# Concurrent validity and test–retest reliability of VALD ForceDecks' strength, balance, and movement assessment tests

## Author

Collings, TJ, Lima, YL, Dutaillis, B, Bourne, MN

## Published

2024

## Journal Title

Journal of Science and Medicine in Sport

## Version

Version of Record (VoR)

## DOI

10.1016/j.jsams.2024.04.014

## Rights statement

## Downloaded from

## Griffith Research Online

Journal of Science and Medicine in Sport xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Journal of Science and Medicine in Sport

journal homepage: www.elsevier.com/locate/jsams

Original research

# Concurrent validity and test–retest reliability of VALD *ForceDecks*' strength, balance, and movement assessment tests

Tyler J. Collings *, Yuri Lopes Lima, Benjamin Dutaillis, Matthew N. Bourne

*School of Health Sciences and Social Work, Griffith University, Gold Coast Campus, Australia*
*Griffith Centre of Biomedical and Rehabilitation Engineering (GCORE), Menzies Health Institute Queensland, Griffith University, Gold Coast Campus, Australia*

ARTICLE INFO

*Article history:*
Received 25 January 2024
Received in revised form 26 April 2024
Accepted 28 April 2024
Available online xxxx

*Keywords:*
Force plate
Ground reaction force
Center of pressure
Jump testing
Isometric strength
Measurement properties

ABSTRACT

*Objectives:* To evaluate the concurrent validity and test–retest reliability of common movement, strength, and balance tests using portable uniaxial dual force plates.
*Design:* Repeated measures cross-sectional study.
*Methods:* Sixteen healthy individuals participated in two testing sessions, where they performed 12 different movement, strength, and balance tests. Vertical ground reaction force and centre of pressure data were collected using the VALD *ForceDecks* simultaneously with ground-embedded laboratory force plates. Concurrent validity was assessed using root mean square error for raw time-series data and Bland–Altman plots for discrete metrics. Test–retest reliability was assessed using intraclass correlation coefficients and minimal detectable changes.
*Results: ForceDecks* recorded vertical ground reaction forces and center of pressure with high accuracy compared to laboratory force plates. The mean bias between systems was negligible (<2 N or 0.1 mm), with small limits of agreement (<5 N or 1 mm). Overall, 530/674 (79%) showed good or excellent validity (<10% difference) and 611/773 (79%) had good or excellent reliability (intraclass correlation coefficient >0.75). *ForceDecks* reliability was similar to laboratory force plates (<0.07 intraclass correlation coefficient median difference for all metrics).
*Conclusions:* Portable uniaxial force plates record highly accurate vertical ground reaction forces and center of pressure during a range of movement, strength, and balance tests. The *VALD ForcDecks* are a valid and reliable alternative to laboratory force plates when strict standardized testing and data analysis procedures are followed. Users should be aware of the validity and reliability characteristics of the tests and metrics they choose.

## Practical implications

- Portable uniaxial force plates are a valid and reliable alternative to lab-based force plates for assessing strength, balance, and movement.
- The VALD *ForceDecks* recorded raw vertical ground reaction forces and center of pressure with high accuracy compared to the laboratory force plates.
- Most test metrics had good or excellent validity and test–retest reliability, although this can vary depending on the chosen test metrics.
- Users are encouraged to follow the correct testing protocols in order to obtain the most reliable and valid test measures possible.

## 1. Introduction

The measurement of ground reaction force (GRF) using force plates during movement tests is increasingly popular in many areas of sports science, physiotherapy, strength and conditioning, and other health and exercise-based professions.[1] Although force plate technology dates back over a century, recent developments in portable wireless force plate systems have greatly improved affordability and usability.[2,3] Through easy-to-use software that assists with automatic data collection, analysis, and reporting, practitioners can include objective measurements in their decision-making process more easily than ever before.[4]

In sporting and clinical contexts, common applications of force plates include assessing performance (e.g., jump height and force production strategies),[5–7] monitoring fatigue and readiness to train or play,[9] assessing injury risk,[10,11] and informing rehabilitation.[12,13] Vertical GRF reflects a change in the vertical height of the center of mass, and can therefore be used to derive velocity, acceleration, power, and jump height. Further, dual-force platforms enable between-leg asymmetries to be measured. Force plates also enable rapid assessments of isometric strength for the

---

* Corresponding author.
*E-mail address:* t.collings@griffith.edu.au (T.J. Collings).
*Social media:* @TylerCollings3 (T.J. Collings) @YuriLimaPT (Y.L. Lima) @benji_dutaillis (B. Dutaillis) @MBourne5 (M.N. Bourne).

https://doi.org/10.1016/j.jsams.2024.04.014
1440-2440/

Please cite this article as: T.J. Collings, Y.L. Lima, B. Dutaillis, et al., Concurrent validity and test–retest reliability of VALD *ForceDecks*' strength, balance, and movement assessment tests, Journal of Science and Medicine in Sport, https://doi.org/10.1016/j.jsams.2024.04.014

lower and upper limbs,[11] or during multi-joint movements.[14] Force plates comprising four uniaxial load cells can also record the center of pressure (COP) to assess balance and falls risk in older individuals.[15]

With force plate technology being used for assessing athletic and clinical populations, consideration of the reliability and validity of test metrics is vital. The concurrent validity of a new device or method is generally established by comparing it against the highest-quality criterion measurement available (i.e., "gold standard"). Due to variations in human performance, measurements are best performed with both systems simultaneously to enable direct comparisons. To date, a limited number of studies have assessed the concurrent validity of portable uniaxial force plates compared to laboratory force plates.[2,16] One study has assessed concurrent validity during countermovement jumps[16] and another during both countermovement and drop jumps,[2] reporting high agreement and minimal bias between systems. However, the concurrent validity of portable uniaxial force plates for a wide range of movement, strength, and balance tests, across an extensive range of force–time metrics, including COP, has not been established.

Understanding the test–retest reliability and minimal detectable change of specific tests and metrics is important to determine whether longitudinal changes exceed the expected variation in performance and sampling error.[17] For example, when monitoring a patient's rehabilitation progress or assessing the impact of an exercise intervention, ensuring testing procedures have high test–retest reliability can provide confidence in the effect of the treatment, as opposed to measurement error.[18] Test–retest reliability for several force plate tests, primarily using triaxial laboratory force plate systems, has been reported including the Athletic Shoulder Test,[11] unilateral and bilateral countermovement jumps,[8,19] unilateral isometric squats,[7] and COP during balance tests.[13] Many of these tests display acceptable levels of reliability, however, reliability is specific to the device, testing protocol, and metric of interest. For portable uniaxial force plate systems, between-day reliability of countermovement jump metrics has been compared between different commercial systems (ForceDecks, Hawkin Dynamics, and Sparta Science) showing predominantly high reliability (>0.9 intraclass correlation coefficients).[3] Further, a recent study demonstrated that the analysis methods used by different commercial systems (ForceDecks and Hawkin Dynamics) may produce different metric results from the same force–time input data, predominantly effecting temporal variables and force-derivatives, such as velocity.[4] Differences in analysis procedures should be considered when comparing test results and study findings based on different force plate systems. Other than countermovement jumps, no studies have investigated reliability across a diverse range of tests and metrics using portable uniaxial force plates.

