

# The Validity and Reliability of the My Jump Lab Artificial Intelligence Application

Lee Bridgeman<sup>1\*</sup> , Bailey Cameron<sup>1</sup> , James Steele<sup>1</sup> 

<sup>1</sup>Department of Sport and Health, Solent University, UK

\*Correspondence: lee.bridgeman@solent.ac.uk

Cite as: Lee Bridgeman et al. (2024). The Validity and Reliability of the My Jump Lab Artificial Intelligence Application. *SportRxiv*.

Supplementary Materials: <https://osf.io/z9q2k/>

## Abstract

Jump height (JH) achieved in a countermovement jump (CMJ) has been suggested to allow for the monitoring of neuromuscular fatigue (NMF) and assessment of lower body power. Although force platforms (FP) are considered the gold standard for measuring CMJ height, they are expensive compared to mobile apps such as My Jump Lab (MJL). Therefore, this study aimed to assess the concurrent validity and agreement of the MJL app compared to a FP (ForceDecks [FD]) system and to determine its test-retest reliability. A convenience sample of 26 ( $n = 11$  females and  $n = 15$  males) recreationally active university sport students and staff (mean  $\pm$  SD; age:  $23.08 \pm 6.33$  years; mass:  $72.85 \pm 9.93$  kg; stature:  $176.63 \pm 10.18$  cm) participated in the study. Participants attended the laboratory for testing on two separate occasions, separated by one week. After a standardised warm-up, they completed three CMJs on each occasion, with CMJ height simultaneously assessed by the FD and MJL app. The MJL Artificial Intelligence mode showed a mean bias of 4.32 cm [95% CI: 3.4, 5.26] overestimation with 95% limits of agreement ranging from -3.33 cm [95% CI: -4.96, -0.85] to 11.98 cm [95% CI: 10.13, 13.41]. Both methods demonstrated minimal mean bias (FD = 0.61 cm [95% CI: -0.31, 1.37] and MJL = 0.25 cm [95% CI = -0.48, 0.98]) between sessions, and both showed a similar width to their limits of agreement, ranging  $\sim 7$  cm about the mean bias. In summary, the MLJ overestimated CMJ height in this sample compared to the FD system, but both methods were reliable. Given the significant differences in cost for these two methods, teams on a budget may interested in trialling the MJL app.

Keywords: *Countermovement Jump; Force Decks; My Jump; Impulse-momentum*

## Introduction

The jump height (JH) achieved in a countermovement jump (CMJ) has previously been suggested to allow for the monitoring of neuromuscular fatigue (NMF) (Bishop, Turner, et al., 2022; Claudino et al., 2017; Hughes et al., 2019) and assessment of lower body power (McMaster et al., 2014; Nuzzo et al., 2008). As a result, this test is often utilised in laboratory and field-based testing. While several methods are available to assess CMJ height, force platforms (FP) are typically seen as the gold-standard measure (Balsalobre-Fernández et al., 2015) as they can capture both kinematic and kinetic variables. In particular, it has been proposed that dual FPs utilising the impulse-momentum (IM) relationship to assess CMJ height are desirable (Heredia-Jimenez & Orantes-Gonzalez, 2020; Xu et al., 2023). However, these systems come with a price tag (~£10,000+) that may be beyond the means of some teams and, therefore, is prohibitive. The proliferation of smartphone devices, which most coaches possess, now means they can access more affordable and easy-to-use solutions for assessing CMJ performance (Balsalobre-Fernández et al., 2015) through app-based solutions. One such app is My Jump Lab (MJL), available on Android and Apple devices via a subscription-based model (1 month = £4.99, 1 year = £34.99, Lifetime = £99.99 – prices correct at time of writing). A meta-analysis which investigated the validity of the app reported a correlation using a fixed effect model of 0.994 (95% CI = 0.992 – 0.995,  $p < 0.001$ ) and 0.992 (95% CI = 0.998 – 0.995,  $p < 0.001$ ) when using a random effect model to compare jump performance with the criterion method (FP) (Gençoğlu et al., 2023). When examining the app's reliability for the CMJ using a fixed effects model, an ICC of 0.969 (95% CI = 0.965 – 0.972,  $p < 0.001$ ) was reported and 0.982 (95% CI = 0.961 – 0.992,  $p < 0.001$ ) when using a random effect model (Gençoğlu et al., 2023). The authors concluded that the MJL app was valid and reliable for assessing jump performance (Gençoğlu et al., 2023).

