Physical Therapy in Sport 57 (2022) 46–52

Contents lists available at ScienceDirect

Physical Therapy in Sport

journal homepage: www.elsevier.com/ptsp

# Validity, reliability, minimal detectable change, and methodological considerations for HHD and portable fixed frame isometric hip and groin strength testing: A comparison of unilateral and bilateral testing methods

Cian Dunne, Andrew J. Callaway*, Joanna Thurston, Jonathan M. Williams

*Department of Rehabilitation and Sport Sciences, Bournemouth University, BH12 5BB, UK*

A R T I C L E I N F O

*Article history:*
Received 24 March 2022
Received in revised form
4 July 2022
Accepted 5 July 2022

*Keywords:*
Kangatech
Groinbar
Isometric
Handheld dynamometer
Unilateral testing
Bilateral testing

A B S T R A C T

*Objectives:* Comparative assessment of bilateral (KangaTech) and unilateral (HHD) testing modalities through concurrent validity and test-retest reliability. Methodological considerations explored include minimum repetitions and comparison of average and maximum values.
*Design:* Experimental, observational.
*Setting:* Biomechanics laboratory.
*Participants:* Thirty-three participants.
*Main outcome measures:* Concurrent validity using peak force. Test-retest reliability used Abduction and Adduction using 2 trials, randomised between devices. Maximum peak force and average of both trials were used.
*Results:* HHD and KT360 are concurrently valid ($r = 0.996$); with no significant difference ($z = -0.681$). Excellent HHD reliability (ICC:0.92–0.96) and KT360 (ICC:0.89–0.97). Significant difference between max peak force and average peak force but within the calculated MDC(%). No significant differences between max peak force between trials. Spearman-Brown prophecy predicted excellent reliability for one trial (ICC:0.81–0.95). Bilateral facilitation was demonstrated using the KT360 with 94.6–101.2% increase in force compared to HHD.
*Conclusions:* With no significant difference between first and second max effort, and excellent prophesised reliability, one rep max effort should be acceptable to use. Body positioning within the KT360 seems to elicit bilateral facilitation rather than deficit, therefore unilateral and bilateral force values are not interchangeable.

## 1. Introduction

Between 2 and 5% of all sporting injuries occur in the hip and groin (Morelli & Smith, 2001). Within Rugby Union they are the fourth most common injury (Ryan, Kempton, Pacecca, & Coutts, 2014) and account for 17% of all match play injuries in Australian Rules Football (Orchard & Seward, 2002). In elite professional football, hip and groin injuries contribute to 12–16% of all injuries reported within a season (Werner, Hägglund, Waldén, & Ekstrand, 2009).

Challenges in the management of these injuries arise from the wide variety of possible pathologies in many anatomical structures in the region, in addition to commonly found abnormalities presenting in asymptomatic athletes (Weir et al., 2015). Despite this complexity a critical component to rehabilitation is the restoration of strength. Testing and measurement of strength vary in methodology, however isokinetic dynamometry (IKD) is considered the "gold standard" throughout research and in clinical practice (Kambič, Lainščak, & Hadžić, 2020). However, IKD's are neither generally affordable nor easily transportable for field-based assessment. Therefore, hand-held dynamometry (HHD) provides a more accessible and portable method of measurement of isometric strength and therefore is commonly used in the assessment of isometric hip strength in clinical settings. HHD has been shown as a valid method for measuring peak force output compared to IKD

---

* Corresponding author.
*E-mail address:* acallaway@bournemouth.ac.uk (A.J. Callaway).

https://doi.org/10.1016/j.ptsp.2022.07.002

for hip extension (Keep, Luu, Berson, & Berson, 2016), and abduction but not internal or external rotation (Bazett-Jones & Squier, 2020).

Whilst HHD has been shown to be valid in measuring peak force, there are conflicting results in the literature regarding HHD reliability. Fulcher, Hanna, and Raina Elley (2010) reported good to excellent inter-reliability (ICC = 0.66–0.87) and intra-reliability (ICC = 0.70–0.89) when assessing hip strength in athletes. Kemp, Schache, Makdissi, Sims, & Crossely (2013) found HHD to be a reliable method of measuring isometric hip strength in healthy adults. Whilst Thorborg, Bandholm, Schick, Jensen, and Hölmich (2013) showed that HHD are susceptible to inter-tester bias, with tester experience and strength both influencing scores. Thorborg et al. (2013) demonstrated that in their study of 2 testers, the female tester recorded around 10% less force on hip supine abduction and adduction. O'Brien, Bourne, Heerey, Timmins, and Pizzari (2019) tested differences between HHD and a fixed-frame system (GroinBar) for long lever, but did not compare short lever body positions. O'Brien et al. (2019) also suggested that some differences between uni- and bi-lateral testing methods could in part be due to HHD testing being conducted where participants were able to hold onto the side of the plinth.

Fixed-frame dynamometry systems have recently been developed utilising rigid fixation of the force transducers to mitigate inter-tester biases. The KangaTech 360 (KT360) model has been described as a testing platform consisting of a portable, adaptable, fixed frame dynamometry system that allows accurate measurement of isolated neuromuscular strength, endurance, and control. The fixed frame systems use 2 loads cells to simultaneously record the force applied, theoretically this effectively function as 2 unilateral tests, but requires investigation to ensure the system reports this expected output.

