  cms which is approximately equal to the  $SE_m$  in the current study ( $SE_m=3.95$  cms). That the CV% was approximately equal to the  $SE_m$ , is a confirmation that the trial data analyzed in this study was void of bias and/or non-uniformity of error.

As we initially analyzed the data we considered removing what we perceived as an outlier score. However, we choose not to because of the limited number of participants (the results reported in the preceding section include the outlier score). As a point of interest, we did re-run the statistics with the outlier score removed. The interclass correlation coefficient was  $r=0.87$  (0.97, 0.55), and the intraclass correlation coefficient was  $ICC=0.91$  (0.98, 0.71). Both

coefficients were essentially unchanged and confirmed the robustness of the initial statistical analysis.

The trials were essentially split in that of the 9 participants 5 had greater trial 3 scores and 4 had greater trial 4 scores. As such, it would appear that fatigue did not impact the trial scores. A potential indicator that the rest periods between trials and the warm-up selected for this study were appropriate.

The participants in this study were college aged trained athletes who were familiar with the SLJ and the mechanics required to execute the movement. The physical maturity and familiarity of the participants with the SLJ likely added to the extremely reliable results noted in the current study. Future studies should include participants from different age, and sport/gender populations in order to fully enhance the feasibility of use of the SLJ by practitioners.

Another point worthy of discussing has to do with the actual value of collecting a fourth trial of SLJ scores. Generally speaking, individuals typically have better scores during subsequent repeated trials of a physical test due to some type of learning (in the absence of appreciable fatigue). Since that was not the case when comparing the 3<sup>rd</sup> trials to the 4<sup>th</sup> trials of SLJs, and given that the average difference in trial scores was -0.41 cms, then why conduct the 4<sup>th</sup> trials if one is looking for max performance? Conversely, given the ease of collecting multiple SLJ trials and in the absence of fatigue, why not collect multiple trials of SLJ scores to establish horizontal plane power output?

## 5. Conclusions

Within the parameters of this study, the SLJ should be considered a reliable test of horizontal plane muscular power output. Further, it appears that there is no need to collect more than 3 SLJ trials to establish a reliable maximum score. The SLJ is an easily administered test that has a high degree of task specificity to a number of sports. The authors recommend its use to strength and conditioning coaches who work with this population.

## ACKNOWLEDGEMENTS

We would like to thank Coach Sue Dulaney for allowing the facilities and equipment of Salt Lake Community College to be used freely during the study.

## REFERENCES

- [1] Baechle, T.R., and Earle, R.W., 2008, Resistance training and spotting techniques. In *Essentials of Strength and Conditioning*, 3rd ed., Human Kinetics: Champaign, IL, USA, pp. 255.
- [2] Coburn, J. W., 2012, Measuring power. *Strength & Conditioning Journal*, 34(6), 25-28.

- [3] Henriksson, T., Vescovi, J. D., Fjellman-Wiklund, A., and Gilenstam, K., 2016, Laboratory-and field-based testing as predictors of skating performance in competitive-level female ice hockey. *Open Access Journal of Sports Medicine*, 7, 81.
- [4] Ah Sue, R., Adams, K.J., and DeBeliso, M., 2016, Optimal timing for post-activation potentiation in women collegiate volleyball players, *Sports*, 4(27), 1-9.
- [5] Hopkins, W.G., 2007), Understanding statistics by using spreadsheets to generate and analyze samples. *Sportscience*, 11, 23-36.
- [6] White, K., DeBeliso, M., Sevene, T., and Adams, K.J., 2015, The Reliability of the 300-yard shuttle run in high school girls basketball players. *Journal of Sports Science*, 4(5), 214-218.
- [7] Tano, G., Bishop, A., Climstein, M., and DeBeliso, M., 2016, The reliability of the prowler in high school male football players. *Journal of Sports Science*, 4(4), 183-188.
- [8] Bland, J.M., and Altman, D.G., 1986, Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*, 8, 307-310.
- [9] Baumgartner, T.A., Jackson, A.S., Mahar, M.T., and Rowe, D.A., 2007, *Measurements for evaluation in physical education and exercise science* (8th Ed.). New York, NY: McGraw Hill.
- [10] Safrit, M.J., and Wood, T.M., 1995, *Introduction to measurement in physical education and exercise* (3rd Ed.). St. Louis, MO: Mosby.
- [11] Hopkins, W.G., 2013, A new view of statistics. Internet Society for Sport Science, <http://www.sportsci.org/resource/stats/>.
- [12] Hoffman, J., 2006, *Norms for fitness, performance, and health*. Champaign, IL: Human Kinetics.
- [13] Markovic, G., Dizdar, D., Jukic, I., and Cardinale, M., 2004, Reliability and factorial validity of squat and countermovement jump tests. *The Journal of Strength & Conditioning Research*, 18(3), 551-555.
- [14] Bruton, A, Conway, J.H., and Holgate, S.T., 2000, Reliability: What is it and how is it measured? *Physiotherapy*, 86(2), 94-99.