One issue with previous iterations of the MJL app was that it required practitioners to manually identify an athlete's take-off and touch-down from frame-by-frame video once the jumps had been completed. While this may not be an issue when working with individual athletes, it is suggested that when working with squads of 20-plus athletes, this may become a labour-intensive process for time-poor support staff (Şentürk et al., 2024). A newer version of the app has been launched to alleviate these issues. This app now allows real-time jump height (JH) measurement without post-processing (Balsalobre-Fernández et al., 2015). At present, two studies have investigated the reliability and validity of this app (Balsalobre-Fernández et al., 2015; Şentürk et al., 2024). In the 1st proof of concept study (single participant), Balsalobre-Fernandez (2015) completed 400 jumps over 24 consecutive weeks, recording JH simultaneously on a FP (Hawkins Dynamics) and the MJL app. They reported a very high correlation ( $r = 0.971$ , 95% CI = 0.963 – 0.975) and large agreement (ICC 0.969, 95 CI = 0.963 – 0.975) between the measures. Initially, though, large differences were reported between the instruments (mean absolute difference =  $0.06 \pm 0.01$  m,  $d = 4.4$ ,  $p < 0.001$ ).

However, after applying a regression equation to correct the app's raw data, non-significant, trivial differences were reported (mean absolute difference =  $0.01 \pm 0.008$  m,  $d = 0.1$ ,  $p = 1.000$ ) between the devices. In a further study, Sentürk *et al.* (2024) compared the CMJ height recorded simultaneously by the MJL app and a FP (Vald – Force Decks) in 36 recreationally active participants. They reported a nearly perfect correlation ( $r = 0.968$ ,  $p = 0.001$ ) between the devices and very good agreement (mean difference =  $-1.016$  cm, 95% CI =  $-1.229 - -0.803$  cm) (Şentürk *et al.*, 2024). This study also reported high intra-session reliability for the app (SEM =  $0.43$  cm; CV =  $1.23\%$ ). These findings led the authors to conclude that the app was reliable and valid for measuring CMJ height.

However, it should be noted that in this study, all the participants were male and completed five jumps with a two-minute rest period between jumps (Şentürk *et al.*, 2024). The authors suggest that this is different from what actually happens in a team sport environment, where there are often many athletes to test in a short period, and a 30-second rest period, as employed in this study, is more appropriate. Previous research has also noted concerns about replication in sport and exercise science; thus, further research using similar procedures to see whether the previous results are replicated is considered valuable (Mesquida *et al.*, 2022). Therefore, this study aimed to assess the concurrent validity and agreement of the MJL application for CMJ height compared to the ForceDecks system and its test-rest reliability.

## Method

### Experimental Approach to the Study

This cross-sectional study aimed to investigate the validity of the MJL application for measuring JH compared to the gold-standard method (jump height measured by the impulse-momentum calculation using a FP). The secondary aim was to investigate each piece of equipment's reliability. A convenient sample of recreationally active participants with previous CMJ experience was recruited for this study. Participants attended the laboratory for testing on two separate occasions, separated by one week, and completed three CMJs on each occasion, with JH simultaneously assessed by both the FP and MJL.

### Participants

A convenience sample of 26 ( $n = 11$  females and  $n = 15$  males) recreationally active university sport students and staff (mean  $\pm$  SD; age:  $23.08 \pm 6.33$  years; mass:  $72.85 \pm 9.93$  kg; stature:  $176.63 \pm 10.18$  cm) participated in the study. Before commencing testing, all participants were fully informed about the procedures, possible risks, and purpose of the study. All participants also completed a PAR-Q form and provided written informed consent. The Solent University Ethics Committee approved this study.