With this development in technology, there has been some research into the reliability of fixed-frame systems. Fixed frame testing has been shown to be highly repeatable with excellent ICC for both KT360 (ICC = 0.923–0.958; Ransom et al., 2020) and Groinbar (ICC = 0.940; Ryan, Kempton, Pacecca, & Coutts, 2019). The KT360 and the GroinBar are similar fixed frame devices. However, the KT360 has the ability for the sensor pads to be rotated on two axes giving it the ability to test in all planes of motion, something that initial versions of the Groinbar (Ryan et al., 2019) and the KangaTech (Ransom et al., 2020) are unable to do. Therefore, it is essential that the reliability of this device is established for use in an elite clinical environment against the HHD.

There have been a variety of different calculation methodologies utilised throughout the reliability studies conducted on fixed frame systems, displayed in Table 1 testing average values and/or maximum values. Desmyttere, Gaudet, and Begon (2019) used a test-retest approach and used both maximum peak force and average force of three trials to assess the reliability of groin-bar. Their findings indicated the average of three trials to be more reliable with higher ICCs ($ICC_{3,1}$ = ADD:0.92, ABD: 0.90) than a single measure ($ICC_{3,1}$ = ADD: 0.85, ABD: 0.82), although both excellent. Whilst an average has been shown to be reliable, there has not been an attempt to identify whether there is a difference between single and multiple maximum voluntary isometric contractions (MVIC), where a single test would halve the duration of testing the practitioner and participant.

Therefore, the aims of this current study are to 1) establish the concurrent validity of the HHD compared to the KT360, 2) establish the test-retest reliability of the KT360 in assessing short-lever isometric hip and groin strength compared to HHD, 3) investigate any within-method (uni- and bi-lateral testing) differences including use of average and maximum peak force and their interchangeability and minimum number of repeat trials, and 4) to investigate any between-method differences associated with HHD and fixed frame testing methods.

## 2. Methods

### 2.1. Participants

Reliability study size was calculated using the method described in Walter, Eliasziw, and Donner (1998) with a desired r value range of 0.7–0.9 and two repeated tests giving a sample size of at least 18.4. With institutional ethical approval (xxxxxxx) thirty-three participants volunteered for the study (male = 26, female = 7) aged between 18 and 25.

Inclusion criteria for the research were: recreationally active university students who accepted the invitation to participate. Exclusion Criteria were participants under the age of 18 at the time of the study; participants with any lower limb injury or surgeries in the previous 12 months; participants with any significant medical illnesses or diseases; participants with any heart instability or disorders, screened using a health questionnaire.

### 2.2. Concurrent validity data collection

The KT360 utilises 2 separate load cells to qualify force and should therefore effectively function as 2 unilateral tests completed simultaneously. To remove any error that could be attributed to HHD testing methods and ensure the KT360 operated as separate tests and no calculations were conducted within the software processing the end result, concurrent validity was performed.

The HHD pad was placed directly onto one of the 2 pads on the KT360 device to get a direct reading of peak force applied (Fig. 1). 50 repetitions at random different forces were applied and peak force was recorded across a range of resistance from 15 N to 600 N. This method mitigates any differences in the testing method (uni- or bilateral) and removed human rater interference as it is a direct device to device comparison where any differences in measurement are either differences within the sensors themselves, or software associated with the KT360.

### 2.3. Test-retest reliability data collection

Reliability was tested using a randomised repeated-measures design between HHD and KT following similar study designs in previous reliability studies (Fulcher et al., 2010; Ryan et al., 2019; Ransom et al., 2020).

**Table 1**
Previous works using averages and maximums for testing.

| Authors | Purpose | Warmup | Number of maximal efforts | Measurement taken |
|---|---|---|---|---|
| Keep, Luu, Berson, and Garland (2016) | Concurrent validity | Familiarisation | 5 | Average(s) |
| Ryan et al. (2019) | Reliability and weekly variation | 1 × 80% | 1 | Max |
| Jeon et al. (2020) | Reliability of minimum number of tests | 1 effort | 3 | Max |
| Desmyttere et al. (2019) | Test-Retest Reliability | Familiarisation | 3 | Max and Average |
| Ransom et al. (2020) | Reliability | 1 effort | 1 | Max |

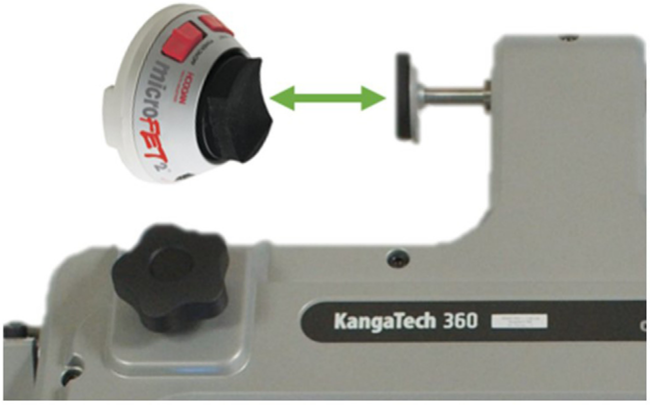

**Fig. 1.** Example of HHD being pressed to a KT360 pad.

Testing staff included two male researchers with physiotherapy/biomechanics backgrounds, and one male physiotherapy intern. All testing staff received training prior to use of equipment.

