Clinical Measurement of Range of Motion
Review of Goniometry Emphasizing Reliability and Validity

RICHARD L. GAJDOSIK
and RICHARD W. BOHANNON

Clinical measurement of range of motion is a fundamental evaluation procedure with ubiquitous application in physical therapy. Objective measurements of ROM and correct interpretation of the measurement results can have a substantial impact on the development of the scientific basis of therapeutic interventions. The purpose of this article is to review the related literature on the reliability and validity of goniometric measurements of the extremities. Special emphasis is placed on how the reliability of goniometry is influenced by instrumentation and procedures, differences among joint actions and body regions, passive versus active measurements, intratester versus intertester measurements, and different patient types. Our discussion of validity encourages objective interpretation of the meaning of ROM measurements in light of the purposes and the limitations of goniometry. We conclude that clinicians should adopt standardized methods of testing and should interpret and report goniometric results as ROM measurements only, not as measurements of factors that may affect ROM.

Key Words: Physical therapy; Tests and measurements, general.

Goniometric measurements are used by physical therapists to quantify baseline limitations of motion, decide on appropriate therapeutic interventions, and document the effectiveness of these interventions. Probably our most widely used evaluation procedure, goniometry, can be considered a fundamental part of the “basic science” of physical therapy.

Historically, goniometry developed over the last 60 years in conjunction with the rapid growth of the field of physical medicine and rehabilitation. A recent article by Smith provides an interesting account of this development. Our review of the literature revealed a surprising number of articles describing different methods of measuring ROM, from simple visual estimation to high-speed cinematography. To most physical therapists, however, the universal goniometer (ie, full-circle manual goniometer) remains the most versatile and widely used instrument in clinical practice. The design of the universal goniometer and the procedures for its use have been described in detail in numerous publications. The articles by Moore are particularly comprehensive. Although the procedures described by the American Academy of Orthopedic Surgeons more than 20 years ago are cited frequently in the literature, we encourage the reader to see the recent publication by Norkin and White for current and complete descriptions.

We believe that physical therapists should master the objective measurement of range of motion and learn to interpret measurement results correctly. By doing so, the potential contribution to goniometry to the development of a scientific basis of physical therapy will be enhanced. The purpose of this article is to discuss what we view as the two most important factors affecting objective goniometric measurements, reliability and validity. The related literature will be reviewed and its implications discussed. The need for additional research to clarify our current knowledge about the reliability and validity of goniometry will be presented. Because a test must be reliable to be valid, we will address the reliability of ROM measurements before considering their validity. For sake of clarity, we will discuss only the reliability and validity of ROM measurements of the extremities. The basic principles presented, however, can be applied to ROM measurements of the trunk. This article should help the reader to gain both a conceptual understanding and a practical appreciation of clinical ROM measurements. Furthermore, we hope that physical therapists will be stimulated to develop further the clinical application of goniometry.

RELIABILITY

Reliability in goniometry simply means the consistency or the repeatability of the ROM measurements, that is, whether the application of the instrument and the procedures produce the same measurements consistently under the same conditions. We wish to emphasize from the outset that because few investigators have used the same study design, comparisons of the results of reliability studies often are difficult. For example, studies in which repeated tests are separated by short time intervals (ie, one hour) may yield very different results than studies in which repeated tests are separated by longer time intervals (ie, days or weeks). The reliability of the measurements expresses their reproducibility or stability only in relation to the time intervals reported. The most accurate evaluation of the reliability of the instrument and procedures is determined when short time intervals separate tests, the classic “test-retest” study design. These results probably will be more reliable than the results of studies with long time intervals between tests because the accuracy of the measurements is increased with few uncontrolled variables. Even though test-retest studies conducted over short time intervals...
best reflect the reliability of the test, reliability studies conducted over days, or over weeks, are important clinically for assessing the stability of the measurement so that comparisons can be made to evaluate patient improvements. Because so many different study designs and indexes for analysis of reliability have been reported, we will defer discussion of this information in this article. We will discuss reliability, but we refer the reader to the recent article by Stratford et al11 for a review of specific methodology issues. The reader also may find the guidelines outlined by Miller9 helpful for establishing reliability in a clinical setting.

