Intrarater and interrater reliability were determined for middle finger range of motion (ROM) using the Rolyan finger goniometer. Seven raters measured ROM 3 times (trials) on 20 healthy adults. Intraclass reliability analysis and analysis of variance was used to assess the consistency and stability of measures. Level of significance was set at $p \leq 0.05$. Significant differences ($p < 0.05$) existed between raters for metacarpophalangeal, proximal interphalangeal, and distal interphalangeal active finger ROM measurements and for metacarpophalangeal and proximal interphalangeal passive range of motion. Intrarater reliability ranged from $0.43$ to $0.99$ (single measure intraclass coefficient). Interrater reliability ranged from $0.24$ to $0.95$, with passive measures lower than active measures. These results provide clinicians with evidence for using this specific goniometer and issues associated with various therapists assessing finger ROM.


Joint range of motion (ROM) measurements are used to assess patients’ rehabilitative status and progress (Norkin & White, 2003) and to determine impairment ratings when a patient is unable to return to his or her prior level of function. Goniometers are used to measure ROM, and reliability of the measurements is important. Patients are likely to be measured repeatedly, often by different therapists. For practical purposes, different therapists should obtain the same ROM results at a given point in time.

A few studies have assessed finger ROM with a focus on specific assessment methods or specific goniometers (Breger-Lee, Voelker, Giurintano, Novick, & Browder, 1993; Dijkstra, de Bont, van der Weele, & Boering, 1994; Ellis & Bruton, 2002; Ellis, Bruton, & Goddard, 1997; Groth, VanDeven, Phillips, & Ehretsman, 2001; Hamilton & Lachenbruch, 1969; Hellebrandt, Duvall, & Moore, 1949). Ellis and Bruton (2002) reported that intrarater reliability of finger goniometry was within $5\%$ of the time and interrater reliability was in the range of $7\%–9\%$ of the time. Hamilton and Lachenbruch (1969) reported that a dorsal goniometer, a universal goniometer, and a special pendulum goniometer all measured within the same degree of accuracy. Groth et al. (2001) found high reliability coefficients when comparing the dorsal approach ($0.99$) to the lateral approach ($0.86$) of two goniometers: the DeVore goniometer (Hand Therapy Devices, Tucson, AZ) and a 6-in. cutoff goniometer. Although a variety of finger goniometers are available, only one study assessed ROM of finger joints using the Rolyan finger goniometer (Patterson Medical Products, Inc., Cedarburg, WI). That study, which used cadavers with wired finger joints, examined differences between the Rolyan goniometer versus a short-arm metal goniometer and a plastic long-arm goniometer. The general conclusion from that study was that goniometers with short arms are considered suitable for the dorsal method (Kato et al., 2007).
The current study examined the Rolyan goniometer because that was the instrument of choice for the participating hand therapy clinic. Clinicians adopted the use of this particular instrument because it is easier to read in its 2° increments rather than the conventional 5° increments. Moreover, ease of reading and ease of use are enhanced with its contrasting colors (blue on white) for its lines and numbers and its freely moving hinge design. The comparatively lower cost of the Rolyan goniometer compared with other goniometers is another positive aspect. Kato et al. (2007) included this instrument in their study, stating that it is a popular device. However, the reliability of this device on healthy hands has not yet been determined. The purposes of this study were to determine the intrarater and interrater reliabilities of finger goniometry using the Rolyan finger goniometer of the metacarpophalangeal (MCP), proximal interphalangeal (PIP), and distal interphalangeal (DIP) joints of the middle finger. The hypothesis was that there would be no significant differences in ROM between trials or across raters using the Rolyan finger goniometer.

Method

Participants

Twenty-seven people participated in the study. Participants were briefed on the study, and written informed consent was obtained before any data were collected. This study was reviewed and approved by the university’s and hospital’s institutional review boards before data collection.

Participants completed a brief survey that included questions about their age, gender, hand dominance, and past medical history related to upper-extremity injury. Participants were included if they were ≥19 years old and had no history of hand or finger injuries. They were excluded if they had a history of arthritis or fractures of the upper extremity, Raynaud’s syndrome, carpal tunnel syndrome or carpal tunnel release, or other neurological problems related to the upper extremity. Using healthy participants in this study controlled for the variations in ROM that can be present with injured digits, depending on the variety and a range of severity of injuries. Three participants were excluded from the study because they had a history of upper-extremity injury. Of the 24 participants enrolled in the study, 4 participants were lost because of schedule conflicts on the second day of data collection. Therefore, a total of 20 participants (14 women and 6 men) completed this study. Mean age of the participants was 45 yr (standard deviation [SD] was 15 yr). Eighteen of the participants were right-hand dominant, and 2 were left-hand dominant.