Data were collected using two systems of isometric strength testing, the KT360 (KangaTech Pty Ltd, Australia) and the MicroFET 2 Wireless HHD (Hogan Scientific LLC, USA). Each testing session consisted of orientation, education and 5-min submaximal self-regulated static bike warm-up (Watt Bike) at a resistance of 3. Participants were randomly assigned to start with either KT testing, or HHD testing. Testing at both stations occurred simultaneously and participants swapped stations once testing had been completed.

Bilateral testing (KT360) used short lever, supine lying with 90-degree knee flexion and 45-degree hip flexion in bare feet and arms across the chest (Fig. 2 a,b). Previous research has shown this position as optimal in eliciting adductor muscle activity and is a position commonly used in clinical practice (Delahunt, Kennelly, McEntee, Coughlan, & Green, 2011; Lovell, Blanch, & Barnes, 2012; O'Brien et al., 2019). O'Brien et al. (2019) tested differences between HHD and a fixed-frame system (GroinBar) for long-lever, but not short-lever. O'Brien et al. (2019) also suggested that some differences in their results between uni- and bi-lateral testing could in part be due to HHD testing being conducted where participants were able to hold onto the side of the plinth. This methodological consideration was used in this method to ensure all participants had arms across the chest.

Unilateral testing (HHD), also used short lever, supine lying with the tested leg 90-degree knee flexion and 45-degree hip flexion in bare feet and arms across the chest, the untested leg was at zero degrees hip and knee flexion, and repeated for each side (Fig. 2 c,d). The force transducers of the KT360 and HHD were positioned at the lateral epicondyle of the femur for abduction and medial epicondyle of the femur for adduction (O'Brien et al., 2019). For HHD, the tester adopted a ‘make’ test positioning (Keep et al., 2016) with bilateral elbow extension and wrist extension for abduction and bilateral elbow extension and wrist flexion for adduction.

Each test was performed three times in total comprising a familiarisation test at 50% effort, followed by two recorded MVICs. Each repetition was held for 5s and separated by 30s rest to minimize the effects of fatigue (Keep et al., 2016).

### 2.4. Statistical analysis

Statistical analyses were conducted using SPSS v.26 (IBM Corp., USA) and MatLab R2021b (The Mathworks Inc., USA). Data were analysed for normality using the Shapiro-Wilk method. Validity data were not normally distributed, reliability data were normally distributed.

Reliability analysis was assessed using Two Way Random effects model using absolute agreement Intraclass Correlation Coefficient (ICC) (Koo & Li, 2016). Mean estimations along with 95% confidence intervals (CI) were reported for each ICC. Interpretation was as follows: <0.50, poor; between 0.50 and 0.75, fair, between 0.75 and 0.90 good; above 0.90, excellent (Koo & Li, 2016).

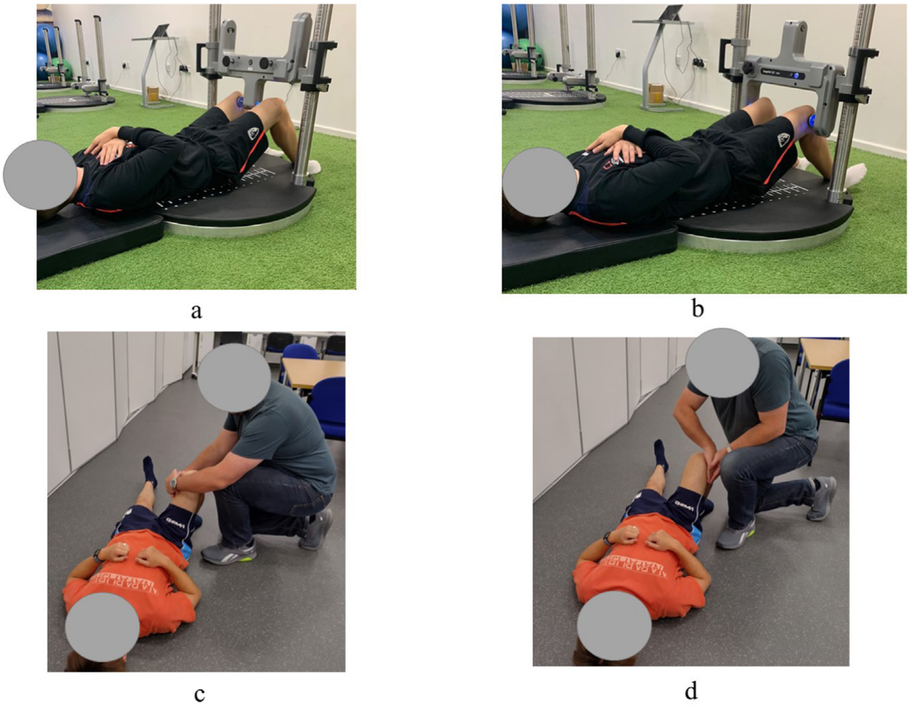

**Fig. 2.** a) Adduction and b) Abduction using KT360. c) Adduction and d) Abduction using HHD.

The standard error of measurement (SEM) was calculated to provide an estimate of error in the unit of interest, using equation (1):