Given that estimating ROM by visual inspection is unreliable when precision and accuracy are needed12 and that measurements with the goniometer are more reliable estimates of ROM,13 we will consider the reliability of the goniometer and procedures for its use. Although we acknowledge that the reliability of goniometry is affected by many factors, we will discuss only the complexity of the movement being measured, variations among the body regions measured, whether the movement is active or passive, and whether the measurement is conducted by the same examiner (intraexaminer reliability) or by different examiners (intertester reliability). We believe that the reliability of ROM measurements for patients may vary with their clinical problems and may differ from that reported for healthy subjects. Accordingly, we also will address the potential differences in reliability among patient types. These differences have important implications for documenting the effects of physical therapy on ROM.

**Instrumentation and Procedures**

The comprehensive articles by Moore published in 1949 set the stage for our understanding of the importance of objective instrumentation and standardized clinical procedures for measuring ROM.23 Since that time, most reports have acknowledged that if the goniometer is reliable, then the reliability of ROM measurements depends primarily on the standardization of procedures. Hellebrant et al found the universal goniometer to be a “more dependable tool than devices designed to measure particular motions in specific joints.”4(p460) Salter assumed that inaccurate measurements were mainly the result of faulty application and concluded that goniometer error is negligible.6 Hamilton and Lachenbruch found no difference in the accuracy and reliability of three different goniometers (dorsal, universal, and pendulum) for measuring finger joint angles when the procedures were standardized.14 They suggested that the application of the goniometers should depend on the ease of operation and on the adaptability of the instrument to the particular clinical problem (edema, joint enlargements, joint deformities). Using a Leighton flexometer (a contact pendulum goniometer), Ekstrand et al compared a method used in clinical practice to measure passive ROM of the lower extremities (Series A) with an “optimal” measurement method (Series B).15 In contrast with the method of Series A, the method of Series B included covering the examining table with a wood board, standardizing the height of the table at each session, marking bony landmarks, and checking for faulty movement of the pendulum dial. They found that the mean intratester coefficient of variation (CV) for Series B measurements (1.9% ± 0.7%) was significantly lower than the mean CV for Series A measurements (7.5% ± 2.9%) and concluded that accurate reproducibility depends on careful measurement technique. Rothstein et al examined the reliability of three different goniometers (large metal, large plastic, and small plastic) for measuring passive elbow and knee positions and reported high interdevice reliability (intraclass correlation coefficient [ICC] > .91) for the three instruments.16 In a more recent study, Fish and Wingate reported that improper alignment of the goniometer, misidentification of bony landmarks, and variations in manual force, all contributed to goniometric error at the elbow.17 Their results indicated that even relatively inexperienced examiners should be able to use goniometers accurately to measure elbow positions if a standardized method is followed carefully. The results of these studies clearly demonstrate the importance of standardizing testing procedures and support the suggestion by Salter that although error is inherent in the construction of goniometers, “inaccuracies occur during their use, mainly from their faulty application.”4(p460)

