Mohsin, Fatma and McGarry, Anthony and Bowers, Roy J. (2015) Factors influencing the reliability of the universal goniometer in measurement of lower limb range of motion : a literature review. Journal of Prosthetics and Orthotics, 27 (4). pp. 140-148. ISSN 1040-8800 , http://dx.doi.org/10.1097/JPO.0000000000000074

Factors influencing the reliability of the universal goniometer in measurement of lower limb range of motion: a literature review

Abstract

Introduction: The universal goniometer (UG) is commonly used in clinical practice to measure lower limb joint range of motion (ROM). Reliability of the UG is essential to ensure consistency of measurement between and within practitioners. Clinically, it is important to understand how reliability may be affected by various factors.

Methods: An electronic and manual literature search was conducted to determine the reliability of the UG. A variety of search terms were used to search between 1980 and July 2015. Papers sourced were graded according to the Scottish Intercollegiate Guideline Network guidelines.
Papers reviewed included both measurements of healthy subjects and those with different pathologies. Active and passive lower limb ROM were studied and intratester and intertester reliability were examined.

Results: Twenty one studies were included and fully reviewed. Most studies indicated that UG reliability was best when used to measure ROM in healthy subjects in comparison to patients. The limited number of studies measuring active motion compromised the ability to make comparisons with measuring of passive ROM. It was reported from the studies investigating both intratester and intertester reliability that intratester reliability was higher than intertester reliability. Reliability of measurements varied depending on the joint measured. Tester training and standardisation of the measurement procedure led to increased reliability and there was a suggestion that involving two testers in the measurement procedure may have a beneficial effect.

Conclusion: This literature review highlights variation in study methodology employed, which reduces the ability to directly compare studies. Clinicians should be aware of the variability of

reliability of the UG and the effect of different factors when interpreting measurements taken with this instrument. Further research is required to investigate the effect different factors may have on the reliability of the UG and the possibility of using protocols and technology to increase reliability when measuring joint ROM.

Keywords:

Reliability, universal goniometer, intratester, intertester

Introduction

Documentation of joint ROM helps the clinician to assess limitations that may be present and hence plan the most appropriate treatment strategy. Assessment of joint ROM may also facilitate evaluation of treatment and development of protocols for appropriate intervention. A variety of measurement tools are currently available, ranging from simple visual estimation to advanced three-dimensional video recording systems.[1-3] The most practical, and most frequently used, clinical tool is the goniometer and several different designs have been developed over the years. The most simple and inexpensive type is the universal goniometer (UG). For measurements obtained using the UG to be clinically useful, results must be accurate and reliable.

Intratester reliability is important as the same clinician may take the same measurement on different occasions to document change. In the clinical setting, however, more than one tester may be involved in the measurement. As this has potential for further error, intertester reliability must also be considered clinically relevant.

Several factors may affect the reliability of the measurements obtained using the UG. Reliability might vary between different joints as each joint has different characteristics. Arguably this may make it easier or harder to obtain reliable measurements depending on the joint measured.[4-9]

The type of movement measured may also affect reliability, and reliability of active movements may differ when compared to passive movements, as if the force applied by the therapist to move the joint vary, this may cause different angles to be obtained each time measurements are taken.[9] In addition, following a standard instruction procedure and prior training may affect the reliability as this minimises the error associated with different procedures.

Different pathologies, such as upper motor neuron disorders may influence the UG reliability. Upper motor neuron disorders may cause altered muscle function leading to variation in muscle tone. Hypertonicity and spasticity may have an effect on the ability to define end range of joint motion which in return may affect the reliability of the measurements. Additionally, the presence of bony deformations may cause difficulties in clearly identifying bony landmarks which may compromise the reliability of measurements.

It is essential to understand how reliability may be affected when such variables are introduced, since incorrect interpretation of measurements obtained may lead to inappropriate treatment. The aim of this review is therefore to investigate the intratester and intertester reliability of the UG and to examine how different factors influence measurement reliability.

Methods

An electronic and manual literature search was conducted to investigate the intratester and intertester reliability of the UG for measurement of lower limb joints. A variety of search terms was used to search different medical and engineering databases such as Medline, EMBASE, NHS Scotland e-library, Science Direct, PubMed and Google Scholar. Search terms included: reliability and/or universal goniometer; universal goniometer; goniome* measurement reliability; intertester reliability of universal goniometer or UG; and intratester reliability of universal goniometer or UG. Identified secondary references from the articles were found and related books were also reviewed.

The review investigated the reliability of the UG in measuring ROM of the lower limb. Inclusion criteria were as follows: studies that evaluated intratester reliability and / or intertester reliability of the UG; studies that included patients and / or healthy subjects; studies that measured hip, knee, ankle, and/ or subtalar joint and studies that used intraclass correlation coefficient (ICC) to calculate reliability. Exclusion criteria included: studies that only investigated measurement of the upper limb, and studies that did not use ICC to calculate the reliability. References from 1980 to present (July 2015) were included to ensure the number of the studies was manageable.

Papers sourced were graded according to the Scottish Intercollegiate Guideline Network (SIGN) guidelines [10] and thematic tables of evidence were constructed for each design of goniometer and the pathology of the subjects tested.

