Concurrent Validity and Intertester Reliability of Universal and Fluid-based Goniometers for Active Elbow Range of Motion

MICHELE PETHERICK,
WENDY RHEAULT,
SHARON KIMBLE,
CONNIE LECHNER,
and VIRGINIA SENEAR

The purposes of this study were 1) to establish the concurrent validity of the universal goniometer and the fluid-based goniometer and 2) to determine the intertester reliability of these two instruments. A correlational study was performed in which two testers used the universal goniometer and the fluid-based goniometer in measuring elbow range of motion in 30 healthy subjects. The fluid-based goniometer had high intertester reliability (R = .92), and the standard goniometer had poor reliability (R = .53). The Pearson product-moment correlation between the two instruments was .83. A significant difference was shown between the standard goniometer and the fluid-based goniometer by the t test (t = 4.4, df = 28, p < .05). The results support the use of the fluid-based goniometer between testers on elbow range of motion; however, the two instruments cannot be used interchangeably.

Key Words: Elbow joint; Physical therapy; Tests and measurements, range of motion; Upper extremity, elbow.

The goniometer long has been recognized as an essential tool for clinical physical therapists. Goniometers are used by physical therapists to assess treatment effectiveness and disease progression and to document this information to third-party payers.1-7 Unfortunately, the homogeneity of this important assessment tool and the technique associated with its use have evolved slowly. This problem was addressed by Moore, who attributed errors to a lack of standardization in nomenclature, numerical expression, patient position, and instrumentation.6,8 Hellebrandt et al suggested that rigid control of these variables would result in greater reliability for measurement of joint range of motion.3

In 1949, Moore advocated the use of an instrument that is currently considered the universal (standard) goniometer (UG).8 This goniometer is a protractor with an extended stationary arm and a fulcrum-mounted moveable arm. Tape measures and trigonometry, in addition to pendulum, bubble, and “over-the-joint” goniometers, have also been used to measure joint ROM.9-11 These other measuring tools, however, have not gained wide clinical acceptance. A fluid-based type of goniometer (FG), which works on the principle of a carpenter’s level, was recently introduced. The fluid in this type of instrument is contained in a circular chamber and is displayed by straight-plane movement in gravity. Although manufacturers of the FG have provided no literature on the instrument, Clarke and associates have described the use of a type of FG for the glenohumeral joint.12,13 and other authors have used it to measure cervical ROM.10

For a measurement instrument to be useful, its reliability and validity must be established. Goniometric measurement has been found to have greater intratester reliability than estimation by observation.4 High intratester reliability has been found by several authors in both clinical and research settings when using the UG.1,7 Boone et al have found that for a given motion no significant intratester variation exists in three repetitions of the same motion.1 Rothstein and associates found that intratester reliability for elbow and knee joint ROM measurements was very high (r = .91-.99).7 The results of that study support the finding of Boone et al that multiple measurements do not increase intratester reliability. These findings, however, disagree with those of Low, who contends that the average of multiple measurements leads to greater accuracy and reliability.4

In general, past research has found intertester reliability to be less than intratester reliability.1,5,11 Hamilton and Lachenbruch measured finger joint ROM and found significant
that intertester reliability was .88, .76, and .91 for spinal lower extremity motions (r = .58). They found that the variation in three upper extremity motions (r = .86) than for Boone et al found less intertester variation in three upper extremity motions (r = .58). They found that the average standard deviation for upper extremity motions was 2 degrees, but for elbow flexion was 3.7 degrees. They concluded that a change of greater than 5 degrees in the upper extremities and greater than 6 degrees in the lower extremities is necessary to determine improvement in ROM when more than one tester measures the same motion.

High validity of the measuring instrument is required for the tool to provide meaningful results. Burdett and associates, in a study measuring lumbar spine movements in healthy subjects, found low validity correlations between four measurement instruments and roentgenographic measurements. Gogia et al, however, found high Pearson product-moment correlation coefficients (r = .97–.98) and intraclass correlation coefficients (ICCs) (ICC = .98–.99) between goniometric measurements and roentgenograms taken of the knee. They concluded that the goniometer was a valid instrument for measuring ROM of the knee joint. Enwemeka also reported no significant difference between goniometric and roentgenographic measurements of the knee, except in the first 15 degrees of ROM.