## Procedures

Participants were asked to refrain from lower body exercise 24 hours before each testing session and not to drink caffeine for two hours before testing. All testing took place at the same time of the day for all participants ( $\pm 1$  hour) during both testing sessions. When the participants arrived at the laboratory, mass (Seca 875, Seca, Hamburg, Germany) and stature (Seca portable stadiometer, Seca, Hamburg, Germany) were recorded. The participant's stature (cm) was recorded with shoes on to ensure measurement accuracy as they completed the jumps with shoes on as the app developer recommended. Before completing the jumps, each participant did a warm-up consisting of five minutes of cycling (Wattbike Pro, Wattbike, UK) at 100 W (80 – 90 rpm) followed by five bodyweight squats and three CMJs (30 seconds between jumps). Upon completion of the warm-up, the participants rested for three minutes before completing three CMJ trials (30 seconds between trials). All CMJ trials took place using the FDLite force plates (ForceDecks [FD], Vald, Brisbane, Australia) sampling at 1000 Hz. These FP have been reported to have high reliability (ICC = 0.93) when using the IM method to calculate CMJ height (Merrigan et al., 2024). A study by Collings et al. (2024) also reported a 5% relative difference between FD and an embedded laboratory FP system and excellent test-retest reliability (ICC = 0.97 [0.92 - 0.99]) for JH. Before jumping, the FD were zeroed, and then the participants were weighed on the FD. After a period of quiet standing ( $\sim 2$  seconds), the participants squatted to a self-selected depth with their hands placed akimbo and were instructed to jump as high as possible for each of the three jumps. Jump height (cm) calculated by the IM relationship was the FD variable of interest during these trials. Simultaneously, a tripod with an iPhone 15 Pro Max (Apple, California, USA) was set up two metres away from the FD and at a height of one metre (to keep the participants in the bounding boxes created at a rate of 60 Hz) to record these jumps (consistent across all trials). These were recorded at 240 frames per second (FPS) using the MJL app (v.4.2.8, 2024) with the artificial intelligence (AI) setting activated, which uses computer vision techniques to detect the participant's movement in each frame of the live video (Balsalobre-Fernández et al., 2015). This allows JH to be measured in real time.

## Statistical Analysis

The present analysis was not pre-registered as we had no a priori hypotheses and, given the limited sample size due to resource constraints, was considered exploratory. Inferential statistics were treated as highly unstable local descriptions of the relations between model assumptions and data in order to acknowledge the inherent uncertainty in drawing generalised inferences from single and small samples (Amrhein et al., 2019). For all analyses we opted to take an estimation-based approach typical when examining validity and reliability. That is, we provide point estimates and the uncertainty in them for the statistical parameters reported. Two sets of models were employed exploring the JHs recorded from CMJ trials; one to examine the agreement between the two methods, and one to explore the test-retest reliability for each method. Given we had nested data whereby each participant provided three trials for each method on two separate testing sessions we adopted a mixed effects limits of agreement approach (Parker et al., 2020). This allowed us to estimate mean bias for each method com-

pared with the gold-standard for agreement utilising both first and second session data, and any test retest bias for reliability, in addition to upper and lower limits of agreement providing 95% coverage probabilities for both agreement and test-retest reliability. Models were fit using the `lme4` package and using Restricted Maximum Likelihood Estimation. For each of the models fit we used nonparametric case based bootstrapping resampling 10000 times at the individual participant level and refitting models in order to construct 95% quantile intervals for both the mean bias and limits of agreement estimates. Bootstrapping was performed using the `lmeresampler` package.

## Agreement

For estimation of bias and limits of agreement regarding the agreement between the gold-standard Force-Decks and impulse momentum method with the My Jump Lab Artificial Intelligence mode we utilised the following mixed effects model structure:

$$\begin{aligned} D_{ist} &= \mu + \alpha_i + \gamma_s + \epsilon_{ist} \\ \alpha_i &\sim N(0, \sigma_\alpha^2) \\ \gamma_s &\sim N(0, \sigma_\gamma^2) \\ \epsilon_{ist} &\sim N(0, \sigma_\epsilon^2) \end{aligned}$$

Where  $D_{ist}$  is the difference between measurements taken between the two methods (i.e.,  $y_{ist2} - y_{ist1}$ ), where the method indexed by 1 is the gold-standard and method indexed by 2 is the comparison method, for participant  $i$  during session  $s$  and for trial  $t$ . Here  $\mu$  is the overall mean of the between method differences (i.e., the mean bias),  $\alpha_i$  is the random effect for the  $i^{th}$  participant,  $\gamma_s$  is the random effect for the  $s^{th}$  session which is nested within participant, and  $\epsilon_{ist}$  is the error term. The 95% limits of agreement can then be calculated as:

$$\mu \pm 1.96 \sqrt{\sigma_\alpha^2 + \sigma_\gamma^2 + \sigma_\epsilon^2}$$

with the square root of the total variance providing an estimate of the standard deviation of the differences for use in the conventional Bland-Altman limits of agreement calculation. A model was fit for each method in comparison o the gold-standard.