$$SEM = \sqrt{1 - ICC} \tag{1}$$

The minimal detectable change (MDC) was computed to provide a level of confidence in the sensitivity of the measure to detect change using equation (2):

$$MDC_{95\%} = 1.96 \cdot SEM \cdot \sqrt{2} \tag{2}$$

Following the same process as Jeon, Miller, Kang, and Ye (2020), the Spearman–Brown prophecy formula was used to calculate the minimum numbers of trials needed to obtain reliable isometric peak force scores (equation (3)):

$$rkk = \frac{k(r11)}{(1 + (k - 1) \cdot (r11))} \tag{3}$$

where k is the number of times the test is changed in length (e.g., number of trials wanted/number of trials have) and r11 is the reliability from the criterion score in the ICC calculation. The acceptable level of reliability was ≥0.8 (Morrow, Mood, Disch, & Kang, 2015).

To assess the data for methodological considerations, comparisons between force from trial 1 and trial 2; average force (from trial 1 and 2) and maximum peak force (between trial 1 and 2); force from trial 1 and maximum peak force, and trial 2 and maximum peak force; and maximum peak force between equipment (KT or HDD), were conducted using paired-samples t-tests for each muscle group. Percentage change between the HHD and KT were calculated for average and maximum values from the 2 trials using equation (4):

$$\%Difference = \left[ \frac{\left(HHD_{Left} + HHD_{Right}\right) - \left(KT_{Left} + KT_{Right}\right)}{\left(HHD_{Left} + HHD_{Right}\right)} \right] \tag{4}$$

## 3. Results

### 3.1. Concurrent validity

The Spearman Correlation coefficient demonstrated a strong positive correlation between score achieved with the HHD and those achieved with the KT360 ($r = 0.996$, $n = 50$, $p=<0.001$).

A Wilcoxon Signed Rank test demonstrated no significant difference between the KT and HHD ($z = -0.681$, $p = 0.496$). Absolute mean difference 7.74 N (95%CI: 1.86–16.40), RMSE of 9.90 N taken from peak values.

### 3.2. Reliability

The peak values for resisted muscle testing and reliability for both instruments of equipment can be seen in Table 2. ICC values were excellent (ICC: 0.89–0.97) with narrow confidence intervals regardless of equipment, side or muscle group demonstrating excellent reliability (Table 2).

SEM (%) had a very small range (0.59–2.26%) demonstrating minimal error regardless of equipment, side or muscle group. MDC (%) was smallest in the HHD with a small range of 2.17–3.56%, whilst the KT ranged from 1.63 to 6.26%, with higher MDC (%) noted for Adductor muscle groups.

Within each piece of equipment, there were no significant differences in force produced between trial 1 and trial 2 ($p = 0.099–0.808$) (Table 3). Using the Spearman-Brown prophecy formula to reduce the repat number of trials from 2 to 1, reliability for a single measurement was estimated producing ICC values 0.81–0.95 for all muscles and equipment demonstrating a single trial to be reliable (Table 3). Significant differences were found between force from either trial and the maximum force for both trials in all conditions.

Within each piece of equipment, there was a significant difference between the average force of two trials, and the max force from the same two trials ($p < 0.001$), consistently across all sides and muscle groups (Table 4). Also of note, 54.5% of the population achieved their maximum force of the two trials in their first trial.

Between each piece of equipment, there were significant differences shown between equipment, the KT produced significantly more force using the bilateral testing method, than the HHD using a unilateral testing method, in all sides and muscle groups (Table 5). Using the KT as a bilateral testing method produced averaged increases of 94.6–101.2%, with very large ranges of percentage increases from 11.2 to 247.5%.

## 4. Discussion

The aim of the study was to establish the test-retest reliability of the newly developed fixed frame isometric dynamometry system, the KT360, in comparison to the more standard method of isometric strength testing using handheld dynamometry. Furthermore, this study sort to provide clarity over some methodological ambiguity evident in the literature.

Overall, the results demonstrate excellent agreement between the HHD and the KT360, no significant difference between the devices illustrated that the HHD is a valid device across the measurements tested. This agrees with a large body of literature demonstrating excellent validity of HHD across a range of anatomical regions and populations (Chamorro, Armijo-Olivo, De la Fuente, Fuentes, & Chirosa, 2017; Karabay, Yesilyaprak, & Picak, 2020; Katoh, Hiiragi, & Uchida, 2011; Sullivan, Chesley, Hebert, McFaull, & Scullion, 1988). However, previous research has focussed on exploring the entire muscle testing procedure, likely to include human variation, operator variation and device variation. Our approach concentrated specifically on the validity of the device by applying the load through the HHD to the KT360 device. Such an approach overcomes the challenge of variability within setup usually concerning the individual being tested, tester's ability, body positioning or support, or positioning of the sensors as it is common in the literature to test strength with one device and then another, rather than device to device. This may explain the differences in measured correlation between the current study and previous work reporting lower correlations (Keep et al., 2016). Therefore, our approach enables us to conclude that, having removed human variability, the mean absolute differences were small (<8 N) across a testing range up to 617 N, demonstrates that this device has a high degree of concurrent validity.

