

Original Research

Reliability and Agreement of Hand-Held Dynamometry Using Three Standard Rater Test Positions

Frank Aerts¹, Holly Sheets^{1a}, Chance Anderson¹, Natalie Bussie¹, Rose Hoskins¹, Amanda Maninga¹, Emily Novak¹

¹ Doctor of Physical Therapy Program, Manchester University

Keywords: dynamometry, hand-held, mechanical force, reliability, test position, agreement

<https://doi.org/10.26603/001c.128286>

International Journal of Sports Physical Therapy

Vol. 20, Issue 2, 2025

Background

The use of portable hand-held dynamometers is increasing in popularity due to their ease of use in different clinical settings, convenient size, portability, and overall affordability. Reported reliability for external fixation and rater-stabilized hand-held dynamometry (HHD) strength measurements have been found to be 'good' to 'excellent'. Inconsistent agreement has been found between the two stabilization methods and isokinetic HHD testing.

Purpose / Hypothesis

Determine the reliability and agreement of HHD measurements in three different rater test positions against three different mechanically produced force magnitudes. The study compared measurements obtained by rater-stabilization to external fixation methods.

Study Design

Intra-rater and inter-rater reliability study.

Methods

Ten raters took measurements in three different rater test positions against three different force magnitudes created by an external force. Raters were blinded to the randomized force magnitudes. The rater's measurements were compared to measurements taken against an external fixation stabilization device. To establish reliability, Intraclass Correlation Coefficient (ICC), and Standard Error of Measurement (SEM), and Minimal Detectable Change (MDC) were used. To establish agreement, error rates between the rater-stabilized and external fixation stabilization measurements were calculated.

Results

ICC's were found to be 'excellent' at .97 and above. The relative SEM ranged from 0.2% to 0.9 % and the relative MDC ranged from 0.7% to 2.8%. The overall error rate was 15.5% and was influenced by force magnitude.

Conclusion

The use of standardized rater test positions resulted in 'excellent' intra-rater, inter-rater reliability, low SEM, and low MDC for rater-stabilized HHD measurements. A systematic error was observed, with rater-stabilized measurements resulting in higher values compared with values obtained with the external fixation method.

^a Corresponding Author:

Holly Sheets
10627 Diebold Road, Fort Wayne, IN, 46845
Phone: 260-470-4069
Fax: 260-470-4404
Email: hasheets@manchester.edu

Level of Evidence

3

INTRODUCTION

Measuring and restoring strength to aid in rehabilitating a patient's function is considered a key aspect of physical therapist practice.¹⁻³ The demand for objective quantitative measurements has facilitated the increased utilization of dynamometry in clinical practice.⁴ Isokinetic dynamometry provides highly reproducible results but is limited in clinical practice.⁵ The use of portable or hand-held dynamometers (HHDs) is increasing in popularity due to their affordability, operational and clinical efficiency. Hand-held dynamometry (HHD) is easy to use in different clinical settings with different populations.^{5,6} HHDs are commonly used with rater stabilization or with external fixation stabilization.

Reported psychometric properties for HHD measurements have been reported to be good to excellent.^{3,5,7} A recent systematic review with meta-analysis for hip muscle strength measures with portable dynamometers found moderate to high-quality evidence for sufficient intra-rater and inter-rater reliability for some positions regardless of fixation method.⁷ Florencio et al. report that studies show HHD reliability with rater-stabilized measurements with ICC between 0.70 and 0.98, and measurements with external fixation stabilization with ICC between 0.49 and 0.99 for measurements related to hip and knee.² Maximal isometric force values obtained by HHD are comparable to values obtained with isokinetic dynamometry, but inconsistent validity has been found with both rater-stabilized and external fixation methods.^{3,7}

