

Original Research

Reliability of Goniometric Techniques for Measuring Hip Flexor Length Using the Modified Thomas Test

Kira Eimiller^{1a}, Emma Stoddard¹, Briana Janes¹, Mason Smith¹, Andrew Vincek¹

¹ Physical Therapy, Daemen University

Keywords: Hip extension, hip flexor length, intra-rater reliability, inter-rater reliability, modified Thomas test, iliopsoas

<https://doi.org/10.26603/001c.120899>

International Journal of Sports Physical Therapy

Vol. 19, Issue 8, 2024

Background

The modified Thomas test (MTT) is commonly used to assess the flexibility of hip musculature, including the iliopsoas, rectus femoris, and tensor fascia latae. This measurement is important to include in a comprehensive musculoskeletal examination. However, existing research shows conflicting results regarding its reliability, particularly due to variations in controlling pelvic tilt during testing, which may lead to inaccurate measurements of hip extension when quantifying the test outcomes.

Purpose/Hypothesis

This study aimed to evaluate the intra- and inter-rater reliability of the Modified Thomas Test (MTT) in assessing hip flexor length using a goniometer. It was hypothesized that controlling for pelvic tilt would enhance the reliability of these measurements.

Study Design

Intra- and inter-rater reliability study

Methods

Sixty-four healthy individuals were recruited to participate in this study. The MTT was performed twice on each leg by both an experienced and a student physical therapist. Blinded goniometric measurements for hip extension range of motion (ROM) in the MTT position were taken with neutral pelvic tilt being enforced via palpation. A double-blind protocol was used where both examiners were unaware of each other's measurements and the goniometer was covered to blind the measuring therapist to the values as well. ROM values were entered into a Microsoft Excel spreadsheet and quantified using SPSS software. Statistical analysis included calculating Intraclass Correlation Coefficients (ICCs) and Standard Errors of Measurement (SEMs) using SPSS software.

Results

The study included 64 participants (mean age = 23.7 ± 4.34 years). The MTT demonstrated high intra-rater reliability (ICC = 0.911) and inter-rater reliability (ICC = 0.851). The SEMs indicated minimal variability around the mean scores. The average hip extension ROM measured was 5.43 ± 9.73 degrees.

^a Corresponding author:

Kira Eimiller, PT, DPT, Cert. MDT
Board Certified in Orthopedics
Physical Therapy Department, Daemen University, Amherst, NY 14226, USA
Telephone: (716) 930-5989
Fax: (716) 839-8537
Email: keimille@daemen.edu

Conclusion

These results suggest that the MTT is a reliable tool for assessing hip flexor length in clinical practice, particularly when pelvic tilt is controlled. These results have important implications for accurately testing orthopedic limitations that can contribute to low back, hip, and knee pain.

Level of Evidence

3

INTRODUCTION

The Modified Thomas Test (MTT) is an orthopedic assessment employed by clinicians to examine hip flexor length, specifically targeting the iliopsoas and rectus femoris muscles.^{1,2} Goniometers, commonly utilized in clinical settings, provide quantitative measurements in angular (degrees) units of motion for various joints in the upper and lower extremities.³ They can also be used to quantify outcomes of muscle length tests. In this study, “hip flexor length” refers to the overall assessment of the MTT, quantified by “hip extension ROM” measured using a goniometer during the MTT. Despite the generally recognized good-to-excellent reliability of goniometric measurements for muscle length in the lower extremity, a limitation reported is the need for two hands on the tool, posing challenges in stabilizing other body segments, particularly when measuring isolated hip range of motion that may be affected by lumbopelvic contributions.^{4,5}

Intra-rater reliability measures the consistency of one individual’s measurements, while inter-rater reliability assesses consistency between different individuals measuring the same phenomenon. The MTT can be used to determine hip flexor length using goniometric measurements of hip extension. The MTT can also be scored using a pass/fail method, where a “pass” indicates that the subject’s ROM meets or exceeds 0 degrees of hip extension in the testing position, and a “fail” signifies that the subject does not achieve this position and remains in varying degrees of hip flexion while in the testing position.

The MTT is used in clinical practice to assess hip flexor length in patients with conditions such as lower back pain, knee dysfunction, and hip pain.⁶⁻¹⁰ However, the limited evidence regarding the MTT has shown conflicting reliability due to confounding variables including the lack of pelvic stabilization and varying positions of the contralateral hip.¹ Researchers have found that uncontrolled pelvic tilt during the MTT measurement contributes to the overestimation of hip extension, leading to poor reliability.^{11,12} The purpose of this study was to evaluate the intra- and inter-rater reliability of the MTT to assess hip flexor length using a goniometer. It was hypothesized that controlling for pelvic tilt would result in high inter- and intra-rater reliability of these measurements.

