

# Inter-Rater Reliability and Intra-Rater Reliability Testing of My Jump 2 Mobile Application in Measuring Countermovement Jump

Muhammad Faiz Alias<sup>1</sup>, Hosni Hasan<sup>1\*</sup>, Mohd Syafiq Miswan<sup>2</sup>, Daniyal Donny Man<sup>3</sup>

<sup>1</sup>Faculty of Sports Science and Recreation, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia

<sup>2</sup>Faculty of Sports Science and Recreation, Universiti Teknologi MARA Perlis, 40450 Arau, Perlis, Malaysia

<sup>3</sup>Faculty of Sports Science and Recreation, Universiti Teknologi MARA Sarawak, 40450 Samarahan, Sarawak, Malaysia

Received August 4, 2021; Revised October 5, 2021; Accepted October 17, 2021

## Cite This Paper in the following Citation Styles

(a): [1] Muhammad Faiz Alias, Hosni Hasan, Mohd Syafiq Miswan, Daniyal Donny Man, "Inter-Rater Reliability and Intra-Rater Reliability Testing of My Jump 2 Mobile Application in Measuring Countermovement Jump," *International Journal of Human Movement and Sports Sciences*, Vol. 9, No. 6, pp. 1319 - 1323, 2021. DOI: 10.13189/saj.2021.090628.

(b): Muhammad Faiz Alias, Hosni Hasan, Mohd Syafiq Miswan, Daniyal Donny Man (2021). *Inter-Rater Reliability and Intra-Rater Reliability Testing of My Jump 2 Mobile Application in Measuring Countermovement Jump*. *International Journal of Human Movement and Sports Sciences*, 9(6), 1319 - 1323. DOI: 10.13189/saj.2021.090628.

Copyright©2021 by authors, all rights reserved. Authors agree that this article remains permanently open access under the terms of the Creative Commons Attribution License 4.0 International License

**Abstract** My Jump 2 is a mobile application that is objectively valid and reliable to measure vertical jumping height. The objectives of this study are to find the inter-rater and intra-rater reliability of the My Jump 2 mobile application in measuring countermovement jump (CMJ) height. A total of 25 male recreational athletes performed CMJ for five attempts and was recorded by using iPhone 7 Plus in 240 frames per second. The videos were then rated by three raters and rated again seven days later using the My Jump 2 mobile application. An excellent degree of reliability was found between rater measurements. The average measure ICC was 1.00 with a 95% confidence interval from 1.00 to 1.000 ( $F(124,248) = 18867201.171, p < .001$ ). An excellent degree of reliability was found within rater measurements. The average measure ICC was 1.00 with a 95% confidence interval from 1.00 to 1.00 ( $F(124,124) = 44750598.291, p < .001$ ). This study provides evidence in supporting the use of the My Jump 2 mobile application to measure the height of the countermovement jump in a research setting. Future research is needed to study the validity and reliability of other parameters possibly measured by the My Jump 2 mobile application like force-velocity profile, jumping launch force, and power.

**Keywords** My Jump 2 Mobile Application, Countermovement Jump, Inter-Rater Reliability,

Intra-Rater Reliability

## 1. Introduction

The ability of modern mobile phones to record videos as high as 240 frames per second is considered very useful for analyzing the movement of participants [1, 2]. One of the rising mobile applications recognized as valid and reliable today is the My Jump 2 mobile application [3]. This mobile application was developed ad hoc by a researcher to find a new valid, reliable, and simple way to measure jumping height in the sample. During its initial release, this mobile application could only measure jumping height in Counter Movement Jump.

The My Jump 2 mobile application developer claimed that this mobile application was rebuilt from scratch and used a more complicated algorithm. After several updates, this mobile application could now calculate and project more complex tasks such as the force-velocity profile. However, this mobile application still has one weakness. It still needs to rely on the human eyes to pick take-off and landing points during testing. Human factors during testing lead to systematic bias [4].