HHD is susceptible to many sources of error and lacks standardization. Sources of error may be grouped into subject attributes, testing procedures, instrument characteristics, and rater attributes. Studies found the reliability of HHD measurements to be reliable in various patient populations.⁸ For example, Vaz et al. found excellent (≥ 0.90) intra-rater and inter-rater intraclass correlation coefficient (ICC) for hip measurements in individuals with symptomatic hip osteoarthritis.⁹ Koblaue et al. found similar excellent intra-rater and inter-rater ICC for knee extensor strength for patients awaiting total knee replacement.⁸ However, they did find a high error rate, expressed in the smallest detectable difference (SDD), ranging from 19.0% to 57.5%, concluding that the use of HHD is not advised for clinical practice. The lack of standardized testing protocols and testing positions for different muscle groups is another source of error.^{3,5,7} Instruments characteristics might be another source. Du et al. examined the variability between different HHD instruments for measuring muscle strength and found differences between 0.2% and 16% between dynamometers.¹⁰ Errors associated with the rater include experience with HHD and rater strength.^{3,5,11,12} The rater's inability to maintain a stable base against higher torque outputs, has led to the recommendation to use external fixation methods.⁷

Operational considerations might not make it feasible to use external fixation methods, such as in inpatient acute care setting. The lack of consensus on using external fixation stabilization versus rater-stabilized methods for stabilization warrants further investigation. In addition, there is no consensus on the most appropriate rater testing positions to enable proper stabilization against an external force.¹³ A wide variety of methods have been described in research and clinical practice.⁷

The primary aim of this study was to assess the inter-rater and intra-rater reliability of rater-stabilized HHD measurements from a mechanically produced force in three different standardized rater test positions. Reliability is defined as the degree of consistency with which a rater measures a variable.¹⁴ A second aim of the study was to assess the agreement between external fixation stabilization and rater-stabilized measurements.

METHODS

DESIGN

This is an intra-rater and inter-rater reliability study approved by the Manchester University Institutional Review Board.

INSTRUMENTATION AND RATER TEST POSITIONS

A novel device was designed to provide a mechanical force. A mannequin leg was securely attached to the base of a height-adjustable plinth. Through a rope, the mannequin leg was attached to a pulley with a 60 kg weight stack (Speed Pulley 702600, STEENS, Norway). The external leverage arm, the point of rotation to the location of the instrument placement, was constant at 43.5 cm. The set-up was configured to allow for three standardized rater test positions commonly used in clinical practice. The positions were selected to allow for optimal rater stabilization and are commonly used to measure hip flexion, hip abduction and knee extension. The selected positions are described by Aerts and Alwood¹⁵ and in [Figure 1](#).

A newly obtained Lafayette Hand-Held Dynamometer (Lafayette Instrument Company, Lafayette IN, USA, Model 01165A) was used to obtain and record the measurements. The hand-held dynamometer was not altered in any way. The device was accompanied by a certificate of calibration with absolute errors ranging from -3.1 newton (N) to +0.6 N, and relative errors ranging from -0.25% to +0.65% between the HHD measurement and actual value. An external fixation stabilization device (Hand-Held Dynamometer Support Stand Model 01166, Lafayette Instrument Company, Lafayette IN, USA) was used to stabilize the HHD measurements obtained by external fixation ([Figure 2](#)).

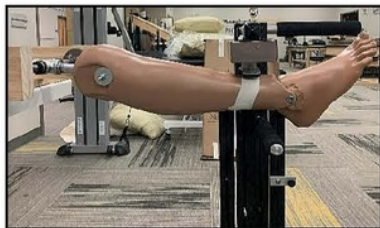
Test Position 1 – Hip Abduction		
External Fixation	Rater-Stabilized	RATER POSTURE Staggered stance, with the lower extremity ipsilateral to the instrument arm forward. POSITION INSTRUMENT ARM (arm/hand holding the HHD): the forearm and hand of the rater's instrument arm should be in the same plane as the movement being tested. The elbow is flexed ninety degrees and stabilized against the rater's pelvis.
		
Test Position 2 – Hip Flexion		
External Fixation	Rater-Stabilized	RATER POSTURE Wide base straddle stance. POSITION INSTRUMENT ARM (arm/hand holding the HHD): The rater's instrument arm, including the shoulder, is in the same plane as the movement being tested. The elbow is extended and is in a neutral pronation / supination. The wrist is in extension.
		
Test Position 3 – Knee Extension		
External Fixation	Rater-Stabilized	POSTURE Kneeling. POSTURE INSTRUMENT ARM (arm/hand holding the HHD): The forearm and hand of the rater's instrument arm should be in the same plane as the movement being tested. The elbow is flexed ninety degrees and stabilized against the rater's pelvis. The non-instrument arm holds the table for stabilization.
		