METHODS

Subjects consisted of 64 volunteers (128 limbs). A power analysis estimation revealed that a sample size of approxi-

mately 93 limbs would provide a 95% confidence level when analyzing data at a $p < 0.05$ level of significance. Inclusion criteria was willingness to participate and being at least 20 years old. Exclusion criteria included recent surgery (within the prior six months) of the lumbar, hip, or knee region, recent physical trauma to the lumbar region or lower extremity (within the prior six months), or if the participant was under the care of a clinician for low back pain. Participants, primarily university students and staff, completed a detailed pre-participation questionnaire covering weekly activity level, height, weight, average sleep duration, dominant leg, prior injuries or surgeries, and back pain intensity (0-10 scale). This study adhered to ethical standards for human research and was approved by the Institutional Review Board (IRB). Informed consent was obtained from all individual participants included in the study.

The primary researcher was a physical therapist, board certified in Orthopedic physical therapy with 12 years of experience in an outpatient setting. The additional four researchers were Doctor of Physical Therapy (DPT) students in the final year of their program. Following completion of all required paperwork, participants proceeded to the first station for MTT measurements of hip flexor length.

PROCEDURES

Participants were instructed to sit at the edge of the mat table, pull one knee towards the chest and gradually roll back to the table. Subjects were instructed to allow the opposite leg to hang off the table. Once the participant was in supine on the mat table, the examiner assisted the participant in further flexing the hip not being measured with one hand while palpating under the lumbar spine for a neutral lumbopelvic tilt with the other. Neutral lumbopelvic tilt was operationally defined as the natural lordosis of the lumbar spine without excessive arching or flattening. Once neutral was found, the participant maintained their neutral position (confirmed by examiner palpation) for completion of the measurement. This allowed for maintaining a neutral lumbopelvic tilt and avoiding compensatory excessive lumbar lordosis during testing. Upon achieving the testing position, examiners utilized a standard plastic goniometer, positioning the fulcrum at the greater trochanter, the distal arm at the lateral midline of the femur, and the proximal arm at the lateral midline of the trunk. Stickers were placed on bony landmarks via palpation to encourage accuracy during measurements. Degrees of goniometric measurements were blinded by a piece of construction paper placed over the goniometer face to prevent the measurer from seeing the results.

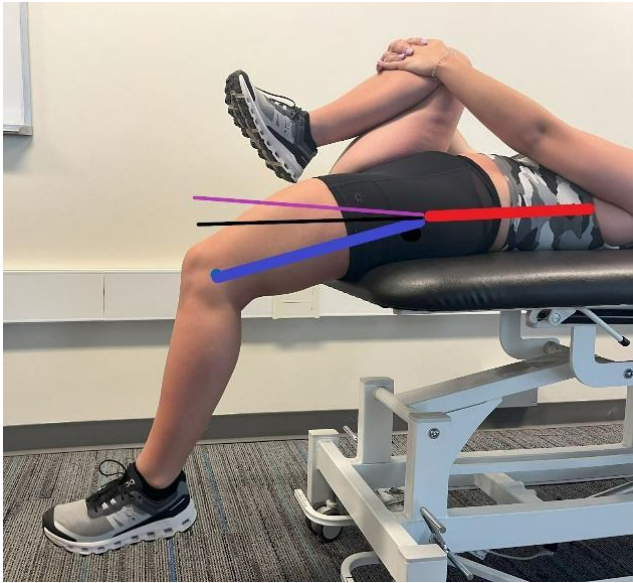


Figure 1. Modified Thomas Test with Pelvic Neutral

Red Line: Represents the stationary goniometer line at the lateral midline of the trunk, indicating pelvic neutral position. **Purple Line:** Demonstrates a tight hip flexor, with the thigh above 0 degrees of hip extension, indicating a failed MTT. **Black Line:** Indicates the pass line where the thigh is parallel to the table, showing 0 degrees of hip extension. **Blue Line:** Shows the placement of the moving arm of the goniometer, aligned with the lateral midline of the femur and fulcrum at the greater trochanter, also representing a passing test.