Therefore, this study aimed to assess the concurrent validity and test–retest reliability of common movement, strength, and balance tests performed using portable uniaxial dual force plates and create a comprehensive database of results for practitioners to reference.

## 2. Methods

### 2.1. Participants

The sample size for this study (n = 8–19) was determined based on the upper range of expected intraclass correlation coefficient values (>0.9–0.95)[3] for 3 trials (k) with an 80 % probability of obtaining a confidence interval width of 0.2.[20] Participants included 16 healthy individuals (10 males and six females) recruited from an Australian University student population during 2023. Mean ± one standard deviation: age = 25.9 ± 2.9 years, body mass = 71.8 ± 14.9 kg, height = 174.2 ± 12.4 cm, and time between tests = 7.2 ± 0.7 days. To be included, participants were required to: 1) be aged between 18 and 40 years old, 2) be physically active for 30+ min at least three times per week, 3) not have any current upper or lower body injuries, pain, or neurological conditions, and 4) not have any major injuries in the previous 1–2 months that may influence their ability to perform the tests. All participants were encouraged to refrain from performing resistance training or intense physical activity for 24 h before testing. All participants provided written informed consent before data collection. Study ethics were approved by the University Human Research Ethics Committee (reference number: 2023/208).

### 2.2. Equipment setup

Vertical GRF and COP data were collected using portable, uniaxial, dual force plates (*ForceDecks*, FDLite V.2, VALD, Brisbane, Australia) placed on top of two tri-axial force plates embedded in the laboratory floor (AMTI, MA, United States). A stacked force plate arrangement enabled vertical GRF and center of pressure to be simultaneously measured from the *ForceDecks* and laboratory force plates.[2] Both force plates recorded data at 1000 Hz, except for balance tests which *ForceDecks* recorded at 200 Hz.[21] The *ForceDecks* were operated using the iOS application (V1.8.9) on an iPad (Apple, CA, United States) and the laboratory force plates were connected to a desktop computer running the VALD *ForceDecks* software for Windows (V2.0.8587). Force plates were switched on a minimum of 1 h before testing. To assess potential systematic bias in force measurements due to the stacked arrangement of the force plates, a range of static loads from 0 to 100 kg (10, 20, 30, 40, 60, 80, and 100 kg) were recorded using both force plates concurrently. The systematic error ranged from 0 to 3 N (0.3–0.5 % of load), which was considered a negligible difference at all levels of load measured. However, 0–3 N should be considered as the minimum level of agreement expected for concurrent validity analyses.

### 2.3. Data collection

Participants attended two identically structured testing sessions approximately 7 days apart and at the same time of day, to limit the effect of a weekly schedule (e.g. training on a set night of the week), task familiarization/practice, and fatigue from day one to day two. A 5-minute warm-up was performed on a stationary cycle ergometry at a self-selected moderate intensity. Twelve standard tests available in the *ForceDecks* software were performed in a pre-defined order from least to most fatiguing: balance tests first, high rate of force development tests second, and strength tests third. Tests in the order of completion were: *Quiet Stand*, *Single Leg Range of Stability*, *Squat Assessment*, *Countermovement Jump*, *Squat Jump*, *Drop Jump*, *Hop Test*, *Single Leg Hop Test*, *Shoulder ISO-I*, *Shoulder ISO-Y*, *Shoulder ISO-T*, and *Isometric Mid-Thigh Pull*. All tests were performed in accordance with published protocols[5,11,14] as per the VALD Knowledge Base website and are outlined in full detail in Supplementary A. Participants were cued to jump as high as possible during the countermovement jump and as high as possible while minimizing ground contact time for the drop jump and hop test. The drop jump was performed from a box that was 30 cm higher than the force plates.[22] The isometric mid-thigh pull was performed using a portable rig placed under the *ForceDecks* with a short bar attached by a chain. The length of the chain was adjusted to ensure that the knee was at approximately 135° and hip was at 145°, which was verified via a handheld goniometer.[14]

For each test, participants were provided with a demonstration and given a chance to practice until they demonstrated satisfactory performance. Satisfactory performance was based on adherence to strict testing protocols outlined in Supplementary A and was judged by one of two researchers overseeing testing who were experienced force plate users. The signals from both force plates were zeroed before the beginning of each test. The participant was then weighed in a stationary and relaxed standing position looking straight ahead. Three trials of each test were performed, with the unilateral test performed with three trials on the right leg followed by three trials on the left leg. Each repetition began and ended with 2–3 s of quiet standing before performing the next repetition. Approximately 30 s of rest was provided between trials and 1–2 min of rest between tests.

### 2.4. Data processing

Data from both force plates were analyzed with the *ForceDecks* software using default settings. This analysis involved 1) auto-detection of trial repetitions and analysis windows, 2) calculation of force-derived metrics, and 3) exporting of raw force–time data and discrete metrics for further analysis in R studio (v2023.3.0.386, Posit team, Boston, US). The start of the countermovement/squat jump was identified as a 20 N deviation from body weight force. For isometrics, the start of the movement was identified using a yank threshold of 40 N/s for low-force tests (peak < 250 N) and 350 N/s for high-force tests (peak > 250 N).[23] Integration of force–time data was performed from the start of movement for each trial. Trials that were not executed in accordance with the test protocol, such as not landing entirely in the force plate or placing the contralateral leg down during single-leg tests, were removed. Due to the considerable number of possible metrics, a small selection of commonly used

