SPORTLOGIA
Scientific-Expert Journal of Anthropological Aspects of Sports, Physical Education and Recreation

Submitted: 23/04/2024
Accepted: 15/05/2024
Published: 02/08/2024
UDK: to be registered

Original article

Validity and Reliability of an Adapted Leg/Back Isometric Strength Testing Device

[1]Robert G. Lockie, [2,3]J. Jay Dawes, [1]David A. Alvarez

[1]Department of Kinesiology, California State University, Fullerton, Fullerton, CA, USA.
[2]Department of Health and Human Performance, Oklahoma State University, Stillwater, OK, USA.
[3]Tactical Fitness and Nutrition Lab, Oklahoma State University, Stillwater, OK, USA.

*Correspondence: rlockie@fullerton.edu

Abstract

The leg/back dynamometer (LBD) is a valid lower-body strength test. For people stronger than the typical device, a crane scale could provide an adapted leg/back dynamometer (ALBD), although validity and reliability analyses are needed. Fifty participants completed three testing sessions, consisting of three LBD and ALBD trials each. One-sample t-tests determined whether LBD and ALBD mean differences were significant compared to 0. Paired samples t-tests and effect sizes (*d*) compared average and maximum LBD and ALBD. Pearson's correlations and regression derived LBD and ALBD relationships. Intra-class correlation coefficients (ICC) and coefficient of variation (CV) assessed reliability. A three-way repeated measures ANOVA compared between-session LBD and ALBD. Test usefulness was analyzed by typical error (TE) and smallest worthwhile change (SWC). The mean differences analyses indicated a fixed bias ($p \leq 0.04$); the LBD recorded greater values than the ALBD ($p<0.01$; $d=0.43$-0.89). Relationships between the LBD and ALBD across all sessions were significant ($p<0.001$; $r=0.902$-0.985), with 94% explained variance. The ICCs and CVs for all sessions were acceptable (ICC≥0.97; CV≤6%). There were no significant LBD or ALBD differences between Sessions 1 or 2. However, average and maximum LBD for Session 3 was greater than Sessions 1 and 2; Session 3 average ALBD was greater than Sessions 1 and 2 ($p \leq 0.002$); and maximum ALBD was greater than Session 1 ($p \leq 0.042$). The LBD and ALBD were useful (TE=5.13-5.65kg; SWC=7.26-7.99kg). The ALBD can measure strength, although the data was lower than the LBD. The LBD and ALBD are reliable and useful; two practice sessions could enhance reliability.

Keywords: Cane scale, Familiarization, Intra-class correlation coefficient, Lower-body strength, Leg/back dynamometer

Introduction

Strength is an essential quality for many different athletic (Suchomel et al., 2016) and occupational (Orr et al., 2022) activities. As a result, strength testing is a common tool used by practitioners. The data derived from strength tests can highlight the limitations for a particular individual and help drive appropriate training programs (McGuigan, 2015). Repetition-maximum (RM) tests with dynamic exercises (e.g., bench press, squat, deadlift) are often used to measure strength. While valuable, this type of testing may not be appropriate for all individuals (e.g., those with limited resistance training experience). An ideal strength test would be an assessment that is easy to administer on large groups, limits fatigue that could influence other aspects of testing or training, and is safe to perform regardless of age, sex, and relative skill level (Sheppard et al., 2013; West et al., 2011). A leg/back dynamometer (LBD) may provide a test that is easy to administer, while also being an affordable option for many practitioners.

The LBD has been used to assess strength in high school athletes (Lockie et al., 2023a; Lockie et al., 2023b; Wakely et al., 2022), university athletes (Najiah et al., 2021), and tactical personnel such as first responders (Dawes et al., 2019; Dawes et al., 2017; Lockie et al., 2020a; Lockie et al., 2020b). This test allows for maximal force exertion against an external resistance and dynamometry provides an efficient way to measure strength (Najiah et al., 2021; Ten Hoor et al., 2016). As a measure of validity, Najiah et al. (2021) found very large-to-near perfect relationships between the LBD (measured in kg) with the 1RM back squat ($p < 0.001$, $r = 0.877$) and deadlift ($p < 0.001$, $r = 0.984$) in male university athletes. Thus, the LBD provides a lower-body strength test that is simple to administer, while also being applicable to highly-trained or lesser-trained individuals.

However, some individuals may have a pulling strength that exceeds that of the typical leg/back dynamometer, which can be up to 299.37 kg (660 lbs) (Fabrication Enterprises Inc., 2023). A commercial crane scale could provide an affordable option while also having the value of a greater capacity than the typical LBD (up to 907 kg, or 2000 lbs) (Global Industrial™, 2023). However, there is no research that has investigated whether an adapted leg/back dynamometer (ALBD) that uses a commercial crane scale provides a valid and reliable measure of isometric leg/back strength.

Therefore, the purpose of this study was to determine the validity and reliability of an ALBD that used a commercial crane scale with a greater capacity than a typical LBD. The ALBD was compared to and correlated with the LBD to determine validity. A test-retest analysis was used to determine reliability of the ALBD metric. It was hypothesized that the ALBD would provide a valid and reliable measure of isometric leg/back maximal strength in physically-active individuals.

Methods

Experimental Approach

This study was a prospective, validation and test-retest reliability study, with procedures adapted from previous studies (Lockie et al., 2013; Najiah et al., 2021; Ten Hoor et al., 2016). Subjects completed three testing sessions. The first session was originally intended as a familiarization session (Courel-Ibáñez et al., 2020); sessions two and three were to be the test-retest sessions for the reliability analysis. The ALBD was compared to and correlated with the LBD to determine validity. Reliability was assessed by intra-class correlation coefficients, coefficient of variation (CV), and a three-way repeated measure analysis of variance (ANOVA). Both the LBD and ALBD were measured in kilograms (kg).

