

Inter-rater reliability of trained and untrained raters for measuring jump height with the MyJump app

BASILIO PUEO, JOSE MANUEL JIMENEZ-OLMEDO, ALFONSO PENICHER-TOMAS, MARI CARMEN BERNAL-SORIANO

Sport Sciences, Faculty of Education, University of Alicante, SPAIN
Alicante University General Hospital, SPAIN

Published online: June 30, 2018

(Accepted for publication May 22, 2018)

DOI:10.7752/jpes.2018.02121

Abstract:

Smartphone apps have been very popular among researchers, coaches and clinicians. The aim of this study was to analyze the inter-rater reliability of trained and untrained raters of the iOS app *MyJump*. Eighteen participants completed eight counter-movement jumps (CMJ) each (n=144) which were recorded using the app. In order to examine inter-rater reliability derived from raters' experience, independent assessments were obtained by an untrained rater and two trained raters of the same recordings. Results showed very high agreement between trained raters (ICC=0.982, α =0.991) and between trained and untrained raters (ICC=0.984, α =0.992). Bland-Altman plots depicted negligible bias between both types of raters. Finally, very good internal consistency was found, regardless of experience (α from 0.995 to 0.997 and CV from 1.1 to 1.4%). This study has shown no significant difference in performance of trained and untrained raters of *MyJump*, and therefore no special training must be observed for an accurate use of the app.

Key words: instrumentation, agreement, consistency, smartphone, measure, observation

Introduction

Vertical jump tests are commonly used in sport sciences to measure physical performance of lower limbs. Both for athletic and non-athletic populations, an improvement in jump height corresponds with better functional performance (Çakir-Atabek, 2014). Therefore, vertical jump tests are used by sport professionals to obtain information about several neuromuscular and performance qualities of individuals (Driller, Tavares, McMaster, & O'Donnell, 2017).

There are a number of protocols and instruments to assess lower body power through vertical jump tests, although all of them lie in the classification of one of the following three methods (Baca, 1999). First, force plates measure vertical ground reaction forces and perform numerical integration to obtain jump height (Whitmer et al., 2014). Second, Motion Capture allows for the tracking of kinematics of center of gravity of the body to follow excursion during jump executions (Barris & Button, 2008). Third, flight-time based devices measure the time span between take-off and landing to calculate jump height through a basic kinematic equation (Hatze, 1998).

The latter instruments are very popular among physical trainers and sport scientists due to the portability, ease of use and low cost compared to laboratory-based equipment. The first instrument within this category was the jump mat, consisting of a mat serving as an electric switch operated by athlete weight (Pueo, Lipinska, Jiménez-Olmedo, Zmijewski, & Hopkins, 2017). Then, photocell mats were introduced comprising two parallel bars of emitting and receiving IR diodes that scan any break in optical barrier made by athletes jumping (Castagna et al., 2013; García-López et al., 2005). More recently, the smartphone app *MyJump* has been developed for iOS devices that use the high frame rate video camera for the assessment of vertical jumps. The validity and inter- and intra-reliability of the instrument have been verified (Balsalobre-Fernandez, Glaister, & Lockey, 2015; Driller et al., 2017; Gallardo-Fuentes et al., 2016).

However, data on inter-reliability from trained and untrained raters remain absent in the literature. Inter-reliability is important in such assessments since it reflects coaching and clinical practice where different raters make observations of the same jump execution. Therefore, the present study aims to compare the assessments of different raters with or without prior experience in the use of *MyJump*.

Material & methods

Participants

Eighteen healthy sport sciences students (10 male and 8 female; mean \pm SD: age 21.2 \pm 4.9 years, height 171.8 \pm 8.1 cm, body mass 74.4 \pm 14.1 kg) volunteered to participate in the present study. Participants are

a sample of recreational active students with a broad range of abilities and training. None of them showed any lower-limb injury nor they were medicated. This study was approved by the Human Research Ethics Committee of the University of Alicante. Each participant signed a written informed consent before participation.