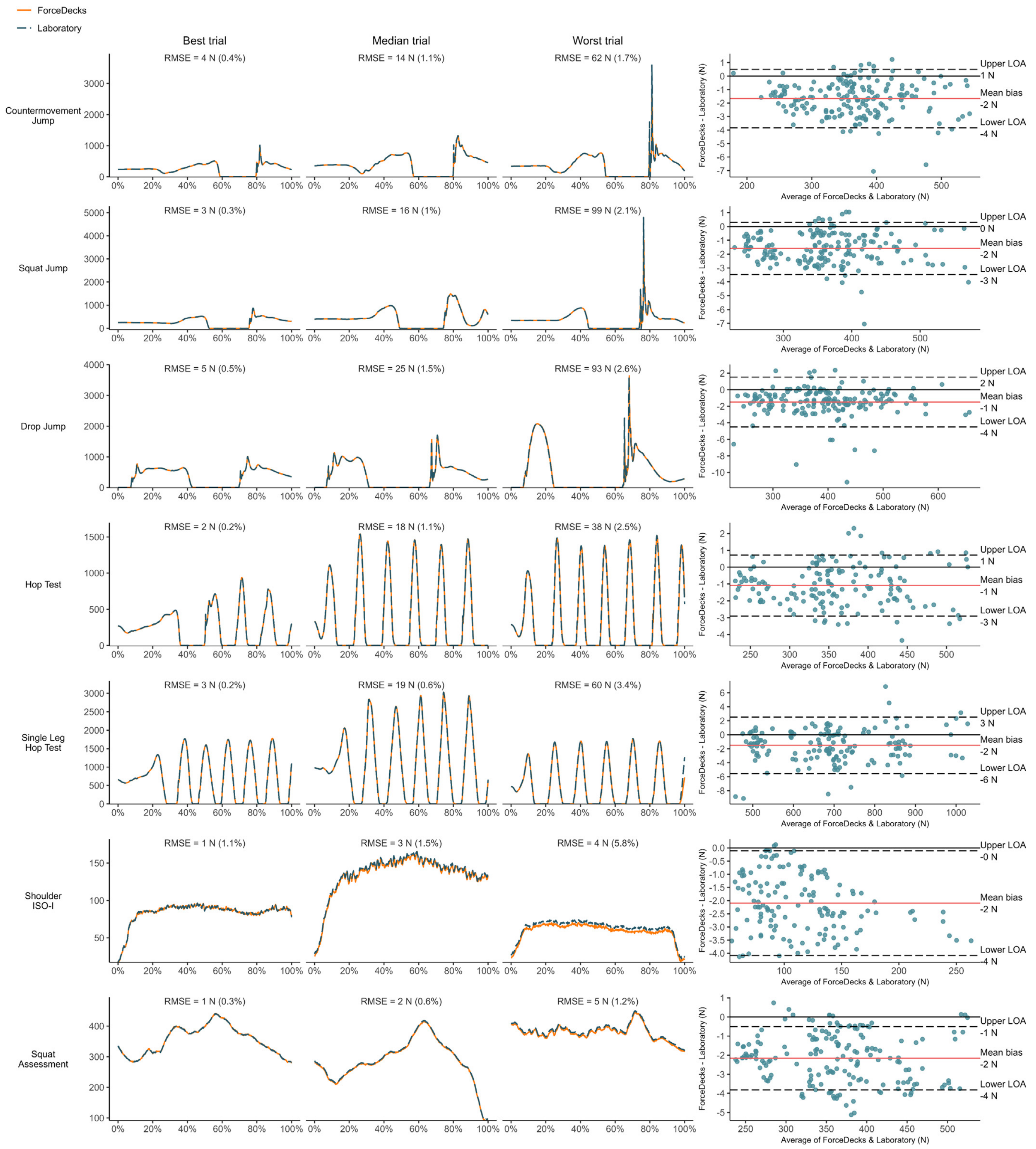

**Fig. 1.** Comparison of vertical ground reaction force time-series data between *ForceDecks* and laboratory force plates for the best, median, and worst individual trials. Bland–Altman plots for the difference between force plates (y-axis) and the average of both force plates (x-axis), with mean bias, and 95 % limits of agreement (LOA).

metrics based on previous studies[11,13,14] and VALD documentation (Supplementary B) is shown in the text. For a complete list of results with all possible metrics, see Supplementary C.

### 2.5. Validity: raw time-series

Force and COP time-series data from the *ForceDecks* and laboratory force plates were cross-correlated removing any temporal delay due to the systems starting to record at different times. Each trial was time-normalized from 0 to 100 % of the automatically detected analysis window. Agreement between systems over the full analysis window was assessed using root mean square error (RMSE), where $\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}(ForceDecks_i - Laboratory_i)^2}{N}}$. To visualize the agreement between force plates for individual trials, one trial with the best/lowest RMSE, the typical/median RMSE, and worst/highest RMSE are shown. The mean force and COP were also calculated from the full analysis window and compared between force plates using Bland–Altman plots, mean bias, and mixed effects limits of agreement to account for multiple trials per participant.[24]

### 2.6. Validity: discrete metrics

The validity of *ForceDecks* discrete metrics was determined by comparing to metrics derived concurrently from laboratory force plates. The difference between force plates (i.e. measurement error) was expressed as a percentage ($\% = \frac{ForceDecks\ -\ Laboratory}{Laboratory} * 100$) and interpreted as poor (>20 %), moderate (10–20 %), good (5–10 %), or excellent (<5 %), based on commonly used thresholds for decision-making in clinical practice.[25] Please note that variables with small units may inflate relative errors, while variables with large units may reduce relative errors. For unilateral tests, validity statistics were averaged for the left and right legs. Due to time delays in the start of data recording between systems, quiet stand and single-leg range of stability tests were first cross-correlated, adjusted for time-lag, cropped to equal size windows, and reuploaded to the *ForceDecks* software for analysis.

### 2.7. Reliability

Test–retest reliability of force plate metrics between days (~1 week apart) was assessed with intraclass coefficients (ICCs) using a two-way random effects model, absolute agreement, and multiple measurements (ICC2,k, where k is the mean of 3 repetitions)[17] ICC values were interpreted as poor (<0.5), moderate (0.5–0.75), good (0.75–0.9), or excellent (>0.9), based on the lower to upper bounds of the 95 % confidence interval where possible.[26] Measurement variability within a participant was quantified using standard error of measurement (SEM), where $SEM = total\ SD\sqrt{1 - ICC}$.[17,27] The minimal detectable change (MDC) representing the smallest real difference between two testing time points was calculated using the SEM and the 95 % confidence interval for detecting a change in scores beyond 0, where $MDC = 1.96 \times \sqrt{2} \times SEM$.[17,27] Reliability was calculated for both force plate systems across testing days using, enabling a comparison of reliability values between *ForceDecks* and laboratory systems. The reliability measured using ICCs was compared for all possible metrics between force plate systems using median absolute difference due to the skewed distribution with a center close to zero. For unilateral tests, reliability statistics were averaged across the left and right legs to summarize results into a representative value regardless of the leg assessed. To ensure consistency in body weight normalized metrics between systems, the body weight recorded from the laboratory force plates was modified to match the *ForceDecks* measured body weight.

## 3. Results

A total of 12 tests, 1255 trials, and 773 unique metrics were analyzed. For a comprehensive report and lookup table of specific metrics, please see Supplementary C.

### 3.1. Validity: raw time-series

Force and COP data recorded using *ForceDecks* and laboratory force plates were highly consistent in terms of pattern and amplitude (Figs. 1, 2).