Figure 1: Test Positions

RATERS

A convenience sample of raters was recruited by the Manchester University faculty from the local geographical area (Fort Wayne, Indiana – USA). After signing a consent form to participate, raters were asked to complete a survey inquiring about their practice experience and familiarity with using HHD. Age, sex, anthropometric information, and hand grip strength were collected. Each rater was assigned a random number between 1 through 10. The raters underwent a thirty-minute training session and performed several practice measurements until they reported being comfortable with the measurement techniques and rater test positions.

TESTING PROCEDURES

The mechanical set-up was calibrated to produce predetermined external forces. The predetermined external forces were based on forces that a rater may encounter during clinical practice when measuring hip abduction, hip flexion, and knee extension using the selected test positions. The force values were based on a data set obtained in clinical practice (n=800) by the main investigator (unpublished clinical data). The force values were categorized into three groups: 1. low (-1SD), 2. medium (mean), and 3. high (+1SD). The testing was performed over three days. Before each testing session, the investigators performed three measurements against each external force magnitude, by using the external fixation method to ensure that the me-



Figure 2: External Fixation Stabilization Device

chanical setup produced the expected predetermined forces. The force values are presented in [Table 2](#).

Each rater was then asked to complete nine measurements in each of the three different test positions, i.e. hip abduction, hip flexion, and knee extension. Three different force magnitudes (i.e. low, medium, high) were randomized across the nine measurements so that each rater performed three measurements against each force magnitude in random order. To mimic the clinic situation where therapists do not exactly know how much force the patient or client will produce, the raters were blinded to the preset external forces. The investigators, recording the rater-stabilized HHD measurements, were also blinded to the preset external forces. The reading of the peak force (N) of each measurement was recorded. The raters used a “make” technique where the raters matched the external mechanical force for a duration of five seconds. With a “make” technique the rater holds the hand-held dynamometer stationary matching the external torque. This would mimic the measurements taken when the patient / client produces force through an isometric muscle contraction. In contrast with a “break” technique, the rater must create enough force to break the isometric muscle contraction. This would mimic measurements taken when the patient / client produces force through an eccentric muscle contraction. Both techniques have been used in clinical practice and research. A “make” technique might be favored during early rehabilitation as a “break” technique may increase patients / client risk for injury.⁶ Additionally, a “make” technique may have better reliability and provide more accurate measurements.^{3,8}

STATISTICAL ANALYSIS

Statistical analysis was performed using the Statistical program for the Social Sciences version 28 (SPSS, IBM corp., Armonk, NY) and significance was set at $\alpha < 0.05$.

Raters’ demographics were analyzed and reported using descriptive statistics. Quantitative variables are expressed in mean (N) and standard deviation (SD). Based on the sam-

ple calculator presented by Bonett, using 10 raters, sample size calculation was conducted for reliability data with an ICC estimated at 0.80, an amplitude-based confidence interval of 0.3, that is, $0.5 < ICC < 1.0$, and a confidence coefficient $\alpha < 0.05$, resulted in 11 different tests (subjects).^{2,16} The intraclass correlation coefficient (ICC) is the most common statistic used to assess intra-rater and inter-rater reliability.⁸ Intra-rater reliability for each rater was assessed by using the three measurements of each force magnitude obtained in each test position. The two-way mixed intra-class correlation coefficient model 3 (ICC3,k)¹⁴ and the 95% confidence intervals were calculated. Inter-rater reliability was assessed using the average of the three trials for each measurement. The random effects absolute agreement inter-class correlation coefficient model 2 (ICC2,k) and the 95% confidence intervals were calculated.¹⁴ The following guidelines were used to interpret the ICC: below 0.75 as poor to moderate, above 0.75 as ‘good’, and above 0.90 as ‘excellent’ which ensures reasonable reliability.¹⁴

The clinical utility of the ICC is minimal as it provides a measure of reliability within a study group and not individual measures. Therefore, the Standard Error of Measurement (SEM) and Minimal Detectable Change (MDC) are used as additional measures of reliability.

The (SEM) and (MDC) were calculated as an additional assessment of inter-rater reliability. Lower values of SEM and MDC values are indicative of lower measurement error and better reliability.⁸ SEM was calculated using the following formula: $SEM = SD \times \sqrt{(1 - ICC)}$. MDC was calculated as follows: $MDC = 1.96 \times \sqrt{2} \times SEM$. All SEM and MDC values are presented in absolute values and as a percentage of the mean maximal strength. For each rater-stabilized test, the Mean and SD were calculated using the average of the three measurements from each rater. For the SEM and MDC, the inter-rater ICC for each test position was used.