Participants

Fifty physically active people (age: 23.02 ± 3,43 years; height: 1.68 ± 0.11 m; body mass: 78.46 ± 14.63 kg), including 34 men (age: 22.76 ± 2.79 years; height: 1.72 ± 0.10 m; body mass: 82.03 ± 13.07 kg) and 16 women (age: 23.56 ± 4.56 years; height: 1.60 ± 0.05 m; body mass: 70.88 ± 15.27 kg) were recruited for this study. Participants self-reported whether they completed the minimum recommended physical activity for cardiorespiratory and musculoskeletal fitness as detailed by the American College of Sports Medicine

(Garber et al., 2011), and were free from injuries that could influence study participation. G*Power software was also used to confirm post hoc that the sample size of 50 was sufficient for a correlation, point biserial model, and ensured the data could be interpreted with a moderate effect level of 0.40 (Hopkins, 2004b), and a power level of 0.86 when significance was set at 0.05 (Faul et al., 2007). The institutional review board approved the study (HSR-22-23-334), all Participants received a clear explanation of the procedures. This included the risks and benefits of participation, and written informed consent was obtained. The study followed the recommendations of the Declaration of Helsinki (World Medical Association, 1997).

Measurements and Procedures

Participants completed three testing sessions. The first session was originally intended as a familiarization session (Courel-Ibáñez et al., 2020); sessions two and three were to be the test-retest sessions for the reliability analysis. Data was collected in three approximate 30-40-minute testing sessions which were separated by 48-72 hours depending on participant availability. Participants were informed to wear athletic clothing and shoes that they would typically use for training (i.e., sneakers). The same shoes and similar clothes were to be worn at all sessions. No supportive garments (e.g., wrist wraps, weight belts) were worn during any of the tests. All testing was conducted in the university's Human Performance Lab, which had rubberized flooring. Prior to testing in session one, participants signed an informed consent form and completed a physical activity readiness questionnaire. Height was measured using a stadiometer (Detecto, Webb City, MO, USA), and body mass was measured by an electronic digital scale (Model HBF-510, Omron Healthcare, Kyoto, Japan). After this, participants completed a dynamic warm-up that lasted approximately 8-10 minutes. All participants completed the same dynamic warm-up prior to all testing sessions, which comprised of skips with arm swing, side jacks with arm swing, lunge to rotation with hamstring stretch, pigeon stretch, A-Skips, cariocas, and five bilateral base drops with arm swing.

The procedures for this study were adapted from previous research (Lockie et al., 2013; Najiah et al., 2021; Ten Hoor et al., 2016). Participants alternated whether they completed the LBD or the ALBD first, such that it was evenly divided amongst the sample. The order of the LBD and ALBD was kept consistent across all sessions. As noted, the first session was originally intended to be a familiarization session (Courel-Ibáñez et al., 2020), where participants completed both the LBD and ALBD with the same procedures as that for the second and third sessions. For both the LBD and ALBD, participants completed two warm-up or practice pulls using the specific procedures for each device. This helped prepare the participants for the required maximal effort pulls, and also allowed the researchers to correct any flaws in technique (Nuzzo et al., 2011). Following this, participants completed three trials with either the LBD or ALBD, with rest times of 2 minutes allocated between attempts. The tests will be presented here as if the standard LBD was performed first. Nevertheless, the procedures used to set-up the participant for the strength test with either device was the same.

Standard Leg/Back Dynamometer

The standard LBD had an oversize base with a capacity of 299.37 kg (Fabrication Enterprises, Inc., New York, USA). The methods were adapted from previous research (Dawes et al., 2019; Dawes et al., 2017; Lockie et al., 2020a; Lockie et al., 2020b; Lockie et al., 2023a; Lockie et al., 2023b; Wakely et al., 2022). Participants were positioned so their arms were extended and both hands were on the handle positioned at the mid-thigh, with a knee flexion angle of approximately 110° (Figure 1). The knee angle was measured with a goniometer so that participants would be positioned in the same way for the second strength test. A countdown of "3, 2, 1, pull" was given to the participants before they initiated the pull. Similar instructions have been shown to produce optimal results for maximum force development in isometric pulls (Haff et al., 1997). Participants were to maintain proper spinal alignment and their feet flat on the base and pulled the handle upward as hard as possible by attempting to extend the hips and knees. The pull was held for approximately 5 seconds, with data recorded to the nearest kg.

Figure 1. Frontal (A) and sagittal (B) view of the set-up for the leg/back dynamometer test.

Adapted Leg/Back Dynamometer
The ALBD used a commercial crane scale (Global Industrial™, New York, USA) connected to a cable straight bar attachment and custom base. The crane scale had a capacity of 907 kg. Participants were set-up in the same position for the ALBD as for the LBD, with the same knee angle in the pull position (Figure 2). The exact same procedures for the LBD were also used for the ALBD. Three, 5-second pulls were completed for the ALBD, with data also recorded to the nearest kg.

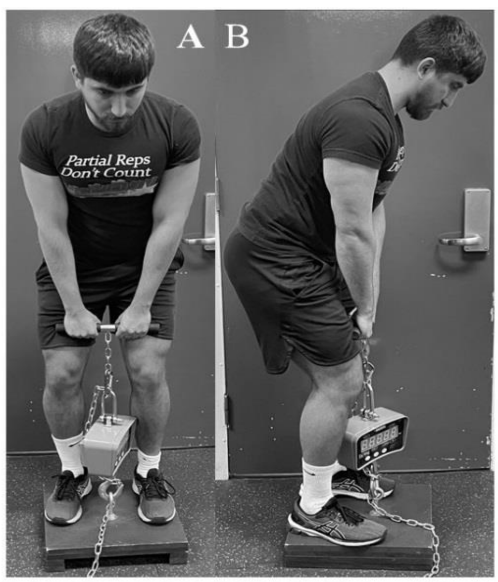

Figure 2. Frontal (A) and sagittal (B) view of the set-up for the adapted leg/back dynamometer test.