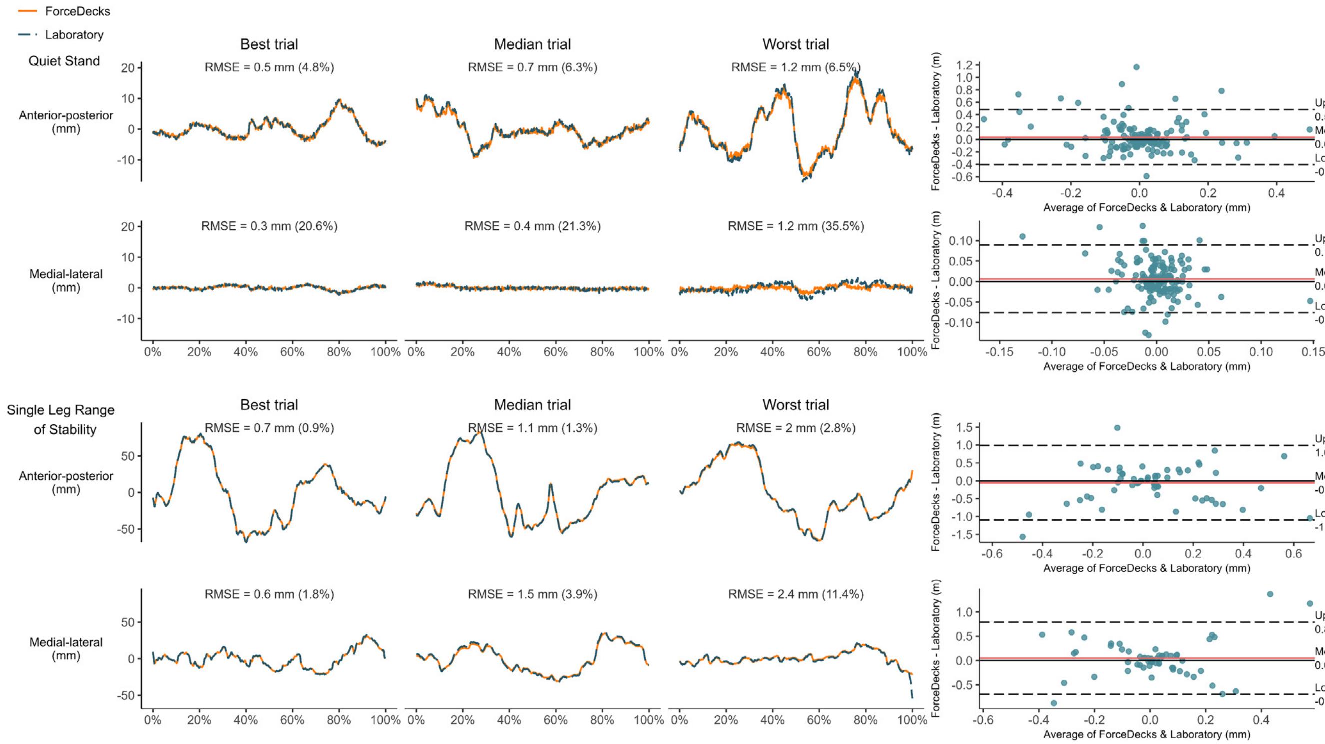

**Fig. 2.** Comparison of center of pressure (COP) time-series data between *ForceDecks* and laboratory force plates for the best, median, and worst individual trials. Bland–Altman plots for the difference in COP between force plates (y-axis) and the average COP of both force plates (x-axis), with mean bias, and 95 % limits of agreement (LOA).

For vertical GRF across the duration of a trial, the median RMSE between systems ranged from 2.5 to 25.2 N, or 0.6 to 1.5 % error relative to the peak amplitude. The mean bias between systems ranged from −2 to −1 N lower for *ForceDecks*, with limits of agreement between −6 and 3 N (Fig. 1).

For COP across the duration of a trial, the median RMSE between systems ranged from 0.4 to 1.5 mm, or 1.3 to 21.3 % error relative to the peak amplitude. The mean bias between systems ranged from −0.1 to 0.0 mm lower for *ForceDecks*, with limits of agreement between −1.1 and 1.0 mm (Fig. 2).

### 3.2. Validity: discrete metrics

Overall, 67.8 % (457/674) of metrics had *excellent* validity, 10.7 % (72/674) of metrics showed *good* validity, 9.4 % (63/674) of metrics showed *moderate* validity, and 12.2 % (82/674) of metrics showed *poor* validity (Fig. 3A).

For a selection of common metrics (Table 1), *excellent* validity was found for squat jump, drop jump, hop test, and squat assessment (relative difference = 0–4 %), *good* to *excellent* for countermovement jump and single-leg hop test (relative difference = 0–9 %), *moderate* to *excellent* for shoulder isometric 'I', 'T', and 'Y' and quiet stand (relative difference = 2–18 %) and *good* for single-leg range of stability (7–9 %).

### 3.3. Reliability

Overall, 49.8 % (385/773) of *ForceDecks* metrics had *excellent* reliability, 29.5 % (228/773) of metrics had *good* reliability, 12.8 % (99/773) of metrics had *moderate* reliability, and 7.9 % (61/773) of metrics had *poor* reliability (Fig. 3B).

For a selection of common metrics (Table 1), *excellent* reliability was found for countermovement jump, squat jump, and drop jump (ICC = 0.91–0.99), *good* to *excellent* for single-leg hop test, isometric mid-thigh pull, shoulder isometric 'Y', squat assessment, and single-leg range of stability (ICC = 0.77–0.89), *good* for hop test (ICC = 0.77–0.89), and *poor* for quiet stand (ICC = 0.21–0.43). Minimal detectable changes between days ranged widely depending on the unit of measurement and metric reliability (Table 1).

Test–retest reliability was highly consistent between *ForceDecks* and laboratory force plates (Fig. 4A). For specific tests, the median absolute difference in ICCs between force plate systems ranged from <0.001 for squat assessment to 0.07 for shoulder isometric 'Y' (Fig. 4B).

## 4. Discussion

This study provides a comprehensive evaluation of the validity and reliability of strength, balance, and movement assessments performed with VALD *ForceDecks* compared to laboratory force plates. The *ForceDecks* recorded vertical GRF and COP with high accuracy compared to the laboratory force plates for all tests. The systematic bias for mean force and COP data between systems was negligible (2 N) with narrow limits of agreement (less than ± 5 N and ± 1 mm, respectively). Overall, most *ForceDecks* tests and metrics displayed excellent (ICC > 0.9) or good test–retest reliability (ICC > 0.75) which closely matched the reliability of laboratory force plate data. Some test metrics displayed poor reliability due to natural variation in task performance between days, and users need to be aware of the reliability properties of the tests and metrics selected. When strict standardized testing and data analysis procedures are followed, portable uniaxial force plates are a highly valid and reliable alternative to laboratory-based force plates for assessing vertical GRF and COP during a wide range of tests.