Procedures

Inter-rater reliability of *MyJump* smartphone app to measure flight height by means of flight time observation through video digitizing was determined by comparing observations with three raters of the same set of jump executions. Participants followed a standardized 10-min warm-up including hip rotations, leg swings, jogging and submaximal jumps. Then, each participant performed eight maximal CMJs with one minute passive resting period. In all executions, participants kept their hands on hips with legs straight during flight phase to achieve a proper, uniform technique. All trials were recorded with an iPhone 6s (Apple Inc., USA) high-speed camera facing the frontal plane of participants' feet at approximately 1.5 m. Once all executions were recorded, the first frames displaying both feet off the ground and at least one foot touching the ground were selected as start and end, respectively, of flight time. Later, in order to analyze inter-rater reliability, three independent observers identified these threshold frames for each jump by means of the *MyJump* app. Following the equation described in the literature (Bosco, Luhtanen, & Komi, 1983),

$$h = t^2 \frac{g}{8}$$

where h is jump height, t is flight time and g is the gravity acceleration (9.81 m/s^2), the app computes and displays jump height directly on screen. Rater 1 did not have experience in assessing jump heights through video observation and received instructions from the app help alone with no feedback from the remainder observers before assessment. On the other hand, raters 2 and 3 were trained observers who agreed in take-off and landing identification as training phase prior to observations. In all cases, raters visualized video recordings independently to avoid cross influences.

Statistical analysis

Descriptive statistics were presented as mean \pm SD to show participant characteristics for each rater. To examine inter-reliability of trained and untrained raters for measuring jump heights with the app, the Intra-class Correlation Coefficient (ICC) (2,1), 2-way random single measurements (consistency and absolute agreement) and Cronbach's α were used pairwise. Furthermore, paired samples t -tests and mean differences with 95% CI were calculated to compare outcomes from raters. Also, Bland-Altman plots were used as a representation of the agreement between two measures (Bland & Altman, 1986). They allow for the identification of any potential systematic bias between raters through the analysis of mean bias, standard deviations and regression lines of differences of observed pairs against mean values. Finally, Cronbach's α and the Coefficient of Variation (CV) were calculated to analyze the stability of the app for each rater when measuring the eight executions of each participant. All statistical analyses were performed using SPSS v. 22 (IBM Corp, Armonk, NY), including Bland-Altman plots. The level of statistical significance was set at $p < 0.05$.

Results

All participants completed jump executions and therefore, a total of 144 jumps were considered. Descriptive statistics show the following jump height values for raters 1, 2 and 3, respectively: 35.87 ± 4.56 , 35.46 ± 4.69 and 36.42 ± 4.82 cm.

There was good reliability between observers, regardless of their experience in operating the app, as it can be seen in ICC results both for consistency and absolute agreement (Table 1). Also, Cronbach's α values in the vicinity of unity demonstrated high reliability between experienced and non-experienced raters. Mean differences between ratings were less than 1 cm for all cases.

Table 1. Pairwise reliability of trained vs untrained raters

	Rater 1 vs Rater 2	Rater 2 vs Rater 3	Rater 1 vs Rater 3
ICC (2,1) [#] (95% CI)	0.984 (0.972 – 0.990)	0.982 (0.969 – 0.989)	0.984 (0.973 – 0.991)
ICC (2,1) [§] (95% CI)	0.962 (0.692 – 0.988)	0.980 (0.956 – 0.990)	0.978 (0.929 – 0.990)
Cronbach's α	0.992	0.991	0.992
Mean difference (95% CI)	0.41* (0.18 – 0.64)	-0.55* (-0.78 – -0.33)	-0.96* (-1.21 – -0.72)
Bland-Altman regression	$y = 1.40 - 0.03x$	$y = 2.91 \cdot 10^{-3} - 0.03x$	$y = 1.42 - 0.05x$

Intra-class correlation coefficient (ICC) showing consistency([#]) and absolute agreement([§]) for the comparison between raters; 95% CI = 95% confidence interval; * $p < 0.01$

Bland-Altman plots displaying limits of agreement for jump height between raters show that most of the observations lie within the 95% CI's, depicted as dashed lines in Fig. 1. Observations made by rater 1 against raters 2 and 3 were more scattered than those made by experienced raters (1 and 2), as to judge by the number of data points outside 95% limits of agreement. The plots also show systematic bias between raters of < 1 cm for all

cases (0.41, -0.96 and -0.55 cm), demonstrating that little over- or underestimation in jump height can be expected for both types of raters. The analysis of regression lines of paired differences against mean values showed little bias between raters over the range of jump height values, where $R^2 \approx 0.02$ for all paired differences (Table 1 for regression equation).