Different statistical methods are used in order to measure the reliability, such as ICC, Pearson product moment correlation coefficient, analysis of variance reliability, coefficient of variation and generalisability theory. However, this review concentrated on papers that used ICC values to calculate reliability. This method was considered the most appropriate method for reliability measurement as data is centred and scaled using a pooled mean and standard. Additionally as the correlation line between the values is drawn at a 45 $^{\circ}$ angle, this was considered to reflect the most accurate reliability value.[11, 12] Most studies did not report the reason for the chosen method of analysis although one paper stated that ICC best reflected errors associated with measurements.[13] It has been suggested that the Pearson product moment correlation coefficient may produce high reliability values even when large inconsistency between paired scores is found.[14] This statistical method may overestimate reliability as each variable is centred and scaled by its own mean and standard deviation. Additionally the correlation line is drawn at its best position without specifying location.[11]

Results

Search results

The initial search yielded 71 papers, of which 21 matched the inclusion criteria and were fully reviewed.[4-7, 9, 13, 15-29] All the studies were case series (SIGN grade 3).

Statistical analysis

Intraclass correlation coefficient (ICC or reliability) values were rated as weak (0-0.60); good (0.60-0.80); or excellent (above 0.80).[12, 15, 16, 30, 31] Three papers reported on intratester reliability only.[4, 6, 21] Five studies reported on intertester reliability only.[5, 7, 20, 25, 29] Thirteen papers reported on both intratester and intertester reliability (Table 1, 2, 3).[9, 13, 15-19, 22-24, 26-28]

Motion measured and measurement procedure:

Eighteen studies examined passive motion [4-7, 9, 13, 15-18, 21, 23-29] and three studies active motion.[19, 20, 22] Four studies did not give testers standard instructions to follow or prior training.[13, 15, 16, 19]

Participants

Seven studies included healthy subjects [7, 20, 21, 25-28] while eleven studies included patients with various pathologies including diabetes,[17] neurological conditions,[4, 5, 9, 22-24] orthopaedic conditions [19] and neurological and orthopaedic conditions.[13, 16] One study stated only that the participants were nursing home residents.[18] Two studies included both subjects with neurological conditions and healthy subjects.[6, 29] Sample sizes varied widely (range: 6-150).

Reliability of universal goniometer

Hip joint

Healthy subjects

Active motion. One study was found which concluded that intertester reliability for measuring internal and external rotation was excellent (0.90-0.94) (Table 1).[20]

Passive motion. One study reported good to excellent intratester reliability for measurement of hip extension (0.70-0.96).[21]Two studies reported weak to excellent intratester reliability for measurement of hip extension (0.09-0.92).[6, 26] Two studies found weak to good intertester reliability for measurement of hip extension (0.10-0.65).[26, 29] In contrast, another study found excellent intertester reliability for measurement of hip extension (0.92).[25]
A single study reported on intratester reliability for hip flexion and found weak to excellent reliability (0.52-0.99) (Table 1).[6]

Patients

Active motion. No study was found investigating measurement of active hip motion.

Passive motion. Three studies reported good to excellent intratester and intertester reliability (0.61-0.981) for measurements of hip extension, flexion, abduction and lateral rotation amongst patients with neurological conditions.[4, 9, 24] Two studies also found excellent intratester reliability for measurement of hip abduction (0.82-0.95) and hip extension (0.98) but weak intertester reliability for measurement of hip extension (0.24) and hip abduction (0.37-0.47) amongst the same patient group.[18, 23] Another study reported weak intertester reliability for measurement of hip extension (0.19-0.50).[29] One study found inconsistent results and significant variation in intratester reliability within one session and between sessions for measurement of hip extension (0.17-0.91) and flexion (0.55-0.80) (Table 1).[6]

Knee joint

Healthy subjects

Active motion. No study was found investigating measurement of active knee motion.

Passive motion. Two studies reported excellent intratester reliability for measurement of knee flexion (0.96-0.99) and knee extension (0.83-0.97).[15, 27] One study reported good intratester reliability for measurement of knee flexion (0.65-0.72) [28] while, another study found weak to excellent intratester reliability for knee extension measurement (0.34-0.99).[6] Other studies, found good to excellent intertester reliability during measurement of flexion (0.88-0.99) and extension (0.64-0.71) (Table 2).[7, 15, 27] On the other hand, three studies found weak to good intertester reliability for measurement of knee extension (0.21-0.68) and flexion (0.44-0.59).[27-29]

Patients

Active motion. A single study was found which reported excellent intratester and intertester reliability of flexion and extension amongst patients with orthopaedic conditions (0.89-0.99) (Table 2).[22]

Passive motion. Intratester reliability was found to be excellent in four studies investigating measurements of knee flexion (0.99) and extension (0.81-0.98) amongst subjects with neurological and orthopaedic conditions.[4, 9, 13, 18] However, one study reported weak to excellent intratester reliability for measurement of knee extension (0.57-0.92) amongst subjects with neurological conditions.[6] Two studies reported weak intertester reliability for measurement of knee extension (0.26 [29] and 0.58 [9]) amongst subjects with neurological conditions. On the other hand, three studies reported good to excellent intertester reliability for measurement of knee extension (0.78-0.96) (Table 2) amongst subjects with neurological and orthopaedic conditions.[5, 13, 18] One study reported excellent intertester reliability (0.90) for measuring knee flexion amongst patients with neurological and orthopaedic conditions.[13]

Ankle joint

Healthy subjects

Active motion. No study was found to determine measurement of active ankle motion.