### 4.1. Validity: raw time-series

Over the full duration of a trial, *ForceDecks* recorded raw vertical GRF with high accuracy compared to the laboratory force plates. Across all tests, the agreement between force plate systems for a typical/median trial ranged from 2 to 25 N. For jump landing tests with high impact forces (up to 4000 N), the agreement between force plates was lower than the strength and balance tests due to momentary spikes in force. However, when considering the agreement between force plates relative to the amplitude of the signal, all trials were within 0.3 to 5.8 %. When considering the mean force recorded during a trial, the mean bias between force plates was consistently between −1 N and −2 N lower for the *ForceDecks* compared to the laboratory force plates. This mean bias is a negligible amount, given that a −2 N difference between

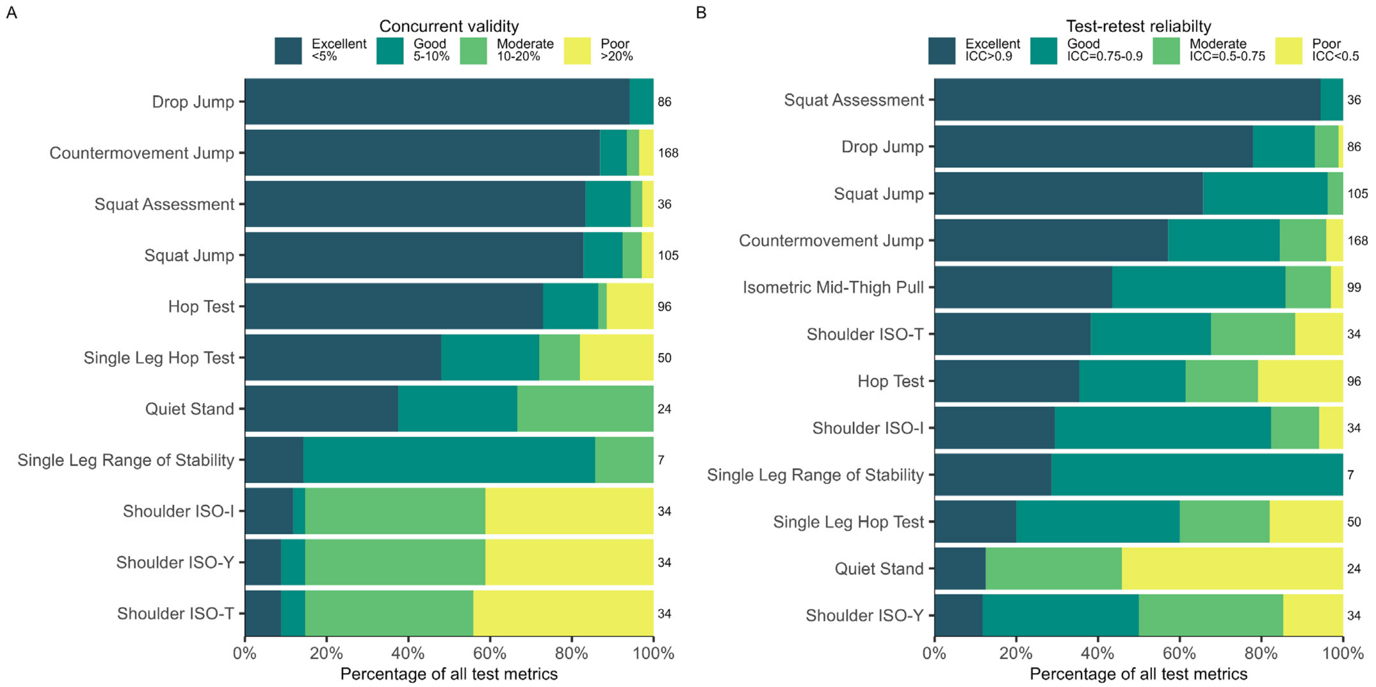

**Fig. 3.** Interpretation summary of concurrent validity (A) and test–retest reliability based on ICC point estimates (B) for all VALD *ForceDecks* tests and metrics. Numbers at the end of each bar represent the total number of unique metrics for each test.

**Table 1**

Test–retest and concurrent validity statistics for a selection of common test metrics for movement, strength, and balance tests using *ForceDecks* and laboratory force plates.