Because little literature is available on the FG, we felt that a study exploring the validity and reliability of this instrument was necessary. The purposes of this study were 1) to establish the concurrent validity of the UG and the FG and 2) to determine intertester reliability of these two instruments in measuring elbow joint ROM. We hypothesized 1) that no significant difference would exist between the UG and the FG in measuring elbow flexion ROM and 2) that the intertester reliability of these two instruments would be high.

METHOD

Subjects

Subjects were 10 male and 20 female volunteers with no previous history of musculoskeletal or neurological problems of the right upper extremity (Tab. 1). The mean age of the subjects was 24 years (s = 4.2 years). All subjects were informed of the testing procedure, and a signed statement of consent was obtained from each subject.

| Physical Characteristics of Subjects (N = 30) |
|-----------------|--------|--------|
| Characteristic  | X      | s      |
| Age             | 24.00  | 4.20   |
| Height (m)      | 1.68   | .08    |
| Weight (kg)     | 63.62  | 11.25  |
| Right arm girth (cm)* | 26.40  | 3.60   |
| Right arm length (cm)* | 74.26  | 4.45   |

* Girth measured 10 cm proximal to olecranon process.
  b Arm length measured from acromion process to distal end of third digit.

Four goniometers were used for comparison of their validity and intertester reliability, a UG and a FG.* The UG was a full-circle plastic goniometer with 25.4-cm moveable arms, marked in 1-degree increments. The FG had a fluid-filled chamber that responded to position changes in space. It had a 360-degree scale divided into 1-degree increments and was read from the bottom of the fluid meniscus. The FG weighed about 11 g, had a 0.7-cm base, and was 10 cm in height. Each goniometer was calibrated before each testing session to known angles of 0, 90, and 180 degrees as determined from a standard protractor.

Procedure

Subjects were placed supine in the anatomical position on a treatment table. The subject's right arm was positioned with a 5.08-cm thick towel under the distal humerus and with the forearm off the edge of the plinth. This arm position allowed full elbow extension ROM. The right shoulder was stabilized with a 2.3-kg weight over the deltopectoral groove to prevent shoulder protraction. The subject's right arm was exposed for easy identification of bony landmarks. The UG was aligned according to the procedure outlined by Norkin and White. The stationary arm of the goniometer paralleled the lateral midline of the humerus, pointing at the acromion process. The goniometer's moveable arm was aligned with the lateral midline of the radius and the styloid process. The lateral epicondyle was used to center the fulcrum of the goniometer.

The FG was placed on the dorsal aspect of the subject's forearm, 6 cm distal to the olecranon process. The scale of the liquid level was set at 0 degrees when the limb was at the starting position of full extension. The scale was read when the subject's elbow was in full flexion.

Two testers (S.K. and C.L.) each took three measurements on every subject with both instruments. This testing procedure was followed because the literature disagrees on whether multiple measurements lead to greater intratester reliability. The order of measurement was randomized to eliminate sampling bias. The researchers practiced the measurement procedure before the experiment to ensure testing reliability.

Subjects actively flexed and extended their right elbow five times before the first measurement to ensure constant ROM throughout the testing session. The goniometer was aligned with the arm in full extension. The subjects then actively flexed their right elbow through the full ROM, avoiding forceful contraction. The full arc of motion was recorded by separate researchers (M.P. and V.S.) to blind the testers to previous measurements and thus prevent tester bias. All data were collected in the same laboratory at the university.

Data Analysis

Two statistical methods were used to assess the concurrent validity of the goniometers. We used a Pearson product-moment correlation to determine the covariance of measurements taken with both goniometers. Because the correlation coefficient can be high even if large differences exist between covarying paired measurements, the paired-data t test was also used. This method measured the agreement between

---

*a Chattanooga Corp, 101 Memorial Dr, PO Box 4287, Chattanooga, TN 37405.
testers to establish the concurrent validity of the goniometers. To compute these statistics, we used the mean of all six measurements per subject taken with each instrument.