Passive motion. A single study was found which reported good to excellent intratester and intertester reliability for dorsiflexion (0.63-0.99) (Table 3).[6]

Patients

Active motion. A single study reported weak to excellent intratester reliability (0.47-0.93) and weak intertester reliability (0.25-0.28) for the measurement of plantarflexion and dorsiflexion amongst patients with orthopaedic conditions (Table 3).[19]

Passive motion. A single study examined measurement of dorsiflexion and plantarflexion amongst patients with diabetes and found excellent intratester reliability (0.89-0.96) while the intertester reliability varied between good to excellent (0.74-0.89).[17] Five studies reported excellent intratester reliability (0.81-0.99) during measurement of dorsiflexion and plantarflexion amongst patients with neurological and orthopaedic conditions.[4, 6, 9, 16, 24] By contrast, two other studies reported weak to good intertester reliability (0.12-0.73) during measurement of dorsiflexion and plantarflexion amongst the same patient group [9, 16] excluding two studies where excellent intertester reliability was found for measurements of plantarflexion and dorsiflexion (0.87-0.88).[5, 24]

Discussion

Several different designs of goniometers have been developed over the years including the UG, electrical goniometer (EG) and gravity-dependant goniometer (inclinometers).[25, 32, 33] The UG is the most frequently used tool in the clinical environment. However, the disadvantage of using a UG is the requirement of using two hands to move the joint while simultaneously aligning the UG with bony landmarks, which may compromise reliability.[16, 34] New technologies are recently emerging for joint ROM measurements, such as dimensional and 3 dimensional video recording systems.[35, 36]

This review included 21 studies investigating intratester and intertester reliability of the UG in measuring active or passive ROM of lower limb joints amongst patients or healthy subjects.

In general, reliability of the UG varied across different pathologies, proving to be most reliable amongst healthy subjects. A number of studies stated that the presence of spasticity is a major cause of error [23, 30, 37, 38] and concluded that care should be taken when using the measurements obtained using the UG for assisting in clinical judgment.[30, 37] However, Kilgour et al [6] compared measurement reliability of healthy subjects to those with spastic diplegia, and found equal reliability.[6] This study concluded that a major cause of error was in defining the end range of the joint ROM rather than the presence of spasticity. Furthermore, Lee et al [29] also compared measurement reliability of healthy subjects to those with cerebral palsy (CP) and found higher reliability amongst subjects with CP. Elveru et al [16] reported a higher intertester reliability for ankle plantarflexion ROM in patients with general orthopaedic conditions in comparison with patients with neurological conditions.

Some studies included more than one form of pathology and grouped results without reporting on each pathology individually.[15, 18, 19, 22] Three studies included more than one form of CP and did not report on each group separately hence, reducing the ability to interpret results.[5, 23 29]

Watkins et al [13] investigated intratester and intertester reliability of knee joint ROM amongst patients with different pathologies with knee joint problems and reported excellent reliability (Table 2). A posterior analysis was performed in the study to determine the effect of different pathologies on the reliability. Overall, pathology did not have an effect on the intratester and intertester reliability. However, intertester reliability for knee extension was found to be weak amongst below knee amputees which may be explained due to short distal limb segment causing difficulties in aligning the UG.

Passive ROM is the motion mostly measured in clinical environment, and only 3 studies included in this review reporting on active motion. Two studies reported excellent intratester and intertester reliability for measuring hip and knee joint active ROM (Table 1, 2).[20, 22] Another study reported weak to excellent intratester reliability and weak intertester reliability for measuring active ankle

ROM (Table 3).[19] The limited number of studies found measuring active motion compromised the ability to make comparisons with measuring passive ROM. However, one study stated that the low intertester reliability could be explained due in part to the difference in the force applied by therapists during assessments of passive motion, causing different angles to be obtained during each session.[9]

It was noted that reliability varied across the joints measured due to the different joint characteristics and ease of identifying bony landmarks. Overall, reliability varied from weak to excellent across the hip, knee and ankle joint. Despite the fact that the knee joint is a polycentric joint where the centre of rotation changes with motion, it was found to be a reliable joint to measure and this is supported with the high ICC values found (Table 2).[7, 13, 15, 22] Measurement of knee flexion appears more reliable than measurement of knee extension ROM (Table 2).[13, 15, 22] Similar results were found for measurement of hip joint ROM, as some studies reported excellent intratester and intertester reliability for the measurement of hip extension, abduction and external rotation (Table 1).[4, 9, 18, 23-25] This suggests that although joint characteristics are a factor affecting reliability, the length of lever arms may have more effect. Aligning the arms of the UG to follow the long bones in the thigh and calf and the mid-lateral trunk, may assist knee and hip joint ROM measurements making this more reliable. Excellent reliability was also reported for measuring the ankle joint ROM despite the short lever arm of the foot. Furthermore, it has been reported that even complex motions can be measured reliably when strict standard position is applied.[39, 40]