Statistical analyses
Statistical analyses were computed using the Statistics Package for Social Sciences (Version 29.0; IBM Corporation, New York, USA) and Microsoft Excel (Microsoft Corporation, Redmond, Washington, USA). Descriptive statistics (mean ± standard deviation [SD]; 95% confidence limits [CL]) were calculated for the LBD and ALBD data in each session. Males and females were combined in the sample. Normality of the data

was evaluated by the Kolmogorov-Smirnov test and visual evaluation of Q-Q plots. Analysis was conducted on the average from the three trials for the LBD and LABD within each session, as well as the best trial (i.e., the trial with the maximum strength value). To assess agreement, the difference and average between the LBD and ALBD mean for each session were calculated with the intention of creating Bland-Altman plots (Doğan, 2018). One-sample t-tests were used to ascertain whether the difference between the LBD and ALBD means was statistically significant ($p < 0.05$) compared to 0 (i.e., no difference between the means). Further, paired samples t-tests ($p < 0.05$) were used to compare the average and maximum values for the LBD and ALBD to provide a measure of concurrent validity (Aung et al., 2020; Lockie et al., 2013). Effect sizes (*d*) were also calculated for the between-strength test comparisons, where the difference between the means was divided by the pooled SD (Cohen, 1988). A *d* less than 0.2 was considered a trivial effect; 0.2 to 0.6 a small effect; 0.6 to 1.2 a moderate effect; 1.2 to 2.0 a large effect; 2.0 to 4.0 a very large effect; and 4.0 and above an extremely large effect (Hopkins, 2004b). To further analyze validity, Pearson's correlations ($p < 0.05$) and regression were used to derive relationships between the LBD and ALBD. The correlation strength was designated as: an *r* between 0 to ±0.3 was small; ±0.31 to ±0.49, moderate; ±0.5 to ±0.69, large; ±0.7 to ±0.89, very large; and ±0.9 to ±1 near perfect for relationship prediction (Hopkins, 2006). Regression have been recommended for use in validity analyses (Hopkins, 2004a), and thus were included in this research.

Intra-class correlation coefficients (ICC) and CV were used to analyze reliability across the trials (single and average) within each session. The CV for the LBD and ALBD in each session was calculated as the standard deviation of the dataset divided by the mean, before being converted into a percentage (%). An ICC equal to or above 0.70 and a CV of less than 5% was acceptable (Baumgartner & Chung, 2001; Lockie et al., 2013). The differences between the sessions for the LBD and ALBD across the three testing sessions was assessed by a three-way repeated measures ANOVA ($p < 0.05$). If a significant between-session interaction was found, a Bonferroni post hoc adjustment for pairwise comparisons was implemented.

The usefulness of the LBD and ALBD was determined by comparing the typical error (TE) to the smallest worthwhile change (SWC) in kg for each test (Hopkins, 2004b). Within each session, TE was calculated for each variable via the formula $TE = SD \div (\sqrt{N})$, where *N* was the sample size of 50. The SWC was determined by multiplying the between-participant SD by 0.2, which is the typical small effect (Hopkins, 2004b), If the TE for either the LBD or ALBD was below the SWC, the test was 'good'; if the TE was similar to the SWC, the test was 'OK'; and if the TE was higher than the SWC, the test was 'marginal' (Hopkins, 2004b).

Results

The Kolmogorov-Smirnov data indicated all variables were normally distributed ($p = 0.053$-0.200), and visual analysis of the Q-Q plots confirmed this analysis. Table 1 displays the within-session comparisons between the average and maximum values for the LBD and ALBD. With regards to the one-sample t-test, all LBD and ALBD mean differences were significant ($p \leq 0.04$), which indicated a fixed bias between the LBD and ALBD. Thus, Bland-Altman plots were not created. The paired-samples t-tests confirmed these results, as the LBD recorded significantly higher average and maximum values when compared to the ALBD. In Sessions 1 and 2, all the differences between the LBD and ALBD had small effects ($d = 0.43$-0.58). In Session 3, the differences between the average and maximum LBD and ALBD values had moderate effects ($d = 0.89$ and 0.79, respectively). Nonetheless, the relationships between the LBD and ALBD, regardless of session, for both the average (Table 2) and maximum (Table 3) values were all significant ($p < 0.001$) and had an *r* above 0.9, indicating near perfect relationships. To this end, data from all three testing sessions was used to produce regression equations for both the average and maximum values. For both the average (Figure 3) and maximum (Figure 4) values, there was 94% explained variance.

Table 1. Comparisons between the leg/back dynamometer (LBD) and adapted leg/back dynamometer (ALBD) within the three testing sessions. Data reported as mean ± SD (95% CI). Intra-class correlation coefficients (ICC) for average and single measures and coefficient of variation (CV) for the session trials are also displayed.