| Test | Metric | Test–retest reliability[a] | | | | | | Concurrent validity | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Day 1 Mean (SD) | Day 2 Mean (SD) | ICC (95 % CI) | SEM | MDC | Interpretation | ForceDecks Mean (SD) | Laboratory Mean (SD) | Relative difference | Interpretation |
| Countermovement jump | Jump height (Imp-Mom) [cm] | 34 (7) | 34 (8) | 0.97 (0.92–0.99) | 1.2 | 3.4 | Excellent | 34 (8) | 35 (8) | 5 % | Good |
| Countermovement jump | Positive takeoff impulse [N s] | 284 (74) | 279 (68) | 0.96 (0.90–0.99) | 13 | 37 | Excellent | 281 (70) | 283 (71) | 1 % | Excellent |
| Countermovement jump | Takeoff peak force [N] | 1568 (392) | 1612 (405) | 0.99 (0.96–1.00) | 42 | 117 | Excellent | 1590 (397) | 1594 (398) | 0 % | Excellent |
| Squat jump | Jump height (Imp-Mom) [cm] | 31 (6) | 31 (7) | 0.94 (0.82–0.98) | 1.6 | 4.3 | Good–excellent | 31 (6) | 32 (7) | 3 % | Excellent |
| Squat jump | Takeoff peak force [N] | 1700 (386) | 1717 (422) | 0.98 (0.96–0.99) | 50 | 139 | Excellent | 1708 (401) | 1712 (403) | 0 % | Excellent |
| Drop jump | Peak landing force [N] | 3378 (849) | 3552 (904) | 0.94 (0.81–0.98) | 215 | 596 | Good–excellent | 3470 (856) | 3556 (898) | 2 % | Excellent |
| Drop jump | RSI (flight time/contact time) | 1.2 (0.4) | 1.3 (0.4) | 0.97 (0.91–0.99) | 0.1 | 0.2 | Excellent | 1.2 (0.4) | 1.2 (0.4) | 0 % | Excellent |
| Hop test | Best impulse [N s] | 338 (103) | 311 (70) | 0.89 (0.70–0.96) | 28 | 79 | Moderate–excellent | 325 (85) | 326 (85) | 1 % | Excellent |
| Hop test | Contact time [ms] | 192 (30) | 193 (36) | 0.77 (0.37–0.92) | 15 | 43 | Poor–excellent | 193 (32) | 194 (34) | 1 % | Excellent |
| Single leg hop test | Active stiffness [N/m] | 22,649 (6472) | 23,131 (7817) | 0.87 (0.65–0.96) | 2397 | 6645 | Moderate–excellent | 22,860 (6530) | 22,147 (6717) | 6 % | Good |
| Single leg hop test | Best peak force [N] | 2179 (411) | 2158 (412) | 0.97 (0.91–0.99) | 67 | 186 | Excellent | 2153 (396) | 2170 (411) | 1 % | Excellent |
| Isometric mid-thigh pull | Peak vertical force [N] | 2063 (503) | 2064 (505) | 0.99 (0.97–1.00) | 52 | 143 | Excellent | [b] | [b] | [b] | [b] |
| Isometric mid-thigh pull | RFD – 200 ms [N/s] | 2629 (1233) | 2610 (1328) | 0.82 (0.50–0.94) | 531 | 1471 | Moderate–excellent | [b] | [b] | [b] | [b] |
| Shoulder ISO-I | Peak vertical force [N] | 148 (58) | 135 (52) | 0.95 (0.82–0.99) | 12 | 32 | Good–excellent | 144 (55) | 147 (56) | 2 % | Excellent |
| Shoulder ISO-I | RFD – 200 ms [N/s] | 254 (151) | 222 (112) | 0.92 (0.75–0.98) | 36 | 100 | Good–excellent | 245 (135) | 260 (155) | 18 % | Moderate |
| Shoulder ISO-T | Peak vertical force [N] | 108 (34) | 116 (49) | 0.92 (0.67–0.98) | 11 | 32 | Moderate–excellent | 111 (40) | 114 (40) | 2 % | Excellent |
| Shoulder ISO-T | RFD – 200 ms [N/s] | 199 (129) | 234 (166) | 0.91 (0.67–0.98) | 43 | 119 | Moderate–excellent | 215 (140) | 224 (134) | 15 % | Moderate |
| Shoulder ISO-Y | Peak vertical force [N] | 121 (53) | 114 (40) | 0.98 (0.92–0.99) | 6.4 | 18 | Excellent | 121 (50) | 123 (50) | 3 % | Excellent |
| Shoulder ISO-Y | RFD – 200 ms [N/s] | 229 (145) | 166 (101) | 0.87 (0.52–0.96) | 45 | 125 | Moderate–excellent | 217 (144) | 218 (133) | 18 % | Moderate |
| Squat assessment | Concentric peak force asymmetry [% L:R] | 3.1 (8.8) | 3.4 (7.3) | 0.81 (0.46–0.94) | 3.4 | 9.5 | Poor–excellent | 3.3 (7.4) | 3.7 (7.4) | −4 % | Excellent |
| Squat assessment | Maximum negative displacement [cm] | −27.16 (10.50) | −26.71 (9.38) | 0.91 (0.75–0.97) | 2.9 | 7.9 | Good–excellent | −27.03 (9.42) | −26.47 (9.36) | −3 % | Excellent |
| Quiet stand | Area of CoP ellipse [mm$^2$] | 25 (19) | 20 (17) | 0.21 (0.00–0.72) | 16 | 45 | Poor–moderate | 23 (14) | 28 (16) | 17 % | Moderate |
| Quiet stand | CoP range – anterior–posterior [mm] | 17 (7) | 16 (4) | 0.43 (0.00–0.80) | 4.4 | 12 | Poor–good | 16 (5) | 17 (5) | 4 % | Excellent |
| Single leg range of stability | Area of CoP ellipse [mm$^2$] | 6823 (3150) | 7268 (3351) | 0.91 (0.73–0.97) | 950 | 2634 | Moderate–excellent | 6748 (3069) | 7283 (3219) | 9 % | Good |
| Single leg range of stability | Mean velocity [mm/s] | 80 (23) | 79 (18) | 0.89 (0.67–0.96) | 6.8 | 19 | Moderate–excellent | 77 (21) | 83 (23) | 7 % | Good |

95 % CI = 95 % confidence interval; ICC = intraclass coefficient; L = left; LOA = limits of agreement; MDC = minimal detectable change; R = right; SD = standard deviation; SEM = standard error of measurement.

[a] Based on the mean of 3 trials per participant.

[b] No laboratory ground reaction forces were able to be measured due to the IMTP platform.

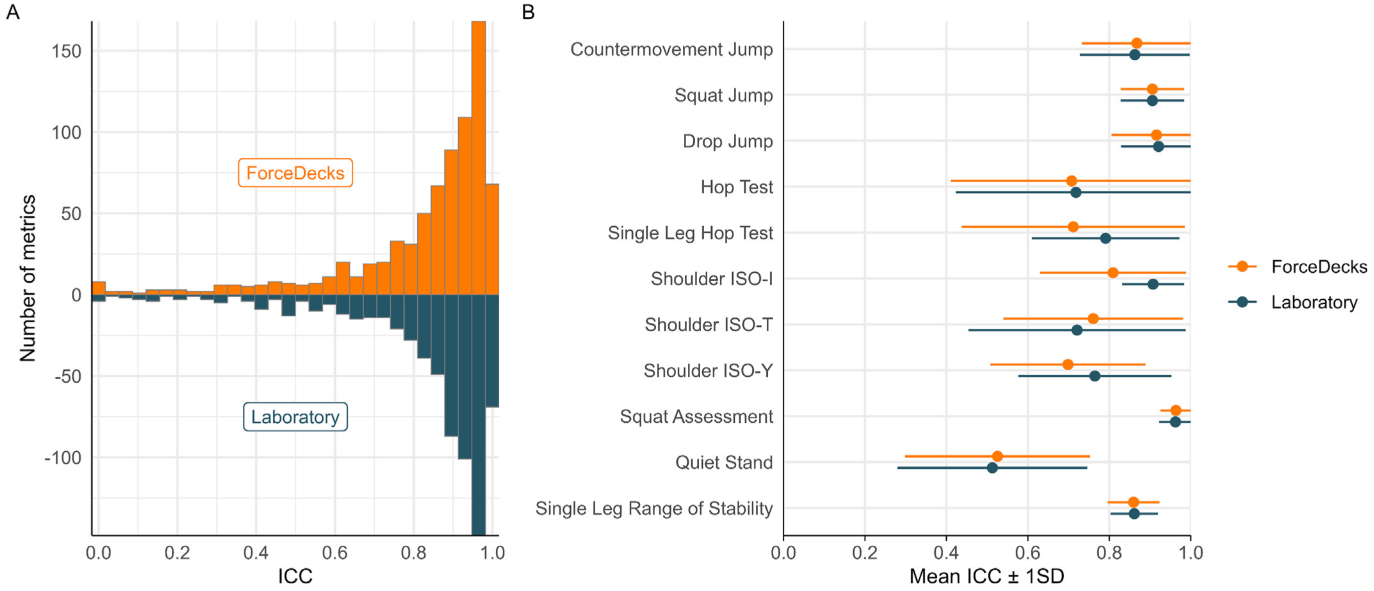

**Fig. 4.** Comparison of test–retest reliability between *ForceDecks* and laboratory force plates using intraclass correlation coefficients (ICCs). A) shows the distribution of ICC values for all metrics and tests for *ForceDecks* (top, orange) and laboratory force plates (bottom, navy). B) shows the mean ICC ± 1 standard deviation (SD) for all metrics of each test. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

force plates was observed due to slight differences in calibration assessed before testing using a static mass. The minimal bias between force plate systems is consistent with previous portable force plate validation studies.[2,16,28] Errors as small as 2 N are not likely to influence decision-making in an applied setting, where differences between injured and uninjured populations or limbs are much greater.[29]

