Intrarater and Interrater Reliability of First Metatarsophalangeal Joint Dorsiflexion

Goniometry versus Visual Estimation

Angela M. Jones, BSc(Hons)*
Sarah A. Curran, PhD, BSc(Hons)*

Background: Visual estimation (VE) and goniometric measurement (GM) are commonly used to assess first metatarsophalangeal joint dorsiflexion. The purposes of this study were to determine the intrarater and interrater reliability of VE and GM and to establish whether reliability was influenced by the experience of the examiner.

Methods: Ten experienced and ten inexperienced examiners evaluated three real-size photographs of a first metatarsophalangeal joint positioned in various degrees of dorsiflexion on two separate occasions.

Results: Experienced examiners demonstrated excellent intrarater and interrater reliability for GM (intraclass correlation coefficient [ICC], >0.953; standard error of measurement [SEM], 1.8°–2.5°) compared with inexperienced examiners, who showed fair-to-good intrarater and interrater reliability (ICC, 0.322–0.597; SEM, 2.0°–3.0°). For VE, inexperienced examiners demonstrated fair-to-good interrater and excellent intrarater reliability (ICC, 0.666–0.808), which was higher compared with experienced examiners (ICC, 0.167–0.672). The SEM (2.8°–4.4°) was less varied than that of experienced examiners (SEM, 3.8°–6.4°) for VE, but neither group's SEMs were clinically acceptable.

Conclusions: Although minimal differences between intrarater and interrater reliability of GM and VE are noted, this study suggests that GM is more reliable than VE is when used by experienced examiners. These findings support the continued use of GM for first metatarsophalangeal joint dorsiflexion assessment. (J Am Podiatr Med Assoc 102(4): 290-298, 2012)

Adequate dorsiflexion at the first metatarsophalangeal joint is an important characteristic that facilitates normal, efficient forward progression of the body during gait.[1-3] Assessment of dorsiflexion at this joint, therefore, forms an essential part of a biomechanical examination of the lower limb.[3-6] This information is used to establish baseline measurements, exchange information between colleagues, evaluate the efficacy of medical intervention, and classify hallux rigidus and hallux limitus.[7,8] Dorsiflexion at the first metatarsophalangeal joint is measured in degrees, which is a ratio scale measurement where there is always a nonarbitrary zero value.[9,10] The measurement is used for obtaining the magnitude of a quantity.[6] The reliability of the measurement is the extent to which the result is reproducible either when repeated by a rater (intrarater reliability) or between different raters (interrater reliability).[5,11-13]

The degree of dorsiflexion at the first metatarsophalangeal joint is usually measured statically, with either passive dorsiflexion of the hallux applied by the clinician in a nonweightbearing position or active dorsiflexion as the patient dorsiflexes the hallux in a standing position.[10] Average active and passive dorsiflexion measurements have been reported to be 70° to 90° from a neutral 0° position.[14,15] In contrast, three-dimensional electromagnetic tracking devices allowed Nawoczenski et al[16] to record dynamic function. They reported a mean of 42° of dorsiflexion at the joint during normal walking. The present study however, did not measure motion at the first metatarsophalangeal joint during more demanding dynamic activities,

*Wales Centre for Podiatric Studies, Cardiff Metropolitan University, Cardiff, Wales.

Corresponding author: Sarah A. Curran, PhD, BSc(Hons), Wales Centre for Podiatric Studies, Cardiff Metropolitan University, Cardiff, Western Avenue, Cardiff, CF5 2YB, Wales. (E-mail: scurran@cardiffmet.ac.uk)

such as running or jumping, which may require a greater degree of dorsiflexion.

The gold standard tool for measuring the angle of dorsiflexion at the first metatarsophalangeal joint is radiography.[14,17] Radiographs allow direct visualization of the angle created by longitudinal bisection of the medial midline of the first metatarsal and the proximal phalanx of the hallux, with the first metatarsophalangeal joint acting as the fulcrum.[4,18-20] In contrast, podiatric physicians and other health professionals use clinical methods to establish these values, including goniometric measurement (GM) and visual estimation (VE). These methods require identification of bony surface landmarks of the foot to estimate the previously mentioned midlines. Goniometric measurement is considered by Gajdosik and Bohannon[21] to be a reliable method of assessing joint range of motion. However, they acknowledge that a variety of methodological issues influence its reliability, including joint position, identification of the bony landmarks, and incorrect application of the goniometer. By emphasizing the need for clinicians to adopt standardized methods of testing,[20] reliability improves, and the goniometer remains the preferred instrument for measuring range of motion during clinical assessment. However, in clinical practice, VE of the first metatarsophalangeal joint is frequently used in lieu of a goniometer. These measurements influence clinical decision making and, therefore, must be reliable if they are to have any clinical significance.

