Open Journal of Orthopedics, 2026, 16(5), 252-263
https://www.scirp.org/journal/ojo
ISSN Online: 2164-3016
ISSN Print: 2164-3008

Inter-Rater and Intra-Rater Reliability of Handheld Dynamometry for Hip External Rotation in Seated versus Side-Lying

Jacob P. Irwin, Maggie D. Robinson, Deliz M. Williams, Harrison Riley

Department of Physical Therapy, Georgia State University, Atlanta, USA
Email: jirwin@gsu.edu

How to cite this paper: Irwin, J.P., Robinson, M.D. Williams, D.M. and Riley, H. (2026) Inter-Rater and Intra-Rater Reliability of Handheld Dynamometry for Hip External Rotation in Seated versus Side-Lying. *Open Journal of Orthopedics*, **16**, 252-263.
https://doi.org/10.4236/ojo.2026.165024

Received: April 10, 2026
Accepted: May 23, 2026
Published: May 26, 2026

Abstract

The purpose of this study is to compare the reliability of manual muscle testing (MMT) for hip external rotation (ER) in the side-lying (SL) and seated positions. The SL position aligns hip ER testing with current clinical practice and may allow greater contribution from the piriformis during strength assessment in physical therapy. Hip external rotator strength of 18 subjects was measured using a handheld dynamometer (HHD) placed 3 cm above the medial malleolus in both seated and SL positions. Tibial and femoral lengths were recorded to calculate torque for each position. Inter- and intra-rater reliability were assessed using intraclass correlation coefficients (ICC). Across all three examiners, intra-rater reliability was consistently higher in the SL position, ranging from poor to excellent (ICC = 0.229 [95% CI −0.072, 0.503], 0.811 [0.594, 0.909], 0.942 [0.889, 0.970]) compared to the seated position (ICC = 0.187 [−0.101, 0.463], 0.555 [0.271, 0.747], 0.739 [0.387, 0.881]). Inter-rater reliability was poor in both positions (seated ICC = 0.240 [0.006, 0.563]; SL ICC = 0.234 [0.003, 0.551]). These findings suggest that testing hip ER strength in the SL position poses similar reliability to assessment in the seated position. While inter-rater reliability was determined to be poor for both testing positions, some examiners were found to have increased intra-rater reliability in the SL position. The SL position aligns resistance against gravity, improves objective strength measurement, enhances isolation of the external rotators, increases contribution from all ER muscles, and reduces hip flexor compensation seen in seated testing. Reliable assessment of these muscles is important for injury prevention and rehabilitation, so the SL position can be considered an equally reliable way to assess hip ER strength clinically.

Keywords

Muscle Strength Dynamometer, Hip, Reproducibility of Results, Adult,

Range of Motion

1. Introduction

Assessment of muscle strength is a key factor in the examination of the musculoskeletal system that helps to identify impairments, guide treatment, and monitor progress in clinical practice [1]-[3]. Manual muscle testing remains a common technique used to assess patient strength, yet the reliability of the results varies based on the muscle group and testing position [1] [3]-[5]. As the hip joint angle changes with position for strength testing, strength values produced may vary [5] [6]. When the hip is flexed less than 90 degrees, piriformis acts as an external rotator. However, hip flexion greater than 90 degrees alters the moment arm of the piriformis, changing the muscle action to internal rotation [7].

Hip external rotators play a crucial role in lower extremity alignment, stabilization, and movement efficiency [2] [5] [6] [8]. We defined hip ER MMT as the ability of the hip girdle ER muscles to maintain an isometric hold against outside force applied by an examiner. Weakness in these muscles is associated with various musculoskeletal conditions, including anterior cruciate ligament (ACL) injuries [9], patellofemoral pain syndrome [10], and low back pain [1]. Current approaches to addressing hip external rotator strength may be limited by inter-rater and intra-rater reliability, which may compromise the findings [1] [3] [4]. Previous data collection attempts and analyses performed in preparation for the continuation of this study suggest that differences in physical characteristics of the examiner, such as height, upper body strength, and leverage may influence the reliability of MMTs. To account for variability, we chose to standardize the table height relative to each examiner to minimize discrepancies by optimizing examiner mechanics during testing.