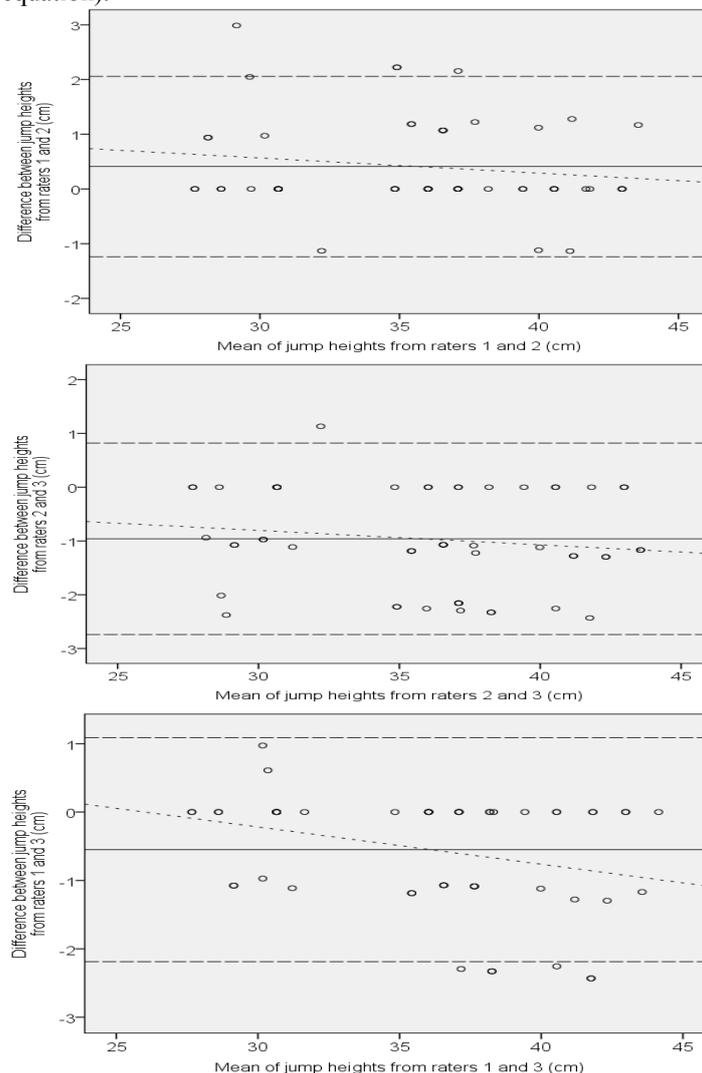


Fig. 1. Level of agreement plots (Bland-Altman) displaying mean differences or bias between raters (solid line), 95% limits of agreement calculated as $\pm 1.96 \cdot SD$ (dashed lines) and regression line (dotted line).

Finally, raters show very good internal consistency when measuring each session of eight jump executions. Table 2 shows that Cronbach's alpha was virtually unity for both trained and untrained raters. Coefficients of variation were also low, although the larger value was observed for the untrained rater.

Table 2: Internal consistency of raters within a session

	α	CV (%)
Rater 1	0.995	1.38 \pm 0.52
Rater 2	0.996	1.30 \pm 0.45
Rater 3	0.997	1.10 \pm 0.45

* α = Cronbach's alpha, CV = Coefficient of Variation

Discussion

The aim of this study was to analyze the inter-rater reliability of trained and untrained raters of the *MyJump* app. Several studies have addressed validity and reliability of the app. For instance, Balsalobre-Fernandez, Glaister, & Lockey (2015) showed that reliability of the app between two experienced raters were high (ICC=0.997) for both, with mean difference of 1.1 ± 0.5 cm and 1.3 ± 0.5 cm for raters 1 and 2, respectively. Similarly, Gallardo-Fuentes et al., (2016) reported high correlation (ICC=0.99) for both sessions and mean differences of 0.0 ± 0.7 and 0.1 ± 1.0 cm for session 1 and 2, respectively. Stanton, Wintour, & Kean (2016) showed intra-rater reliability of ICC=0.99 and mean difference of 0.43 cm. These values are in agreement with ICC values of this study, ranging from 0.962 for unexperienced raters to 0.984 for trained raters. Absolute mean differences of this study were also similar for both types of raters (0.4 to 0.96 cm).

The level of agreement between raters are of interest for practical purposes since it allows the app to be used regardless of prior experience in identifying take-off and landing frames. The level of scattered data points in Bland-Altman plots of previous studies facing experienced raters are very similar to that of the present study (Balsalobre-Fernandez et al., 2015). This result suggests that no additional uncertainty is assumed by the fact that non-trained raters would use the app.

Internal consistency was very good, regardless of prior training of raters ($\alpha \approx 0.99$ and $CV = 1.1$ to 1.4%). Other studies showed similar reliability values for experienced raters: $\alpha = 0.997$, $CV = 3.4$ and 3.6% (Balsalobre-Fernandez et al., 2015) and $\alpha = 0.99$ and $CV = 4.82 - 5.58\%$ for CMJ (Gallardo-Fuentes et al., 2016).