Previous studies[22-25] examining the assessment of joint angles suggest that GM is more reliable than is VE. Watkins et al[24] concluded that GMs taken by the same therapist were more reliable than those taken by different therapists. Although Somers et al[26] found VE to be more reliable than GM was in forefoot position assessments, the trial is limited owing to a small sample size (four testers and ten healthy individuals). In terms of experience, Rose et al[23] noted improved reliability with greater tester experience. This finding is, however, in contrast to that of Bruton et al[22] and Somers et al,[26] who found no statistically significant differences related to tester experience. Although Bruton et al[22] conducted the only trial to use a standardized protocol for recording GMs, their work shares similarities with other studies by failing to use a protocol for VE. They did, however, conclude that further investigation using a protocol for VE is needed. The aims of this study were to establish the interrater and intrarater reliability of VE and GM when used to assess first metatarsophalangeal joint dorsiflexion



Figure 1. Clinical photographs (scaled down) of the medial aspect of the first metatarsophalangeal joint (right foot) showing dorsiflexion of approximately 0° to 15° (A), 15° to 55° (B), and 55° to 70° (C).

Table 1. Description of the Norkin and White[20] Recommended Testing Position

	Description
Extension	Position the patient sitting, with the ankle and foot at 0° of dorsiflexion, plantarflexion, inversion, and eversion. The metatarsophalangeal joint should be at 0° of abduction and adduction. The IP joints should also be positioned at 0° of flexion and extension.
	Note: If the ankle is dorsiflexed and the IP joints of the toe being tested are extended, tension in the flexor hallucis longus or flexor digitorum longus muscle will restrict motion. If the IP joints of the toe being tested are in extreme flexion, tension in the lumbricals and interosseus muscles may restrict the motion.
Stabilization	Stabilize the metatarsal to prevent dorsiflexion of the ankle and eversion of the foot. Do not hold the metatarsophalangeal joints of the other toes in extreme flexion as tension in the transverse metatarsal ligament will restrict the motion.

Abbreviation: IP, interphalangeal.

and to determine the influence of experience on the reliability of VE and GM when used to assess this joint.

Methods

Twenty examiners were selected by convenience sampling from the Wales Centre for Podiatric Studies, Cardiff Metropolitan University. The sample consisted of ten members of staff with greater than 5 years of podiatric medical experience (experienced examiners) (mean ± SD experience, 23.4 ± 14.0 years; range, 7–40 years) and ten third-year students (inexperienced examiners) (mean ± SD experience, 2.2 ± 0.0 years; range, 2.2–2.2 years). All of the examiners met the inclusion and exclusion criteria according to their degree of experience and availability for the trial. Ethical approval was granted by the Cardiff School of Health Care and Psychology Ethics Panel, Cardiff Metropolitan University. Three real-size clinical photographs (Fig. 1) were produced of one right foot from a healthy individual (a 37-year-old woman) who had no obvious hallux rigidus, hallux abducto valgus, or hyperextension at the interphalangeal joint; no previous surgery; and no discernable features such as skin or nail lesions. The photographs demonstrated dorsiflexion of the first metatarsophalangeal joint approximating 55° to 70° (adequate dorsiflexion for propulsive gait), 15° to 55° (representing intermediate hallux limitus), and 0° to 15° (representing hallux rigidus).[10] Photographs were used rather than live patients to avoid any movement or bias during assessment arising from variations of the examiner obtaining the same foot position or determining the end of range of motion.[20]

The Norkin and White[20] recommended testing position for first metatarsophalangeal joint dorsiflexion was used while taking the photographs (Table 1). This provided a standardized protocol for not only the GMs but also VE, as suggested by Bruton et al.[22] To ensure that the photographed positions represented the three defined categories, baseline measurements were established by an impartial staff member (9 years of experience) but were not made known until trial end. Using the recommended testing position, a task protocol was developed to provide the examiner with clear task instructions (Fig. 2).