	LBD	**ALBD**	***p***	***d***
Session 1				
Average (kg)	128.59 ± 38.33 (117.70-139.48)	122.87 ± 38.47* (111.94-133.80)	<0.01	0.58
Maximum (kg)	133.85 ± 38.72 (122.84-144.85)	129.28 ± 38.40* (118.20-140.36)	<0.01	0.43
$ICC_{Average}$	0.99	0.99		
ICC_{Single}	0.97	0.96		
CV (%)	4.64	5.94		
Session 2				
Average (kg)	129.71 ± 38.64 (118.73-140.69)	123.32 ± 36.29* (113.01-133.63)	<0.01	0.57
Maximum (kg)	134.36 ± 39.96 (123.00-145.72)	128.94 ± 37.28* (118.35-139.53)	<0.01	0.50
$ICC_{Average}$	0.99	0.99		
ICC_{Single}	0.98	0.97		
CV (%)	3.87	5.02		
Session 3				
Average (kg)	135.01 ± 39.29 (123.85-146.18)	128.83 ± 38.06* (118.02-139.65)	<0.01	0.89
Maximum (kg)	137.66 ± 39.72 (125.37-148.95)	132.22 ± 38.78* (121.20-143.24)	<0.01	0.79
$ICC_{Average}$	1.00	1.00		
ICC_{Single}	1.00	0.99		
CV (%)	2.11	2.82		

Table 2. Correlations between the average leg/back dynamometer (LBD) and adapted leg/back dynamometer (ALBD) trials recorded in Sessions 1-3 in college-aged men and women. All relationships were significant at p < 0.001.

	LBD Session 1	**LBD Session 2**	**LBD Session 3**
ALBD Session 1	0.97		
ALBD Session 2	0.90	0.96	
ALBD Session 3	0.95	0.96	0.98

Table 3. Correlations between the maximum leg/back dynamometer (LBD) and adapted leg/back dynamometer (ALBD) trials recorded in sessions 1-3 in college-aged men and women. All relationships were significant at p < 0.001.

	LBD Session 1	**LBD Session 2**	**LBD Session 3**
ALBD Session 1	0.96		
ALBD Session 2	0.91	0.96	
ALBD Session 3	0.96	0.97	0.99

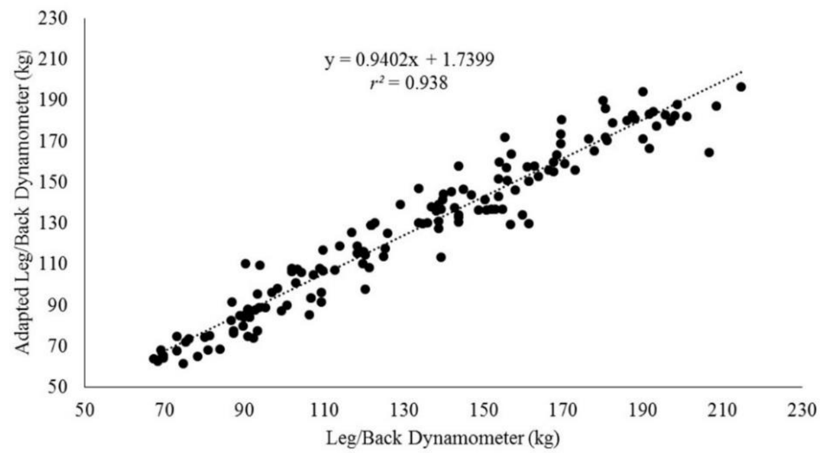

Figure 3. Regression for the average values recorded from 50 participants across three sessions from the leg/back dynamometer and adapted leg/back dynamometer.

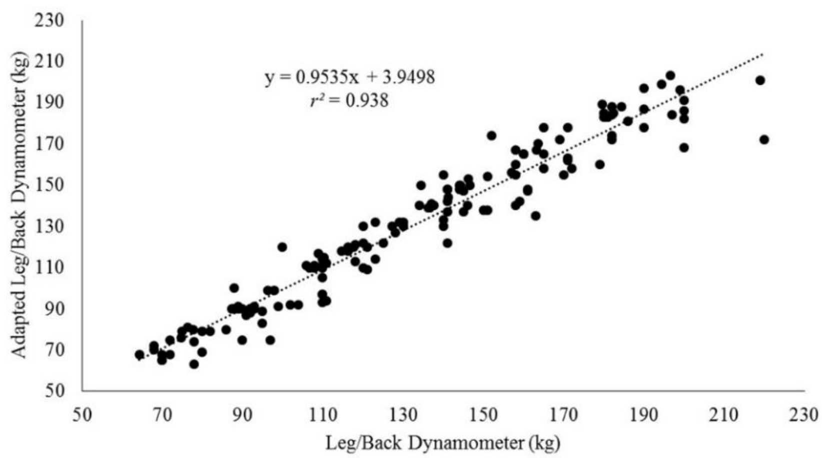

Figure 4. Regression for the maximum values recorded from 50 participants across three sessions from the leg/back dynamometer and adapted leg/back dynamometer.

In all testing sessions for both the LBD and ALBD, the ICCs for single or average trials were high (ICC ≥ 0.97; Table 1). For the CV, the LBD was below 5% in all testing sessions. For the ALBD, CV was approximately 6% in session 1, 5% in Session 2, and 3% in Session 3. There were notable results when comparing the average (Figure 5) and maximum (Figure 6) LBD and ALBD mean data. For the LBD, the main effects for session for the average ($F_{(2,48)}$ = 28.727, p < 0.001, η_p^2 = 0.545) and maximum ($F_{(2,48)}$ = 11.457, p < 0.001, η_p^2 = 0.323) LBD were significant. The average LBD for Session 3 was 4-5% significantly greater than that recorded in Sessions 1 and 2 (p < 0.001). This was also the case for the maximum LBD from Session 3; the value from these sessions were 2-3% significantly greater than Sessions 1 (p < 0.003) and 2 (p = 0.008). Similar results were observed for the ALBD, where the main effect for session was significant for average ($F_{(2,48)}$ = 22.365, p < 0.001, η_p^2 = 0.482) and maximum ($F_{(2,48)}$ = 7.548, p = 0.001, η_p^2 = 0.239) values. The Session 3 average ALBD value was 4-5% significantly greater than that recorded in Sessions 1 (p < 0.001) and 2 (p = 0.002). The Session 3 maximum ALBD value was 2% significantly greater than Session 1 (p = 0.042), but not Session 2 (p = 0.089). There were no significant differences between any LBD or ALBD values recorded in Sessions 1 or 2. The test usefulness data for all sessions is displayed in Table 4. Both the LBD and ALBD were deemed to be good tests regardless of session. All SWC values exceeded the TE.