Each participant underwent a total of eight measurements: two trials on each leg by each of the two examiners at two separate stations. While one examiner aligned the goniometer, blinded to the measurement by a piece of construction paper, the other examiner removed the paper, read and recorded the goniometric measurement. The stations were separated by a curtain to ensure independent measurements without vocal communication. Intra-rater reliability was assessed by comparing the two measurements on the same leg taken by the same examiner at each station, while inter-rater reliability was determined by comparing measurements on the same legs taken by different examiners across the two stations.

In this study, hip flexor length was measured in the MTT position with hip extension ROM being recorded in degrees. Data were entered into a Microsoft Excel spreadsheet for initial organization and verification before being analyzed in SPSS.

STATISTICAL ANALYSIS

Statistical analysis was performed using SPSS software. ICCs were calculated to assess both intra-rater and inter-rater reliability. SEMs were also computed to quantify the variability around the mean scores across trials, providing a comprehensive understanding of measurement consistency and reliability. Descriptive statistics, including means and standard deviations, were calculated to summarize the central tendency and dispersion of the hip ROM data.

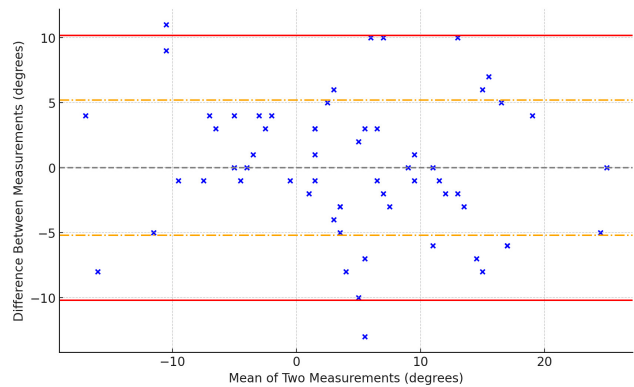


Figure 2. Bland-Altman Plot for Examiner 1 - Left Side

This plot shows the agreement between two goniometric measurements on the left side taken by Examiner 1. The blue dots represent the differences between measurements plotted against their mean. The gray dashed line indicates the mean difference between measurements. The red solid lines show the 95% limits of agreement (mean difference \pm 1.96 times the standard deviation), and the orange dash-dotted lines represent one standard deviation from the mean difference.

RESULTS

The convenience sample of 64 volunteers, primarily physical therapy students, was between 20 and 43 years of age, and comprised 42 females and 22 males. On average, participants were moderately active, engaging in physical activities for about 5.61 hours weekly (SD \pm 3.63). The average sleep duration was approximately 6.77 \pm 0.88 hours per night, indicating consistent sleep patterns. Additionally, the average Body Mass Index (BMI) was 25.55 \pm 3.66 kg/m². The average hip flexor length measured using the MTT among the participants was 5.43 \pm 9.73 degrees, suggesting a moderate level of variability across this sample population.

INTRA-RATER RELIABILITY

The intra-rater reliability was high, indicated by mean ICC's of 0.899 and 0.923, suggesting repeatable measurement outcomes and good levels of agreement across multiple trials by the same rater. (Figures 2-5, Table 1)

INTER-RATER RELIABILITY

High inter-rater reliability was also demonstrated, with mean ICC's ranging between 0.831-0.871. (Table 2)

Additionally, the SEMs for the hip flexor length using ROM data from the MTT further validated the precision of the measurements. The overall average SEM across both examiners and both sides (left and right) is 2.85 degrees. Overall, these results underscore the high reliability and measurement precision of the MTT, affirming its utility in both clinical and research settings for evaluating hip flexor length.

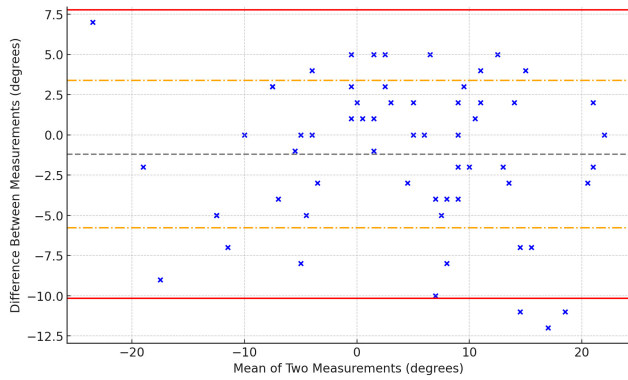


Figure 3. Bland-Altman Plot for Examiner 1 - Right Side

The blue dots represent the differences between measurements plotted against their mean. The gray dashed line indicates the mean difference between measurements. The red solid lines show the 95% limits of agreement (mean difference \pm 1.96 times the standard deviation), and the orange dash-dotted lines represent one standard deviation from the mean difference.