The purpose of this study is to compare the reliability of manual muscle testing for hip ER in two positions. In standard muscle testing textbooks such as Muscles: Testing and Function with Posture and Pain [11] and Daniels and Worthingham's Muscle Testing: Techniques of Manual Examination and Performance Testing [12], hip ER MMT is performed seated. Our goal is to compare the reliability of the standard measurement method to the SL position. This would potentially improve the standardization of results and bring the test for hip ER strength into line with the vast majority of manual muscle testing, which are performed against gravity, as this will reflect hip ER strength in a more functional position [1] [5] [13]. If our method is determined to be equally as reliable as the results gathered using standard MMT measurements in seated position, it may be implemented in a clinical setting as a more functional way to demonstrate hip ER strength in patient populations.

2. Materials and Methods

2.1. Ethics

The Georgia State University Institutional Review Board (IRB) reviewed and approved this study in accordance with 45 CFR 46.111 (IRB Number H23121). The IRB has reviewed and approved the research protocol and any informed consent forms, recruitment materials, and other research materials that are marked as approved in the application.

2.2. Participants

A convenience sample of 18 asymptomatic individuals, with no prior hip injury, were recruited to participate in this study. Participants were selected from the local university population, with ages ranging from 18 to 65 years old. Recruitment was conducted primarily by word of mouth, targeting students that reported no past medical history that would influence results or participant health. Participants were included if they met the following criteria: 18 - 65 years of age, healthy and asymptomatic, absence of musculoskeletal pain or dysfunction affecting the hips, demonstration of understanding study procedures, and providing consent for participation. Exclusion criteria consisted of reporting current pain in the lumbopelvic region, having a neurologic condition that affects strength, having a systemic condition that impacts ability to participate or poses additional risk, and being unable to provide consent (**Table 1**).

Table 1. Inclusion criteria exclusion criteria.

Inclusion Criteria	Exclusion Criteria
1. 18 - 65 years old	1. Current hip, pelvis, or low back pain
2. Healthy and asymptomatic	2. Neurological conditions affecting strength
3. No current MSK pain or dysfunction affecting the hip	3. Systemic conditions interfering with ability to participate or pose additional risk
4. Ability to understand study procedures and provide consent	4. Inability to provide consent

Participation was voluntary, and no compensation was provided. Prior to enrollment in the study, all participants provided written informed consent and a medical history form, confirming that they understood the purpose, procedures, risks, and benefits of the study. Participants were informed of their right to withdrawal at any time.

The study was conducted under controlled conditions, with the sample population including only healthy, asymptomatic participants and the applied methods being designed to pose minimal risk.

The mean age of participants was 26.3 years (21 to 41 years of age), 9 females were included, 9 males were included, and all participants reported having a moderate to high weekly physical activity level. Due to each hip on all participants being assessed and only compared to the values obtained by each examiner for that same limb, limb dominance was determined not to be relevant to the statistical analysis. Therefore, limb dominance was not assessed, and each hip for each participant was analyzed separately (36 hips assessed across 18 participants).

2.3. Testing Set up and Procedures

The testing procedure began with a warmup of walking laps on flat ground to properly prepare the hip musculature for exertion. The distance determined to constitute a proper warm up was 0.25 miles, and the condition of the participants' hip musculature and joints were monitored through subjective reporting before performing the trial measurements. The participants' leg lengths were measured starting from the greater trochanter of the femur to lateral epicondyle, and then from lateral epicondyle to 3 cm proximal to the lateral malleoli to ensure proper torque calculations for each position. Each participant performed four trials for each leg on their given testing day: two in the seated position and two in the SL position against gravity. A rest period of 2 minutes was given between each trial to ensure maximum muscle recovery [1]. Each participant was cycled through a randomized pairing of participants to researchers for each trial day, ensuring that each participant was measured at least once by each researcher. The participants were measured a total of three times and given at least a 24-hour rest period between trial dates to give their muscles time to recover for the next trial day. The interval between trials ranged from 24 to 72 hours.

Each participant was assessed in the seated and SL positions, with both lower extremities being individually tested twice per trial date in each testing position. For each participant on a single trial date, the left leg was assessed twice in the SL position and twice in the seated position, and the right leg was assessed twice in the SL position and twice in the seated position. One examiner performed all seated measurements for both attempts on each lower extremity during a single trial date, and a different examiner assessed the participant in the SL position on that same trial date for an equivalent number of trials. ICC values pertaining to Inter-rater Reliability were based on ratings collected across different trial dates (for performing the same test position on the same limb). ICC values pertaining to Intra-rater Reliability were based on measurements collected by the same examiner across seated and SL assessments for the same limb. This occurred across multiple trial dates, as each examiner only performed two measurements for each position on each limb. Therefore, intra-rater reliability is reflective of the consistency within each examiner to perform the strength assessment in both test positions.