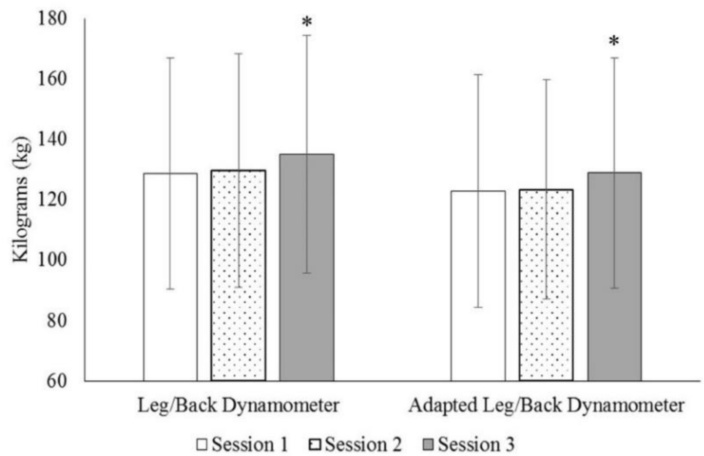

Figure 5. Between-session comparisons for average values recorded across three sessions from the leg/back dynamometer and adapted leg/back dynamometer (N = 50). *Significantly different from Sessions 1 and 2.

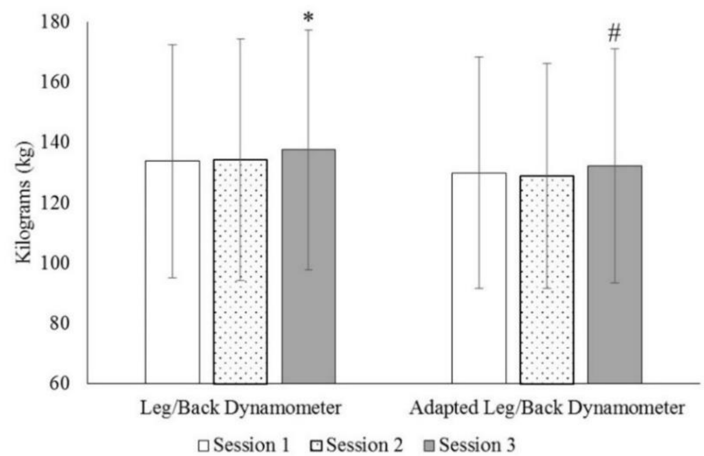

Figure 6. Between-session comparisons for maximum values recorded across three sessions from the leg/back dynamometer and adapted leg/back dynamometer (N = 50). *Significantly different from Sessions 1 and 2; #Significantly different from Session 1.

Table 4. Usefulness of the average and maximum leg/back dynamometer (LBD) and adapted leg/back dynamometer (ALBD) values from the three testing sessions when considering typical error (TE) and smallest worthwhile change (SWC).

	LBD		ALBD	
	Average	**Maximum**	**Average**	**Maximum**
Session 1				
TE (kg)	5.42	5.48	5.44	5.52
SWC (kg)	7.67	7.74	7.69	7.80
Test Rating	Good	Good	Good	Good
Session 2				
TE (kg)	5.46	5.65	5.13	5.27
SWC (kg)	7.73	7.99	7.26	7.46
Test Rating	Good	Good	Good	Good
Session 3				
TE (kg)	5.56	5.62	5.38	5.48
SWC (kg)	7.86	7.94	7.71	7.76
Test Rating	Good	Good	Good	Good

Discussion

This study determined the validity and reliability of an ALBD that used a commercial crane scale with a greater capacity than a typical LBD. Najiah et al. (2021) has previously acknowledged that the LBD provided a valid measure of lower-body maximal strength, with near perfect correlations shown with the 1RM back squat (r = 0.877) and deadlift (r = 0.984). Additionally, the LBD has been used as a strength testing tool within numerous populations (Dawes et al., 2019; Dawes et al., 2017; Lockie et al., 2020a; Lockie et al., 2020b; Lockie et al., 2023a; Lockie et al., 2023b; Najiah et al., 2021; Wakely et al., 2022). These studies helped support the use of the LBD as the standard for comparisons with the ALBD. The results indicated that there was a fixed bias which affected agreement between the LBD and ALBD, in that the values recorded from the ALBD were significantly lower than that for the LBD. The was shown by both the one-sample and paired samples t-tests. It is not surprising that ALBD tended to record lower values than the LBD, given the capacity for each device. The standard LBD had a limit of 299.37 kg, while the crane scale in the ALBD had a limit of 907 kg. Even a small crane scale is generally designed to tolerate much heavier loads than that could be exerted by a person (Eilon Engineering, 2022; Global Industrial™, 2023). The crane scale design could have affected the data recorded from the participants in this study, which resulted in smaller average and maximum strength metrics when compared to the LBD. Nevertheless, the results suggested that assuming the LBD was a valid measure of maximal strength (Najiah et al., 2021), the ALBD provided data different to the LBD.