In recent years, the mobile phone considered a very

versatile tool that is affordable and accessible to anyone. The ability of the modern mobile phone to record as high as 240 frames per second is considered very useful for analyzing the movement of participants [3]. This advancement is a blessing for the researcher because the tool for research has become cheap and portable [5]. One of the rising mobile applications that are valid and reliable today is the My Jump 2 mobile application [3]. This app is used widely by coaches and athletes in assessing their jumping height. However, some researchers are a bit skeptical about the ability of this app [6]. Therefore, the purpose of this study was to investigate the inter-rater and intra-rater reliability of My Jump 2 application in measuring countermovement jump height.

## 2. Materials and Methods

### 2.1. Participants

The recruiting of a total sample of 25 healthy male recreational athletes in this study, by adapting the research by Balsalobre-Fernández, Tejero-González, del Campo-Vecino and Bavaresco [1], conducted with the use of the  $\alpha = 0.05$  and  $\beta = 0.2$ . The researcher used the same value in the equation to calculate sample size for the reliability study formulated by Walter, Eliasziw and Donner [7]. Therefore, according to this equation, the total sample is 25 ( $n = 25$ ) and each participant performing five ( $k = 5$ ) attempts. Each attempt counted as a single measurement ( $nk = 125$ ). Three raters were involved in measuring the intra-reliability and inter-reliability of this mobile application.

### 2.2. Instrumentations

The instrument used in this study was the My Jump 2 mobile application readily installed on iPhone 7 Plus (iOS 13.0, Apple Inc. USA). This app has already been proven valid and reliable by a study conducted by Carlos-Vivas, Martin-Martinez, Hernandez-Mocholi and Perez-Gomez [2]. Several measures and/ or actions took to provide the best condition for the test. The conditions were:

- i. Carried outdoor to get optimum lighting condition.
- ii. On an anti-slip red mat to provide contrast during video analysis.
- iii. Perform on a bracket and use a wireless remote to get a stable video.

### 2.3. Procedures

For countermovement jump height, the participants first performed a warm-up procedure, through 5 minutes on a stationary bike without any resistance. Then, the participants proceed with the instruction on the countermovement jump procedure [2]:

- i. Standing with feet shoulder-width apart.
- ii. Hand placed at the hip throughout the test.
- iii. The selection of depth of countermovement is on the participant.
- iv. Jump as high as possible.
- v. Land as close as possible to the jumping point.

The participants performed a countermovement jump with their timing. If any mistake or foul presented during the jumping procedure, a reattempt would take place. The researcher records this video on iPhone 7 plus at 240 frames per second at 1280 x 720-pixel resolution in portrait orientation. The total number of recorded video clips for this study was 125 videos, whereby each of the 25 participants, had five attempts.

For analyzing the videos, the study made an appointment with three raters. Each rater used the same model smartphone (iPhone 7 Plus). Each rater analyzed the jumping on a different occasion to avoid bias. Besides, the video order was analyzed and randomized to minimize the risks of prejudice against the participant in question. A week later, all raters are called to rate the video again.

All raters underwent the same training to equalize their knowledge about this mobile application. The first training was theory training by watching a tutorial video officially produced by the application developer. This video is available on the application official website. After theory training, they were given hands-on training on the mobile application by the researcher. The instructions made for the raters were as follow:

- i. Please choose the final frame in which the participant's feet touch the ground during take-off.
- ii. Please choose the first frame in which the participant's feet touch the ground during landing.

### 2.4. Data Analysis

A record of the details identified after the assessment finishes. All data received are analyzed using the Statistical Package for Social Science version 21.0 (SPSS 21.0). Inferential and descriptive statistics functioned to obtain quantitative metrics for the finding. As for the descriptive data, the analysis includes frequency distribution, mean, standard deviation and rank.

By referring to a study conducted by Koo and Li [8] where they used the same specific rater for all subjects, this study chose a two-way mixed effect agreement for the inferential statistics. This study used the mean of three raters for an intra-rater, while inter-rater is based on each rater. Both tests aim to find an absolute agreement.

For both inter-rater and intra-rater reliability analyses, ICC values were interpreted as 0.90–1.00 = very high correlation, 0.70–0.89 = high correlation, 0.50–0.69 = moderate correlation, 0.26–0.49 = low correlation, and 0.00–0.25 = little, if any correlation [9]. All data were visually analyzed by a Bland Altman Plot and examined for heteroscedasticity.