Separating the data by instrument and tester, we calculated the mean of each subject's measurements. These values were used to compute the ICC (formula 2,3), which was used to determine the intertester reliability of each instrument. The subjects acted as their own control, thus increasing the sensitivity of the experiment. The use of ICCs allowed the determination of the source of variability between and within testers. All statistical tests were two-tailed at an alpha level of .05.

RESULTS

The Pearson product-moment correlation between the two goniometers equaled .83, which is significant at an alpha level of .05 (Tab. 2). The paired-data t test revealed a significant difference between the two instruments (t = 4.4, p < .05) (Tab. 2). Intertester reliability was determined by the ICC as .53 for the UG and .92 for the FG (Tab. 3).

DISCUSSION

The results of this study, although conducted on a small group of subjects, showed high correlation between the FG and the UG. Because the instruments were measuring the same joint angle, a high correlation would be expected. The instruments were not shown to be concurrently valid by the t test.

The intertester reliability was found to be higher with the FG than with the UG. Our results do not support those of Rothstein and associates, who found high intertester reliability (r = .95 for elbow flexion and extension) with the UG. Twelve testers took measurements in their study. Because standard deviations and degrees of freedom were not published with their study results, we could not determine how their statistics were calculated. The subjects in their study included patients with elbow disorders, which produced wider variability in the data. Statistically, this variability, if large, will produce higher correlation values. In our study, subjects with normal elbow mobility were measured, producing a restricted statistical range, which leads to attenuated correlation values.

We found the FG to be more reliable than the UG, perhaps because of its simple technique of use. The UG, in contrast to the FG, requires careful and sustained alignment to bony landmarks during movement. Other advantages of the FG include its light weight, small size, and speed of application. Its ease of application allows the therapist to obtain ROM measurements with one hand, leaving the other hand free to passively move the patient's extremity.

The FG, however, does have several limitations. It must be used perpendicular to the gravitational field, which necessitates frequent patient position changes. Its size precludes use on smaller joints such as the interphalangeal joints. Another disadvantage is that the FG measures only total excursion as opposed to movement from neutral. This factor may account for the significant difference we found between the two measurement instruments with the t test. Because the FG measures only total movement, it cannot specify the part of the ROM where movement occurs. Lastly, the cost of the FG is about three times that of the UG.

Future research involving the FG is necessary to fully understand its clinical applicability. This research could be in the areas of 1) protocol establishment for standardization, 2) determining reliability for other joints, and 3) clinical trials. Further study of both the FG and the UG is required to determine their validity.

CONCLUSION

This study compared the UG and the FG and found that concurrent validity did not exist between these two instruments. It also addressed intertester reliability for the two instruments and found that the FG had greater intertester reliability than the UG. Although the UG appears to be a mechanically sound measurement instrument, even the use of a standardized testing procedure allowed excessive variability between therapists. Although more research is needed, we foresee potential usefulness of the FG for improving goniometric measurement in physical therapy.

REFERENCES


**TABLE 2**

<table>
<thead>
<tr>
<th>Goniometer</th>
<th>X</th>
<th>s</th>
<th>SEM</th>
<th>r*a</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Universal</td>
<td>145.8</td>
<td>6.3</td>
<td>1.2</td>
<td>.83</td>
<td>4.4b</td>
</tr>
<tr>
<td>Fluid-based</td>
<td>149.4</td>
<td>7.9</td>
<td>1.5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*a* Pearson product-moment correlation. 
*b* p < .05 using the paired-data t test.

**TABLE 3**

<table>
<thead>
<tr>
<th>Source</th>
<th>Universal Goniometer</th>
<th>Fluid-based Goniometer</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>df</td>
<td>SS</td>
</tr>
<tr>
<td>Between subjects</td>
<td>29</td>
<td>2292.9</td>
</tr>
<tr>
<td>Within subjects</td>
<td>30</td>
<td>711.1</td>
</tr>
</tbody>
</table>

PHYSICAL THERAPY