## Reliability

For estimation of bias and limits of agreement regarding the test-retest reliability between each test session for each method we utilised the following mixed effects model structure:

$$\begin{aligned} D_{it}^* &= \mu^* + \alpha_i^* + \epsilon_{it}^* \\ \alpha_i^* &\sim N(0, \sigma_{\alpha^*}^2) \\ \epsilon_{it}^* &\sim N(0, \sigma_{\epsilon^*}^2) \end{aligned}$$

Where  $D_{it}$  is the difference between measurements taken between the two sessions for a given method (i.e.,  $y_{it2} - y_{it1}$ ), where the session indexed by 1 is the first test session and the session indexed by 2 is the second test session, for participant  $i$  and for trial  $t$  (note we use the superscript  $*$  to distinguish this from the agreement model. Here  $\mu^*$  is the overall mean of the between session differences (i.e., the mean bias),  $\alpha_i^*$  is the random effect for the  $i^{th}$  participant, and  $\epsilon_{it}^*$  is the error term. The 95% limits of agreement can then be calculated as:

$$\mu^* \pm 1.96\sqrt{\sigma_{\alpha^*}^2 + \sigma_{\epsilon^*}^2}$$

with the square root of the total variance providing an estimate of the standard deviation of the differences for use in the conventional Bland-Altman limits of agreement calculation. A model was fit for each method in order to examine its test-retest reliability.

## Open data, code, and materials

All data and code is presented in the supplementary materials (<https://osf.io/z9q2k/>). The `renv` package was used for package version reproducibility and a function based analysis pipeline using the `targets` package was employed (the analysis pipeline can be viewed by downloading the R Project and running the function `targets::tar_visnetwork()`). We cite all packages used in the analysis pipeline below using the `grateful` package (Rodriguez-Sanchez et al., 2023):

We used R version 4.3.3 (R Core Team, 2024) and the following R packages: `glue` v. 1.7.0 (Hester & Bryan, 2024), here v. 1.0.1 ([here?](#)), `janitor` v. 2.2.0 ([janitor?](#)), `knitr` v. 1.47 (Xie, 2014, 2015, 2024), `lme4` v. 1.1.35.4 (Bates et al., 2015), `lmeresampler` v. 0.2.4 (Loy et al., 2024), `patchwork` v. 1.2.0 (Pedersen, 2024), `quarto` v. 1.4 ([quarto?](#)), `renv` v. 1.0.5 (Ushey & Wickham, 2023), `rmarkdown` v. 2.27 (Allaire et al., 2024; Xie et al., 2018, 2020), `tarchetypes` v. 0.9.0 (Landau, 2021a), `targets` v. 1.7.1 (Landau, 2021b), `tidyverse` v. 2.0.0 (Wickham et al., 2019).

## Results

Table 1: Mean  $\pm$  SD jump height (cm) for each piece of equipment in session 1 and 2

Equipment	Session 1 Jump Height (mean $\pm$ SD)	Session 2 Jump Height (mean $\pm$ SD)
ForceDecks	29.68 $\pm$ 7.03 cm	30.44 $\pm$ 6.50 cm
My Jump Lab	34.13 $\pm$ 7.63 cm	34.70 $\pm$ 6.73 cm

Table 1 shows the mean  $\pm$  SD CMJ heights for each piece of equipment in both sessions.

## Agreement

The mean bias and limits of agreement along with their bootstrapped interval estimates can be seen for the agreement between methods in Figure 1 along with the raw data. The My Jump Lab Artificial Intelligence mode showed a mean bias of 4.32 cm [95% CI: 3.4, 5.26] overestimation with 95% limits of agreement ranging from -3.33 cm [95% CI: -4.96, -0.85] to 11.98 cm [95% CI: 10.13, 13.41].