The test table height was standardized for each examiner. Table height was adjusted to be level with the fingertips of the examiner, while the examiner stood erect with a neutral spine and arms at their side. Standardizing the table height was intended to allow for greater inter-rater reliability by to provide consistency when applying break force in attempts to account for height and mass differences. Maintaining this line of force allows each examiner to apply a similar break force [1] [14]. This is important as height and strength differences of the examiners could influence reliability of participant strength results. Participants' ER strength was assessed in the seated and SL positions. For the SL position, the bottom leg was tested with the hip and pelvis neutral in the sagittal plane and the knee flexed to 90 degrees with the joint line positioned at the edge of the table (see **Figure 1**). The top leg rested in approximately 45 degrees of hip flexion for participant comfort. Examiners applied a stabilizing force with the cephalic hand on the pelvis during measurement to avoid pelvic rotation. Participants were instructed to hold on to the table for stability. Each examiner applied a downward perpendicular force toward the floor with a handheld dynamometer (HHD) placed 3 cm above the medial malleolus and attempted to push until the participant's ankle moved downward, indicating a proper break force was applied [14]. In the seated testing position, the participants' hips and knees were flexed to 90 degrees with a towel roll under the knee tested, promoting optimal alignment (see **Figure 2**) [1] [4] [11] [12] [15]. The HHD was placed 3 cm above the medial malleolus, and the examiner applied a lateral force until the ankle moved laterally, indicating that a proper break force was applied [14]. Participants were randomly assigned on the trial day to begin in either the seated or SL positions, and starting positions alternated at subsequent data collection dates to minimize potential order effects. All three examiners performed this protocol on each participant during each session.

Figure 1. SL hip ER MMT position.

Figure 2. Seated hip ER MMT position.

2.4. Calculation of Strength-Torque

The strength values for both seated and SL positions were determined by calculating torque [1] with the following formula: $\tau = rF\sin(\theta)$. In this formula, "r" represents the distance from the axis of rotation, also known as the lever arm. "θ" represents the angle from the hip axis to where the force is being applied at the medial malleolus. This was measured in centimeters and then converted into meters. For the seated position, the lever arm consisted of the tibia length, [1] measured from the medial joint line to 3 cm above the medial malleolus, where the HHD was placed. The SL position utilized the length of the femur, measured from the greater trochanter to the lateral femoral condyle. "F" represents the force produced by the participants in joules (later converted into newtons) and was measured using a handheld dynamometer (HHD) [16]. The force was applied perpendicular to the lever arm, resulting in an angle (θ) of 90˚, for maximal torque production. With these variables applied, the final unit to perform calculations and data analysis is the newton-meter (N·m).

2.5. Statistical Analysis

Intraclass correlation coefficients were calculated using IBM SPSS Statistics (Version 29). Average torque values across three trials were used for analysis. Intra-rater reliability was assessed using a two-way mixed-effects model with absolute agreement for single measures (ICC [3, 1]), and inter-rater reliability was assessed using a two-way random-effects model with absolute agreement for single measures (ICC [2, 1]). ICC reliability was classified as poor (<0.50), moderate (0.5 - 0.75), good (0.75 - 0.90), and excellent (>0.90) [17]. As explained above, ER torque output was calculated for both hips on every participant. Calculated torque measurements were then compared to one another using ICC to assess intra-rater reliability across repeated trials conducted by the same examiner and inter-rater reliability between measurements obtained by different examiners on separate testing days.

3. Results

Across data collection, all participants completed the testing procedures as instructed and attended their three allotted trial dates. Few adverse events were noted among participants, including intense muscle soreness in the hip external rotators lasting two days or less and proximal ankle discomfort, likely from HHD placement.