A 12-cm plastic goniometer with moveable distal and proximal arms and measurement increments of 2° was used (Nova tractograph; Nova Medical Inc, Wilmington, Massachusetts) (Fig. 3). The measurement increments were blinded with a card on the side facing the examiners so that they could not read their measurement and, thus, bias further measurements. The same goniometer was used for all of the measurements in this study. A pilot study performed before the trial revealed successful easy use of the photographs, but the task protocol was found to be difficult to follow. Easy-to-follow diagrams were added to illustrate the measurement protocol.

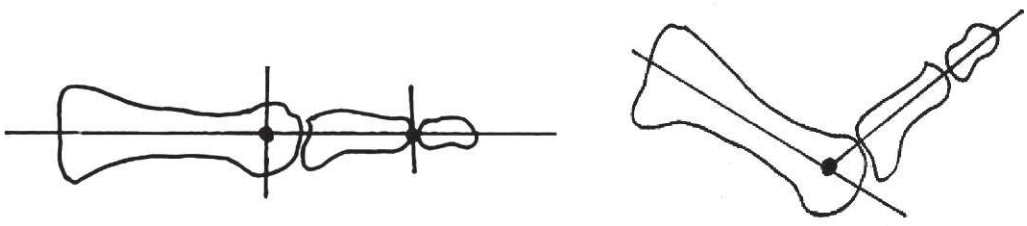
Procedure

The examiner randomly selected a photograph and, in accordance with the task protocol, visually estimated the degree of dorsiflexion at the first metatarsophalangeal joint. To allow comparison of the two assessment methods, they were asked to use the blinded goniometer to measure the same angle on the same photograph. Once the examiner had satisfactorily placed the proximal and distal arms of the blinded goniometer, the assessor read the measurement from the unblinded side, recorded the data, and returned the goniometer to 0° before

Please read the following protocol for performing the research task:

1. Visual Estimation

- For each of the photographs labeled A, B, and C, estimate the degree of dorsiflexion at the first MTP joint using the medial midline of the first metatarsal and the medial midline of the proximal phalanx of the hallux as your reference points. The following illustrations will guide you.
- Give your answer for each photograph to the assessor.



2. Goniometric Measurement

- For each of the photographs labeled A, B, and C and in conjunction with the goniometry protocol outlined at the bottom of this page (section 3), use the goniometer to measure the degree of first MTP joint dorsiflexion.
- Once you have positioned the arms of the goniometer pass it to the assessor.
- The assessor will read and record your result, reposition the goniometer to 0°, and ask you to repeat the exercise three times.

DO NOT ATTEMPT TO READ THE GONIOMETER MEASUREMENT YOURSELF AS THIS WILL INVALIDATE THE TEST

3. Goniometry Protocol: Alignment for first MTP joint

a) Center the fulcrum of the goniometer over the medial aspect of the first MTP joint.
b) Align the proximal arm with the medial midline of the first metatarsal.
c) Align the distal arm with the medial midline of the proximal phalanx of the hallux.
d) The order of each photograph will be randomized.

Figure 2. Task protocol. MTP indicates metatarsophalangeal.

further measurements. One VE and three GMs were taken for each of the three photographs, permitting interrater reliability assessment. To prevent the examiners from being influenced by their GMs, they were asked to complete their VE first. The mean of three GMs was calculated for each position. Two separate testing sessions were conducted 14 days apart to assess intrarater reliability. The examiners were blinded to their measurements to eliminate bias. All of the measurements were recorded to the nearest whole degree. The effect of clinical experience was measured by selecting inexperienced and experienced examiners according to the inclusion and exclusion criteria.