Most studies provided the testers with a standard measurement procedure and prior training in order to minimise associated error. Rothstein et al [15] deliberately did not standardise the measurement procedure (measuring technique and patient's position) to mimic the clinical setting and stated that "measurement technique will often vary between the therapist, partially because of their training and preferences and partially because of adaptation, such as positioning which are necessary with different patients" (p 1611). The study reported high intratester reliability during measurement of passive knee flexion and extension ROM and high intertester for knee flexion ROM but lower intertester reliability for knee extension ROM amongst patients (Table 2).[15] To examine the effect of the different patient positions used in the study, a posterior analysis of the results was carried out and an increase in intertester reliability for knee extension was reported from (0.20-0.69) to (0.74-0.84) when the same position was used. It was suggested that using different patient positions while measuring causes variability due to the bi-articular muscles (hamstrings) affecting the knee extension.[15] The hamstrings muscles cross both hip and knee joint limiting knee extension ROM when the hip is flexed, hence variation in position of the hip joint during measuring knee joint extension can cause differences in the measurements obtained. Subject position varied across a number of studies. The positions used to measure hip joint ROM were Thomas test, modified Thomas test, prone hip extension test, supine position with knee maintained in different degrees of flexion, prone position and seated position. Kilgour et al [6] reported higher intratester ICC values for the Staheli test (0.78-0.91) in comparison to the Thomas test (0.17-0.66) in subjects with CP. Furthermore, intratester ICC values for the prone hip extension test (0.80-0.92) were found to be higher than intratester ICC values for the Thomas test (0.09-0.91) amongst healthy subjects in the previous study. A further study found weak intertester reliability when using the Thomas test (0.58).[5] In contrast, Lee et al [29] reported higher ICC values for the Thomas test (0.20-0.50) in comparison to prone hip extension test (0.10-0.19) amongst subjects with CP and healthy subjects. Van Dillen et al [21] compared 4 positions for measuring hip extension, which included femur maintained in 0 degrees abduction with knee maintained in 80 degrees flexion, femur maintained in 0 degrees of abduction with knee fully extended, femur fully abducted with knee maintained in 80 degrees of flexion and femur in full abduction and knee fully extended. In this study the higher intratester reliability was achieved for the position where the femur was fully abducted and knee fully extended (0.96). Simoneau et al [20] measured hip external and internal rotation using prone and seated position and concluded higher ICC values were achieved for internal rotation (0.94) and external rotation (0.93) in the prone position. The study recommended documentation of the position of the hip to allow repeated reliable measurements. It has been reported that proper aligning of the UG when measuring using the Thomas test or prone hip extension test can be difficult, as one hand is used to ensure the lumbar spine is flat and other hand is used to align the UG while maintaining the position of both arms of the UG.[5] For knee ROM measurements, the following positions were used; popliteal angle and supine position with hip extended. Kilgour et al [6] compared measuring knee extension with the hip in neutral and in 90 degrees flexion and found higher ICC values with the hip in a neutral position in subjects with CP and healthy subjects. On the other hand Cadenhead et al [4] found equal reliability when measuring knee extension while maintaining the hip in neutral or 90 degrees flexion. For ankle joint ROM measurements, the following positions were used; supine position with knee extended, supine position with knee flexed and prone position. A study found equal intratester reliability when measuring ankle dorsiflexion with knee extended and knee flexed.[6] Diamond et al [17] measured ankle dorsiflexion in prone position and reported good to excellent intratester and intertester reliability.

Another source of error could be due to the discrepancies in identification of bony landmarks and goniometric alignment between the testers. Peeler and Anderson [28] carried out a pilot testing with 3 testers and found that differences were reported between testers when identifying the lateral epicondyle of femur (used to align the axis of the UG) especially in patients with pathological changes at the knee.[15] In addition, they reported that difficulties were found in maintaining the position of the axis of the UG when trying to align the two arms. Additionally, Watkins et al [13] followed non-standard measuring procedure and reported excellent intratester and intertester reliability for measurements of knee flexion and extension in patients with knee pathologies. A posterior analysis of the results showed that non-standardisation of the measurement procedure contributed slightly to measurement error but still suggested that standard procedures be applied to minimise this error (when the same patient position was used, ICC for flexion increased by 0.02 and for extension by 0.01).[13] A further study found weak intertester reliability when measuring active ankle ROM in patients when the position was not standardised suggesting that a standard protocol should be established and followed.[19] The weak intertester reliability reported in this study may be explained due to the variation in the UG alignment using bony landmarks. In measuring the ankle joint, the fixed arm is aligned over the long axis of fibula however, the moveable arm could be aligned with the heel, fifth metatarsal or plantar surface of the foot causing variation in measurements amongst testers.[19] In addition, the variation found in the study by Youdas et al [19] may be explained due to the effect of the gastrocnemius muscle which crosses the knee and ankle joint limiting the available ankle dorsiflexion ROM when the knee is extended. Different knee joint positions used in the previous study may have caused different ankle dorsiflexion ROM to be recorded leading to variation in results (wide range of intratester and intertester ICC reported). In contrast, another study stated that lack of standardisation was not a significant factor for difference amongst the testers (intertester) during measurements of passive ankle plantarflexion ROM.[16] However, the opposite was reported for ankle dorsiflexion ROM (ICC increased by 0.09 when using different position and decreased by 0.10 when using same position) but still rated as weak intertester reliability (Table 3).[16] Furthermore, it was stated that involving two testers in the measurement procedure may increase the reliability of the UG amongst CP patients as one tester stabilises the limb and the second tester takes the measurements.[5]

Generally, it was found from the studies included in this review that intratester reliability was higher than intertester reliability (Table 1, 2, 3).[9, 15-19, 23, 27, 29] One study suggested that averaging two measurements each session increases the reliability of the measurements obtained [19] agreeing with the findings of Low [41]. However, a study by Rothstein et al [15] found that no greater reliability is obtained when mean of measurements is used suggesting that reliability can be achieved by taking a single measurement in clinical settings. In addition, a posterior analysis of the results by Elveru et al [16] suggested that no increase in reliability was achieved when using the mean of two measurements, agreeing with the finding of Boone et al.[42] The results of Kilgour et al [6] showed no increase in reliability when averaging two measurements but stated that "taking two duplicate measures in clinical practice could help therapists to identify measurements within on session that might need to be repeated" (p 399).