The agreement refers to the ability of a measurement tool to produce the same exact values.¹⁷ Agreement was assessed by calculating the error rate between external fixation stabilization and rater-stabilized measurements. Both external fixation stabilization and rater-stabilized HHD measurements were obtained from a mechanical force therefore controlling for patient / client variability. By controlling errors associated with subjects, testing procedures, and instrument variability, the researchers aimed to obtain an estimate of error associated solely with the rater.

The agreement was estimated by calculating the error rate as follows¹⁸:

$$\text{Error Rate (\%)} = \frac{(\text{non} - \text{fixed dynamometer reading} - \text{fixed dynamometer reading})}{\text{fixed dynamometer reading}} \times 100$$

The error rates of the average measurements of each rater’s tests were used to estimate overall accuracy. A one-way ANOVA was conducted to determine if the error rate (%) between the rater-stabilization and external fixation stabilization measurements were different by rater, test position, or by force magnitude.

Table 1. Rater Characteristics

Sex (Female / Male)	4 / 6
Age (years)	33.7 (7.6)
Height (m)	1.77 (0.9)
Weight (Kg)	91.7 (23.2)
Handedness (Right / Left)	9 / 1
Maximum Grip Strength (N)	456 (140)
Years of Practice	Less than 10 years: 6 raters Between 10 and 19 years: 4 raters
Practice Setting	Collegiate Sport Facility: 1 rater Hospital Based Outpatient: 3 raters Private Practice: 6 raters
Specialization	Board-Certified Orthopaedic Clinical Specialist: 3 raters
Practical HHD Experience (Estimated by the number of patients / clients a rater uses HHD)	50-90%: 6 raters 10-50%: 2 raters Never used: 2 raters

*Continuous variables presented in mean (SD)

RESULTS

RATERS

Ten participants consented to be a part of the study. All raters were licensed physical therapists and worked in various clinical settings. Raters' characteristics are presented in [Table 1](#). Six raters reported using HHD with 50 to 90% of their patient / clients, two raters with 10 to 50% of their patient / clients, and two raters stated they never used HHD. The overall strength of the raters was assessed through grip strength. The mean grip strength for all raters was 456N $\pm$ 140. All 10 raters performed all measurements in Test Position 1 - Hip Abduction and Test Position 2 - Hip Flexion. Two raters did not complete measurements in Test Position 3 - Knee Extension against the highest external force magnitude. The measurement attempts were halted as the raters were unsure if they could maintain a stable base. The raters who were unable to maintain a stable base were both females who never or rarely uses HHD and had the lowest mean dominant grip strength measurements of 275N and 314N.

RELIABILITY

Intra-rater reliability for each rater was assessed using the three trials for each measurement. Two providers did not perform measurements in Test Position 3 - Knee Extension against the highest resistance. The intra-rater reliability (ICC3,k) for the different raters across the measurements ranged from 0.97 to 1.00. Inter-rater reliability was assessed using the means of the three trials from each rater for each measurement. After examining box plots, one extreme outlier (three box lengths away from the edge of the box), classified as (Rater 3, Test Position 2 - Hip Flexion, low force magnitude) was identified but was not removed before further analysis. The inter-rater reliability (ICC2,k) between the different raters in the three test positions was 0.99 (CI 95% 0.93; 1.00). The absolute SEM ranged from 0.5

to 3.0 N and the relative SEM from 0.2% to 0.9 % respectively. The absolute MDC ranged from 1.4 to 8.3N and the relative MDC from 0.7% to 2.8%. [Table 2](#) provides a summary of the measurements by test position and force magnitude (nine tests) for the external fixation and rater-stabilized measurements.