A commercially available software program (SPSS, version 17 for Windows; SPSS Inc, Chicago, Illinois) was used to analyze the data produced. Initial analysis of intrarater and interrater reliability was summarized by the following descriptive statistics: mean, standard deviation (SD), and range

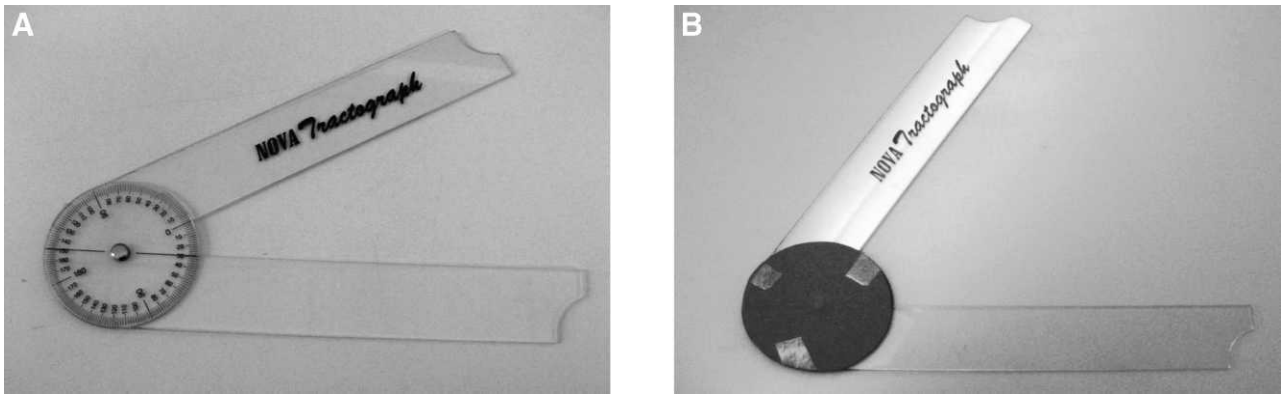


Figure 3. Goniometer unblinded (A) and blinded (B).

in degrees. Intraclass correlation coefficients (ICCs) were also calculated to analyze intrarater and interrater reliability. Bench marks endorsed by Fleiss[27] were used to interpret ICC values: greater than 0.75 indicates excellent; 0.4 to 0.75, fair-to-good reliability; and less than 0.4, poor reliability. The standard error of measurement (SEM; in degrees [statistical equivalent to the SD]) and 95% confidence intervals were calculated to provide an estimate of the amount of error associated with the measurement.

Results

Table 2 illustrates the descriptive statistics and the SEM of each of the measurements performed. The SEM had a wider range of 2.5° to 5.2° for VE across both groups and sessions compared with a narrower range of 1.8° to 3.0° for GMs. Experienced examiners had greater variation in the VE SEM (3.8°–6.4°) compared with inexperienced examiners (2.8°–4.4°); however, the variation was marginally less for their GMs (1.8°–2.5°) compared with inexperienced examiners (SEM, 2.0°–3.0°).

Visual Estimation

The results for VE reliability are summarized in Table 3. Interrater and intrarater reliability for VE for inexperienced examiners ranged from fair-to-good to excellent (ICCs, 0.666–0.794 and 0.673–0.808, respectively). Interrater and intrarater reliability for VE for experienced examiners ranged from poor to fair-to-good (ICCs, 0.187–0.672 and 0.167–0.643, respectively). Inexperienced examiners demonstrated better reliability for VE than did experienced examiners (Fig. 4).

Goniometric Measurement

Experienced examiners demonstrated excellent interrater and intrarater reliability for GM, with ICCs greater than 0.953 (Table 4). However, inexperienced examiners were not so reliable, with their GMs achieving ICCs of 0.322 to 0.597,

Table 2. Descriptive Statistics and SEM of VE and GM

	Session 1				Session 2	
	Inexperienced Examiner		Experienced Examiner		Inexperienced Examiner	
	Mean ± SD (Range)	SEM	Mean ± SD (Range)	SEM	Mean ± SD (Range)	SEM
VE image A	8.9 ± 8.8 (0–30)	2.8	15.6 ± 17.9 (3–60)	5.7	11.3 ± 8 (2–30)	2.5
VE image B	28.1 ± 13.7 (10–45)	4.3	29.3 ± 20.2 (10–80)	6.4	32 ± 11.4 (10–45)	3.6
VE image C	72 ± 13 (50–90)	4.1	67.2 ± 16.4 (45–90)	5.2	75.6 ± 4.1 (50–90)	4.4
GM image A	20.3 ± 8.6 (8.3–31)	2.7	20.7 ± 6.6 (10.3–29.3)	2.1	20.8 ± 7 (10.7–30.7)	2.2
GM image B	39 ± 6.4 (30.3–50)	2.0	34.5 ± 6.42 (23.7–43.7)	2.0	36.4 ± 7.25 (26–48)	2.3
GM image C	80.7 ± 9.3 (74.3–92.7)	3.0	74.8 ± 6.6 (63–82)	2.1	75.8 ± 6.9 (66.7–83.7)	2.1

Note: Data are given in degrees.
Abbreviations: GM, goniometric measurement; SEM, standard error of measurement; VE, visual estimation.