The reliability of ROM measurements also may be influenced by changes in ROM that result from repeated testing trials. Cobe, in 1928, studied the active ROM at the wrist and recommended using the average of several measurements because individual fluctuations increased the variability of the measurements.18 Low reported that the average of several measurements of elbow flexion and wrist extension was more reliable than one measurement.19 Boone et al, however, studied the reliability of measuring the active ROM of six joint motions of the upper and lower limbs and concluded that “one measurement per measurement session is as reliable as the average of repeated measurements in one session.”19(p360) Rothstein et al assessed the reliability of goniometric measurements for passive elbow and knee positions and found only slight improvement in the intertester reliability by taking the mean score of multiple measurements.16 These conflicting reports may be explained at least partially by inherent differences in the actions measured (simple or complex) or by the influence of active versus passive measurements. Atha and Wheatley, for example, studied the effect of repeated trials of passive straight leg raising (SLR) and reported a systematic increase in the angle of SLR over the first four to five trials when measured on the same day.20 The SLR angle also increased over two consecutive days. Bohannon passively loaded SLR and found systematic increases in the SLR angle over a three-day period.21 These results indicate that recording the average of several measurements may increase the reliability of some ROM measurements, such as those showing wide variations or systematic increases, although some ROM measurements may be very stable and therefore reliable without averaging. Physical therapists should understand, however, that posttreatment increases in some measurements may be a normal function of increased compliance of tissues from repeated testing over time. More research is needed to examine the effects of repeated measurements, particularly in relation to the specific actions measured and the methods used.

**Differences Among Joint Actions and Among Structure and Function of Body Regions**

The reliability of measuring ROM of the extremities is affected by the complexity of the actions measured and by the inherent structural and functional differences of the action. Hellebrandt et al found that wrist flexion, medial rotation of the shoulder, and shoulder abduction “did not lend themselves to reliable repetitive measurement.”4(p306) Low also reported that the reliability of goniometric measurements varies in different joints.13 Boone et al reported that the intertester reliability was higher for upper extremity motions (r = .86) than for lower extremity motions (r = .58).19 These
results suggest that reliability of measuring ROM is specific to the action measured and to regional structure and function. For example, measurements of the elbow, generally considered a simple hinge joint, show less day-to-day variation in ROM than measurements of the wrist, the movement of which is affected by multiple joints and numerous muscles crossing these joints. Although the ROM measurements of movements that are influenced by movement of adjacent joints or by multiple joint muscles might be expected to be less reliable than the ROM measurements of simple hinge joints, strict standardization of procedures should increase reliability. Gajdosik and Lusin and Gajdosik et al have demonstrated that even complex movements can be measured reliably when the measurement procedures are controlled strictly. The importance of applying standardized procedures was emphasized in these reports. Physical therapists should realize, however, that measuring complex actions reliably may be more difficult than measuring simple actions because of the greater potential for error resulting from large fluctuations within the same individual or because of systematic changes over time.

### Intratester Versus Intertester Measurements

In 1949, Hellebrandt et al studied the reliability of active movements of the upper extremity. Even though the results of their study may have been influenced by the broad variety of patients examined, they did show that a well-trained therapist can measure ROM with a high degree of reliability. The authors stated, however, the following: "Where the inter-reliability of different physical therapists has not been established, different observers should not be used interchangeably to obtain measurements on the same patient." Since this study by Hellebrandt et al, many investigations have demonstrated in healthy subjects that intratester reliability is higher than intertester reliability. Hamilton and Lachenbruch reported greater intertester variance than intratester variance for measuring finger joint angles. Low measured full elbow flexion and full wrist flexion and showed that the error between observers was much larger than the error within the same observer. In a well-documented study that is cited frequently in the literature on goniometry, Boone et al reported that the intertester reliability was lower than the intratester reliability for weekly measurements of active ROM of the upper and lower extremities. As stated earlier in this article, they found higher intertester reliability for upper extremity measurements than for lower extremity measurements. Based on their results, they suggested that when more than one tester measures the same motion, changes in ROM should exceed 5 degrees in upper extremity measurements and 6 degrees in lower extremity measurements to document "improvement." In light of the difficulty of measuring the ROMs of different actions with similar reliability, we believe that these guidelines should be reconsidered. These guidelines may be particularly difficult to apply to different patient types. Even studies of healthy subjects have demonstrated different outcomes. Ellis and Stowe found 1% to 5% error for intratester measurements, but as much as 5% to 10% error for intertester measurements of the ROM of the hip. Thus, if hip flexion ROM was about 130 degrees, then 5% to 10% error would equal about 7 degrees to 13 degrees. Solgaard et al reported that intratester CVs were as much as 5% to 8% and that intertester CVs were 6% to 10% for ROM measurements of the wrist. We encourage physical therapists to be cautious accepting generalizations about the reliability of ROM measurements, such as those suggested by Boone et al., in light of our current knowledge about the potential for variations among different ROM measurements.