ICC values for intra-rater reliability varied across examiners and positions. In the seated position, examiner one demonstrated poor reliability ICC = 0.187 (95% CI [−0.101, 0.463]), examiner two demonstrated moderate reliability ICC = 0.739 (95% CI [0.387, 0.881]), and examiner three demonstrated fair reliability ICC = 0.555 (95% CI [0.271, 0.747]). Across all three examiners, intra-rater reliability was higher in the SL position than in the seated position. However, only one examiner's ICCs demonstrated non-overlapping 95% confidence intervals as illustrated in **Figure 3**. For the SL position, examiner one demonstrated poor reliability ICC = 0.229 (95% CI [−0.72, 0.503]), examiner two demonstrated excellent reliability ICC = 0.942 (95% CI [0.889, 0.970]) and examiner three demonstrated good reliability ICC = 0.811(95% CI [0.594, 0.909]) (Figure 4).

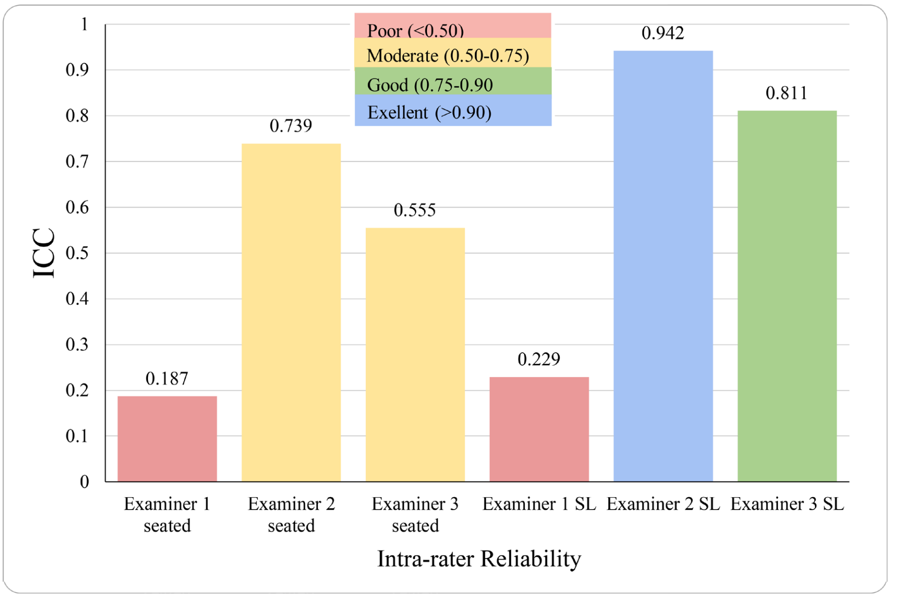

Figure 3. Intra-rater reliability among examiners for each testing position.

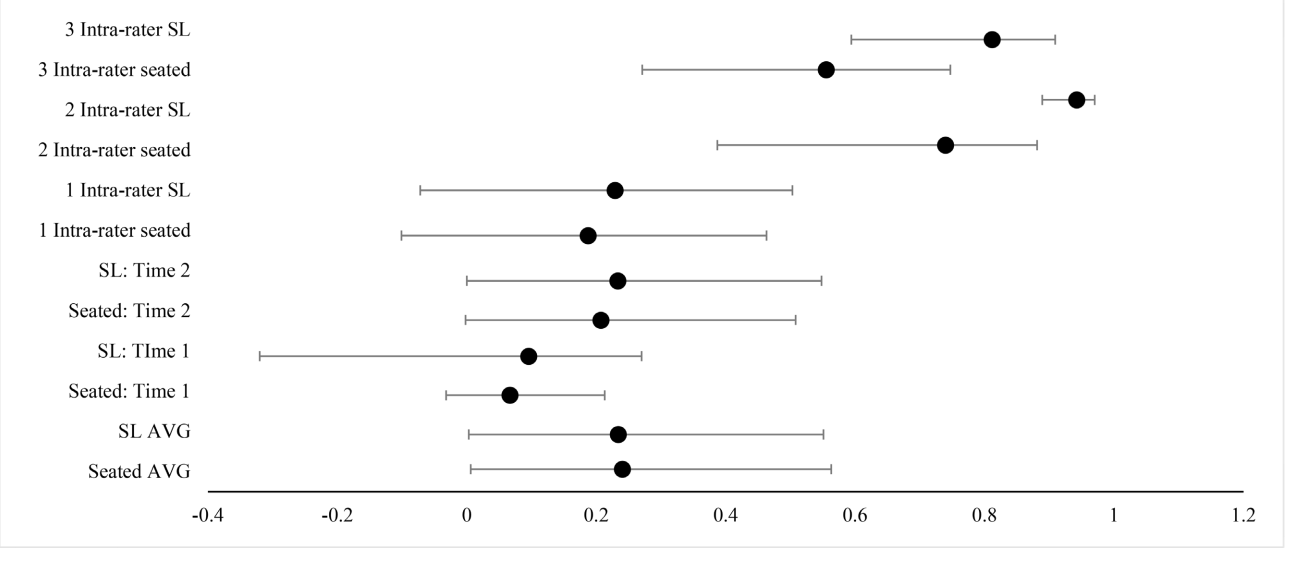

Figure 4. Comparison of intra-rater and inter-rater reliability for hip ER strength for the SL and seated MMT positions.