Overall, the results demonstrate excellent reliability for HHD in support of that previously reported in the literature (Fulcher et al., 2010; Kemp at al., 2013; Thorborg, Petersen, Magnusson, & Hölmich, 2010). However, the ICC value in isolation is of limited use for the clinician trying to determine if a testing method is reliable enough. In this case the SEM or MDC is fundamental for the clinician in interpreting the sensitivity of a test. Any test-retest protocol has many sources of variability which all contribute to the reliability statistic. For example, natural human variation of the person completing the test will be an important source of variance in testing. Understanding this variability is critical for the clinician wishing to have confidence in the change of values over time. To

**Table 2**
Reliability ICC results of HDD and KT, Left and Right.

| Equipment | Side and Muscle group | ICC | 95%CI | Peak (N) | SEM (N) | SEM (%) | MDC (N) | MDC (%) |
|---|---|---|---|---|---|---|---|---|
| HHD | L Abductors | 0.960 | 0.920–0.980 | 189.1 | 1.5 | 0.79 | 4.1 | 2.17 |
| | R Abductors | 0.921 | 0.847–0.960 | 188.1 | 2.4 | 1.28 | 6.7 | 3.56 |
| | L Adductors | 0.937 | 0.878–0.968 | 173.9 | 2 | 1.15 | 5.5 | 3.16 |
| | R Adductors | 0.943 | 0.888–0.971 | 172.6 | 1.9 | 1.1 | 5.2 | 3.01 |
| KT | L Abductors | 0.974 | 0.945–0.986 | 356.9 | 2.1 | 0.59 | 5.8 | 1.63 |
| | R Abductors | 0.944 | 0.891–0.972 | 367.1 | 3.6 | 0.98 | 10.1 | 2.75 |
| | L Adductors | 0.915 | 0.837–0.957 | 351.1 | 6.2 | 1.77 | 17.1 | 4.87 |
| | R Adductors | 0.895 | 0.799–0.946 | 354.5 | 8 | 2.26 | 22.2 | 6.26 |

Notes: HHD: hand-held dynamometry; KT: KT360; L: left, R: right; ICC: Intraclass Correlation Coefficient; CI: 95% Confidence Intervals; N: Newtons; SEM: standard error of measurement; MDC: minimal detectable change.

**Table 3**
Differences within equipment in peak force production between trial 1 and trial 2.

| Equipment | Side and Muscle Group | Trial 1 Force: N(SD) | Trial 2 Force: N(SD) | Difference between trial 1 and trial 2 | Maximum peak force from either trial vs trial 1 (p) | Maximum peak force from either trial vs trial 2 (p) | Spearman-Brown ICC for 1 trial |
|---|---|---|---|---|---|---|---|
| HHD | L Abductors | 185.6 (48.9) | 182.2 (47.2) | 0.161 | .007 | <.000 | 0.92 |
| | R Abductors | 180.9 (46.9) | 183.2 (46.7) | 0.477 | .010 | <.000 | 0.85 |
| | L Adductors | 170.3 (42.0) | 166.3 (37.5) | 0.099 | .002 | <.000 | 0.88 |
| | R Adductors | 166.7 (39.0) | 167.5 (41.9) | 0.767 | .001 | <.000 | 0.89 |
| KT | L Abductors | 350.4 (100.5) | 345.7 (95.7) | 0.808 | .004 | <.000 | 0.95 |
| | R Abductors | 359.4 (98.6) | 353.2 (98.2) | 0.532 | .005 | .003 | 0.89 |
| | L Adductors | 335.2 (96.2) | 336.9 (100.2) | 0.251 | .001 | .002 | 0.84 |
| | R Adductors | 334.5 (98.9) | 339.6 (103.6) | 0.284 | <.000 | .006 | 0.81 |

Notes: HHD: hand-held dynamometry; KT: KT360; L: left, R: right; N: Newtons; SD: standard deviation.

**Table 4**
Differences within equipment for average force and maximum force calculations.

| Equipment | Side and Muscle Group | Average Force: N(SD) | Max Force: N(SD) | p | Percentage of population who achieved maximum force in Trial 1 |
|---|---|---|---|---|---|
| HHD | L Abductors | 183.9 (47.6) | 189.1 (48.1) | <.0001 | 66.6 |
| | R Abductors | 182.0 (45.9) | 188.1 (45.5) | <.0001 | 51.5 |
| | L Adductors | 168.3 (39.2) | 173.9 (41.9) | <.0001 | 57.6 |
| | R Adductors | 167.1 (39.9) | 172.6 (42.0) | <.0001 | 48.5 |
| KT | L Abductors | 348.0 (97.5) | 356.9 (99.9) | <.0001 | 54.5 |
| | R Abductors | 356.3 (97.0) | 367.1 (99.0) | <.0001 | 60.6 |
| | L Adductors | 336.1 (96.0) | 351.1 (96.8) | <.0001 | 51.5 |
| | R Adductors | 337.0 (98.5) | 354.5 (103.0) | <.0001 | 45.5 |
| | | | | | Average = 54.5 |

Notes: HHD: hand-held dynamometry; KT: KT360; L: left, R: right; N: Newtons; SD: standard deviation.