AGREEMENT

After examining boxplots, two outliers were identified. The authors did not remove this data from the analysis as it did not impact the analysis outcome. The means obtained by rater-stabilized measurements compared to the means obtained by external fixation measurements were higher in all nine tests. The error rate between the external fixation and rater-stabilized measurements ranged from 6.9% in Test Position 3 - Knee Extension, low force magnitude to 31.2% in Test Position 3 - Knee Extension, high force magnitude. Data was normally distributed for all groups, as assessed by Shapiro-Wilk's test ($p > 0.05$), except for one group i.e. "Test Position 2 - Hip Flexion". After reviewing the "normal Q-Q plot" and considering the robustness of one-way ANOVA against Type I error with equal group sizes, the analysis proceeded.

For "Rater", variances were homogeneous by Levene's for equality of variances ($p = 0.443$). There was no statistically significant difference in error rate (%) between the different raters ($F(9,77) = 1.358$, $p = 0.222$). Error rates by rater are presented in [Table 3](#).

For "Test Position", there was no homogeneity of variances by Levene's for equality of variances ($p < 0.001$). There was no statistically significant difference in error rate (%) between the different test positions (Welch's $F(2,47.691) = 1.583$, $p = .216$). Error rates by test position are presented in [Table 4](#).

For "Force Magnitude", there was no homogeneity of variances by Levene's for equality of variances ($p < 0.001$). There was a statistically significant difference in error rate (%) between the different force magnitudes (Welch's F

Table 2. Agreement of Peak Strength Measurements Using External Fixation and Rater-Stabilization of Mechanical Produced Force (n=9)

EXTERNAL FIXATION TESTS						RATER-STABILIZED TESTS									DIFFERENCE		
	n	Mean (N)	SD (N)	SEM (N)	SEM (%)	n	Mean (N)	SD (N)	ICC (2,k)	CI 95%	SEM (N)	SEM (%)	MDC (N)	MDC (%)	DIFFERENCE (N)	SD (N)	Error Rate (%)
Test Position 1 - Hip Abduction									0.99	(0.98; 1.00)							
Low Resistance	9	80.4	5.3	0.3	0.4	10	95.4	5.2			0.5	0.5	1.4	1.5	15.0	5.2	18.7
Medium Resistance	9	146.1	8.1	0.4	0.3	10	166.2	6.7			0.7	0.4	1.9	1.2	20.1	6.7	13.8
High Resistance	9	215.5	5.8	0.3	0.1	10	234.7	5.8			0.6	0.2	1.7	0.7	19.2	5.8	8.9
Test Position 2 - Hip Flexion									0.99	(0.93; 1.00)							
Low Resistance	9	60.2	7.0	0.4	0.7	10	75.3	7.1			0.7	0.9	1.9	2.6	15.1	4.8	25.1
Medium Resistance	9	119.3	4.5	0.2	0.2	10	129.7	12.6			1.3	1.0	3.6	2.8	10.4	12.6	8.7
High Resistance	9	178.5	6.1	0.3	0.2	10	194.1	8.7			0.9	0.4	2.5	1.3	15.6	8.7	8.7
Test Position 3 - Knee Extension									0.99	(0.93; 1.00)							
Low Resistance	9	66.6	11.2	0.6	0.9	10	87.4	6.6			0.7	0.8	1.9	2.2	20.8	6.6	31.2
Medium Resistance	9	272.4	26.5	1.5	0.6	10	315.4	26.0			2.6	0.8	7.2	2.3	43.0	26.0	15.8
High Resistance	9	464.2	33.8	1.8	0.4	8	496.2	29.6			3.0	0.6	8.3	1.7	32.0	29.6	6.9

Table 3. Error rate by rater

Rater	Error Rate Mean $\pm$ SD (%)
Rater 1 (n=9)	12.9 $\pm$ 9.6
Rater 2 (n=9)	13.3 $\pm$ 7.6
Rater 3 (n=9)	25.4 $\pm$ 15.4
Rater 4 (n=9)	15.7 $\pm$ 7.4
Rater 5 (n=9)	13.5 $\pm$ 9.0
Rater 6 (n=9)	11.3 $\pm$ 8.4
Rater 7 (n=8)	16.4 $\pm$ 12.5
Rater 8 (n=9)	14.6 $\pm$ 9.6
Rater 9 (n=8)	17.1 $\pm$ 9.6
Rater 10 (n=9)	14.8 $\pm$ 9.9

n, number; SD, Standard Deviation

Table 4. Error rate by test position

Position	Error Rate Mean $\pm$ SD (%)
Position 1 – Hip Abduction (n=30)	13.7 $\pm$ 5.3
Position 2 – Hip Flexion (n=30)	14.3 $\pm$ 11.2
Position 3 – Knee Extension (n= 28)	18.7 $\pm$ 12.7