Raw COP data recorded using the *ForceDecks* also had close agreement with the laboratory force plates throughout both balance tests. For a typical trial, anterior–posterior and medial–lateral COP was within 0.7 mm of the laboratory force plates for quiet stand and 1.5 mm for single leg range of stability. Although the absolute difference is small, the relative error appears large (up to 21.3 % due to the small deviations in COP that occur during quiet standing). To the best of our knowledge, this is the first study to report the accuracy of raw COP time-series data using a portable uniaxial force plate. The high agreement between systems indicates users can be confident in the COP data acquired during *ForceDecks* balance tests and can be considered a highly valid alternative to laboratory systems.

### *4.2. Validity: discrete metrics*

Of the common metrics shown in Table 1, the majority had excellent agreement with laboratory force plates (<5 % relative difference). Given that the *ForceDecks* raw force–time data showed high accuracy compared to laboratory force plates, metrics such as peak force will have equally high validity. However, where metrics are time dependent, such as rate of force development and impulses, concurrent validity between force plate systems is dependent on the identification of movement start and phases. The phases during jump-landing tests (drop jump, countermovement jump, squat jump) were identified with high accuracy resulting in excellent test validity across the majority of possible metrics (Fig. 3). Conversely, shoulder isometric 'Y', 'T', and 'I' metrics had the lowest agreement with laboratory force plates. Low validity was due to small differences in start of movement detection between force plates that greatly influence rate of force development or impulse over short periods (e.g., 50–200 ms). These results are comparable to previous validity studies, which reported high agreement compared to laboratory force plates, with lower validity in phase-dependent metrics.[2,16,28]

Balance testing COP metrics showed moderate to excellent validity with a 4 to 17 % relative difference between force plate systems. Due to the small measurement units of COP, small absolute differences between systems can appear as large relative errors. For metrics such as the area of COP ellipse, small errors between systems may accumulate throughout the trial increasing the total metric error relative to the laboratory force plate. Compared to other portable balance testing devices, *ForceDecks* appeared to be more accurate, however, insufficient reporting of data or not collecting COP simultaneously limits the ability to compare findings between studies.[30,31]

### *4.3. Reliability*

For the selection of commonly used metrics presented in Table 1, most tests showed excellent reliability (ICC > 0.9) indicating highly repeatable test results between days. Notably, the reliability of all *ForceDecks* test metrics was highly consistent with the reliability of the laboratory force plates, regardless of being poor or excellent reliability (less than 0.1 ICC difference between systems) (Fig. 4). A strength of this study is the ability to compare the concurrent reliability of force plate systems, which is a major limitation of previous studies comparing devices.[3] The results suggest that the portable uniaxial force plates used in this study may therefore be used as an equally reliable alternative to laboratory force plates.

High reliability and low expected measurement variability provide confidence that users can detect small changes in test performance between testing sessions, such as examining training-induced changes in strength or performance after an intervention. The high reliability observed in this study supports previous findings for force plate testing.[8,11,19] Where poor reliability was found, for example, during COP of quiet standing or single-leg range of stability, test metrics generally had small absolute values (e.g., 25 $mm^2$) and large variability across individuals (e.g., 19 $mm^2$ standard deviations), as seen in a previous study.[2]

It is important to note that test–retest reliability contains multiple sources of measurement error beyond the device's accuracy, such as variation in human performance on separate days. Therefore, test–retest reliability is predominantly a property of the test metric and the consistency of the performer. Familiarization with testing protocols, training history, or events that occur between testing days may all impact test–retest reliability. Performing more trials and taking the mean or peak of multiple trials may help improve reliability.

To maximize the validity and reliability of data using *ForceDecks*, it is important to recognize the significant impact users have in obtaining high-quality results. The accuracy of calculated metrics depends on following strict testing protocols, such as maintaining a steady 3 s period before initiating movement to ensure the correct start of movement is detected, ensuring feet land entirely within the force plate during jumping tests, or avoiding placing the contralateral leg down before the end of a single-leg test. If any errors are suspected in testing protocols, it is recommended that the test is repeated. Further, it is recommended to double-check the automated test detection, leg used, and start of movement are accurate.

### *4.4. Study limitations*

A main limitation of this study was the stacked arrangement of force plates to enable GRF to be recorded concurrently. Placing the *ForceDecks* on top of the laboratory force plates may have introduced small errors in force and COP recorded by the laboratory force plates. Nevertheless, this experimental setup was necessary to evaluate concurrent validity as opposed to completing separate trials on each force plate. A second limitation to determining the concurrent validity of metrics was the difficulty in ensuring both systems were calculating metrics from the exact same time-window. Extra efforts were made to ensure metrics were as comparable as possible, such as time-syncing balance test data using cross-correlation before extracting metrics. Further, readers should be aware that the validity and reliability of force plate metrics are specific to the analysis methods used by the ForceDecks software and may not generalize to other force plate systems.[3,4]

## 5. Conclusion

Portable uniaxial force plates recorded vertical GRF and COP with high accuracy compared to laboratory force plates during a range of movement, strength, and balance tests. Overall, most test metrics had excellent or good validity and test–retest reliability. However, validity and reliability are specific to the test type and metrics selected and users need to be aware of these characteristics when planning testing and interpreting test results. Metric validity and reliability are heavily influenced by factors beyond the device, and it is important users conduct testing with strict protocols.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jsams.2024.04.014.

## Funding information

This research was funded by VALD Performance, Brisbane, Australia and Griffith University Health Group.

## Confirmation of ethical compliance

All participants provided written informed consent before data collection. Study ethics were approved by the Griffith University Human Research Ethics Committee (GU reference: 2023/208).