Table 3. Visual Estimation Reliability

	Interrater Reliability (ICC [95% CI])		Intrarater Reliability (ICC [95% CI])	
	Inexperienced Examiner	Experienced Examiner	Inexperienced Examiner	Experienced Examiner
Image A	0.666 (0.138 to 0.904)	0.672 (0.148 to 0.906)	0.673 (0.119 to 0.907)	0.643 (0.066 to 0.897)
Image B	0.671 (0.147 to 0.905)	0.588 (0.009 to 0.877)	0.692 (0.154 to 0.914)	0.553 (−0.074 to 0.866)
Image C	0.794 (0.394 to 0.944)	0.187 (−1.587 to 0.705)	0.808 (0.401 to 0.949)	0.167 (−0.484 to 0.699)

Abbreviations: CI, confidence interval; ICC, intraclass correlation coefficient.

indicating poor to fair-to-good intrarater and interrater reliability. Figure 5 illustrates the increased reliability of the GMs taken by experienced examiners compared with those of inexperienced examiners.

Discussion

Visual estimation is often used as a clinical assessment method for joint range of motion despite research that suggests it is less reliable than is GM. The aims of this study were to establish the intrarater and interrater reliability of VE and GM when used to assess first metatarsophalangeal joint dorsiflexion and to determine the influence of experience on the two methods of assessment.

The excellent intrarater and interrater and reliability results achieved by the experienced examiners for the GMs in this study support the findings of Bruton et al[22] and Watkins et al[24] and indicate that GM is more reliable than is VE when assessing joint angles. The results also suggest that GM reliability is excellent for repeated assessment and for differing examiners when used to assess the first metatarsophalangeal joint. Conversely, the study suggests that experienced examiners should expect poor to fair-to-good reliability when using VE as an assessment method for the same joint.

Table 2. extended

Session 2	
Experienced Examiner	
Mean ± SD (Range)	SEM
14.9 ± 14.3 (0–45)	4.5
29.8 ± 12.2 (8–45)	3.8
72.3 ± 16 (48–90)	5.1
19.6 ± 7.7 (7.7–28.7)	2.5
36 ± 5.8 (26–44.3)	1.8
75.8 ± 6.2 (66–81.3)	2.0

Menz et al[28] state that as a result of variations in the application and interpretation of a clinical test, interrater reliability of clinical measurements is often lower than intrarater reliability. This is supported by the study by Watkins et al,[24] who concluded that the reliability of the same therapist performing the assessment was better than that of differing therapists performing the assessment. This has not been the finding of this study as there is no difference between the intrarater and interrater reliability results for VE (Table 3) or GM (Table 4) regardless of examiner experience. This finding may be attributed to the clinical photographs and the task protocol (Fig. 2), which could have reduced some of the variation in test application and interpretation. Nevertheless, the degree of clinician

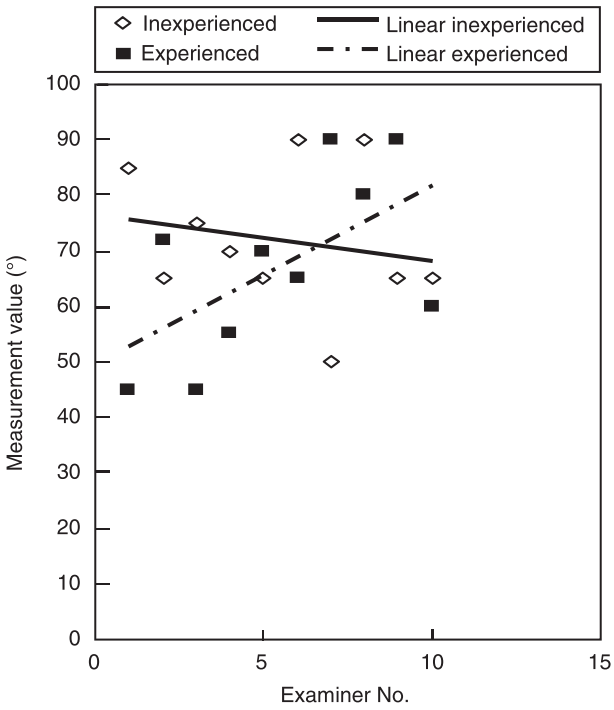


Figure 4. Visual estimation interrater reliability for image C (session 1).