n, number; SD, Standard Deviation

Table 5. Error rate by force magnitude

Force Magnitude	Error Rate Mean $\pm$ SD (%)
Low Force Magnitude (n=30)	25.2 $\pm$ 9.6
Medium Force Magnitude (n=30)	12.6 $\pm$ 7.5
High Force Magnitude (n= 28)	8.2 $\pm$ 3.8

n, number; SD, Standard Deviation

(2,50.798) = 42.938, $p < 0.001$). Error rates by force magnitude are presented in [Table 5](#).

DISCUSSION

This study investigated the variability associated with the rater by assessing the reliability and agreement of HHD measurements. External fixation was compared to rater-stabilized measurements using the same instrument, taken in three standardized rater test positions against three different force magnitudes. The external force was created by a mechanical device eliminating the variability associated with patients or clients.

Both intra-rater and inter-rater reliability were excellent in this study with intra-rater and interclass correlation coefficients of 0.97 and above. These results are consistent with previous studies reporting reliability for HHD measurements for hip measurements.^{5,7,19} The inter-rater relative SEM values obtained in this study ranged from 0.2% to 1.0% and relative MDC values ranged from 0.7% to 2.8%.

The values of the current study compare with the values reported by Morin et al. who reports relative inter-rater SEM values between 1.1% and 3.0% and relative MDC values between 3.1% and 8.3% for hip and knee measurements obtained by using a semi-fixed and pull dynamometer.⁵ The values found in this study for relative SEM were lower than the values reported by other studies.^{2,20} Florencio et al. reported intra-session relative SEM values for HHD hip and knee, using examiner and belt stabilized measurements, between 6% and 15% using a similar instrument.²

Error rates were calculated to assess agreement. The error rates between the external fixation stabilization and rater-stabilized measurements ranged from 6.9% to 31.2% comparable to -4.9% to 27.1% calculated values based on the data provided by Florencio et al.²

The raters consistently, regardless of the force magnitude and test position, measured higher values compared to the external fixation stabilization measurement. This finding is consistent with Florencio et al. who reported that measurements obtained with examiner stabilization were generally greater than those observed for belt-stabilization in 13 of 16 measurement positions.² This systematic error might be related to the raters trying to adjust to the direction and magnitude of the force. Providers were instructed to ensure that the pressure pad stayed perpendicular to the movement arm (artificial limb), which was not assured by the mechanical stabilization. This finding is similar to Florencio et al. who suggested that better control of the dynamometer position is guaranteed when the stabilization is provided by the rater.² Another reason can be related to the increased impulse force created when raters are trying to adjust to the force magnitude. The raters were instructed to maintain the starting position and not to push back. The external force was released in a controlled two-second manner. By adjusting to the unknown force, the rater may create a higher impulse force increasing the force output registered by the hand-held dynamometer. In addition, the results of this study suggest that this systematic relative error is more pronounced with lower force magnitudes.

It has been recognized that assessor strength impacts the ability to provide proper stabilization when confronted with higher forces exceeding the assessor's strength.⁶ In this study, two raters were unable to provide a stable base when confronted with a higher force magnitude in Test Position 3 – Knee Extension.

The major strength of this study is the exclusion of variability related to patients or clients. The investigators could only identify one other study investigating HHD accuracy using a mechanically produced force by a spring-loaded device.¹⁸ Overall accuracy was estimated at 3%. The main difference was the raters used a “break technique” compared to a “make technique” in this study. Another strength of the study was the use of the same instrument for both stabilization methods, eliminating the error associated with the use of different instruments. A third strength of this study was the randomization and blinding of recording investigators and raters to the forces applied by the mechanical device. This mimics the clinical situation where raters do

not know how much resistance to provide to counteract the forces produced by the patient / client.