The correlation data, however, provided some support to how the ALBD was measuring similar qualities to the LBD. There were near perfect, positive relationships between the LBD and ALBD. Although there are limitations to using correlations to assess validity (Doğan, 2018), the data did suggest that those participants who performed well in the LBD also performed well in the ALBD. Regression can also be valuable in validity analyses (Hopkins, 2004a). The regression equation developed between the LBD and ALBD had 94% explained variance (i.e., r^2 = 0.938) for both average and maximum values. A high r^2 value suggests better predictive ability for the regression equation (Chicco et al., 2021). These data suggest that based on the sample from the current study, LBD performance could be predicted from the ALBD, and vice-versa. This was true for both the average ($y = 0.9402x + 1.7399$) and maximum ($y = 0.9535x + 3.9498$) values recorded in this study. It should be noted that the fixed bias between the tests (i.e., the LBD recorded higher values than the ALBD) suggested that each test be considered separate to the other. However, given the relationships between the LBD and ALBD, predictive equations could be used by the practitioner as needed, with the acknowledgement of potential agreement or bias issues.

With regards to the reliability data, single and average ICC values were very high (above 0.90) and above the accepted standard in this study of 0.70 (Baumgartner & Chung, 2001; Lockie et al., 2013) in all sessions, which indicated good trial-to-trial reliability. The was also the case for the LBD when considering CV, where the CV was below 5% in all sessions. The CV for the ALBD was highest in session 1 at 5.95% and slightly above the accepted range for this study. However, the CV for the ALBD decreased in sessions 2 and 3 into the acceptable range of 5% or below. Thus, good inter-trial reliability can occur with either the LBD or ALBD.

The results did indicate some variation in average and maximum values across the sessions. The study featured three testing sessions, and originally the first session was intended to be a familiarization session (Courel-Ibáñez et al., 2020). These procedures followed that for RM strength testing. Following a systematic review of the literature, Grgic et al. (2020) stated that reliability in RM strength testing was similar between trained and untrained individuals, and familiarization may not have as big an impact as previously thought. Accordingly, it was decided that the first session would serve as a familiarization session, with the exact same procedures performed in this session as for the second and third testing sessions. Upon analysis of the data, it was found that the average and maximum data for both the LBD and ALBD were not significantly different between Sessions 1 and 2. However, the average and maximum LBD values from Session 3 were significantly greater than that recorded in Sessions 1 and 2. Similarly, the average ALBD value was significantly greater than Sessions 1 and 2, while the maximum ALBD was greater than Session 1. The actual differences were relatively small – between 2-5% for any of the significant comparisons. Nonetheless, these data may indicate that the participants may have become more familiar with both leg/back dynamometer tests by Session 3 and could produce higher values (Ploutz-Snyder & Giamis, 2001). In physically active men, Dias et al. (2005) suggested 2-3 practice sessions for bench press, back squat, and arm curl 1RM strength testing were needed to produce reliable results. Specific to an isotonic knee extension dynamometer strength test, Ploutz-Snyder & Giamis (2001) recommend 3-4 practice sessions for young women (~23 years of age), and 8-9 practice sessions for older women (~66 years of age). Nonetheless, Grgic et al. (2020) noted that excessive familiarization sessions are not practical, and in research, could lead to excessive dropout rates. Indeed, the study by Ploutz-Snyder & Giamis (2001) only featured 13 participants, far below that for the current study (*N* = 50). It is not known whether the values recorded from the LBD and ALBD would change over subsequent testing sessions. Grgic et al. (2020) did state that familiarization may be required for an exercise if the individual is not experienced with the movement used in a strength test. It is plausible to suggest that implementing two practice sessions for the leg/back dynamometer could allow for greater reliability in strength testing.

It should be noted it may not always be practical to conduct numerous practice testing or familiarization sessions. For example, and as noted, the LBD has been used to assess strength in first responders (Dawes et al., 2019; Dawes et al., 2017; Lockie et al., 2020a; Lockie et al., 2020b). Police officers and firefighters are notably time-poor (i.e., there is often limited time to perform extraneous fitness-related activities due to work demands) (Lockie et al., 2021), so there may be limited opportunities for practitioners to conduct fitness testing. Moreover, due to the risk of injury in these populations (Lockie et al., 2022), staff are often reticent to allow for multiple maximal strength testing sessions, so these may not always be possible. In these situations, what was particularly notable from the study results is that all average and maximum LBD and ALBD values were classified as being 'good' relative to their usefulness (Hopkins, 2004b), as the SWC exceeded the TE from each session. What this means is that the smallest notable performance change for either the LBD and ALBD exceeded the error associated with the test. Accordingly, if a practitioner sees in increase of 7-8 kg in the LBD or ALBD (the approximate SWC across all sessions), they can be confident that it is a real change in performance. Practitioners could use the LBD (or ALBD) with certain populations without specific familiarization, while recognizing the potential for improvements due to individuals becoming better at performing the test. Practitioners should carefully monitor their test results, so as to recognize whether any improvements are due to learning or training effects. Nevertheless, even without familiarization, both the LBD and ALBD are useful tests.

There are study limitations that should be acknowledged. Only one type of commercial crane scale was used for the ALBD, and it is certainly possible that different models would produce different results. This would affect the validity and reliability when compared to a standard LBD. The study utilized college-aged men and women. All participants were physical active, but inclusion criteria did not explicitly state that participants had to be resistance training (although that could be part of the physical activity completed by participants). Strength-trained people could present different results (Dias et al., 2005; Ritti-Dias et al., 2011), especially when considering the reliability measurements over the three sessions. Nonetheless, given that isometric leg/back strength testing is a viable often for populations that may not have long strength training backgrounds, such as high school athletes (Lockie et al., 2023a; Lockie et al., 2023b; Wakely et al., 2022) or first responders (Dawes et al., 2019; Dawes et al., 2017; Lockie et al., 2020a; Lockie et al., 2020b), the study sample is an applicable population. The sexes were combined in the sample for this study. As there can be performance differences in maximal strength tests when comparing men and women (Leyk et al., 2007; Lockie et al., 2020b; Reynolds et al., 2006), it is possible the relationships and predictive equations between the LBD and ALBD could vary by sex. Only three testing sessions were used, and performance in the LBD and ALBD seemed to improve. Despite potential impracticalities when using multiple familiarization or practice sessions (Grgic et al., 2020), It is not known how many testing sessions may be required before performance plateaus using either isometric strength testing device.