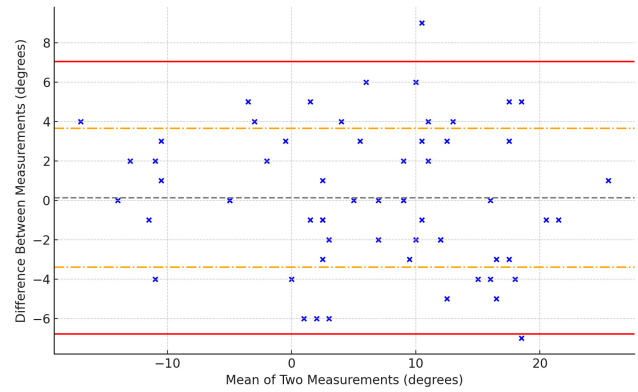


Figure 5. Bland-Altman Plot for Examiner 2 - Right Side

The blue dots represent the differences between measurements plotted against their mean. The gray dashed line indicates the mean difference between measurements. The red solid lines show the 95% limits of agreement (mean difference \pm 1.96 times the standard deviation), and the orange dash-dotted lines represent one standard deviation from the mean difference.

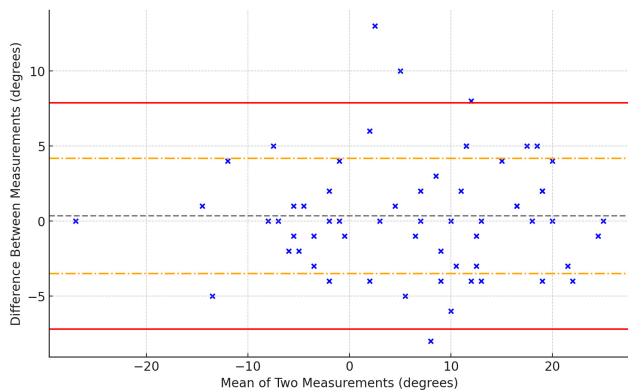


Figure 4. Bland-Altman Plot for Examiner 2 - Left Side

The blue dots represent the differences between measurements plotted against their mean. The gray dashed line indicates the mean difference between measurements. The red solid lines show the 95% limits of agreement (mean difference \pm 1.96 times the standard deviation), and the orange dash-dotted lines represent one standard deviation from the mean difference.

DISCUSSION

Although the MTT is widely used in orthopedic and physical therapy practice, its reference validity and measurement reliability amongst practitioners has been questioned. This study aimed to assess the reliability of the MTT, addressing inconsistencies in previous studies, particularly those with variations in controlling pelvic tilt during testing.^{2,12}

The results of the current study indicate strong intra- and inter rater reliability for the MTT when utilizing goniometric measurements. These findings align with those reported by Vigotsky et al.¹² and Kim and Ha,¹³ underscoring the increased reliability, specificity, and sensitivity of the MTT when accounting for lumbopelvic movement and controlling for pelvic tilt. Prior studies have consistently shown that pelvic tilt significantly affects the differences between MTT measurements of hip flexor length and standard hip extension goniometric measurements taken in the

prone position. This emphasizes the important role pelvic tilt plays in the relationship between hip muscle length and pelvic position, compared to measuring hip joint ROM.

Due to the lack of pelvic control in prior studies, results of the current study contradicted the findings of both Peeler and Anderson¹⁴ and Gabbe et al.¹⁵ who reported poor reliability for this test. However, Neves et al.⁸ suggested that positive results for shortening may also be influenced by an increase in the joint capsule and ligament stiffness, a factor not considered in this study. The improved reliability observed in the current study can likely be attributed to the control of pelvic tilt during the MTT measurements. By attempting to reach and maintain a neutral pelvic position, the potential for measurement error and variability was lessened, leading to more consistent results.