Raters

Each of the 10 therapists used at this hand clinic was invited to participate in this study. Eight raters volunteered; however, one rater did not obtain measurements on many of the participants and was dropped from the analysis. Of the seven raters included in the study, six were occupational therapists and one was a physical therapist. Four of the seven therapists were certified hand therapists. Six currently worked as hand therapists in the same hand clinic, and one was formerly used by the hand clinic but currently worked as a full-time academician. All raters had completed protection of human participants training. One rater had completed her doctoral degree; all others had completed their bachelor’s degree.

Raters’ years of clinical experience were 3–5 yr (n = 1), 11–15 yr (n = 3), 16–20 yr (n = 2), and >20 yr (n = 1). When asked about hand therapy experience (defined as a caseload of >50% of hand/upper extremity experience), only one of the seven therapists reported ≤5 yr of experience, and one therapist reported >20 yr of experience. Five raters had used the Rolyan finger goniometer for >1 yr, one had used it 6 mo to 1 yr, and one rater was using it for the first time in this study.

Instruments

Raters used the white-and-blue plastic Rolyan finger goniometer (Figure 1). The units of measurement are in degrees with a resolution of 2° increments. The long arm of the goniometer is 3.75 in., and the short arm is 0.75 in. The scale of the goniometer ranges from 30° of hyperextension to 120° of flexion.

Position for Measuring ROM

Each participant sat at a hand treatment table with his or her elbow supported on the table. The shoulder was in a position of 45°–60° flexion, neutral rotation, and 0°–15° of adduction. The forearm was in neutral rotation (in a transverse plane) with the wrist in 5°–15° extension. The following bony landmarks were used to measure the middle finger of the dominant hand:

- **MCP joint**: fulcrum: dorsally over MCP joint
- **Long arm**: over dorsal midline of proximal phalanx
- **Short arm**: over dorsal midline of metacarpal
- **PIP joint**: fulcrum: dorsally over PIP joint
- **Long arm**: over dorsal midline of proximal phalanx
- **Short arm**: over dorsal midline of middle phalanx
- **DIP joint**: fulcrum: dorsally over DIP joint
- **Long arm**: over dorsal midline of middle phalanx
- **Short arm**: over dorsal midline of distal phalanx.

Dorsal placement of the goniometer was used as recommended by the American Society for Hand Therapists (Casanova, 1992) and Norkin and White (2003). This
placement was preferred by 73% of hand therapists surveyed by Groth and Ehretsman (2001).

Procedure
An informational session describing the procedure for proper alignment of the goniometer for the three joints was described verbally and demonstrated physically 2 days before the study. A photograph of each joint and the goniometer, along with a brief description, were printed on one page for quick reference during the study. These materials were also presented during the informational session. This information was posted in the department, and a second copy was filed in a specially prepared manual along with a rationale and full proposal of the study. A pilot study was conducted 1 mo before data collection to practice the procedures, assess the amount of time needed, and work out other logistical issues associated with repeated testing of the participants. During the pilot study, raters demonstrated to the investigators their ability to follow the protocol.

Data collection occurred on 2 days scheduled 3 wk apart. Each participant was measured by no more than four raters in 1 day to avoid participant fatigue, soreness, or changes in ROM caused by repeated measurements. On the second day of data collection, the participant was measured by the remaining raters. Participants were scheduled within three time sessions. All raters began measuring at the same time in each session. Measurements for one session took approximately 35–45 min. Raters’ stations were separated from each other so that they could not hear other conversations, and each rater recorded data on a separate sheet of paper. Each participant had at least 5 min of rest between each rater. Order effects were controlled for by using a partial Latin Square Design (Portney & Watkins, 2009). Therefore, each participant was measured by four raters on 1 day and returned 3 wk later to have the remaining raters take measurements. Each rater measured each participant 3 times.