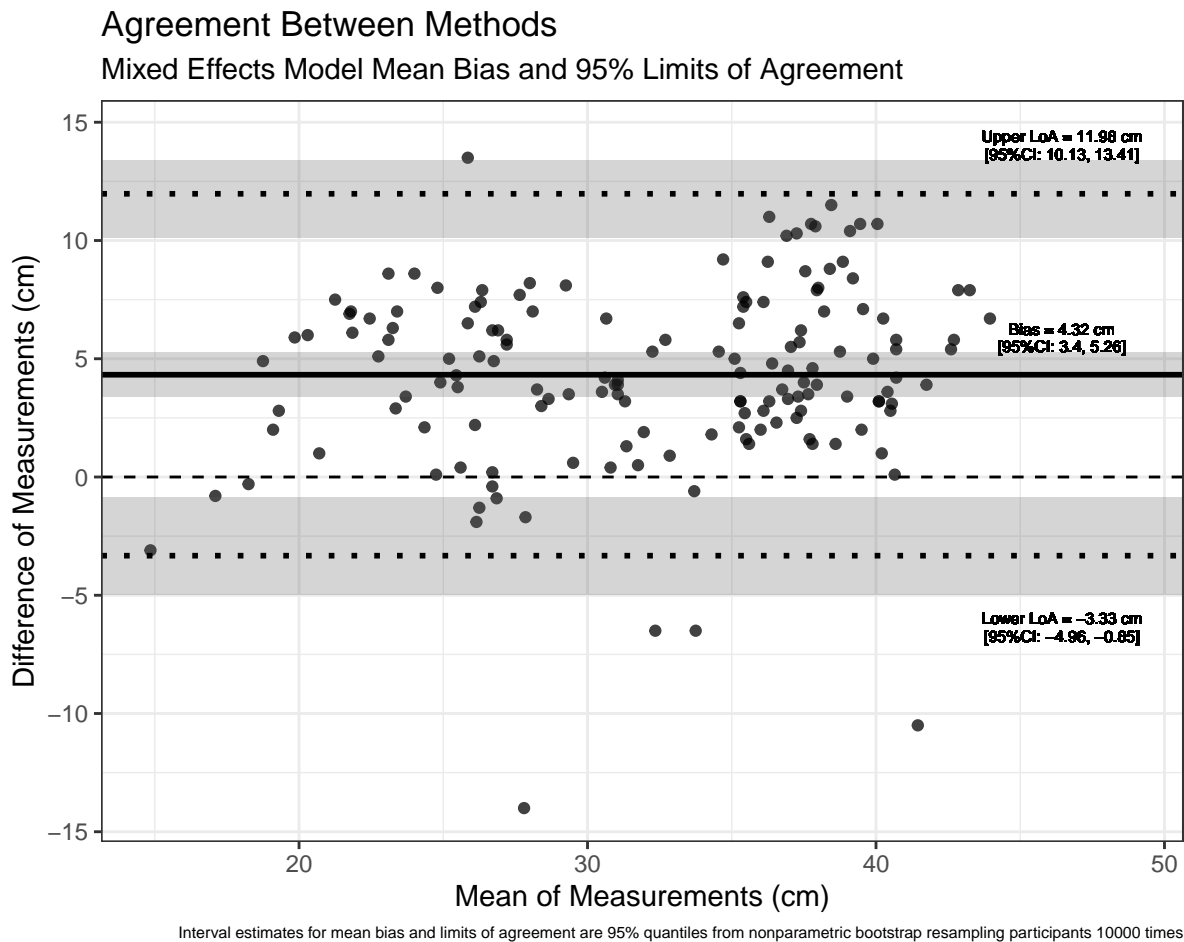


Figure 1: Mean bias (thick horizontal line) and 95% limits of agreement (dotted horizontal lines) along with 95% quantile interval estimates for agreement of My Jump Lab Artificial Intelligence mode in comparison to the gold-standard (Force Decks impulse-momentum method).

## Reliability

The mean bias and limits of agreement along with their bootstrapped interval estimates can be seen for the test-retest reliability of each method between sessions in Figure 2 along with the raw data. Both methods demonstrated minimal mean bias between sessions each typically less than 1 cm, and both demonstrated a similar width to their limits of agreement ranging ~7 cm about the mean bias.