**Table 5**
Differences between devices for average force and maximum force.

| Measure | Side and Muscle Group | HHD | KT | p | Percentage Change (SD); Range |
|---|---|---|---|---|---|
| Average | L Abductors | 183.9 (47.6) | 348.0 (97.5) | <.0001 | 95.6 (46.3); 32.8–247.5 |
| | R Abductors | 182.0 (45.9) | 356.3 (97.0) | <.0001 | |
| | L Adductors | 168.3 (39.2) | 336.1 (96.0) | <.0001 | 101.2 (39.1); 30.9–188.5 |
| | R Adductors | 167.1 (39.9) | 337.0 (98.5) | <.0001 | |
| Max | L Abductors | 189.1 (48.1) | 356.9 (99.9) | <.0001 | 94.6 (43.6); 31.8–230.4 |
| | R Abductors | 188.1 (45.5) | 367.1 (99.0) | <.0001 | |
| | L Adductors | 173.9 (41.9) | 351.1 (96.8) | <.0001 | 96.7 (46.7); 11.2–219.5 |
| | R Adductors | 172.6 (42.0) | 354.5 (103.0) | <.0001 | |

Notes: HHD: hand-held dynamometry; KT: KT360; L: left, R: right; N: Newtons; SD: standard deviation.

this end the MDC value is particularly useful as it provides an estimate, at the 95% confidence level, that a change greater than the MDC value represents true change and not natural variability due to repeat testing.

As such, with 95% confidence, a 6.7 N change (increase or decrease) is indicative of true change in adductor strength as measured with the HHD (Table 2). Previously reported MDC values were higher than the current study (9.4–32.1 N for hip adduction and abduction) perhaps reflecting the different testing protocol between the studies (Thorberg et al., 2010) therefore, as the current study demonstrates small MDC values, clinicians can be confident about the HHD's sensitivity to measure change. The KT360 system, utilising a bilateral testing modality, produced higher force values, which naturally leads to a higher MDC (N). As a percentage however, they are very similar to HHD. HHD presented a range across the different muscle groups and sides of 2.17–3.56% (4.1–6.7 N), and the KT360 of 1.63–6.26% (5.8–22.2 N). Both the KT360 and the HHD system, and these results support a growing body of knowledge with similar findings (Ransom et al., 2020; Fulcher et al., 2010; Kemp et al., 2013).

Maximum force production during bilateral testing is typically inhibited due to the bilateral deficit phenomenon present when both limbs operate maximally, simultaneously (O'Brien et al., 2019). However, there are some inconsistencies in the literature with some studies suggesting a contrasting phenomenon known as “bilateral facilitation” (Howard & Enoka, 1991; Schantz, Moritani, Karlson, Johansson, & Lundh, 1989; Secher, 1975), where the sum of the unilateral forces is less than the bilateral force.

The significant differences found between peak force for HHD and KT360, regardless of muscle group or side, demonstrates a bilateral facilitation of 94.6–101.2% increase in force. It is important to note very large ranges are demonstrated here (e.g. Adductors 11.2–219.5%) which could indicate some individualisation which could be explored further. Part of this result could be due to tester strength where unilateral testing has been shown to be around 10% lower with a female tester compared to a male (Thorborg et al., 2013), however, this was mitigated against in our method design using two male testers who were trained. Furthermore, this work demonstrated lower MDC values than previous work (discussed earlier) which evidences lower variation in the participants and or testers, in combination with excellent ICC results, demonstrating that the testers likely played very little role in the generation of the force data.

O'Brien et al. (2019), testing long lever uni- and bi-lateral testing, attributed differences in their results between unilateral and bilateral testing to be because participants were able to hold onto the plinth during unilateral testing. To mitigate this, we designed the method so that unilateral and bilateral testing used the arms folded across the chest to avoid this potential issue.

To ensure that the KT360 operated as 2 unilateral tests, i.e. each load cell reports a unique force value and not an unknown calculation of the 2 load cells, concurrent validity was conducted by pressing the HHD directly onto one of the KT360 pads. The results help further explain the bilateral facilitation is apparent using bilateral testing in a fixed frame. The results demonstrate high levels of concurrent validity and reveal that the KT360 is recording force as 2 separate load cells. This confirms that the differences shown between KT360 and HHD were due to the testing method (bilateral testing) and not directly to the equipment, helping to demonstrate that bilateral facilitation appears to be present.

Bilateral adductor and abductor testing offers a large contribution of trunk stiffness due to the counterforce from the opposing legs, as opposed to unilateral testing imparting large asymmetrical forces through the trunk with no counter force or stabilising force being able to be applied. Therefore, an abduction force unilaterally with the right leg will result in the body being ‘pushed’ to the left with maximal muscle testing.

Although, the results of this study portray a bilateral facilitation rather than deficit, it is more likely that this novel finding is as a result of force duplication (or overlap) where the sensors are recording the same force twice due to the stabilising nature within a fixed frame system.

Further research is needed to investigate the mechanical explanation behind these findings, but this does question the comparisons of dominant and non-dominant sides until this is determined. It is not clear whether one testing method is superior to the other, but they are not interchangeable as methods, or results.

A further aim of the current work was to make some clinical practice recommendations regarding the metrics used, and number of trials. It is interesting to note that the methods employed for HHD and isometric muscle strength more widely, seem to vary across the literature. Studies have reported single maximum measurements (Ryan et al., 2019; Ransom et al., 2020), the mean of multiple measurements (Desmyttere et al., 2019; Keep et al., 2016) or the maximum force achieved over multiple measurements (Ryan et al., 2019; Desmyttere et al., 2019; Jeon et al., 2019; Ransom et al., 2020).