Smith and Walker demonstrated higher intratester reliability (r = .90) than intertester reliability (r = .70) for knee and elbow ROM in healthy elderly persons. In a later study, Walker et al reported greater variability between testers than within testers for active ROM measurements of the upper and lower extremities of elderly subjects.

Generally, the results of intratester and intertester reliability studies with patients support the findings reported for healthy subjects. Using patients as subjects, Rothstein et al found that intratester reliability for flexion and extension of the knee and elbow joints was high and that intertester reliability, although lower, also was high, except for knee extension (r range, .57–.80). Pandya et al studied seven upper and lower extremity movements and found that the error between observers was much larger than the error within the same observer. In a well-documented study that is cited frequently in the literature on goniometry, Boone et al reported that the intertester reliability was lower than the intratester reliability for weekly measurements of active ROM of the upper and lower extremities. As stated earlier in this article, they found higher intertester reliability for upper extremity measurements than for lower extremity measurements. Based on their results, they suggested that when more than one tester measures the same motion, changes in ROM should exceed 5 degrees in upper extremity measurements and 6 degrees in lower extremity measurements to document "improvement." In light of the difficulty of measuring the ROMs of different actions with similar reliability, we believe that these guidelines should be reconsidered. These guidelines may be particularly difficult to apply to different patient types. Even studies of healthy subjects have demonstrated different outcomes. Ellis and Stowe found 1% to 5% error for intratester measurements, but as much as 5% to 10% error for intertester measurements of the ROM of the hip. Thus, if hip flexion ROM was about 130 degrees, then 5% to 10% error would equal about 7 degrees to 13 degrees. Solgaard et al reported that intratester CVs were as much as 5% to 8% and that intertester CVs were 6% to 10% for ROM measurements of the wrist. We encourage physical therapists to be cautious accepting generalizations about the reliability of ROM measurements, such as those suggested by Boone et al., in light of our current knowledge about the potential for variations among different ROM measurements.

Smith and Walker demonstrated higher intratester reliability (r = .90) than intertester reliability (r = .70) for knee and elbow ROM in healthy elderly persons. In a later study, Walker et al reported greater variability between testers than within testers for active ROM measurements of the upper and lower extremities of elderly subjects.

Generally, the results of intratester and intertester reliability studies with patients support the findings reported for healthy subjects. Using patients as subjects, Rothstein et al found that intratester reliability for flexion and extension of the knee and elbow joints was high and that intertester reliability, although lower, also was high, except for knee extension (r range, .57–.80). Pandya et al studied seven upper and lower extremity movements and found that the error between observers was much larger than the error within the same observer. In a well-documented study that is cited frequently in the literature on goniometry, Boone et al reported that the intertester reliability was lower than the intratester reliability for weekly measurements of active ROM of the upper and lower extremities. As stated earlier in this article, they found higher intertester reliability for upper extremity measurements than for lower extremity measurements. Based on their results, they suggested that when more than one tester measures the same motion, changes in ROM should exceed 5 degrees in upper extremity measurements and 6 degrees in lower extremity measurements to document "improvement." In light of the difficulty of measuring the ROMs of different actions with similar reliability, we believe that these guidelines should be reconsidered. These guidelines may be particularly difficult to apply to different patient types. Even studies of healthy subjects have demonstrated different outcomes. Ellis and Stowe found 1% to 5% error for intratester measurements, but as much as 5% to 10% error for intertester measurements of the ROM of the hip. Thus, if hip flexion ROM was about 130 degrees, then 5% to 10% error would equal about 7 degrees to 13 degrees. Solgaard et al reported that intratester CVs were as much as 5% to 8% and that intertester CVs were 6% to 10% for ROM measurements of the wrist. We encourage physical therapists to be cautious accepting generalizations about the reliability of ROM measurements, such as those suggested by Boone et al., in light of our current knowledge about the potential for variations among different ROM measurements.