## CRediT authorship contribution statement

**Tyler J. Collings:** Conceptualization, Methodology, Software, Validation, Formal analysis, Data curation, Writing – original draft, Visualization, Supervision, Project administration, Funding acquisition. **Yuri Lopes Lima:** Methodology, Investigation, Data curation, Writing – review & editing. **Benjamin Dutaillis:** Methodology, Investigation, Writing – review & editing. **Matthew N. Bourne:** Methodology, Resources, Writing – review & editing, Supervision, Funding acquisition.

## Declaration of interest statement

All co-authors affirm that they have no financial interest in the commercial organization or products used in this study. The authors declare a perceived conflict of interest due to previous and current research unrelated to this study that was/is partly funded or supported by VALD. The authors also declare personal relationships with several members of the VALD organization. Results were reported to VALD as part of the funding arrangement and approval to publish was provided without any adjustments being made to the manuscript. All data collection, analysis, and interpretation were performed independently of VALD staff.

## Acknowledgments

Thank you to Dr Daniel Devaprakash and Dr Gavin Lenton for their assistance, VALD Performance and Griffith Health Group for funding this research, and all participants who volunteered their time.

## References

1. Bishop C, Turner A, Jordan M et al. A framework to guide practitioners for selecting metrics during the countermovement and drop jump tests. *Strength Cond J* 2022;44(4):95-103.
2. Badby AJ, Mundy PD, Comfort P et al. The validity of Hawkin dynamics wireless dual force plates for measuring countermovement jump and drop jump variables. *Sensors* 2023;23(10).
3. Merrigan JJ, Strang A, Eckerle J et al. Countermovement jump force–time curve analyses: reliability and comparability across force plate systems. *J Strength Cond Res* 2022;38(1):30-37.
4. Merrigan JJ, Stone JD, Galster SM et al. Analyzing force-time curves: comparison of commercially available automated software and custom MATLAB analyses. *J Strength Cond Res* 2022;36(9):2387-2402.
5. McMahon JJ, Suchomel TJ, Lake JP et al. Understanding the key phases of the countermovement jump force–time curve. *Strength Cond J* 2018;40(4):96-106.
6. Jarvis P, Turner A, Read P et al. Reactive strength index and its associations with measures of physical and sports performance: a systematic review with meta-analysis. *Sports Med* 2022;52(2):301-330.
7. Bishop C, Read P, Chavda S et al. Using unilateral strength, power and reactive strength tests to detect the magnitude and direction of asymmetry: a test–retest design. *Sports* 2019;7(3).
8. Lombard W, Reid S, Pearson K et al. Reliability of metrics associated with a countermovement jump performed on a force plate. *Meas Phys Educ Exerc Sci* 2017;21(4):235-243.
9. Claudino JG, Cronin J, Mezêncio B et al. The countermovement jump to monitor neuromuscular status: a meta-analysis. *J Sports Sci Med* 2017;20(4):397-402.
10. Collings TJ, Diamond LE, Barrett RS et al. Strength and biomechanical risk factors for non-contact ACL injury in elite female footballers: a prospective study. *Med Sci Sports Exerc* 2022;54(8):1242-1251.
11. Ashworth B, Hogben P, Singh N et al. The Athletic Shoulder (ASH) test: reliability of a novel upper body isometric strength test in elite rugby players. *BMJ Open Sport Exerc Med* 2018;4(1).
12. Dutaillis B, Diamond LE, Lazarczuk SL et al. Vertical jump testing following anterior cruciate ligament reconstruction: a systematic review and meta-analysis. *Med Sci Sports Exerc* 2023;56(2).
13. Aryan R, Inness E, Patterson KK et al. Reliability of force plate-based measures of standing balance in the sub-acute stage of post-stroke recovery. *Heliyon* 2023;9(10).
14. Comfort P, Dos'Santos T, Beckham GK et al. Standardization and methodological considerations for the isometric midthigh pull. *Strength Cond J* 2019;41(2):57-79.
15. Quijoux F, Vienne-Jumeau A, Bertin-Hugault F et al. Center of pressure displacement characteristics differentiate fall risk in older people: a systematic review with meta-analysis. *Ageing Res Rev* 2020;62:101117.
16. Lake J, Mundy P, Comfort P et al. Concurrent validity of a portable force plate using vertical jump force⇓time characteristics. *J Appl Biomech* 2018;34(5):410-413.
17. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *J Strength Cond Res* 2005;19(1):231-240.
18. Lexell JE, Downham DY. How to assess the reliability of measurements in rehabilitation. *Am J Phys Med Rehabil* 2005;84(9):719-723.
19. Pérez-Castilla A, García-Ramos A, Janicijevic D et al. Between-session reliability of performance and asymmetry variables obtained during unilateral and bilateral countermovement jumps in basketball players. *PloS One* 2021;16(7):e0255458.
20. Borg DN, Bach AJ, O'Brien JL et al. Calculating sample size for reliability studies. *PM&R* 2022;14(8):1018-1025.
21. Scoppa F, Capra R, Gallamini M et al. Clinical stabilometry standardization: basic definitions–acquisition interval–sampling frequency. *Gait Posture* 2013;37(2):290-292.
22. Addie CD, Arnett JE, Neltner TJ et al. Effects of drop height on drop jump performance. *Int J Kinesiol Sports Sci* 2019;7(4):28-32.
23. Pinto BL, Callaghan JP. Movement onset detection methods: a comparison using force plate recordings. *J Appl Biomech* 2023;39(2):118-123.

24. Parker RA, Weir CJ, Rubio N et al. Application of mixed effects limits of agreement in the presence of multiple sources of variability: exemplar from the comparison of several devices to measure respiratory rate in COPD patients. *PloS One* 2016;11(12): e0168321.
25. Read PJ, McAuliffe S, Bishop C et al. Asymmetry thresholds for common screening tests and their effects on jump performance in professional soccer players. *J Athl Train* 2020;56(1):46-53.
26. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med* 2016;15(2):155-163.
27. Beckerman H, Roebroeck M, Lankhorst G et al. Smallest real difference, a link between reproducibility and responsiveness. *Qual Life Res* 2001;10(7):571-578.
28. Walsh MS, Ford KR, Bangen KJ et al. The validation of a portable force plate for measuring force–time data during jumping and landing tasks. *J Strength Cond Res* 2006;20 (4):730-734.
29. Collings TJ, Diamond LE, Barrett RS et al. Impact of prior anterior cruciate ligament, hamstring or groin injury on lower limb strength and jump kinetics in elite female footballers. *Phys Ther Sport* 2021;52:297-304.
30. Walsh M, Church C, Hoffmeister A et al. Validation of a portable force plate for evaluating postural sway. *Percept Mot Skills* 2021;128(1):191-199.
31. Golriz S, Hebert JJ, Foreman KB et al. The validity of a portable clinical force plate in assessment of static postural control: concurrent validity study. *Chiropr Man Therap* 2012;20:1-6.