## Test-Retest Reliability Across Methods

Mixed Effects Model Mean Bias and 95% Limits of Agreement

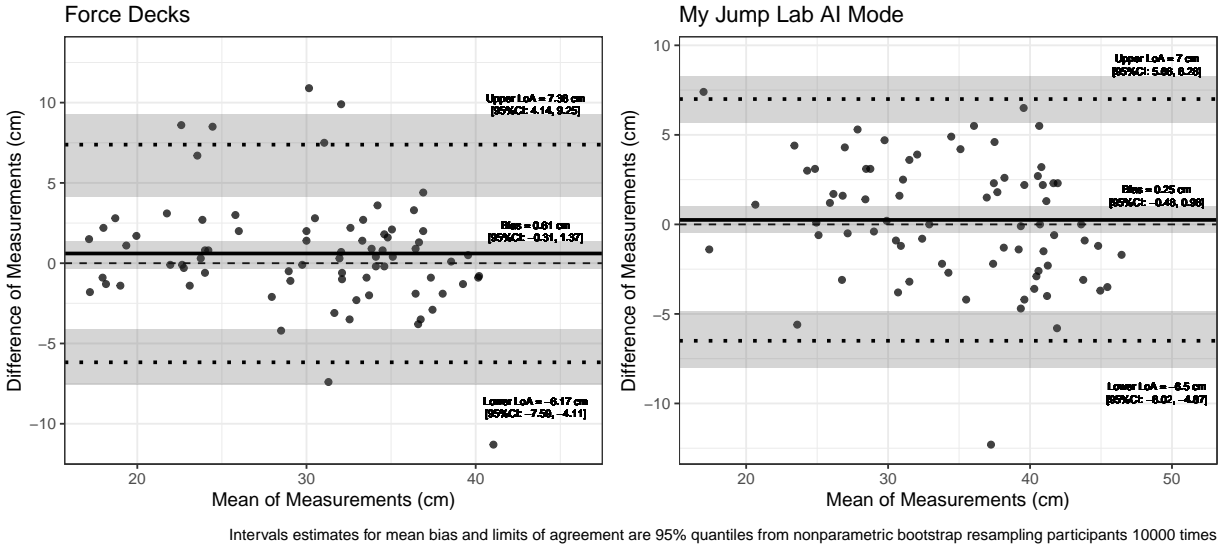


Figure 2: Mean bias (thick horizontal line) and 95% limits of agreement (dotted horizontal lines) along with 95% quantile interval estimates for each the test-retest reliability of method between sessions.

## Discussion

This study aimed to assess the concurrent validity and agreement of the MJL app for CMJ height compared to the FD system and evaluate its test-retest reliability. The main finding was that the MJL app overestimated CMJ height compared to the FD system (Mean bias = 4.32 cm [95% CI: 3.4, 5.26]). Concerning test-retest reliability, the bias for each device was found to be minimal (FD = 0.61 cm [95% CI: -0.31, 1.37] and MJL = 0.25 cm [95% CI = -0.48, 0.98]) and both showed a similar width to their LoA.

The current study agrees with Sentürk et al. (2024), who reported that the FD system produced lower JH than the MJL app (mean difference = -1.016 cm, 95% CI = -1.229 – -0.803 cm). However, it should be noted that the difference is larger in this study, with a mean bias of 4.32 cm in favour of the MJL app. Although the previous study (Şentürk et al., 2024) utilised the same equipment, direct comparisons are difficult as the authors did not report the method used to calculate JH from the FD system. Thus, whether they utilised the IM method or flight time to determine JH is unclear and may account for the difference.

Previous versions of the MJL app, which used manual identification of take-off and touch-down to calculate JH, have been reported to be reliable (Balsalobre-Fernández, 2024; Bishop, Jarvis, et al., 2022; Gençoğlu et al., 2023). However, as stated previously, manually calculating JH for every jump can be time-consuming and does not allow for real-time feedback (Balsalobre-Fernández et al., 2015; Şentürk et al., 2024). In agreement with previous research (Balsalobre-Fernández et al., 2015; Şentürk et al., 2024) investigating the MJL app using AI to detect JH, the current study found minimal bias between the two testing sessions.



Therefore, it is suggested that the MJL app with AI activated can reliably measure JH.

While using a dual FP and the IM calculation to measure JH is considered the gold standard (Heredia-Jimenez & Orantes-Gonzalez, 2020; Xu et al., 2023), a considerable cost is associated with purchasing such a system (~£10,000+), which may make it unfeasible for many individuals and teams. Thus, an app-based system such as MJL may be attractive for its cost, ease of use, and portability. The ability to give athletes real-time feedback as they jump is also an attractive feature of the updated app. However, FPs can provide a wealth of kinetic and kinematic data (Lombard et al., 2017) that is not possible using MJL, allowing practitioners to assess both the outcomes and the strategies used to achieve them. This may be useful when using the CMJ to assess NMF, as it has been suggested that athletes who are still not fully recovered can mitigate reductions in JH by altering their jumping strategy (e.g. longer eccentric duration) (Gathercole et al., 2015). Thus, if you only monitored JH, you may incorrectly decide that the athlete is fully recovered, which could increase their injury risk. Therefore, when considering which system to purchase, ease of use, cost implications, and assessing what metrics you need to monitor CMJ performance is necessary.