Smith and Walker demonstrated higher intratester reliability (r = .90) than intertester reliability (r = .70) for knee and elbow ROM in healthy elderly persons. In a later study, Walker et al reported greater variability between testers than within testers for active ROM measurements of the upper and lower extremities of elderly subjects.

Generally, the results of intratester and intertester reliability studies with patients support the findings reported for healthy subjects. Using patients as subjects, Rothstein et al found that intratester reliability for flexion and extension of the knee and elbow joints was high and that intertester reliability, although lower, also was high, except for knee extension (r range, .57–.80). Pandya et al studied seven upper and lower extremity movements and found that the error between observers was much larger than the error within the same observer. In a well-documented study that is cited frequently in the literature on goniometry, Boone et al reported that the intertester reliability was lower than the intratester reliability for weekly measurements of active ROM of the upper and lower extremities. As stated earlier in this article, they found higher intertester reliability for upper extremity measurements than for lower extremity measurements. Based on their results, they suggested that when more than one tester measures the same motion, changes in ROM should exceed 5 degrees in upper extremity measurements and 6 degrees in lower extremity measurements to document "improvement." In light of the difficulty of measuring the ROMs of different actions with similar reliability, we believe that these guidelines should be reconsidered. These guidelines may be particularly difficult to apply to different patient types. Even studies of healthy subjects have demonstrated different outcomes. Ellis and Stowe found 1% to 5% error for intratester measurements, but as much as 5% to 10% error for intertester measurements of the ROM of the hip. Thus, if hip flexion ROM was about 130 degrees, then 5% to 10% error would equal about 7 degrees to 13 degrees. Solgaard et al reported that intratester CVs were as much as 5% to 8% and that intertester CVs were 6% to 10% for ROM measurements of the wrist. We encourage physical therapists to be cautious accepting generalizations about the reliability of ROM measurements, such as those suggested by Boone et al., in light of our current knowledge about the potential for variations among different ROM measurements.

Smith and Walker demonstrated higher intratester reliability (r = .90) than intertester reliability (r = .70) for knee and elbow ROM in healthy elderly persons. In a later study, Walker et al reported greater variability between testers than within testers for active ROM measurements of the upper and lower extremities of elderly subjects.
joint limitations in children with Duchenne muscular dystrophy and reported that intratester reliability was high (ICC range, .81-.94), but intertester reliability was lower with a wide variation (ICC range, .25-.91).28

Recent evidence suggests that the reliability of ROM measurements may be influenced directly by the type of patient problem. Ashton et al examined the day-to-day reliability of measuring ROM of the hip in children with spastic diplegia and found that the reliability usually was lower for children with a moderate condition than for children with a mild condition. Because the level of reliability was low, they stated that ROM measurements “are not sufficiently reliable to be used in studies of cerebral palsy.”34,39 Harris et al examined further the goniometric reliability for a child with cerebral palsy (ie, spastic quadriplegia).35 They found that intratester reliability exceeded intertester reliability and, as expected, that a wide variation existed in measurements, both within examiners and among examiners. The authors concluded that a difference of ± 10 to 15 degrees in ROM over time does not signify a meaningful change in a child with severe spastic cerebral palsy. Bartlett et al compared four methods of measuring hip extension (prone extension, Thomas test, Mundale, and pelvifemoral angle) in patients with spastic diplegia and meningomyelocele.36 They found that the Thomas test was particularly difficult to apply in patients with spastic diplegia and that improved reliability most likely will result in this patient group by using one of the other techniques. They also reported that the least reliable technique in the meningomyelocele group was that of Mundale, probably because of difficulties associated with identifying bony landmarks in the presence of obesity and deformities. These researchers also found intratester reliability to be superior to intertester reliability. The results of these few studies with patients clearly show that the reliability of ROM measurements is influenced by the specific patient problem, and the results suggest a tremendous need for studies of the reliability of measuring ROM among the different patient types within the specialized areas of physical therapy of orthopedics and neurology.