Maintaining a neutral pelvic tilt helps isolate the hip flexor muscles and provides a more accurate assessment of hip extension ROM.¹² Without controlling for pelvic tilt, compensatory movements in the lumbar spine and pelvis can occur, leading to overestimation or underestimation of actual hip extension.¹³ This discrepancy highlights the importance of standardized testing protocols that account for pelvic positioning to achieve reliable and valid measurements. Additionally, the use of anatomical landmark stickers may have improved accuracy of alignment during goniometer evaluation across data collection by different examiners for each participant. Outcomes of the current study offer valuable information to clinicians, emphasizing the importance of controlling pelvic tilt when performing the MTT, as pelvic tilt appears to contribute to variability in measures.

The current findings contrast with those of Watkins et al.,¹⁶ challenging the assertion that goniometric measurements display high reliability within the same therapist but lacked consistency between different therapists. The current results are consistent with the work of Clapis et al.¹⁷ demonstrating that goniometric measurement of hip flexor length using the MTT displayed both intra and inter-

rater reliability, surpassing that of an inclinometer based measurements. The superior intra-rater reliability observed may be attributed to several factors. Individual examiners tend to develop consistent personal techniques and methods when performing repeated measurements, which reduces variability and leads to higher intra-rater reliability. In contrast, inter-rater reliability involves comparing measurements between different examiners, each of whom may have slight variations in their technique or interpretation, despite standardized training and protocols. The small differences can lead to slightly lower inter-rater reliability compared to intra-rater reliability.

Moreover, the blinding technique used in the study, where the goniometric measurement was obscured by construction paper, helped minimize bias but did not entirely eliminate individual differences in measurement technique. Therefore, while the standardized protocols and training resulted in overall high reliability, intra-rater reliability was higher as each examiner was more consistent with their own methods compared to aligning perfectly with another examiner's methods. Clapis et al.¹⁷ emphasized the importance of consistency in measurement techniques, which inherently supports higher both inter- and intra-rater reliability. The current study's findings align with this, showing that when a single examiner performs repeated measurements, the consistency of their technique leads to more reliable results. However, despite these differences, the study still demonstrated high inter-rater reliability, indicating that standardized training and protocols were effective in achieving reliable measurements across different examiners.

This study's strengths include the inclusion of both sexes, minimizing potential bias through blinding, and involving examiners with varying levels of experience, which demonstrates the reliability of the MTT across different expertise levels.

Although the MTT demonstrated high reliability among examiners, it is crucial to acknowledge certain limitations inherent to the study. The assessment was conducted on a young, healthy population limiting the generalizability of the findings to broader age groups and diverse populations. There is a possibility of hip flexor stretching occurring after repeated measurements, as each participant underwent four goniometric measurements per leg, which could have influenced the hip flexor length and therefore measured ROM. Furthermore, the order in which subjects were tested was not randomized, which could have introduced an order effect, potentially influencing the reliability outcomes. Future studies should consider randomizing the order of testing to eliminate this potential bias.

CONCLUSION

The findings of this study demonstrate that the MTT shows strong inter- and intra-rater reliability when pelvic position is considered, aligning with the results of studies that have implemented similar controls. Previously reported poor reliability of the MTT in some studies may be due to the lack of control of pelvic position. These results support the use of the MTT as a reliable measure of hip flexor length in clinical practice when a neutral lumbar spine is maintained. It is important that physical therapists and medical professionals use reliable tests when assessing hip flexor length.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

Submitted: January 26, 2024 CDT, Accepted: June 17, 2024 CDT
© The Author(s)

Table 1. Intra-rater Reliability Scores for Modified Thomas Test Assessments

Intra-rater Scoring	Examiner 1	Examiner 2	Mean
Goniometer - Left Side	0.939	0.858	0.899
Goniometer - Right Side	0.940	0.907	0.923

Table 2. Comparison of Inter-rater Reliability Across Two Trials for Hip Flexor Length Measurement

Inter-rater Reliability	Trial 1	Trial 2	Mean (Goniometer)
Left Side	0.844	0.898	0.871
Right Side	0.785	0.877	0.831