It was determined that there was no statistically significant difference between trial 1 and trial 2 regardless of the system used. This suggests taking one of these scores as a representation of strength would be possible, and it is commonplace to take the maximum or peak score. However, there was a statistically significant difference between the maximum force and either the first or the second trial (Table 3). This is likely because of the variability in which test (first or second) produced the maximum score, which occurred in the first trial 54.5% of the time (Table 4).

There was also a significant difference between the maximum score and the mean score of the two measurements. This questions the interchangeability of these measurements (average and peak), however the difference between the average and maximum peak force values, whilst significantly different, were within the MDC of each of the devices so likely of little clinical importance (Table 3). This does however challenge the notion of collecting 2 (or more) tests to acquire a mean. If no difference is evident between a single test (maximum value) and the mean of multiple tests, and no difference between trial 1 and trial 2, then taking more than one measurement seems unnecessary. As this type of testing is likely to be part of a suite of isometric tests halving the testing repetitions is likely to result in significant time saving.

It should be noted that most studies reporting measures of reliability have taken multiple measurements in order to facilitate the computation of reliability statistics. The current study is no exception; however, it is the recommendation that a single test will suffice therefore the effect of this on reliability should be estimated. To that end this was achieved in the current study through the use of the Spearman-Brown prophecy formula, demonstrating single measurements are likely to yield high reliability (ICC: 0.81–0.95). Future studies could explore this further through the comparison of scores over two time periods to determine the consistency (as opposed to reliability) of the repeated measurements.

## 5. Conclusion

The KT360 has been demonstrated as a concurrently valid, and reliable device in comparison to handheld dynamometry, they measure the same forces and measure it consistently. Maximum and averages forces show significant differences so are not interchangeable. The results demonstrate that a single maximum effort should be reliable to use.

The KT360, utilising a bilateral testing method, presented results demonstrating bilateral facilitation. The bilateral testing method, using a fixed frame system, offers a reliable, repeatable setup position for rapid testing. Further work is needed to understand the bilateral facilitation suggested here. It seems that due to the body positioning, there is likely to be an overlap (duplication) in force measurement on the opposite sensor. This could indicate that in the same prone position, unilateral and bilateral testing may be testing different muscle groups, or different support structures are present, which requires further investigation.

## Ethical statement

This study was approved by the University Ethical Committee (No. 29269).

## Funding statement

This study received financial support from A.F·C Bournemouth.

## Declaration of competing interest

No conflict of interest. There are no relationships with the equipment providers used in this work.

## References

Bazett-Jones, D. M., & Squier, K. (2020). Measurement properties of hip strength measured by handheld dynamometry: Reliability and validity across the range of motion. *Physical Therapy in Sport, 42*, 100–106. https://doi.org/10.1016/j.ptsp.2020.01.005

Chamorro, C., Armijo-Olivo, S., De la Fuente, C., Fuentes, J., & Chirosa, L. J. (2017). Absolute reliability and concurrent validity of hand held dynamometry and isokinetic dynamometry in the hip, knee and ankle joint: Systematic review and meta-analysis. *Open Medicine, 12*(1), 359–375. https://doi.org/10.1515/med-2017-0052

Delahunt, E., Kennelly, C., McEntee, B. L., Coughlan, G. F., & Green, B. S. (2011). The thigh adductor squeeze test: 45° of hip flexion as the optimal test position for eliciting adductor muscle activity and maximum pressure values. *Manual Therapy, 16*(5), 476–480. https://doi.org/10.1016/j.math.2011.02.014

Desmyttere, G., Gaudet, S., & Begon, M. (2019). Test-retest reliability of a hip strength assessment system in varsity soccer players. *Physical Therapy in Sport, 37*, 138–143. https://doi.org/10.1016/j.ptsp.2019.03.013

Fulcher, M. L., Hanna, C. M., & Raina Elley, C. (2010). Reliability of handheld dynamometry in assessment of hip strength in adult male football players. *Journal of Science and Medicine in Sport, 13*(1), 80–84. https://doi.org/10.1016/j.jsams.2008.11.007

Howard, J., & Enoka, R. (1991). Maximum bilateral contractions are modified by neurally mediated interlimb effects. *Journal of Applied Physiology, 70*(1), 306–316. https://doi.org/10.1152/jappl.1991.70.1.306

Jeon, S., Miller, W. M., Kang, M., & Ye, X. (2020). The minimum number of attempts for a reliable isometric strength test score. *Journal of Science in Sport and Exercise, 2*(1), 89–95. https://doi.org/10.1007/s42978-019-00035-3

Kambič, T., Lainščak, M., & Hadžić, V. (2020). Reproducibility of isokinetic knee testing using the novel isokinetic SMM iMoment dynamometer. *PLoS One, 15*(8), Article e0237842. https://doi.org/10.1371/journal.pone.0237842

Karabay, D., Yesilyaprak, S. S., & Picak, G. S. (2020). Reliability and validity of eccentric strength measurement of the shoulder abductor muscles using a hand-held dynamometer. *Physical Therapy in Sport, 43*, 52–57. https://doi.org/10.1016/j.ptsp.2020.02.002