VALIDITY

Currier has stated that the validity of a measurement “constitutes the degree to which an instrument measures what it is purported to measure; the extent to which it fulfills its purpose.”38,39,40 In goniometry, we must be confident that the goniometer and measurement procedures are accurate and that we understand the meaning of the measurement results. To do so, we must show that the measurement procedures are consistent with our interpretation of the results. Do we understand the purpose of goniometry and the limitations of goniometry? Are we measuring what we say we are measuring?

The primary purpose of goniometry is to measure ROM of the musculoskeletal system of the human body. To fulfill this purpose, the goniometer was designed as a modification of the protractor, a device known to represent accurately the degree intervals of a circle. If the accuracy and, therefore, the validity of goniometers are in question, the degree units can be compared simultaneously against known angles.23,33 A strong relationship establishes concurrent validity, a type of criterion-related validity. This comparison can detect any systematic errors resulting from faulty construction so that proper adjustments can be made. Although small errors in the construction of goniometers may exist, the instruments generally are accepted as valid clinical tools. They do, however, have limitations. For example, the measurements obtained are limited to degree units of a circle and, as a result, we accept that the movements we measure have fixed axes of motion about which movement occurs. Smith, however, demonstrated that this assumption is not true.33 Because of other motions within joints, such as articular sliding and rotation, the axes of motion are not fixed. In a strict sense, the validity of representing movement of body parts by units of a circle can be challenged. Nevertheless, we accept this limitation and are confident that the ROM measured closely approximates movement around a central point, that is, that the ROM measurements are clinically valid.

Another limitation of goniometric measurements is that they are recorded in degree units and, therefore, only ROM can be measured. This limitation is not a problem if the results are interpreted within this restriction, but physical therapists inadvertently may reach beyond this restriction and invalidly expand the meaning of the results. Many factors may affect the outcome of ROM measurements—edema, pain, adhesions, strength deficits, and muscle hypertrophy—but ROM measurements are never measures of factors other than ROM. The validity of ROM measurements is very specific. In the extremities, ROM is recorded in degrees, whereas the factors that may affect ROM must be measured by different methods with equally different measurement units. This point is extremely important and worthy of discussion.

A classic example of how physical therapists may expand the meaning of ROM measurements and invalidly interpret the results can be found by examining the SLR test. Most physical therapists are aware that the maximal angle of SLR in relation to the horizontal plane has been used historically to provide information about hamstring muscle length (HML) and hamstring muscle flexibility (HMF). But can we really use this test to measure HML and HMF? The validity of using the SLR test to assess HML and HMF has been challenged recently, and the results of studies are giving new meaning and validity to the test. Some study results have indicated that during SLR the pelvis moves in conjunction with the lower limb,38-40, thus, the origin of the hamstring muscles is not fixed. The reports also indicated that SLR may be confounded by pain from nerve root irritation and by the position of the ankle and stretching structures other than the hamstring muscles (eg, sciatic nerve, enveloping deep fascia of the lower limb).40 These findings do not mean that the SLR test is invalid for measuring the maximal angle of SLR in relation to the horizontal plane, but that these results are limited to this interpretation alone. Hamstring muscle length may affect the maximal angle of SLR, but the SLR test does not measure HML directly. If the results of SLR were compared with the results of linear measurements of HML and if the two measurements were strongly related, then we could say with confidence that the angle of SLR indirectly represents HML—concurrent validity would be demonstrated.