Each rater asked the participant to actively make a composite fist by saying “make a fist as best you can.” Measurements of the middle finger of the dominant hand were taken at the MCP, PIP, and DIP joints. Raters were then asked to extend actively (A) by saying “straighten your fingers all the way.” Alternating between active fisting and active extension was incorporated into the design to avoid fatigue from actively holding a fist for an extended period of time.

Passive (P) ROM measurements were then taken for flexion and extension for three trials. The participants were instructed, “Now relax your fingers and let me move them. I will measure them passively. Tell me if I am hurting you.” Passive ROM measurements were performed using a joint blocking technique to assess the joint tissues rather than extrinsic tightness that may be present. Hyperextension was not measured passively to simulate the clinical setting and avoid stretching further into an abnormal position.

Raters were surveyed on the specific methods they used for measuring ROM during the study as a follow-up to assess the consistency of the procedures. When assessing MCP flexion, all raters reported aligning the long arm of the goniometer over the dorsal midline of the proximal phalanx. When asked about DIP flexion, all raters reported aligning the long arm of the goniometer over the dorsal midline of the middle phalanx. However, there was some discrepancy with regard to assessing PIP flexion. Four raters aligned the long arm of the goniometer over the dorsal midline of the proximal phalanx, two raters aligned the long arm of the goniometer over the dorsal midline of the middle phalanx, and one rater aligned the long arm of the goniometer over the midline of the proximal phalanx for active motion and used the midline of the middle phalanx for passive motion.

When raters were asked whether they maintained the wrist in neutral flexion or extension when measuring ROM of the fingers (neutral is defined as 5° of flexion to 15° of extension), six raters responded that they did. One rater reported that the wrist was in >15° of extension when she took measurements.

Statistical Analyses
Data were analyzed using SPSS Version 15 Software (SPSS Inc., Chicago). Descriptive statistics are reported as means and SD. Repeated-measures analysis of variance (ANOVA) for each dependent ROM measure by rater (seven raters) and trial (three trials) was conducted. Those analyses with
The ICC Model 3 (Portney & Watkins, 2009), in which raters from a particular clinic were chosen nonrandomly, was used. Model 3 treats raters as a fixed effect and participants as random effects. Intraclass reliability is a measure of reliability obtained through ANOVAs measuring the correspondence and agreement between ratings (Portney & Watkins, 2009). Single-measure ICCs are reported. Coefficient of variation (CV) and 95% confidence intervals (CI) measures were calculated for each rater by type of measurement. Method error (ME) and the coefficient of ME (CVME) were calculated for Trials 2 and 3 for each ROM measurement for each rater. Trials 2 and 3 were used because ME and CVME require the use of difference scores to measure the stability of measures (Portney & Watkins, 2009). Using the latter two trials would likely produce the most stable measurements: The rationale was that in a clinical setting, learning effects would be minimal because many measurements with the