Table 4. Goniometric Measurement Reliability

	Interrater Reliability (ICC [95% CI])		Intrarater Reliability (ICC [95% CI])	
	Inexperienced Examiner	Experienced Examiner	Inexperienced Examiner	Experienced Examiner
Image A	0.395 (−0.242 to 0.803)	0.827 (0.472 to 0.953)	0.351 (−0.318 to 0.787)	0.851 (0.509 to 0.961)
Image B	0.572 (−0.014 to 0.871)	0.951 (0.825 to 0.987)	0.597 (−0.007 to 0.882)	0.975 (0.905 to 0.994)
Image C	0.322 (−0.319 to 0.771)	0.841 (0.509 to 0.957)	0.404 (−0.262 to 0.809)	0.829 (0.452 to 0.955)

Abbreviations: CI, confidence interval; ICC, intraclass correlation coefficient.

experience seems to have influenced the reliability of their results in this study.

It was unsurprising that experienced examiners (mean experience, 23.4 years) were more reliable (excellent reliability) at GM than were inexperienced examiners, who achieved poor to fair-to-good interrater and intrarater reliability. In addition, most of the experienced examiners' SEM values for GM were clinically acceptable (1.8°–2.5°) compared with an increased and slightly more varied SEM (2.0°–3.0°) for nonexperienced examiners. However, inconsistencies arose when the inexperienced examiners returned more reliable VE results (fair-to-good to excellent) compared with their GM results (poor to fair-to-good) for interrater and intrarater reliability (note that their SEMs for VE illustrated a high variation, 2.5° to 4.4° across both sessions, but was closer to being clinically acceptable for GM [2.0°–3.0°]). This finding is in line with the conclusion of Somers et al,[26] who also found inexperienced examiners to be more reliable at VE than at GM. It is, perhaps, not surprising since the students most commonly use VE during their clinical practice at the university. The finding could also be attributed to recent student learning, causing them to inadvertently match the three angulations with categories suggested by Camasta,[10] which were used to produce the images. These assumptions are also supported by the study by Rose et al,[23] who noted that regular goniometer users were less accurate at VE. The influence of experience was also noted in the study by Rose et al,[23] who found that surgeons with a mean of 14.6 years of experience were more reliable at VE than were therapists with a mean of 4.5 years of experience. Studies performed by Bruton et al[22] and Somers et al[26] did not find any significant differences in characteristics such as experience. In contrast, Somers et al[26] provided training for their inexperienced observers (students) and not for their experienced observers (>10 years of experience).

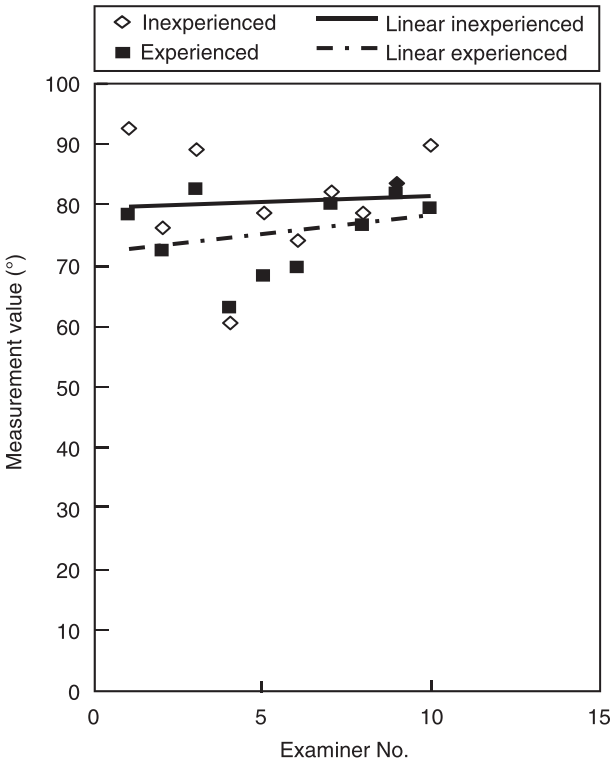


Figure 5. Goniometric measurement interrater reliability for image C (session 1).