The ability to make direct comparison between the studies was compromised due to the differences in methodology adopted in the studies, such as the level of experience of the testers, number of sessions and time between the sessions.

Most studies included testers who were physical therapists with experience level ranging from 1 [15, 23] to 30 [6] years. McWhirk and Glanzman [5] included two therapists with different levels of experience (1and >10 years' experience) to investigate the intertester reliability in subjects with CP. They found good to excellent reliability for all the motions measured excluding hip extension ROM (0.58) (Table 1, 2, 3). Elveru et al [16] preformed a posterior analysis of the results to investigate the effect of experience on reliability and reported an increase in intratester ICC from 0.90 to 0.91 for ankle dorsiflexion and from 0.86 to 0.92 for ankle plantarflexion, and an increase in intertester ICC from 0.50 to 0.54 for ankle dorsiflexion and a decrease from 0.72 to 0.70 for ankle plantarflexion when more experienced testers took the measurements. Although an increase of ICC was reported, this increase did not affect the overall rating of the ICC values. However, limited information about the tester experience was provided. The rest of studies did not provide additional results to show the effect of the testers experience on the reliability obtained.

The number of testing sessions, and the period between each session varied across the studies. Most studies used a test-retest design to calculate intratester reliability of measurements taken by different testers on the same day.[4, 5, 7, 13, 15-22, 25, 29] Kilgour et al [6] investigated intratester reliability within and between sessions (one week apart) for passive ROM of hip, knee and ankle joint amongst subjects with CP and healthy subjects. In this study, all intra-sessional ICC values were found to be higher than inter-sessional ICC values (Table1, 2, 3). Wakefield et al [26] also reported weak intratester inter-sessional (between sessions) ICC values for hip extension amongst healthy subjects (Table 2). In contrast, Mutlu et al [24] and Herrero et al [23] reported high intratester intersessional ICC values for all the motions measured amongst subjects with CP (Table 1, 3). Pandya et al [9] reported excellent intratester intersessional ICC values for all measurements obtained amongst subjects with Duchenne muscular dystrophy, and Peeler and Anderson [28] reported similar results amongst healthy subjects (Table 1, 2, 3).

It is important to consider measurement reliability in the clinical context. It is reported that an error of ±5° in measurement may be clinically acceptable.[43] Hence, clinicians should be cautions when interpreting results of reliability studies and must select studies appropriate to pathology. Although Mutlu et al [24] reported good to excellent intertester reliability, a variation of 0-28 degrees was found in intertester measurements. Additionally, another study reported a variation of 15-20 degrees in the measurements between sessions.[6] The clinical effect of such findings must be considered, especially when using measurements to determine treatment effect.

This review aimed to investigate the intratester and intertester reliability of the most commonly used measurement tool, the UG, and to examine how different factors can influence reliability. Twenty one studies were included which investigated the reliability of measuring hip, knee and ankle joint ROM. This literature review highlights variation in the methodology employed, which reduced the ability to compare studies directly, as the number of testers, experience level, number of sessions, time between the sessions and subject position varied across the studies. Most studies indicated that the UG reliability was best when used to measure ROM in healthy subjects and that reliability may be reduced in the presence of different pathologies. Passive ROM is the motion mostly measured in the clinical environment, and hence a larger number of studies have examined the reliability of measurement of passive motion rather than active motion. The limited number of studies measuring active motion compromised the ability to make comparisons with measurement of passive ROM. It was stated that the low intertester reliability could be explained due in part to the difference in the force applied by therapists during assessments of passive motion, causing different angles to be obtained during each session. Generally, it was found that intratester reliability was higher than intertester reliability. Reliability varied from weak to excellent across the hip, knee and ankle joints due to the different joint characteristics and ease of identifying bony landmarks. It has been reported that even complex motions can be measured reliably when a strict standard position is applied. Standardisation of the measurement procedure and prior training were found to increase measurement reliability and one study suggested that involvement of more than one tester in the measurement procedure may have beneficial effect on reliability. Further research is required to investigate the reliability of the UG and the possibility of using protocols and technology to increase reliability when measuring joint ROM.

References

1. Krause DA, Boyd MS, Hager AN, Smoyer EC, Thompson AT and Hollman JH. Relibility and accuracy of a goniometer mobile device application for video measurment of the functional movement screen deep squat test. *International Journal of Sports Physical Therapy*. 2015; 10: 37-44.

2. Smith DS. Measurement of joint range--an overview. *Clinics in rheumatic diseases*. 1982; 8: 523-31.

3. Lea RD and Gerhardt JJ. *Range-of-motion measurements*. 1995, p.784-98.

4. Cadenhead SL, McEwen IR and Thompson DM. Effect of passive range of motion exercises on lower-extremity goniometric measurements of adults with cerebral palsy: a single-subject design. *Physical therapy*. 2002; 82: 658-69.