Conclusion

The study results indicated that when compared to the LBD, the ALBD recorded lower average and maximum strength values, which indicated a fixed bias relative to agreement. However, the LBD and ALBD were highly correlated. A regression provided prediction equations that explained 94% of the variance between the tests, so even with the fixed bias, a practitioner could predict a LBD value from the ALBD if it was required. The LBD and ALBD data were reliable, especially when considering the ICC and CV values within each testing session. However, the data also suggested that individuals new to the isometric strength testing could exhibit higher strength values when they become more familiar with the test. At least two familiarization or practice sessions could be used to prepare participants for an isometric strength test, whether using the LBD or ALBD. Nonetheless, both the LBD and ALBD were deemed useful tests across all sessions (i.e., SWC greater than the TE), so they both provide viable methods to measure isometric strength.

Acknowledgements: The authors acknowledge the participants for their contribution to the study. This study received no external financial assistance. None of the authors have any conflict of interest.

Conflict of interest: All authors declare that they have no conflict of interest relevant to the content of this article.

References

Aung, N., Bovonsunthonchai, S., Hiengkaew, V., Tretriluxana, J., Rojasavastera, R., & Pheung-Phrarattanatrai, A. (2020). Concurrent validity and intratester reliability of the video-based system for measuring gait poststroke. *Physiotherapy Research International, 25*(1), e1803.

Baumgartner, T.A., & Chung, H. (2001). Confidence limits for intraclass reliability coefficients. *Measurement in Physical Education and Exercise Science, 5*(3), 179-188.

Chicco, D., Warrens, M.J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science, 7*, e623.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, New Jersey: Lawrence Earlbaum Associates.

Courel-Ibáñez, J., Hernández-Belmonte, A., Cava-Martínez, A., & Pallarés, J.G. (2020). Familiarization and reliability of the isometric knee extension test for rapid force production assessment. *Applied Sciences, 10*(13), 4499.

Dawes, J.J., Lockie, R.G., Kornhauser, C.L., Holmes, R.J., & Orr, R.M. (2019). Relationships between absolute and relative strength and power in male police officers of varying strength levels. *Journal of Science in Sport and Exercise, 1*, 281-288.

Dawes, J.J., Orr, R.M., Flores, R.R., Lockie, R.G., Kornhauser, C., & Holmes, R. (2017). A physical fitness profile of state highway patrol officers by gender and age. *Annals of Occupational and Environmental Medicine, 29*(16), 16.

Dias, R.M.R., Cyrino, E.S., Salvador, E.P., Caldeira, L.F.S., Nakamura, F.Y., Papst, R.R., Bruna, N., & Gurjão, A.L.D. (2005). Influence of familiarization process on muscular strength assessment in 1-RM tests. *Revista Brasileira de Medicina do Esporte, 11*(1), 39-42.

Doğan, N.Ö. (2018). Bland-Altman analysis: A paradigm to understand correlation and agreement. *Turkish Journal of Emergency Medicine, 18*(4), 139-141.

Eilon Engineering. (2022). *Crane Scales vs Dynamometers – What's the Difference?* Retrieved February 21, 2024, from https://www.eilon-engineering.com/blog/crane-scales-vs-dynamometers/.

Fabrication Enterprises Inc. (2023). *Baseline Back-Leg-Chest Dynamometers*. Retrieved March 15, 2023, from https://www.fab-ent.com/evaluation/strength/baseline-back-leg-chest-dynamometers/.

Faul, F., Erdfelder, E., Lang, A.G., & Buchner, A. (2007). G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175-191.

Garber, C.E., Blissmer, B., Deschenes, M.R., Franklin, B.A., Lamonte, M.J., Lee, I.M., Nieman, D.C., & Swain, D.P. (2011). American College of Sports Medicine position stand. Quantity and quality of exercise for developing and maintaining cardiorespiratory, musculoskeletal, and neuromotor fitness in apparently healthy adults: Guidance for prescribing exercise. *Medicine and Science in Sports and Exercise, 43*(7), 1334-1359.

Global Industrial™. (2023). *Global Industrial™ Digital Crane Scale with Remote 2000 lbs x 1 lbs*. Retrieved March 15, 2023, from https://www.globalindustrial.com/p/electronic-crane-scale-with-remote-2000-lb-capacity.

Grgic, J., Lazinica, B., Schoenfeld, B.J., & Pedisic, Z. (2020). Test–retest reliability of the one-repetition maximum (1RM) strength assessment: a systematic review. *Sports Medicine - Open, 6*(1), 31.

Haff, G.G., Stone, M., O'Bryant, H.S., Harman, E., Dinan, C., Johnson, R., & Han, K.-H. (1997). Force-time dependent characteristics of dynamic and isometric muscle actions. *Journal of Strength and Conditioning Research, 11*(4), 269-272.

Hopkins, W.G. (2004a). Bias in Bland-Altman but not regression validity analyses. *Sportscience, 8*, 42-46.

Hopkins, W.G. (2004b). How to interpret changes in an athletic performance test. *Sportscience, 8*, 1-7.

Hopkins, W.G. (2006). *A Scale of Magnitudes for Effect Statistics*. Retrieved January 18, 2022, from http://www.sportsci.org/resource/stats/effectmag.html.