Inter-rater reliability ICC values were consistently low for both seated and SL positions, which indicates poor agreement between examiners ICC = 0.240 and ICC = 0.234, respectively. Inter-rater reliability was similar in the seated position for time 1 ICC = 0.066 (95% CI [−0.32, 0.213]) and time 2 ICC = 0.207 (95% CI [−0.002, 0.508]) when compared to SL time 1 ICC = 0.095 (95% CI [−0.32; 0.27]) and time 2 ICC = 0.233 (95% CI [0.000, 0.548]) (Figure 5). Although ICC values were similar in the seated position compared to the SL position, all estimates fell within the range of poor reliability.

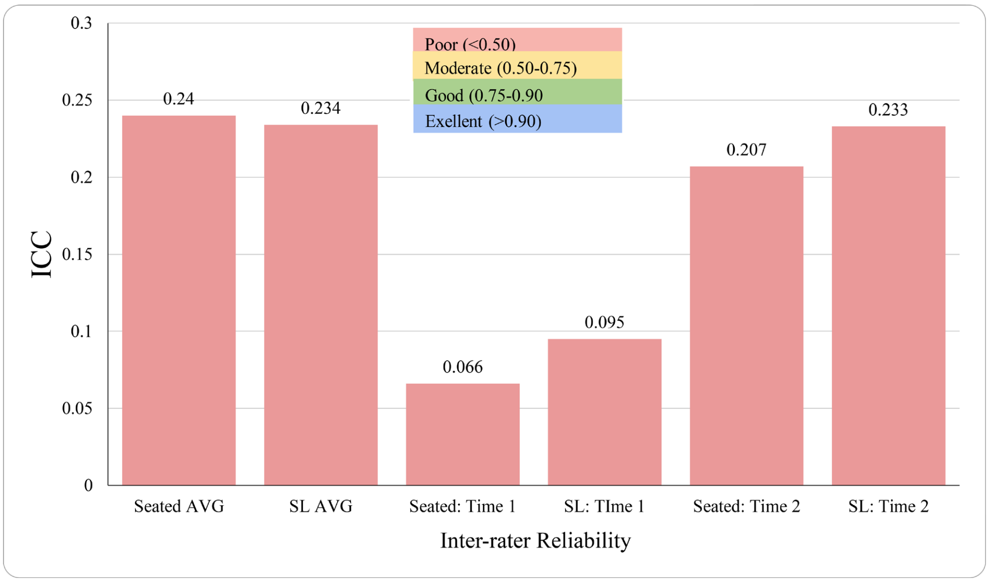

Figure 5. Intra-rater reliability among examiners for each testing position.

4. Discussion

Overall, these findings suggest that testing position influenced the intra-rater measurement of hip ER strength, indicating that positioning may affect the consistency of strength assessments. These results contribute to the understanding of how alternative testing positions may influence the reliability of the differing testing position. However, these results should be interpreted in light of several limitations. Additional methodological constraints also became apparent during data collection and statistical analysis involving both participant performance and examiner-related factors.

Primary limitations with participants included positional compensations observed in the seated hip ER testing position and insufficient protocol addressing preferred physical activity levels of participants between trials. When hip ER muscle strength was tested in the seated position, examiners noted that some participants slightly flexed their working hip as the test was initiated. This could have potentially impacted test results to yield higher force values, due to the additional muscle recruitment of sartorius and iliopsoas when the hips were at 90 degrees of flexion. A small portion of participants, noting performance of a "heavy leg day workout" within one to two days of their trial date, reported hip discomfort during their involvement in this study. Future studies may want to implement activity restrictions during the trial period to prevent participants from overusing their hip ER musculature or exerting force at high intensity between trial dates.