We are unaware of published studies demonstrating a strong relationship between the angle of SLR and HML. Tardieu et al showed that the tibialcalcanal angle corresponds to the length of the triceps surae muscle group in subjects of the same height; therefore, the angles of some joints may be valid, indirect indications of muscle length. Obviously, more research is needed to validate claims that the maximal angle of SLR represents HML.

The SLR test also has been widely accepted as a measure of HMF. Individuals with a small SLR angle are considered
to have poor HMF, and those with a large SLR angle are believed to have good HMF. As with using the SLR test to measure HML, using the test to indicate the flexibility of the muscles may be incorrect and, therefore, invalid. Even though the maximal ROM of a joint or combined motions of multiple joints have been interpreted historically to yield information about the flexibility of the motions examined, no scientific basis exists for these conclusions. Passive flexibility, or extensibility, of skeletal muscles is a physiological property and is defined by the length-tension relationship of the tissues. This relationship can be represented as “stiffness”—the ratio of the change in passive muscle tension (ΔP) to the change in muscle length (ΔL) or as “compliance”—the ratio of the change in muscle length (ΔL) to the change in passive muscle tension (ΔP) (ΔL/ΔP).42 Whether the maximal angle of SLR represents the length-tension relationship of the hamstring muscles is unknown and currently is being investigated by the first author. The same physiological principles apply to ROM measurements throughout the body, and studies are needed to examine the validity of using ROM measurements to represent flexibility.

As we discussed earlier in this article, the standardization of testing procedures improves reliability, and goniometric measurements must be reliable to be valid. Reliable measurements, however, do not ensure that the measurements are valid; invalid measurements can be measured reliably. One obvious example already discussed would be to accept reliable measurements of the angle of SLR as valid measurements of HML or HMF when no evidence exists to indicate that this conclusion is true. This problem results primarily from interpreting the results incorrectly in relation to the measurement procedures. The content validity of the test (ie, how well the test reflects what is being measured), which basically is judgmental, is not established. As we shall see shortly, accurate judgment of the content validity of ROM measurements may vary depending on the complexity of the test. Again, we must ask ourselves, Are we measuring what we say we are measuring?

Physical therapists judge the validity of most ROM measurements based on their anatomical knowledge and their applied skills of visual inspection, palpation of bony landmarks, and accurate alignment of the goniometer. Generally, the accurate application of knowledge and skills, combined with interpreting the results as measurements of ROM only, provide sufficient evidence to ensure content validity. This finding is particularly true with simple measurements, such as those of the elbow and knee. The reliability and the validity of even simple measurements, however, may be decreased because of patient differences that we cannot control. Obesity and variations in bony structures can make accurate visual inspection and bony palpation very difficult. Perhaps one reason Boone et al49 found that the reliability of lower extremity ROM measurements was less than the reliability of upper extremity ROM measurements was because of greater difficulty accurately locating bony landmarks and aligning the goniometer with the landmarks of the lower extremity. Factors such as the size and the weight of the extremity measured also will affect the clinician’s ability to manipulate the part and thus the reliability and the validity of the measurements.

Even though physical therapists may have confidence in the validity of ROM measurements based on judgment, applying more objective methods of motion analyses to study ROM can strengthen our knowledge of the content validity by demonstrating criterion-related validity. Comparing the results of goniometric measurements with those taken from photographs is one method that can be used to demonstrate validity. Fish and Wingate assessed the accuracy and sources of error in goniometry at the elbow using still photography as a reference standard.57 If still photography is used, the photographic procedures must be standardized to ensure their validity. Variations in the position of the camera in relation to the subject and the type of lens used (ie, 35 mm vs 50 mm) are only two of many potential sources of error. Still photography also requires marking anatomical landmarks with skin marks, and movement of the bones under the skin may cause additional error.