Katoh, M., Hiiragi, Y., & Uchida, M. (2011). Validity of isometric muscle strength measurements of the lower limbs using a hand-held dynamometer and belt: A comparison with an isokinetic dynamometer. *Journal of Physical Therapy Science, 23*(4), 553–557. https://doi.org/10.1589/jpts.23.553

Keep, H., Luu, L., Berson, A., & Garland, S. J. (2016). Validity of the handheld dynamometer compared with an isokinetic dynamometer in measuring peak hip extension strength. *Physiotherapie Canada, 68*(1), 15–22. https://doi.org/10.3138/ptc.2014-62

Kemp, J. L., Schache, A. G., Makdissi, M., Sims, K. J., & Crossley, K. M. (2013). Greater understanding of normal hip physical function may guide clinicians in providing targeted rehabilitation programmes. *Journal of Science and Medicine in Sport, 16*(4), 292–296. https://doi.org/10.1016/j.jsams.2012.11.887

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting Intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Lovell, G. A., Blanch, P. D., & Barnes, C. J. (2012). EMG of the hip adductor muscles in six clinical examination tests. *Physical Therapy in Sport, 13*(3), 134–140. https://doi.org/10.1016/j.ptsp.2011.08.004

Morelli, V., & Smith, V. (2001). Groin injuries in athletes. *American Family Physician, 64*(8), 1405.

Morrow, J. R., Jr., Mood, D., Disch, J., & Kang, M. (2015). *Measurement and evaluation in human performance, 5E. Human kinetics*.

O'Brien, M., Bourne, M. N., Heerey, J., Timmins, R. G., & Pizzari, T. (2019). A novel device to assess hip strength: Concurrent validity and normative values in male athletes. *Physical Therapy in Sport, 35*, 63–68. https://doi.org/10.1016/j.ptsp.2018.11.006

Orchard, J., & Seward, H. (2002). Epidemiology of injuries in the Australian football league, seasons 1997–2000. *British Journal of Sports Medicine, 36*(1), 39–44. https://doi.org/10.1136/bjsm.36.1.39

Ransom, M., Saunders, S., Gallo, T., Segal, J., Jones, D., Jones, M., et al. (2020). Reliability of a portable fixed frame dynamometry system used to test lower limb strength in elite Australian Football League players. *Journal of Science and Medicine in Sport, 23*(9), 826–830. https://doi.org/10.1016/j.jsams.2020.03.006

Ryan, J., DeBurca, N., & Mc Creesh, K. (2014). Risk factors for groin/hip injuries in field-based sports: A systematic review. *British Journal of Sports Medicine, 48*(14), 1089–1096. https://doi.org/10.1136/bjsports-2013-092263

Ryan, S., Kempton, T., Pacecca, E., & Coutts, A. J. (2019). Measurement properties of an adductor strength-assessment system in professional Australian footballers. *International Journal of Sports Physiology and Performance, 14*(2), 256–259. https://doi.org/10.1123/ijspp.2018-0264

Schantz, P. G., Moritani, T., Karlson, E., Johansson, E., & Lundh, A. (1989). Maximal voluntary force of bilateral and unilateral leg extension. *Acta Physiologica Scandinavica, 136*(2), 185–192. https://doi.org/10.1111/j.1748-1716.1989.tb08651.x

Secher, N. H. (1975). Isometric rowing strength of experienced and inexperienced oarsmen. *Medicine & Science in Sports, 7*(4), 280–283.

Sullivan, S. J., Chesley, A., Hebert, G., McFaull, S., & Scullion, D. (1988). The validity and reliability of hand-held dynamometry in assessing isometric external rotator performance. *Journal of Orthopaedic & Sports Physical Therapy, 10*(6), 213–217. https://www.jospt.org/doi/10.2519/jospt.1988.10.6.213.

Thorborg, K., Bandholm, T., Schick, M., Jensen, J., & Hölmich, P. (2013). Hip strength assessment using handheld dynamometry is subject to intertester bias when testers are of different sex and strength. *Scandinavian Journal of Medicine & Science in Sports, 23*(4), 487–493. https://doi.org/10.1111/j.1600-0838.2011.01405.x

Thorborg, K., Petersen, J., Magnusson, S. P., & Hölmich, P. (2010). Clinical assessment of hip strength using a hand-held dynamometer is reliable. *Scandinavian Journal of Medicine & Science in Sports, 20*(3), 493–501. https://doi.org/10.1111/j.1600-0838.2009.00958.x

Walter, S. D., Eliasziw, M., & Donner, A. (1998). Sample size and optimal designs for reliability studies. *Statistics in Medicine, 17*(1), 101–110. https://doi.org/10.1002/(SICI)1097-0258(19980115)17:1<101::AID-SIM727>3.0.CO;2-E

Weir, A., Brukner, P., Delahunt, E., Ekstrand, J., Griffin, D., Khan, K. M., et al. (2015). Doha agreement meeting on terminology and definitions in groin pain in athletes. *British Journal of Sports Medicine, 49*(12), 768–774. https://doi.org/10.1136/bjsports-2015-094869

Werner, J., Hägglund, M., Waldén, M., & Ekstrand, J. (2009). UEFA injury study: A prospective study of hip and groin injuries in professional football over seven consecutive seasons. *British Journal of Sports Medicine, 43*(13), 1036–1040. https://doi.org/10.1136/bjsm.2009.066944