**Table 1.** Descriptive data

	n	Minimum	Maximum	Mean	Std. error	Std. deviation
Rater A	125	15.03	39.42	23.6778	.50637	5.66139
Trial 2		15.03	39.42	23.6780	.50637	5.66138
Rater B	125	15.03	39.42	23.6789	.50631	5.66072
Trial 2		15.03	39.42	23.6788	.50630	5.66059
Rater C	125	15.03	39.42	23.6785	.50634	5.66102
Trial 2		15.03	39.42	23.6782	.50634	5.66102

**Table 2.** Intraclass correlation Coefficient for inter-rater reliability

	Intraclass Correlation <sup>b</sup>	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	1.000 <sup>a</sup>	1.000	1.000	18867201.171	124	248	.000
Average Measures	1.000 <sup>c</sup>	1.000	1.000	18867201.171	124	248	.000

**Table 3.** Intraclass correlation Coefficient for intra-rater reliability

	Intraclass Correlation <sup>b</sup>	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	1.000 <sup>a</sup>	1.000	1.000	44750598.291	124	124	.000
Average Measures	1.000 <sup>c</sup>	1.000	1.000	44750598.291	124	124	.000

### 3. Results

An excellent degree of reliability was found between rater measurements (Table 1). The average measure ICC was 1.00 with a 95% confidence interval from 1.00 to 1.000 (F(124,248)= 18867201.171, p<.001). An excellent degree of reliability was found within rater measurements. The average measure ICC was 1.00 with a 95% confidence interval from 1.00 to 1.00 (F (124,124) = 44750598.291, p<.001).

To visualize the difference between session measurements (Table 2), the Bland-Altman plot was utilized with measurement A (mean of the first measurement) and measurement B (mean of the second measurement seven days later). As for the spread of the scores around the zero line, 95% of the different scores fall within two standard deviations above and below the mean difference scores (Table 3). This finding suggests that the data was unbiased, homoscedastic [9], and there was a good agreement between the two sessions.

### 4. Discussions

#### Inter-rater reliability of My Jump 2 mobile application in measuring countermovement jump

One of the purposes of this study was to analyze the reliability of the My Jump 2 mobile application between raters. Several studies tackled the application validity and reliability. For example, [6] demonstrate that the consistency of the application among highly experienced

raters was excellent (ICC=0.997) for both, with a mean difference of  $1.1 \pm 0.5$  cm and  $1.3 \pm 0.5$  cm respectively for raters 1 and 2. Equally, both sessions and mean differences of  $0.0 \pm 0.7$  and  $0.1 \pm 1.0$  cm for sessions 1 and 2 [10] with high correlation reported (ICC=0.99) respectively. Stanton, Wintour and Kean [11] demonstrated intrarater reliability of ICC=0.99 and a mean difference of 0.43 cm. Such values follow this study's ICC values, ranging from 0.962 for inexperienced raters to 0.984 for qualified raters. Absolute mean differences for both types of raters (0.4 to 0.96 cm) were also similar in this study.

The level of agreement between the raters is also significant for practical purposes, as it needs the specification to be used to identify take-off and landing sites, regardless of prior experience. The amount of scattered data points in Bland-Altman plots of previous experiments with experienced raters is very close to that of the current sample [6]. This finding indicates that non-trained raters' utilizing the software does not imply any potential ambiguity. The internal performance was strong, regardless of the raters' prior experience ( $\alpha=0.99$  and CV=1.1 to 1.4%). Related reliability levels for experienced raters observed in other studies:  $\alpha=0.997$ , CV=3.4, and 3.6% [6] and  $\alpha=0.99$  and CV=4.82–5.58% for countermovement jump [10].

In the current study, an excellent degree of reliability was found between rater measurements. The average measure ICC was 1.00 with a 95% confidence interval from 1.00 to 1.000 (F (124,248) = 18867201.171, p<.001). This study showed a countermovement jump height

measurement reliably performed by different raters with varying experience by using this mobile application. This result gives an answer to the research question as proposed by the researcher and accepts the hypothesis. The excellent level of intraclass correlation coefficient might originate from the extra step taken by the researcher to ensure the high quality of the video. The steps taken were a part of the procedure carried out outdoor to get optimum lighting conditions. The participants also performed on an anti-slip red mat to show contrast during video analysis performed on a bracket and use a wireless remote to get a stable video.