## Limitations

The key limitation of this study was that we only used sports students and staff as participants. Therefore, the results may not be as applicable to elite athletes who, in general, will jump higher. However, this study did include female participants, which previous research had not done. More research is required on male and female elite athletes.

## Conclusion

In conclusion, the MJL overestimated CMJ height in this study compared to the FD system. However, the app's test-retest reliability was found to be good. Given the relatively cheap cost of the app in comparison to FP systems, teams with smaller budgets may wish to investigate the MJL app if JH is the primary outcome of interest.

## Contributions

LB, BC and JS conceived and designed the study. LB and BC acquired the data, interpreted the results, and drafted and revised the manuscript. JS completed the data analysis and contributed to interpreting the results and manuscript revisions. All authors provided final approval of the version to be published.

## Acknowledgements

The authors wish to thank the participants for taking part in this study.

## Data and Supplementary Material Accessibility

All data and code is presented in the supplementary materials (<https://osf.io/z9q2k/>).

## References

- Allaire, J., Xie, Y., Dervieux, C., McPherson, J., Luraschi, J., Ushey, K., Atkins, A., Wickham, H., Cheng, J., Chang, W., & Iannone, R. (2024). *rmarkdown: Dynamic documents for r*. <https://github.com/rstudio/rmarkdown>
- Amrhein, V., Trafimow, D., & Greenland, S. (2019). Inferential Statistics as Descriptive Statistics: There Is No Replication Crisis if We Don't Expect Replication. *The American Statistician*, 73(sup1), 262–270. <https://doi.org/10.1080/00031305.2018.1543137>
- Balsalobre-Fernández, C. (2024). Real time estimation of vertical jump height with a markerless motion capture smartphone app: A proof-of-concept case study [Journal {{Article}}]. *Proceedings of the Institution of Mechanical Engineers, Part P: Journal of Sports Engineering and Technology*, 17543371241227817.
- Balsalobre-Fernández, C., Glaister, M., & Lockey, R. A. (2015). The validity and reliability of an iPhone app for measuring vertical jump performance [Journal {{Article}}]. *Journal of Sports Sciences*, 33(15), 1574–1579.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Bishop, C., Jarvis, P., Turner, A., & Balsalobre-Fernandez, C. (2022). Validity and reliability of strategy metrics to assess countermovement jump performance using the newly developed smartphone application [Journal {{Article}}]. *Journal of Human Kinetics*, 83(1), 185–195.
- Bishop, C., Turner, A., Jordan, M., Harry, J., Loturco, I., Lake, J., & Comfort, P. (2022). A framework to guide practitioners for selecting metrics during the countermovement and drop jump tests [Journal {{Article}}]. *Strength & Conditioning Journal*, 44(4), 95–103.
- Claudino, J. G., Cronin, J., Mezêncio, B., McMaster, D. T., McGuigan, M., Tricoli, V., Amadio, A. C., & Serrão, J. C. (2017). The countermovement jump to monitor neuromuscular status: A meta-analysis [Journal {{Article}}]. *Journal of Science and Medicine in Sport*, 20(4), 397–402.
- Collings, T. J., Lima, Y. L., Dutailis, B., & Bourne, M. N. (2024). Concurrent validity and test-retest reliability of VALD ForceDecks' strength, balance, and movement assessment tests [Journal {{Article}}]. *Journal of Science and Medicine in Sport*.
- Gathercole, R., Sporer, B., Stellingwerff, T., & Sleivert, G. (2015). Alternative countermovement-jump analysis to quantify acute neuromuscular fatigue [Journal {{Article}}]. *International Journal of Sports Physiology and Performance*, 10(1), 84–92.
- Gençoğlu, C., Ulupınar, S., Özbay, S., Turan, M., Savaş, B. Ç., Asan, S., & İnce, İ. (2023). Validity and reliability of "My Jump app" to assess vertical jump performance: A meta-analytic review [Journal {{Article}}]. *Scientific Reports*, 13(1), 20137.
- Heredia-Jimenez, J., & Orantes-Gonzalez, E. (2020). Comparison of three different measurement systems to assess the vertical jump height [Journal {{Article}}]. *Revista Brasileira de Medicina Do Esporte*, 26, 143–146.