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. Cady K, Powis M, Hopgood K. Intrarater and interrater reliability of the modified Thomas Test. *J Bodyw Mov Ther.* 2022;29:86-91. [doi:10.1016/j.jbmt.2021.09.014](https://doi.org/10.1016/j.jbmt.2021.09.014)
2. Peeler JD, Anderson JE. Reliability limits of the modified Thomas Test for assessing rectus femoris muscle flexibility about the knee joint. *J Athl Train.* 2008;43(5):470-476. [doi:10.4085/1062-6050-43.5.470](https://doi.org/10.4085/1062-6050-43.5.470)
3. Lea RD, Gerhardt JJ. Range-of-motion measurements. *J Bone Joint Surg Am.* 1995;77(5):784-798. [doi:10.2106/00004623-199505000-00017](https://doi.org/10.2106/00004623-199505000-00017)
4. Peeler J, Leiter J. Using digital photography to document rectus femoris flexibility: A reliability study of the modified Thomas Test. *Physiother Theory Pract.* 2012;29(4):319-327. [doi:10.3109/09593985.2012.731140](https://doi.org/10.3109/09593985.2012.731140)
5. Roach SM, San Juan JG, Suprak DN, Lyda M. Concurrent validity of digital inclinometer and universal goniometer in assessing passive hip mobility in healthy subjects. *Int J Sports Phys Ther.* 2013;8(5):680-688.
6. Borges NF, Borges BS, Sanchez EG, Sanchez HM. Correlation between anterior knee pain with flexibility muscles hip. *Man Ther Postural Rehabil J.* 2020;1-5. [doi:10.17784/mtprehabjournal.2016.14.408](https://doi.org/10.17784/mtprehabjournal.2016.14.408)
7. Roach SM, San Juan JG, Suprak DN, Lyda M, Boydston C. Patellofemoral pain in subjects exhibit decreased passive hip range of motion compared to controls. *Int J Sports Phys Ther.* 2014;9(4):468-475.
8. Neves RP, Oliveira D, Fanasca MA, Vechin FC. Shortening of hip flexor muscles and chronic low-back pain among resistance training practitioners: Applications of the modified Thomas Test. *Sport Sci Health.* Published online 2022. [doi:10.1007/s11332-022-00969-2](https://doi.org/10.1007/s11332-022-00969-2)
9. Kim WD, Shin D. Correlations between hip extension range of motion, hip extension asymmetry, and compensatory lumbar movement in patients with nonspecific chronic low back pain. *NMed Sci Monit.* 2020;26:e925080. [doi:10.12659/MSM.925080](https://doi.org/10.12659/MSM.925080)
10. Wu A, March L, Zheng X, et al. Global low back pain prevalence and years lived with disability from 1990 to 2017: estimates from the Global Burden of Disease Study 2017. *Ann Transl Med.* 2020;8(6):299. [doi:10.21037/atm.2020.02.175](https://doi.org/10.21037/atm.2020.02.175)
11. Nussbaumer S, Leunig M, Glatthorn JF, et al. Validity and test-retest reliability of manual goniometers for measuring passive hip range of motion in femoroacetabular impingement patients. *BMC Musculoskelet Disord.* 2010;11:194. [doi:10.1186/1471-2474-11-194](https://doi.org/10.1186/1471-2474-11-194)
12. Vigotsky AD, Lehman GJ, Beardsley C, Contreras B, Chung B, Feser EH. The modified Thomas test is not a valid measure of hip extension unless pelvic tilt is controlled. *PeerJ.* 4(2016):e2325. [doi:10.7717/peerj.2325](https://doi.org/10.7717/peerj.2325)
13. Kim GM, Ha SM. Reliability of the modified Thomas test using a lumbo-pelvic stabilization. *J Phys Ther Sci.* 2015;27:447-449. [doi:10.1589/jpts.27.447](https://doi.org/10.1589/jpts.27.447)
14. Peeler J, Anderson JE. Reliability of the Thomas test for assessing range of motion about the hip. *Phys Ther Sport.* 2007;8(1):14-21. [doi:10.1016/j.ptsp.2006.09.023](https://doi.org/10.1016/j.ptsp.2006.09.023)
15. Gabbe B. Reliability of common lower extremity musculoskeletal screening tests. *Phys Ther Sport.* 2004;5(2):90-97. [doi:10.1016/s1466-853x\(04\)00022-7](https://doi.org/10.1016/s1466-853x(04)00022-7)
16. Watkins MA, Riddle DL, Lamb RL, Personius WJ. Reliability of goniometric measurements and visual estimates of knee range of motion obtained in a clinical setting. *Phys Ther.* 1991;71(2):90-96. [doi:10.1093/ptj/71.2.90](https://doi.org/10.1093/ptj/71.2.90)
17. Clapis PA, Davis SM, Davis RO. Reliability of inclinometer and goniometric measurements of hip extension flexibility using the modified Thomas test. *Physiother Theory Pract.* 2008;24(2):135-141. [doi:10.1080/09593980701378256](https://doi.org/10.1080/09593980701378256)