Another reason that leads the excellent degree of inter-rater reliability is from the training and tutorial given to the rater. Compared to the study by Pueo, Jimenez-Olmedo, Penichet-Tomás and Bernal-Soriano [12] and Stanton, Wintour and Kean [11], there was no training given to the raters. In this study, the researcher records a different video for different raters. Meanwhile, by using the same video for different raters for the current study, the researcher could eliminate other factors that might cause a distraction in video analysis of the different angles of view and video quality. However, a study by Gallardo-Fuentes, Gallardo-Fuentes, Ramírez-Campillo, Balsalobre-Fernández, Martínez, Caniuqueo, Cañas, Banzer, Loturco and Nakamura [10] replicates a real-world measurement while another study that was close to the current research is [12]. There was excellent reliability amongst observers, irrespective of their involvement in mobile application service. The only difference was that these papers focused more on experienced raters compared to the inexperienced ones. The current study focuses more on the technical site and human error of the application. Besides, the researcher could not find the number of framerate (frame per second) used in the study [12]. The framerate influenced the quality of the video because of the feature of having a high frame rate (240 frames per second) in such a small camera sensor that would cause low light to reach the sensor.

#### **Intra-rater reliability of My Jump 2 mobile application in measuring countermovement jump**

An excellent degree of reliability was found between rater measurements. The average measure ICC was 1.00 with a 95% confidence interval from 1.00 to 1.00 ( $F(124,124) = 44750598.291, p < .001$ ). This study showed that a countermovement jump height measurement could be performed reliably by the same rater and produces a consistent result. Similarly, as stated before, the excellent degree of reliability might be caused by the extra effort taken by the researcher to ensure the quality of the video.

Intra-rater reliability shows a very high level of similarity between Day 1 and Day 7 measurements. This finding suggests that the frequency of random error is

minimal when the equivalent rater scores measurements and systemic error is a significant drawback. A systematic error that occurs when using the My Jump 2 mobile application may be due to the difficulty of correctly selecting the take-off and landing frames because of the video capture frequency used (240 Hz). However, as suggested in the previous section, the systemic error becomes smaller as the researcher uses a higher frequency video sampling.

One similar study with the current research is by [11] which found that the intra-rater reliability was excellent. The methodologies are very similar between these different studies. The only difference is that the past researcher used a well-trained athlete and both genders, while the current study used a recreational athlete and male-only. However, as mentioned earlier, the results in this study were almost in perfect agreement due to the extra precaution taken by the researcher to ensure the high quality of the high-speed video.

Excellent reliability in both inter-rater and intra-rater allows the researcher to use this mobile application without a doubt in conducting research. Although this research proceeded in a controlled manner (excellent lighting, contrast colour and use of a tripod), a similar result from field testing is to be expected [11, 12].

## **5. Conclusions**

This study showed that in a healthy, recreational athlete, the countermovement jump measurement could be performed reliably by the same rater, as well as different rater with varying clinical experience. This study gives evidence to support the My Jump 2 mobile application in measuring the height of the countermovement jump in a research setting. Further research is needed to confirm the result in a different type of jumping, such as drop jump and countermovement jump (freehand). One of the limitations of this study was the smartphone used. The current study uses an iPhone 7 plus, which is not the latest phone from the manufacturer. A relatively old phone is easy to heat up after an extended usage for high framerate videos. This limitation causes the measurement process by the rater to become slightly slower, and the video playback became somewhat sluggish and lagged. Overall, this might cause the process to pick a frame for lift-off and landings to become a bit hard. Future research is needed to study the validity and reliability of another aspect of the My Jump 2 mobile application as it can measure parameters such as jumping height. Besides this, it can also measure force-velocity profile, jumping launch force, and power. Fellow researchers pay lesser attention to this aspect of the study.

Besides that, this research requires a repeat in the future with this study sets in a field-testing environment. As mentioned earlier, this study has both high intra and inter

reliability due to controlled conditions during data collection because of the high amount of light and usage of tripod to ensure recorded videos are of very high quality and not blurry. In addition, the My Jump 2 mobile application is also developed for coaches to monitor the performance and their athletes.