5. McWhirk LB and Glanzman AM. Within-session inter-rater realiability of goniometric measures in patients with spastic cerebral palsy. *Pediatric physical therapy : the official publication of the Section on Pediatrics of the American Physical Therapy Association*. 2006; 18: 262-5.

6. Kilgour G, McNair P and Stott NS. Intrarater reliability of lower limb sagittal range-of-motion measures in children with spastic diplegia. *Developmental medicine and child neurology*. 2003; 45: 391-9.

7. Gogia PP, Braatz JH, Rose SJ and Norton BJ. Reliability and validity of goniometric measurements at the knee. *Physical therapy*. 1987; 67: 192-5.

8. Nicol A. Measurement of joint motion. *Clinical rehabilitation*. 1989; 3: 1-9.

9. Pandya S, Florence JM, King WM, Robison JD, Oxman M and Province MA. Reliability of goniometric measurements in patients with Duchenne muscular dystrophy. *Physical therapy*. 1985; 65: 1339-42.

10. Scottish Intercollegiate Guidelines Network.

11. Denegar C and Ball D. Assessing Reliability and Precision of Measurement: An Introduction to Intraclass Correlation and Standard Error of Measurement. *Journal of Sport Rehabilitation*. 1993; 2: 35-42.

12. Shrout PE and Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychological bulletin*. 1979; 86: 420-8.

13. Watkins MA, Riddle DL, Lamb RL and Personius WJ. Reliability of goniometric measurements and visual estimates of knee range of motion obtained in a clinical setting. *Physical therapy*. 1991; 71: 90-6; discussion 6-7.

14. Bartko JJ and Carpenter WT, Jr. On the methods and theory of reliability. *The Journal of nervous and mental disease*. 1976; 163: 307-17.

15. Rothstein JM, Miller PJ and Roettger RF. Goniometric reliability in a clinical setting. Elbow and knee measurements. *Physical therapy*. 1983; 63: 1611-5.

16. Elveru RA, Rothstein JM and Lamb RL. Goniometric reliability in a clinical setting. Subtalar and ankle joint measurements. *Physical Therapy*. 1988; 68: 672-7.

17. Diamond JE, Mueller MJ, Delitto A and Sinacore DR. Reliability of a diabetic foot evaluation. *Physical therapy*. 1989; 69: 797-802.

18. Mollinger LA and Steffen TM. Knee flexion contractures in institutionalized elderly: prevalence, severity, stability, and related variables. *Physical therapy*. 1993; 73: 437-44; discussion 44-6.

19. Youdas JW, Bogard CL and Suman VJ. Reliability of goniometric measurements and visual estimates of ankle joint active range of motion obtained in a clinical setting. *Archives of physical medicine and rehabilitation*. 1993; 74: 1113-8.

20. Simoneau GG, Hoenig KJ, Lepley JE and Papanek PE. Influence of hip position and gender on active hip internal and external rotation. *The Journal of orthopaedic and sports physical therapy*. 1998; 28: 158-64.

21. Van Dillen LR, McDonnell MK, Fleming DA and Sahrmann SA. Effect of knee and hip position on hip extension range of motion in individuals with and without low back pain. *The Journal of orthopaedic and sports physical therapy*. 2000; 30: 307-16.

22. Brosseau L, Balmer S, Tousignant M, et al. Intra- and intertester reliability and criterion validity of the parallelogram and universal goniometers for measuring maximum active knee flexion and extension of patients with knee restrictions. *Archives of physical medicine and rehabilitation*. 2001; 82: 396-402.

23. Herrero P, Carrera P, Garcia E, Gomez-Trullen EM and Olivan-Blazquez B. Reliability of goniometric measurements in children with cerebral palsy: a comparative analysis of universal goniometer and electronic inclinometer. A pilot study. *BMC musculoskeletal disorders*. 2011; 12: 155.

24. Mutlu A, Livanelioglu A and Gunel MK. Reliability of goniometric measurements in children with spastic cerebral palsy. *Medical science monitor : international medical journal of experimental and clinical research*. 2007; 13: CR323-9.

25. Clapis PA, Davis SM and Davis RO. Reliability of inclinometer and goniometric measurements of hip extension flexibility using the modified Thomas test. *Physiotherapy theory and practice*. 2008; 24: 135-41.

26. Wakefield CB, Halls A, Difilippo N and Cottrell GT. Reliability of goniometric and trigonometric techniques for measuring hip-extension range of motion using the modified Thomas test. *Journal of athletic training*. 2015; 50: 460-6.

27. Peters PG, Herbenick MA, Anloague PA, Markert RJ and Rubino LJ, 3rd. Knee range of motion: reliability and agreement of 3 measurement methods. *American journal of orthopedics (Belle Mead, NJ)*. 2011; 40: E249-52.

28. Peeler JD and Anderson JE. Reliability limits of the modified Thomas test for assessing rectus femoris muscle flexibility about the knee joint. *Journal of athletic training*. 2008; 43: 470-6.

29. Lee KM, Chung CY, Kwon DG, Han HS, Choi IH and Park MS. Reliability of physical examination in the measurement of hip flexion contracture and correlation with gait parameters in cerebral palsy. *The Journal of Bone and Joint Surgery, American Volume*. 2011; 93: 150-8.