This study has several limitations. The limited sample size ($n = 9$ tests) was smaller than the calculated sample size required ($n=11$), which may have resulted in an underpowered study. The mechanical setup did not accommodate an additional test position. In retrospect, an additional force magnitude could have been included. The limited sample size precluded the investigation of interaction effects between rater, test position, and force magnitude. Further research should explore these possible interaction effects. A second limitation was the minimal training provided to the participating raters, approximately 30 minutes of instruction and practice. In the study of Morin et al., trainers received three[days of training followed by 20 hours of practice.⁵ This limited training may have an impact on the raters' ability to provide the proper stabilization. A third limitation might be related to the external fixation stabilization device. The relative SD and relative SEM for our external fixation stabilization device measurements ranged from 2.7 % to 16.8% and from 0.1% to 0.9% respectively. Florencio et al. reported data for knee and hip strength measurements from healthy young individuals using a similar HHD LaFayette instrument.² Based on the data from Florencio et al., the relative SD which ranges from 31.6% to 37.3%. were calculated² Florencio et al. reported relative SEM from 7% to 15%.² The relative SEM includes the error associated with the instrument itself (-0.25% - 0.65%).² Both studies used fixed stabilization technique, but the external mechanical produced torque in our study resulted in lower relative SD and relative SEM compared to the subject produced torque in the study from Florencio et al.²

CONCLUSION

This research study provides evidence that rater-stabilized HHD measurements have excellent intra-rater and inter-rater reliability, with low SEM and MDC. The maximal ex-

ternal force should not exceed the rater capability and raters should use standardized test positions. In clinical practice, providers should be aware that HHD may have a degree of inaccuracy. The values obtained with rater-stabilized HHD were consistently higher compared to the values obtained by external fixation HHD. Further research should investigate if this overestimation is influenced by the force magnitude.

ACKNOWLEDGMENTS

The authors would like to thank all raters for their contributions. We would like to thank Adam Fischer for constructing the mechanical setup and Ed Ball for soliciting the participating raters.

FUNDING

This work was supported in part by the Manchester University Health Sciences and Pharmacy Programs Internal Research Grant. The funding source had no involvement in study design, data collection, analysis, interpretation, manuscript writing, or decision to submit the article for publication.

CONFLICT OF INTEREST

The authors report no conflicts of interest.

Submitted: June 25, 2024 CST. Accepted: December 14, 2024 CST. Published: February 01, 2025 CST.

© The Author(s)



This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CCBY-NC-4.0). View this license's legal deed at <https://creativecommons.org/licenses/by-nc/4.0> and legal code at <https://creativecommons.org/licenses/by-nc/4.0/legalcode> for more information.