Limitations of the study design may have contributed to the results observed in this study. The card used to blank the degrees on the goniometer covered its pivot, making it more difficult for the examiners to place it directly over the fulcrum of the joint. Menadue et al[12] used a similar method and cut a smaller hole in the center of the card that revealed the pivot and solved the problem. There also seemed to be a degree of learning by the examiners as they performed the task for the first time. This learning was demonstrated by hesitance and lack of confidence in task performance during session 1 that was not as evident in session 2. This was illustrated by a reduction of the SEM in session 2 for ten of the 12 tests. In view of this finding, the validity of the results may have been further improved by including a third session after another 2 weeks and disregarding the data from session 1.

This study used clinical photographs to eliminate

any movement or bias during assessment. Youdas et al[25] reported therapists' variability in locating such landmarks to which they attributed much of the unreliability demonstrated in their study. However, the use of photographs limited the ability of the examiner to identify bony landmarks vital to perform the assessment of dorsiflexion at the first metatarsophalangeal joint as outlined by Norkin and White,[20] through lack of palpation and 3-dimensional observations. This limitation may have compromised the study's clinical relevance. The most significant limitation of this study was the small sample size, which was governed by the timeframe of the data collection period and the availability of the experienced and inexperienced examiners of the sample.

Consistent with the findings of previous studies,[22,24] this study showed some clear indications that GM is more reliable than is VE and that greater experience of the examiner improves the reliability of GM. To further elucidate the usefulness and clinical relevance of the results, a larger sample of live participants could be used in a clinical setting. Future research could include the use of parallel-forms reliability, as demonstrated by Watkins et al,[22,24] to establish the interchangeability of GM and VE. Finally, this study investigated only the reliability of VE and GM and did not address the issue of the validity of the two assessment methods raised by Beeson et al,[29] which should be considered for future research.

Conclusions

Assessment of first metatarsophalangeal joint dorsiflexion forms an integral component of a podiatric medical biomechanical evaluation of the foot. For the purposes outlined at the beginning of the present article, an objective numerical value must be provided for the portion of dorsiflexion at this joint. The results of this research suggest that in line with similar reliability studies of other joints, GM is more reliable than is VE when performed by experienced clinicians in conjunction with a standardized task protocol. Contrary to some previous research, perhaps as a result of the standardized task protocol, this study found no difference between interrater and intrarater reliability. This study indicated that the reliability of VE and GM was influenced by the experience of the examiner, with excellent reliability demonstrated by experienced examiners for GM. Inexperienced examiners were more reliable than were experienced examiners at VE; however, most of their ICCs suggested only fair-to-good reliability. This study, therefore, concludes that VE is not as reliable as is GM to evaluate the degree of dorsiflexion at the first metatarsophalangeal joint and, therefore, supports the recommendation that goniometers should be used for this assessment. Further studies using patients with hallux rigidus and hallux limitus to examine the reliability of the two methods in conjunction with a standardized task protocol would increase the clinical application of the results.

Financial Disclosure: None reported.
Conflict of Interest: None reported.