---

## REFERENCES

- [1] Balsalobre-Fernández, C., C.M. Tejero-González, J. del Campo-Vecino, N. Bavaresco, "The concurrent validity and reliability of a low-cost, high-speed camera-based method for measuring the flight time of vertical jumps", *The Journal of Strength & Conditioning Research*, vol. 28, no. 2, pp. 528-533, 2014. DOI: 10.1519/JSC.0b013e318299a52e
- [2] Carlos-Vivas, J., J.P. Martin-Martinez, M.A. Hernandez-Mocholi, J. Perez-Gomez, "Validation of the iPhone app using the force platform to estimate vertical jump height", *The Journal of Sports Medicine and Physical Fitness*, vol. 58, no. 3, pp. 227-232, 2016. DOI: 10.23736/s0022-4707.16.06664-0
- [3] Haynes, T., C. Bishop, M. Antrobus, J. Brazier, "The validity and reliability of the My Jump 2 app for measuring the reactive strength index and drop jump performance", *The Journal of Sports Medicine and Physical Fitness*, vol. 59, no. 2, pp. 253-258, 2019. DOI: 10.23736/S0022-4707.18.08195-1
- [4] Stanton, R., C.O. Kean, A.T. Scanlan, "My Jump for vertical jump assessment", *British Journal of Sports Medicine*, vol. 49, no. 17, pp. 1157-1158, 2015. DOI: <http://dx.doi.org/10.1136/bjsports-2015-094831>
- [5] Vaisman, A., R. Guiloff, J. Rojas, I. Delgado, D. Figueroa, R. Calvo, "Lower limb symmetry: comparison of muscular power between dominant and nondominant legs in healthy young adults associated with single-leg-dominant sports", *Orthopaedic Journal of Sports Medicine*, vol. 5, no. 12, 2017. DOI: <https://doi.org/10.1177/2325967117744240>
- [6] Balsalobre-Fernández, C., M. Glaister, R.A. Lockey, "The validity and reliability of an iPhone app for measuring vertical jump performance", *Journal of Sports Sciences*, vol. 33, no. 15, pp. 1574-1579, 2015. <https://doi.org/10.1080/02640414.2014.996184>
- [7] Walter, S., M. Eliasziw, A. Donner, "Sample size and optimal designs for reliability studies", *Statistics in Medicine*, vol. 17, no. 1, pp. 101-110, 1998. DOI: [https://doi.org/10.1002/\(SICI\)1097-0258\(19980115\)17:1<101::AID-SIM727>3.0.CO;2-E](https://doi.org/10.1002/(SICI)1097-0258(19980115)17:1<101::AID-SIM727>3.0.CO;2-E)
- [8] Koo, T., M. Li, "A guideline of selecting and reporting intraclass correlation coefficients for reliability research", *Journal of Chiropractic Medicine*, vol. 15, no., pp. 155-163, 2016. DOI: <https://doi.org/10.1016/j.jcm.2016.02.012>
- [9] Portney, L.G., M.P. Watkins, *Foundations of clinical research: applications to practice*. Vol. 892. Pearson/Prentice Hall Upper Saddle River, NJ, 2009.
- [10] Gallardo-Fuentes, F., J. Gallardo-Fuentes, R. Ramirez-Campillo, C. Balsalobre-Fernández, C. Martínez, A. Caniuqueo, R. Cañas, W. Banzer, I. Loturco, F.Y. Nakamura, "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*, vol. 30, no. 7, pp. 2049-2056, 2016. DOI: <https://doi.org/10.1519/JSC.00000000001304>
- [11] Stanton, R., S.-A. Wintour, C.O. Kean, "Validity and intra-rater reliability of MyJump app on iPhone 6s in jump performance", *Journal of Science and Medicine in Sport*, vol. 20, no. 5, pp. 518-523, 2017. DOI: <https://doi.org/10.1016/j.jsams.2016.09.016>
- [12] Pueo, B., J.M. Jimenez-Olmedo, A. Penichet-Tomás, M.C. Bernal-Soriano, "Inter-rater reliability of trained and untrained raters for measuring jump height with the MyJump app", *Journal of Physical Education and Sport*, vol. 18, no. 2, pp. 821-824, 2018. DOI: <https://doi.org/10.7752/jpes.2018.02121>