The primary issue during data collection for the examiners was the presence of wrist discomfort when performing the SL hip ER test on stronger participants (reported by two of the three examiners). The two examiners who reported wrist discomfort were also the two examiners who had the lowest intra-rater reliability.

Differences in the size and strength of the examiners may have contributed to the presence or absence of wrist discomfort, which may in turn have affected reliability.

When performing statistical analysis, researchers also noticed that only one examiner exhibited the most reliable results, failing to support the researchers' previous proposition that standardizing the table height would account for physical discrepancies between examiners, including height and mass. This finding is consistent with the claims that earlier analyses indicated perceived examiner differences, which may have contributed to initial limitations in reliability.

Despite attempts to control examiner variability in physical characteristics, these differences may have influenced measured strength values when testing hip ER in both sitting and SL. The discrepancies seen in reliability could have resulted from differences in each examiner's physical ability to provide break force. Additional studies to examine whether strength discrepancies between examiners can be controlled with a different method of standardization may be necessary to improve the reliability of MMT of hip ER in the SL position. Additionally, future studies may benefit from making alterations to the testing protocol. Researchers recognized that only allotting three trial days may have been insufficient, particularly due to the first trial day serving as both a first point of data collection and a learning session for each participant. A designated first day of participation is recommended for each participant to learn how to perform the testing protocol, without including these values in data analysis. Some participants demonstrated large improvements in strength between their first and second trial dates, and several participants verbalized feeling that they were better able to perform the protocol on their second and third trial days, due to learning it on the first trial day. Although participants were given one attempt to perform the technique before, final strength results were recorded on each of the three trial days, including an additional day for participants to practice a full sequence (without using that data in analysis), which may help reduce discrepancies caused by participants' initial difficulty in performing the protocol.

The present findings support that SL hip ER MMT is a viable alternative testing position with similar inter-rater reliability to the seated position, as well as potentially greater intra-rater reliability. Although examiner-related and methodological factors contributed to variability, the SL position demonstrated promise as a clinically relevant alternative that may better reflect functional hip external rotator performance when standardized procedures are implemented. These results underscore the need for greater protocol refinement, including examiner stabilization and participant familiarization to improve measurement consistency. Continued investigation into methods that minimize examiner influence and enhance standardization is warranted to strengthen the reliability and clinical utility of hip ER strength assessment.

5. Perspective

A reliable, functional test for hip external rotation (ER) strength is essential for improving orthopedic physical therapy evaluation and treatment. The side-lying (SL) position may better assess hip ER strength against gravity while reducing contributions from hip flexors such as the iliopsoas, sartorius, and rectus femoris. Because hip angle influences muscle function, testing position can affect strength measurements. Although most manual muscle tests are performed against gravity, hip ER is traditionally tested seated. The SL position allows for gravity-resisted testing, offering a more functional assessment of strength deficits. Hip dysfunction often presents as pain in the knee, ankle, or lumbar spine, making accurate strength assessment critical for targeted rehabilitation. In athletes, hip ER strength can help identify deficits and predict risk for injuries such as ACL tears. Consistent screening and strengthening of hip external rotators may reduce non-contact injury risk. Our results demonstrate no statistically significant difference in the reliability between the SL and seated methods. Therefore, given that the SL position may provide greater functional carryover by testing against gravity, it could serve as a practical alternative for assessing hip strength as it improves muscle isolation and reduces compensatory movements.

Acknowledgements

The authors gratefully acknowledge the support of Georgia State Physical Therapy for providing the facilities and resources necessary to perform this study. We extend our appreciation to Dr. Jacob Irwin, principal investigator, for his guidance, mentorship, and oversight throughout all stages of this study. The authors thank Dr. David Smelser for their thoughtful review and constructive comments on the methodology, which improved the quality of this manuscript.

Data Availability

The datasets generated and analyzed during the current study are available from the corresponding author on request.

Ethics Approval

The Georgia State University Institutional Review Board (IRB) reviewed and approved this study in accordance with 45 CFR 46.111 (IRB Number H23121).