Leyk, D., Gorges, W., Ridder, D., Wunderlich, M., Rüther, T., Sievert, A., & Essfeld, D. (2007). Hand-grip strength of young men, women and highly trained female athletes. *European Journal of Applied Physiology, 99*(4), 415-421.

Lockie, R., Dulla, J., Orr, R., & Dawes, J. (2021). Importance of ability-based training for law enforcement recruits. *Strength and Conditioning Journal, 43*(3), 80-90.

Lockie, R.G., Dulla, J.M., Higuera, D., Ross, K.A., Orr, R.M., Dawes, J.J., & Ruvalcaba, T.J. (2022). Body composition and fitness characteristics of firefighters participating in a health and wellness program: Relationships and descriptive data. *International Journal of Environmental Research and Public Health, 19*(23), 15758.

Lockie, R.G., MacLean, N.D., Dawes, J.J., Pope, R.P., Holmes, R.J., Kornhauser, C.L., & Orr, R.M. (2020a). The impact of formal strength and conditioning on the fitness of police recruits: A retrospective cohort study. *International Journal of Exercise Science, 13*(4), 1615-1629.

Lockie, R.G., Moreno, M.R., McGuire, M.B., Ruvalcaba, T.R., Bloodgood, A.M., Dulla, J.M., Orr, R.M., & Dawes, J.J. (2020b). Relationships between isometric strength and the 74.84-kg (165-lb) body drag test in law enforcement recruits *Journal of Human Kinetics, 74*, 5-13.

Lockie, R.G., Schultz, A.B., Callaghan, S.J., Jeffriess, M.D., & Berry, S.P. (2013). Reliability and validity of a new test of change-of-direction speed for field-based sports: the Change-of-Direction and Acceleration Test (CODAT). *Journal of Sports Science and Medicine, 12*(1), 88-96.

Lockie, R.G., Wakely, A.M., Dawes, J.J., Hernandez, E., & Patron, J.D. (2023a). The relationships between muscular strength and power with throwing velocity in high school boys water polo players following a strength training block. *Facta Universitatis – Series: Physical Education and Sport, 21*(2), 99-110.

Lockie, R.G., Wakely, A.M., Viramontes, E., & Dawes, J.J. (2023b). A research note on relationships between age, body size, strength, and power with throwing velocity in high school water polo players. *Journal of Strength and Conditioning Research, 37*(8), e466-e469.

McGuigan, M.R. (2015). Administration, Scoring, and Interpretation of Selected Tests. In G. G. Haff & N. T. Triplett (Eds.), *Essentials of Strength Training and Conditioning* (4th ed., pp. 259-316). Champaign, IL: Human Kinetics.

Najiah, W.N., Lee, E.L.Y., Malek, N.F.A., Malik, Z.A., Chan, E.W.M., Ghazali, N., & Nadzalan, A.M. (2021). Relationship between leg dynamometer with squat and deadlift 1RM score among university athletes. *Journal of Physics: Conference Series, 1874*(1), 012089.

Nuzzo, J.L., Anning, J.H., & Scharfenberg, J.M. (2011). The reliability of three devices used for measuring vertical jump height. *Journal of Strength and Conditioning Research, 25*(9), 2580-2590.

Orr, R.M., Lockie, R., Milligan, G., Lim, C., & Dawes, J. (2022). Use of physical fitness assessments in tactical populations. *Strength and Conditioning Journal, 44*(2), 106-113.

Ploutz-Snyder, L.L., & Giamis, E.L. (2001). Orientation and familiarization to 1RM strength testing in old and young women. *Journal of Strength and Conditioning Research, 15*(4), 519-523.

Reynolds, J.M., Gordon, T.J., & Robergs, R.A. (2006). Prediction of one repetition maximum strength from multiple repetition maximum testing and anthropometry. *Journal of Strength and Conditioning Research, 20*(3), 584-592.

Ritti-Dias, R.M., Avelar, A., Salvador, E.P., & Cyrino, E.S. (2011). Influence of previous experience on resistance training on reliability of one-repetition maximum test. *Journal of Strength and Conditioning Research, 25*(5), 1418-1422.

Sheppard, J.M., Nimphius, S., Haff, G.G., Tran, T.T., Spiteri, T., Brooks, H., Slater, G., & Newton, R.U. (2013). Development of a comprehensive performance-testing protocol for competitive surfers. *International Journal of Sports Physiology & Performance, 8*(5), 490-495.

Suchomel, T.J., Nimphius, S., & Stone, M.H. (2016). The importance of muscular strength in athletic performance. *Sports Medicine, 46*(10), 1419-1449.

Ten Hoor, G.A., Musch, K., Meijer, K., & Plasqui, G. (2016). Test-retest reproducibility and validity of the back-leg-chest strength measurements. *Isokinetics and Exercise Science, 24*(3) 209-216.

Wakely, A.M., Dawes, J.J., Hernandez, E., & Lockie, R.G. (2022). The effects of a 4-week strength and conditioning program on strength, power, and throwing velocity for junior varsity and varsity high school water polo players. *Facta Universitatis – Series: Physical Education and Sport, 20*(3), 199-216.

West, D.J., Owen, N.J., Jones, M.R., Bracken, R.M., Cook, C.J., Cunningham, D.J., Shearer, D.A., Finn, C.V., Newton, R.U., Crewther, B.T., & Kilduff, L.P. (2011). Relationships between force-time characteristics of the isometric midthigh pull and dynamic performance in professional rugby league players. *Journal of Strength and Conditioning Research, 25*(11), 3070-3075.

World Medical Association. (1997). World Medical Association Declaration of Helsinki. Recommendations guiding physicians in biomedical research involving human subjects. *Journal of the American Medical Association, 277*(11), 925-926.