REFERENCES

1. APTA Guide to Physical Therapist Practice 4.0. American Physical Therapy Association. 2023. Accessed April 19, 2024. <https://guide.apta.org>
2. Florencio LL, Martins J, da Silva MRB, da Silva JR, Bellizzi GL, Bevilaqua-Grossi D. Knee and hip strength measurements obtained by a hand-held dynamometer stabilized by a belt and an examiner demonstrate parallel reliability but not agreement. *Phys Ther Sport*. 2019;38:115-122. doi:[10.1016/j.ptsp.2019.04.011](https://doi.org/10.1016/j.ptsp.2019.04.011)
3. Morin M, Duchesne E, Bernier J, Blanchette P, Langlois D, Hébert LJ. What is known about muscle strength reference values for adults measured by hand-held dynamometry: A scoping review. *Arch Rehabil Res Clin Transl*. 2021;4(1):100172. doi:[10.1016/j.arrct.2021.100172](https://doi.org/10.1016/j.arrct.2021.100172)
4. Loria K. A look ahead: Technology. *APTA Magazine*. 2021;13(8).
5. Morin M, Hébert LJ, Perron M, Petitclerc É, Lake SR, Duchesne E. Psychometric properties of a standardized protocol of muscle strength assessment by hand-held dynamometry in healthy adults: a reliability study. *BMC Musculoskelet Disord*. 2023;24(1):294. doi:[10.1186/s12891-023-06400-2](https://doi.org/10.1186/s12891-023-06400-2)
6. Garcia MAC, Souza VH. The (un)standardized use of handheld dynamometers on the evaluation of muscle force output. *Braz J Phys Ther*. 2020;24(1):88-89. doi:[10.1016/j.bjpt.2019.10.004](https://doi.org/10.1016/j.bjpt.2019.10.004)
7. Waiteman MC, Garcia MC, Briani RV, et al. Can clinicians trust objective measures of hip muscle strength from portable dynamometers? A systematic review with meta-analysis and evidence gap map of 107 studies of reliability and criterion validity using the COSMIN methodology. *J Orthop Sports Phys Ther*. 2023;53(11):655-672. doi:[10.2519/jospt.2023.12045](https://doi.org/10.2519/jospt.2023.12045)
8. Koblbauer IF, Lambrecht Y, van der Hulst ML, et al. Reliability of maximal isometric knee strength testing with modified hand-held dynamometry in patients awaiting total knee arthroplasty: useful in research and individual patient settings? A reliability study. *BMC Musculoskelet Disord*. 2011;12:249. doi:[10.1186/1471-2474-12-249](https://doi.org/10.1186/1471-2474-12-249)
9. Vaz GF, Freire FF, Gonçalves HM, de Aviz MAB, Martins WR, Durigan JLQ. Intra- and inter-rater reliability, agreement, and minimal detectable change of the handheld dynamometer in individuals with symptomatic hip osteoarthritis. *PLoS One*. 2023;18(6):e0278086. doi:[10.1371/journal.pone.0278086](https://doi.org/10.1371/journal.pone.0278086)
10. Du W, Cornett KMD, Donlevy GA, Burns J, McKay MJ. Variability between different hand-held dynamometers for measuring muscle strength. *Sensors*. 2024;24(6):1861. doi:[10.3390/s24061861](https://doi.org/10.3390/s24061861)
11. Ieiri A, Tushima E, Ishida K, Inoue M, Kanno T, Masuda T. Reliability of measurements of hip abduction strength obtained with a hand-held dynamometer. *Physiother Theory Pract*. 2015;31(2):146-152. doi:[10.3109/09593985.2014.960539](https://doi.org/10.3109/09593985.2014.960539)
12. Krause DA, Neuger MD, Lambert KA, Johnson AE, DeViny HA, Hollman JH. Effects of examiner strength on reliability of hip-strength testing using a handheld dynamometer. *J Sport Rehabil*. 2014;23(1):56-64. doi:[10.1123/jsr.2012-0070](https://doi.org/10.1123/jsr.2012-0070)
13. Mentiplay BF, Perraton LG, Bower KJ, et al. Assessment of lower limb muscle strength and power using hand-held and fixed dynamometry: A reliability and validity study. *PLoS One*. 2015;10(10):e0140822. doi:[10.1371/journal.pone.0140822](https://doi.org/10.1371/journal.pone.0140822)
14. Portney L, Watkins M. *Foundations of Clinical Research: Applications to Practice*. 3rd ed. F.A. Davis Company; 2015.
15. Aerts F, Alwood B. *Hand-Held Dynamometry: Guidelines for Daily Clinical Practice*. 1st ed. MET Seminars; 2018.
16. Bonett DG. Sample size requirements for estimating intraclass correlations with desired precision. *Statistics Med*. 2002;21(9):1331-1335. doi:[10.1002/sim.1108](https://doi.org/10.1002/sim.1108)
17. Berchtold A. Test-retest: Agreement or reliability? *Methodological Innovations*. 2016;9. doi:[10.1177/2059799116672875](https://doi.org/10.1177/2059799116672875)
18. Goonetilleke A, Modarres-Sadeghi H, Guilloff RJ. Accuracy, reproducibility, and variability of hand-held dynamometry in motor neuron disease. *J Neurol Neurosurg Psychiatry*. 1994;57(3):326-332. doi:[10.1136/jnnp.57.3.326](https://doi.org/10.1136/jnnp.57.3.326)
19. González-Rosalén J, Benítez-Martínez JC, Medina-Mirapeix F, Cuerda-Del Pino A, Cervelló A, Martín-San Agustín R. Intra- and inter-rater reliability of strength measurements using a pull hand-held dynamometer fixed to the examiner's body and comparison with push dynamometry. *Diagnostics*. 2021;11(7):1230. doi:[10.3390/diagnostics11071230](https://doi.org/10.3390/diagnostics11071230)

20. Ishøi L, Hölmich P, Thorborg K. Measures of hip muscle strength and rate of force development using a fixated handheld dynamometer: Intra-tester intra-day reliability of a clinical set-up. *Int J Sports Phys Ther.* 2019;14(5):715-723. doi:[10.26603/ijsp20190715](https://doi.org/10.26603/ijsp20190715)