Conclusions

MyJump can be used by trained and novel users in assessing jump heights. According to the lack of substantial differences between outcomes from trained and non-trained raters, both types of users can give similar observations of take-off and landing frames, as the key instants from where to compute jump height. The results of the present study also suggest that the source of errors in observational measures based on non-continuous digital video can be due to the lack of high temporal resolution rather than observation issues.

Conflict of interests

The authors declare that there is no conflict of interests connected with other studies or authors.

References

- Baca, A. (1999). A comparison of methods for analyzing drop jump performance. *Medicine and Science in Sports and Exercise*, 31(3), 437–442. <https://doi.org/10.1097/00005768-199903000-00013>
- Balsalobre-Fernandez, C., Glaister, M., & Lockey, R. A. (2015). The validity and reliability of an iPhone app for measuring vertical jump performance. *Journal of Sports Sciences*, 1–6. <https://doi.org/10.1080/02640414.2014.996184>
- Barris, S., & Button, C. (2008). A review of vision-based motion analysis in sport. *Sports Medicine*, 38(12), 1025–1043. <https://doi.org/10.2165/00007256-200838120-00006>
- Bland, J. M., & Altman, D. G. (1986). Statistical Methods for Assessing Agreement Between Two Methods of Clinical Measurement. *Lancet*, 327, 307–310. [https://doi.org/10.1016/S0140-6736\(86\)90837-8](https://doi.org/10.1016/S0140-6736(86)90837-8)
- Bosco, C., Luhtanen, P., & Komi, P. V. (1983). A Simple Method for Measurement of Mechanical Power in Jumping. *European Journal of Applied Physiology and Occupational Physiology*, 50, 273–282.
- Çakir-Atabek, H. (2014). Relationship between anaerobic power, vertical jump and aerobic performance in adolescent track and field athletes. *Journal of Physical Education and Sport*, 14(4), 643–648. <https://doi.org/10.7752/jpes.2014.04100>
- Castagna, C., Ganzetti, M., Ditroilo, M., Giovannelli, M., Rocchetti, A., & Mazi, V. (2013). Concurrent Validity of Vertical Jump Performance Assessment Systems. *The Journal of Strength & Conditioning Research*, 27(3), 761–768. Retrieved from http://journals.lww.com/nsca-jscr/Abstract/2013/03000/Concurrent_VValidity_of_Vertical_Jump_Performance.28.aspx
- Driller, M., Tavares, F., McMaster, D., & O'Donnell, S. (2017). Assessing a smartphone application to measure counter-movement jumps in recreational athletes. *International Journal of Sports Science and Coaching*, 12(5), 661–664. <https://doi.org/10.1177/1747954117727846>
- Gallardo-Fuentes, F., Gallardo-Fuentes, J., Ramírez-Campillo, R., Balsalobre-Fernández, C., Martínez, C., Caniuqueo, A., ... Izquierdo, M. (2016). Intersession and Intrasession Reliability and Validity of the My Jump App for Measuring Different Jump Actions in Trained Male and Female Athletes. *Journal of Strength and Conditioning Research / National Strength & Conditioning Association*, 30(7), 2049–2056. <https://doi.org/10.1519/JSC.0000000000001304>
- García-López, J., Peleteiro, J., Rodríguez-Marroyo, J. A., Morante, J. C., Herrero, J. A., & Villa, J. G. (2005). The validation of a new method that measures contact and flight times during vertical jump. *International Journal of Sports Medicine*, 26(4), 294–302. <https://doi.org/10.1055/s-2004-820962>
- Hatze, H. (1998). Validity and reliability of methods for testing vertical jumping performance. *Journal of Applied Biomechanics*, 14(2), 127–140.
- Pueo, B., Lipinska, P., Jiménez-Olmedo, J. M., Zmijewski, P., & Hopkins, W. G. (2017). Accuracy of jump-mat systems for measuring jump height. *International Journal of Sports Physiology and Performance*, 12(7). <https://doi.org/10.1123/ijsp.2016-0511>
- Stanton, R., Wintour, S.-A., & Kean, C. O. (2016). Validity and intra-rater reliability of MyJump app on iPhone 6s in jump performance. *Journal of Science and Medicine in Sport*. <https://doi.org/10.1016/j.jsams.2016.09.016>
- Whitmer, T., Fry, A. C., Forsythe, C., Andre, M. J., Lane, M. T., Hudy, A., & Honnold, D. (2014). Accuracy of a Vertical Jump Contact Mat for Determining Jump Height and Flight Time. *Journal of Strength and Conditioning Research*, 29(4), 877–881. <https://doi.org/10.1519/JSC.0000000000000542>