30. Stuberg WA, Fuchs RH and Miedaner JA. Reliability of goniometric measurements of children with cerebral palsy. *Developmental medicine and child neurology*. 1988; 30: 657-66.

31. Eliasziw M, Young SL, Woodbury MG and Fryday-Field K. Statistical methodology for the concurrent assessment of interrater and intrarater reliability: using goniometric measurements as an example. *Physical therapy*. 1994; 74: 777-88.

32. Rome K and Cowieson F. A reliability study of the universal goniometer, fluid goniometer, and electrogoniometer for the measurement of ankle dorsiflexion. *Foot & ankle international / American Orthopaedic Foot and Ankle Society [and] Swiss Foot and Ankle Society*. 1996; 17: 28-32.

33. Torburn L, Perry J and Gronley JK. Assessment of rearfoot motion: passive positioning, one-legged standing, gait. *Foot & ankle international / American Orthopaedic Foot and Ankle Society [and] Swiss Foot and Ankle Society*. 1998; 19: 688-93.

34. Bierma-Zeinstra SM, Bohnen AM, Ramlal R, Ridderikhoff J, Verhaar JA and Prins A. Comparison between two devices for measuring hip joint motions. *Clinical rehabilitation*. 1998; 12: 497-505.

35. Nicolas R, Nicolas B, Francois V, Michel T and Nathaly G. Comparison of knee kinematics between meniscal tear and normal control during a step-down task. *Clinical Biomechanics (Bristol, Avon)*

2015.

36. Cronin J, Nash M and Whatman C. Assessing dynamic knee joint range of motion using siliconcoach. *Physical therapy in sport : official journal of the Association of Chartered Physiotherapists in Sports Medicine*. 2006; 7: 191-4.

37. Ashton BB, Pickles B and Roll JW. Reliability of goniometric measurements of hip motion in spastic cerebral palsy. *Developmental medicine and child neurology*. 1978; 20: 87-94.

38. McDowell BC, Hewitt V, Nurse A, Weston T and Baker R. The variability of goniometric measurements in ambulatory children with spastic cerebral palsy. *Gait & posture*. 2000; 12: 114-21.

39. Gajdosik R and Lusin G. Hamstring muscle tightness. Reliability of an active-knee-extension test. *Physical therapy*. 1983; 63: 1085-90.

40. Gajdosik R, Simpson R, Smith R and DonTigny RL. Pelvic tilt. Intratester reliability of measuring the standing position and range of motion. *Physical therapy*. 1985; 65: 169-74.

41. Low JL. The reliability of joint measurement. *Physiotherapy*. 1976; 62: 227-9.

42. Boone DC, Azen SP, Lin CM, Spence C, Baron C and Lee L. Reliability of goniometric measurements. *Physical therapy*. 1978; 58: 1355-60.

43. Bruton A, Conway JH and Holgate ST. Reliability: What is it, and how is it measured? *Physiotherapy*. 2000; 86: 94-9.

Tables:

Table 1: Summary of the papers which studied the reliability of UG for measuring hip joint amongst healthy subjects and patients.

Source	Reliability / condition	Movement	Results
Healthy subjects			
Active motion			
Simoneau et al. (1998) [22]	Inter	External, internal rotation	-Internal rotation = 0.82-0.97 (G/E), external rotation = 0.76-0.98 (E)
Passive motion			
Van Dillen et al. (2000) [23]	Intra	Extension	0.70-0.96 (G/E)
Kilgour, McNair and Stott. (2007) [6]	Intra	Flexion, extension	-Extension = 0.09-0.92 (W/E), flexion = 0.52-0.99 (W/E)
Clapis, S Davis and R Davis. (2008) [27]	Inter	Extension	0.92 (E)
Lee et al. (2011) [31]	Inter	Extension	0.10-0.27 (W)
Wakefield et al. (2015) [28]	Intra + inter	Extension	-Intratester = 0.51-0.54 (W) -Intertester = 0.30-0.65 (W/G)
Patients			
Passive motion			
Pandya et al. (1985) [9]	Intra + inter / Duchene muscular dystrophy	Extension	-Intratester = 0.85 (E) -Intertester = 0.74 (G)
Mollinger, and Steffen. (1993) [20]	Intra + inter / nursing home residents	Extension	-Intratester = 0.98 (E) -Intertester = 0.24 (W)
Cadenhead, McEwen and Thompson. (2002) [4]	Inter / neurological conditions	Extension, abduction, external rotation	-Extension = 0.94-0.98 (E), abduction = 0.96 (E), external rotation = 0.78-0.86 (G/E)
McWhirk and Glanzman. (2006) [5]	Inter / neurological conditions	Extension, abduction	-Extension = 0.58 (W), abduction = 0.90 (E)
Kilgour, McNair and Stott. (2007) [6]	Intra / neurological conditions	Flexion, extension	-Extension = 0.17-0.91(W/E), flexion = 0.62-0.98 (G/E)
Mutlu, Livanelioglu and Gunnel. (2007) [26]	Intra + inter / neurological conditions	Flexion, extension, abduction, external rotation	-Intratester (extension = 0.73- 0.99 (G/E), abduction = 0.48-0.70 (W/G), external rotation = 0.80-0.84 (G/E), flexion = 0.60-0.86 (G/E)) - Intertester (extension = 0.92-0.95 (E), abduction = 0.61-0.77 (G), external rotation = 0.91-0.92 (E), flexion = 0.77-0.83 (G/E))
Herrero et al. (2011) [25]	Intra + inter / neurological conditions	Abduction	-Intratester =0.82-0.95 (E) -Intertester = 0.37 - 0.47 (W)
Lee et al. (2011) [31]	Inter / neurological conditions	Extension	0.19-0.50 (W)

Reliability: E: excellent reliability (>0.80), G: good reliability (0.60-0.80) and W: weak reliability (0.00-0.60)

Table 2: Summary of the papers which studied the reliability of UG for measuring knee joint amongst healthy subjects and patients.