Patient Consent

Written informed consent was obtained from all individual participants included in the study.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Bazett-Jones, D.M. and 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

[2] Soomro, R.R., Karimi, H. and Gillani, S.A. (2022) Reliability of Hand-Held Dynamometer in Measuring Gluteus Medius Isometric Muscle Strength in Healthy Population. *Pakistan Journal of Medical Sciences*, **38**, 1238-1242. https://doi.org/10.12669/pjms.38.5.5207

[3] Waiteman, M.C., Garcia, M.C., Briani, R.V., Norte, G., Glaviano, N.R., De Azevedo, F.M., *et al*. (2023) 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. *Journal of Orthopaedic & Sports Physical Therapy*, **53**, 655-672. https://doi.org/10.2519/jospt.2023.12045

[4] Byrne, A., Lodge, C. and Wallace, J. (2020) Intrarater Test-Retest Reliability of Hip Abduction, Internal Rotation, and External Rotation Strength Measurements in a Healthy Cohort Using a Handheld Dynamometer and a Portable Stabilization Device: A Pilot Study. *Archives of Rehabilitation Research and Clinical Translation*, **2**, Article ID: 100050. https://doi.org/10.1016/j.arrct.2020.100050

[5] Glaviano, N.R. and Bazett-Jones, D.M. (2021) The Influence of Sagittal Plane Hip Position on Lower-Extremity Muscle Activity and Torque Output. *Journal of Sport Rehabilitation*, **30**, 573-581. https://doi.org/10.1123/jsr.2020-0268

[6] Bloom, N. and Cornbleet, S.L. (2014) Hip Rotator Strength in Healthy Young Adults Measured in Hip Flexion and Extension by Using a Hand-Held Dynamometer. *PM&R*, **6**, 1137-1142. https://doi.org/10.1016/j.pmrj.2014.06.002

[7] Delp, S.L., Hess, W.E., Hungerford, D.S. and Jones, L.C. (1999) Variation of Rotation Moment Arms with Hip Flexion. *Journal of Biomechanics*, **32**, 493-501. https://doi.org/10.1016/s0021-9290(99)00032-9

[8] Yoo, T., Kim, S., Burland, J.P. and Glaviano, N.R. (2025) Reliability of Hand-Held Dynamometer in Measuring Gluteal Muscle Rate of Torque Development and Peak Torque: Push and Pull Configurations. *International Journal of Sports Physical Therapy*, **20**, 595-605. https://doi.org/10.26603/001c.133550

[9] Khayambashi, K., Ghoddosi, N., Straub, R.K. and Powers, C.M. (2016) Hip Muscle Strength Predicts Noncontact Anterior Cruciate Ligament Injury in Male and Female Athletes. *The American Journal of Sports Medicine*, **44**, 355-361. https://doi.org/10.1177/0363546515616237

[10] Xie, P., István, B. and Liang, M. (2022) The Relationship between Patellofemoral Pain Syndrome and Hip Biomechanics: A Systematic Review with Meta-Analysis. *Healthcare*, **11**, Article No. 99. https://doi.org/10.3390/healthcare11010099

[11] Reese, N.B. (2012) Muscle and Sensory Testing. 3rd Edition, Elsevier.

[12] Hislop, H.J., Avers, D., Brown, M. and Daniels, L. (2014) Daniels and Worthingham's Muscle Testing: Techniques of Manual Examination and Performance Testing. 9th Edition, Elsevier.

[13] Naqvi, U., Margetis, K. and Sherman, A.L. (2025) Muscle Strength Grading. StatPearls Publishing. http://www.ncbi.nlm.nih.gov/books/NBK436008/

[14] Conable, K.M. and Rosner, A.L. (2011) A Narrative Review of Manual Muscle Testing and Implications for Muscle Testing Research. *Journal of Chiropractic Medicine*, **10**, 157-165. https://doi.org/10.1016/j.jcm.2011.04.001

[15] Thorborg, K., Petersen, J., Magnusson, S.P. and Hölmich, P. (2010) Clinical Assessment of Hip Strength Using a Hand-Held Dynamometer Is Reliable. *Scandinavian Journal of Medicine & Science in Sports*, **20**, 493-501.

https://doi.org/10.1111/j.1600-0838.2009.00958.x

[16] McNabb, K., Sánchez, M.B., Selfe, J., Reeves, N.D. and Callaghan, M. (2024) Handheld Dynamometry: Validity and Reliability of Measuring Hip Joint Rate of Torque Development and Peak Torque. *PLOS ONE*, **19**, e0308956. https://doi.org/10.1371/journal.pone.0308956

[17] Koo, T.K. and Li, M.Y. (2016) A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, **15**, 155-163. https://doi.org/10.1016/j.jcm.2016.02.012