- Hester, J., & Bryan, J. (2024). *glue: Interpreted string literals* [Manual].
- Hughes, S., Chapman, D. W., Haff, G. G., & Nimphius, S. (2019). The use of a functional test battery as a non-invasive method of fatigue assessment [Journal Article]. *PLoS One*, *14*(2), e0212870.
- Landau, W. M. (2021a). *tarchetypes: Archetypes for targets* [Manual].
- Landau, W. M. (2021b). The targets r package: A dynamic make-like function-oriented pipeline toolkit for reproducibility and high-performance computing. *Journal of Open Source Software*, *6*(57), 2959. <https://doi.org/10.21105/joss.02959>
- Lombard, W., Reid, S., Pearson, K., & Lambert, M. (2017). Reliability of metrics associated with a counter-movement jump performed on a force plate [Journal Article]. *Measurement in Physical Education and Exercise Science*, *21*(4), 235–243.
- Loy, A., Steele, S., & Korobova, J. (2024). *Imeresampler: Bootstrap methods for nested linear mixed-effects models* [Manual].
- McMaster, D. T., Gill, N., Cronin, J., & McGuigan, M. (2014). A brief review of strength and ballistic assessment methodologies in sport [Journal Article]. *Sports Medicine*, *44*, 603–623.
- Merrigan, J. J., Strang, A., Eckerle, J., Mackowski, N., Hierholzer, K., Ray, N. T., Smith, R., Hagen, J. A., & Briggs, R. A. (2024). Countermovement jump force-time curve analyses: Reliability and comparability across force plate systems [Journal Article]. *The Journal of Strength & Conditioning Research*, *38*(1), 30–37.
- Mesquida, C., Murphy, J., Lakens, D., & Warne, J. (2022). Replication concerns in sports and exercise science: A narrative review of selected methodological issues in the field [Journal Article]. *Royal Society Open Science*, *9*(12), 220946.
- Nuzzo, J. L., McBride, J. M., Cormie, P., & McCaulley, G. O. (2008). Relationship between countermovement jump performance and multijoint isometric and dynamic tests of strength [Journal Article]. *The Journal of Strength & Conditioning Research*, *22*(3), 699–707.
- Parker, R. A., Scott, C., Inácio, V., & Stevens, N. T. (2020). Using multiple agreement methods for continuous repeated measures data: A tutorial for practitioners. *BMC Medical Research Methodology*, *20*(1), 154. <https://doi.org/10.1186/s12874-020-01022-x>
- Pedersen, T. L. (2024). *patchwork: The composer of plots*. <https://patchwork.data-imaginist.com>
- R Core Team. (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Rodriguez-Sanchez, F., cre, cph, Jackson, C. P., Hutchins, S. D., & Clawson, J. M. (2023). *Grateful: Facilitate Citation of R Packages*.
- Şentürk, D., Yüksel, O., & Akyıldız, Z. (2024). The concurrent validity and reliability of the My Jump Lab smartphone app for the real-time measurement of vertical jump performance [Journal Article]. *Proceedings of the Institution of Mechanical Engineers, Part P: Journal of Sports Engineering and Technology*, 17543371241246439.
- Ushey, K., & Wickham, H. (2023). *renv: Project environments* [Manual].
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Golemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., ... Yutani, H. (2019). Welcome to the

- tidyverse. *Journal of Open Source Software*, 4(43), 1686. <https://doi.org/10.21105/joss.01686>
- Xie, Y. (2014). knitr: A comprehensive tool for reproducible research in R. In V. Stodden, F. Leisch, & R. D. Peng (Eds.), *Implementing reproducible computational research*. Chapman; Hall/CRC.
- Xie, Y. (2015). *Dynamic documents with R and knitr* (2nd ed.). Chapman; Hall/CRC. <https://yihui.org/knitr/>
- Xie, Y. (2024). *knitr: A general-purpose package for dynamic report generation in r*. <https://yihui.org/knitr/>
- Xie, Y., Allaire, J. J., & Golemund, G. (2018). *R markdown: The definitive guide*. Chapman; Hall/CRC. <https://bookdown.org/yihui/rmarkdown>
- Xie, Y., Dervieux, C., & Riederer, E. (2020). *R markdown cookbook*. Chapman; Hall/CRC. <https://bookdown.org/yihui/rmarkdown-cookbook>
- Xu, J., Turner, A., Comfort, P., Harry, J. R., McMahon, J. J., Chavda, S., & Bishop, C. (2023). A systematic review of the different calculation methods for measuring jump height during the countermovement and drop jump tests [Journal {{Article}}]. *Sports Medicine*, 53(5), 1055–1072.