Cinematography also has been used to study the validity of motions routinely measured by goniometry.38-40,44 The results of these studies have provided new insights into the content validity of ROM measurements of the hip. The implications of the results of studies of the SLR test have been discussed. In addition to these results, when active and passive unilateral and bilateral hip flexion with knee flexion was examined by cinematography,44 the displacement of skin marks placed over bony landmarks indicated that the hip flexion movement is composed of two movements: 1) flexion of the thigh on the pelvis and 2) pelvic rotation. The implication of these results is that the traditional method of measuring hip flexion probably includes movement of both the femur and the pelvis, just as shoulder abduction incorporates movement of both the humerus and the scapula.45,46 Consequently, our knowledge of ROM measurements of the hip flexion movement takes on new and more valid meaning.

The application of cinematography and other forms of motion analysis (eg, electromyography, electrogoniometry, videotape recording) have the potential to add substantially to our understanding of ROM measurements. These techniques may be particularly helpful for studying the relationship of goniometric measurements to the functional ROM needed for patients to perform various activities of daily living. As with still photography, cinematography is limited by similar procedural problems, such as the positions of the camera and the subject and documenting movement with skin marks over bony landmarks that may move under the skin. Even so, careful application of these techniques can contribute to our knowledge of ROM measurements.

The most powerful method by which the validity of ROM measurements can be studied is radiography. Lawrence confirmed the validity of goniometry at the knee by comparing goniometric measurements with measurements from radiographs of cadavers and patients.47 More recently, however, Enwemeka compared goniometric knee measurements with radiographic bone angle measurements in healthy adults and reported that “within the first 15 degrees of knee flexion, goniometric measurement of joint excursion may be remarkably wrong.”48,49 He indicated that the error resulted from knee rotation, which changes the relative position of the bony landmarks used in goniometry.

In a classic study, Freedman and Munro demonstrated through radiographs that shoulder abduction results from the relative contributions of both scapular movements and glenohumeral movements.48 The results of this study were reproduced by Doody and Waterland using special goniometric procedures.46 The results of these two studies confirmed that measurements of total shoulder abduction are valid only for the combined movements of the scapula on the thorax and the humerus on the scapula. As a result, physical therapists consider these movements when measuring ROM of the...
shoulder, and valid clinical procedures have been developed to isolate the two component motions. Objective methods of motion analysis other than radiography also have contributed to our knowledge of the complex movements that occur at the shoulder. Blakely and Palmer recently showed that medial rotation of the humerus, not lateral rotation as generally believed, accompanies active and passive shoulder flexion. This finding was reported first for healthy men and women, and subsequently confirmed using a fresh cadaver. The therapeutic implications then were studied by analyzing the flexion-abduction-lateral rotation pattern of proprioceptive neuromuscular facilitation. We commend the authors for their systematic and thorough methods of inquiry and encourage similar endeavors. As the body of knowledge of physical therapy and movement analysis advances, we expect that findings similar to these will continue to change our understanding of movement and the validity of ROM measurements.

CONCLUSIONS

For most applications to the extremities, the universal fullcircle goniometer probably is the preferred instrument for measuring ROM. Although the literature abounds with studies about the reliability of ROM measurements, applying their results to clinical practice may be difficult because of differences in study designs. Because the reliability of goniometry is dependent on a host of factors, such as differences among the motions measured, methods of application, and variations among different patient types, clinicians working in the same setting should adopt standardized methods of testing. These methods would require a thorough orientation for both students and staff. Physical therapists should be careful in the interpretation and reporting of the results of goniometric findings. As a rule, ROM measurements are just that, not measurements of muscle “tightness,” the length of specific structures, or other factors that may affect ROM. We recommend continued research so that physical therapists may gain a complete and objective understanding of the meaning of goniometric measurements.

REFERENCES

42. Walker SM: Tension and extensibility changes in muscles suddenly stretched during tetanus. Am J Physiol 172:37-41, 1953