Source	**Reliability / condition**	**Movement**	**Results**
Healthy subjects			
Passive motion			
Rothstein, Miller, Roettger. (1983) [17]	Intra + inter	Flexion, extension	-Intratester (extension = 0.91-0.97 (E), flexion = 0.97-0.99 (E)) -Intertester (flexion = 0.91-0.99 (E), extension = 0.64-0.71 (G))
Gogia et al. (1987) [7]	Inter	Flexion	0.99 (E)
Kilgour, McNair and Stott. (2007) [6]	Intra	Extension	0.34-0.99 (W/E)
Peeler and Anderson (2008) [30]	Intra + inter	Flexion	-Intratester = 0.65-0.72 (G) -Intertester = 0.44-0.59 (W)
Lee et al. (2011) [31]	Inter	Extension	0.20 (W)
Peters et al. (2011) [29]	Intra + inter	Flexion, extension	-Intratester (flexion = 0.96-0.98 (E), extension = 0.83-0.87 (E)) -Intertester (flexion = 0.88 (E), extension = 0.21(W))
Patients			
Active motion			
Brosseau et al. (2001) [24]	Intra + inter / neurological conditions	Flexion, extension	-Intratester (flexion = 0.99 (E), extension = 0.97-0.98 (E)) -Intertester (flexion = 0.97-0.98 (E), extension = 0.89-0.92 (E))
Passive motion			
Pandya et al. (1985) [9]	Intra + inter / Duchene muscular dystrophy	Extension	-Intratester = 0.93 (E) -Intertester = 0.58 (W)
Watkins et al.(1991) [15]	Intra + inter / neurological and orthopaedic conditions	Flexion, extension	-Intratester (flexion = 0.99 (E), extension = 0.98 (E)) -Intertester (extension = 0.86 (E), flexion = 0.90 (E))
Mollinger, and Steffen. (1993) [20]	Intra + inter / nursing home residents	Extension	-Intratester = 0.99 (E) -Intertester = 0.96 (E)
Cadenhead, McEwen andThompson. (2002) [4]	Intra / neurological conditions	Extension	0.81-0.98 (E)
McWhirk and Glanzman. (2006) [5]	Inter / neurological conditions	Extension	0.78-0.92 (G/E)
Kilgour, McNair and Stott. (2007) [6]	Intra / neurological conditions	Extension	0.57-0.99 (W/E)
Lee et al. (2011) [31]	Inter / neurological conditions	Extension	0.26 (W)

Reliability: E: excellent reliability (>0.80), G: good reliability (0.60-0.80) and W: weak reliability (0.00-0.60)

Table 3: Summary of the papers which studied the reliability of UG for measuring ankle joint amongst healthy subjects and patients.

Source	Reliability	Movement	Results
Healthy subjects			
Passive motion			
Kilgour, McNair and Stott. (2007) (6)	Intra	Dorsiflexion	0.63-0.99 (G/E)
Patients			
Active motion			
Youdos, Bogard, Suman. (1993) (21)	Intra + inter / orthopaedic conditions	Dorsiflexion, plantarflexion	-Intratester dorsiflexion = 0.64-0.92 (G/E), plantarflexion = 0.47-0.96 (W/E) -Intertester dorsiflexion = 0.28 (W), plantarflexion = 0.25 (W)
Passive motion			
Pandya et al. (1985) (9)	Intra + inter / Duchene muscular dystrophy	Dorsiflexion	-Intratester = 0.90 (E) -Intertester = 0.73 (G)
Elveru, Rothstein amd Lamb. (1988) (18)	Intra + inter / neurological and orthopaedic conditions.	Dorsiflexion, plantarflexion	-Intratester dorsiflexion= 0.90 (E), plantarflexion= 0.86 (E) -Intertester dorsiflexion= 0.50 (W), plantarflexion= 0.72 (G)
Diamond et al. (1989) (19)	Intra + inter / diabetes	Dorsiflexion	-Intratester = 0.89-0.96 (E) -Intertester = 0.74-0.87 (G/E)
Cadenhead, McEwen and Thompson. (2002) (4)	Intra / neurological conditions	Dorsiflexion	0.81-0.98 (E)
McWhirk and Glanzman. (2006) (5)	Inter / neurological conditions	Dorsiflexion	0.87 (E)
Kilgour, McNair and Stott. (2007) (6)	Intra / neurological conditions	Dorsiflexion	0.63-0.99 (G/E)
Mutlu, Livanelioglu and Gunnel. (2007) (26)	Intra + inter / neurological conditions	Dorsiflexion	-Intratester = 0.81-0.90 (E) -Intertester = 0.88 (E)

Reliability: E: excellent reliability (>0.80), G: good reliability (0.60-0.80) and W: weak reliability (0.00-0.60)