References

1. Perry J: *Gait Analysis: Normal and Pathological Function*, Yeong Mun Publishing Co, Seoul, 1992.
2. Root M, Weed J, Orien W: *Clinical Biomechanics: Normal and Abnormal Function of the Foot*, Vol 2, Clinical Biomechanics Corp, Los Angeles, 1977.
3. Hetherington V, Carnett J, Patterson B: Motion of the first metatarsophalangeal joint. J Foot Surg **28:** 13, 1989.
4. Coughlin M, Shurna P: Hallux rigidus: grading and long-term results of operative treatment. J Bone Joint Surg Am **85:** 2072, 2003.
5. Gastwirth B: "Biomechanical Examination of the Foot and Lower Extremity," in *Clinical Biomechanics*, edited by R Valmassy, Mosby, St. Louis, 1996.
6. Vanore J, Christenesn J, Kravitz S, et al: Clinical Practice Guideline First Metatarsophalangeal Joint Disorders Panel of the American College of Foot and Ankle Surgeons: Diagnosis and treatment of first metatarsophalangeal joint disorders: section 2. Hallux rigidus. J Foot Ankle Surg **42:** 124, 2003.
7. Roukis T, Scherer P, Anderson C: Position of the first ray and motion of the first metatarsophalangeal joint. JAPMA **86:** 538, 1996.
8. Shereff M, Baumhauer J: Hallux rigidus and osteoarthritis of the first metatarsophalangeal joint. J Bone Joint Surg Am **80:** 898, 1998.
9. Roukis T, Jacobs P, Dawson DM, et al: A prospective comparison of clinical, radiographic, and intraoperative features of hallux rigidus. J Foot Ankle Surg **41:**76, 2002.
10. Camasta C: Hallux limitus and hallux rigidus: clinical examinations, radiographic findings, and natural history. Clin Podiatr Med Surg **13:** 423, 1996.
11. Martin RL, McPoil TG: Reliability of ankle goniometric measurements: a literature review. JAPMA **95:** 564, 2005.
12. Menadue C, Raymond J, Kilbreath S, et al: Reliability of two goniometric methods of measuring active inversion and eversion range of motion at the ankle. BMC Musculoskelet Disord **7:** 60, 2006.
13. Van Gheluwe B, Kirby K, Roosen P, et al: Reliability and

accuracy of biomechanical measurements of the lower extremities. JAPMA **92:** 317, 2002.
14. Buell T, Green D, Risser J: Measurement of the first metatarosphalangeal joint range of motion. JAPMA **78:** 439, 1988.
15. Chang T: Stepwise approach to hallux limitus: a surgical perspective. Clin Podiatr Med Surg **13:** 449, 1996.
16. Nawoczenski D, Baumhauer J, Umberger B: Relationship between clinical measurements and motion of the first metatarsophalangeal joint during gait. J Bone Joint Surg Am **81:** 370, 1999.
17. Menz HB, Munteanu S: Radiographic validation of the Manchester scale for the classification of hallux valgus deformity. Rheumatology **44:** 1061, 2005.
18. Evans J: "Diagnostic Imaging," in *Neale's Disorders of the Foot*, edited by D Lorimer, G French, M O'Donnell, et al, 7th Ed, Churchill Livingstone, Philadelphia, 2006.
19. Gilheany M, Landorf K, Robinson P: Hallux valgus and hallux rigidus: a comparision of impact on health-related quality of life in patients presenting to foot surgeons in Australia. J Foot Ankle Res **1:** 14, 2008.
20. Norkin C, White D: *Measurement of Joint Motion: A Guide to Goniometry*, 2nd Ed, FA Davis Co, Philadelphia, 1995.
21. Gajdosik R, Bohannon R: Clinical measurements of range of motion: review of goniometry emphasizing reliability and validity. Phys Ther **67:** 1867, 1987.
22. Bruton A, Ellis B, Goddard J: Comparison of visual estimation and goniometry for assessment of metacarpophalangeal joint angle. Physiotherapy **85:** 1867, 1999.
23. Rose V, Nduka C, Pereira J, et al: Visual estimation of finger angles: do we need goniometers? J Hand Surg **27:** 382, 2002.
24. Watkins M, Riddle D, Lamb R, et al: Reliability of goniometric measurements and visual estimates of knee range of motion obtained in a clinical setting. Phys Ther **71:** 90, 1991.
25. Youdas J, Bogard C, Suman V: Reliability of goniometric measurements and visual estimates of ankle joint active range of Motion obtained in a clinical setting. Arch Phys Med Rehab **74:** 1113, 1993.
26. Somers DL, Hanson JA, Kedzierski CM, et al: The influence of experience on the reliability of goniometric and visual measurements of forefoot position. J Orthop Sports Phys Ther **25:** 192, 1997.
27. Fleiss J: *Statistical Methods For Rates and Proportions*, 2nd Ed, John Wiley, New York, 1981.
28. Menz HB, Tiedemann A, Kwan MM, et al: Reliability of clinical tests of foot and ankle characteristics in older people. JAPMA **93:** 380, 2003.
29. Beeson P, Phillips C, Corr S, et al: Classification systems for hallux rigidus: a review of the literature. Foot Ankle Int **29:** 407, 2008.