<table>
<thead>
<tr>
<th>Rater and CI</th>
<th>MCP Flexion (A) Mean (SD) [CV]</th>
<th>PIP Flexion (A) Mean (SD) [CV]</th>
<th>DIP Flexion (A) Mean (SD) [CV]</th>
<th>MCP Flexion (P) Mean (SD) [CV]</th>
<th>PIP Flexion (P) Mean (SD) [CV]</th>
<th>DIP Flexion (P) Mean (SD) [CV]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rater 1</td>
<td>85.6 (5.6) [6.5%]</td>
<td>100.1 (5.6) [5.6%]</td>
<td>72.0 (10.1) [14.0%]</td>
<td>92.1 (5.2) [5.6%]</td>
<td>107.4 (7.2) [6.7%]</td>
<td>81.9 (22.6) [27.6%]</td>
</tr>
<tr>
<td>CI</td>
<td>83.0–88.2</td>
<td>97.5–102.8</td>
<td>67.3–76.7</td>
<td>89.6–94.5</td>
<td>104.0–110.8</td>
<td>71.3–92.5</td>
</tr>
<tr>
<td>Rater 2</td>
<td>93.4 (5.5) [5.9%]</td>
<td>102.0 (5.1) [5.9%]</td>
<td>76.0 (9.7) [12.8%]</td>
<td>101.6 (6.1) [6.0%]</td>
<td>108.7 (4.3) [4.0%]</td>
<td>83.5 (10.0) [12.0%]</td>
</tr>
<tr>
<td>CI</td>
<td>90.8–96.0</td>
<td>99.6–104.4</td>
<td>71.5–80.5</td>
<td>98.7–104.5</td>
<td>106.6–110.7</td>
<td>78.9–98.2</td>
</tr>
<tr>
<td>Rater 3</td>
<td>86.6 (5.1) [5.9%]</td>
<td>98.3 (6.2) [6.3%]</td>
<td>71.8 (11.7) [16.3%]</td>
<td>92.4 (5.9) [6.4%]</td>
<td>104.8 (8.0) [7.6%]</td>
<td>74.6 (12.8) [17.2%]</td>
</tr>
<tr>
<td>CI</td>
<td>84.2–88.9</td>
<td>95.3–101.2</td>
<td>66.3–77.3</td>
<td>97.6–98.5</td>
<td>101.1–108.5</td>
<td>68.6–80.6</td>
</tr>
<tr>
<td>Rater 4</td>
<td>90.5 (4.8) [5.3%]</td>
<td>102.5 (4.1) [4.0%]</td>
<td>73.8 (11.4) [15.4%]</td>
<td>95.6 (5.7) [6.0%]</td>
<td>105.7 (5.2) [4.9%]</td>
<td>77.6 (12.0) [15.5%]</td>
</tr>
<tr>
<td>CI</td>
<td>88.3–92.7</td>
<td>100.6–104.4</td>
<td>68.4–79.1</td>
<td>92.9–98.2</td>
<td>103.2–108.1</td>
<td>72.0–83.2</td>
</tr>
<tr>
<td>Rater 5</td>
<td>88.2 (4.6) [5.2%]</td>
<td>102.6 (5.0) [4.9%]</td>
<td>71.3 (10.6) [14.9%]</td>
<td>95.6 (4.9) [5.1%]</td>
<td>109.6 (5.5) [5.0%]</td>
<td>90.6 (34.5) [38.0%]</td>
</tr>
<tr>
<td>CI</td>
<td>86.0–90.3</td>
<td>100.2–104.9</td>
<td>66.8–77.6</td>
<td>93.3–97.9</td>
<td>107.0–112.2</td>
<td>74.5–106.8</td>
</tr>
<tr>
<td>Rater 6</td>
<td>87.9 (4.9) [5.6%]</td>
<td>99.1 (4.4) [4.4%]</td>
<td>71.8 (9.9) [13.8%]</td>
<td>95.3 (3.9) [4.1%]</td>
<td>105.2 (4.9) [4.7%]</td>
<td>80.6 (9.3) [11.5%]</td>
</tr>
<tr>
<td>CI</td>
<td>85.6–90.2</td>
<td>97.0–101.1</td>
<td>67.1–76.4</td>
<td>93.9–96.6</td>
<td>102.9–107.5</td>
<td>76.2–84.9</td>
</tr>
<tr>
<td>Rater 7</td>
<td>85.3 (5.4) [6.3%]</td>
<td>97.5 (4.7) [4.8%]</td>
<td>71.3 (10.6) [14.0%]</td>
<td>91.1 (4.3) [4.3%]</td>
<td>101.1 (4.3) [4.3%]</td>
<td>80.3 (12.2) [15.2%]</td>
</tr>
<tr>
<td>CI</td>
<td>82.8–87.8</td>
<td>95.3–99.7</td>
<td>67.3–77.2</td>
<td>89.0–93.1</td>
<td>99.1–103.1</td>
<td>74.6–86.0</td>
</tr>
</tbody>
</table>

Note. Mean differences with different letters (a, b, c, and d) across raters (within a column) denote significant differences as assessed by least significant difference tests between raters. A = active; P = passive; SD = standard deviation; CV = coefficient of variation; CI = confidence interval (95%); MCP = metacarpophalangeal; PIP = proximal interphalangeal; DIP = distal interphalangeal.
Results

Mean differences between raters existed for every measurement except for passive DIP Flexion (P; Table 1). Relatively large ranges in mean ROM were found over the various measurements when comparing the rater with the smallest ROM assessed to the rater with the largest ROM assessed. The differences were less for the active movements (MCP[A]: 8.1°, PIP[A]: 5.1°, and DIP[A]: 4.7°) than for the passive movements (MCP[P]: 10.5°, PIP[P]: 8.5°, and DIP[P]: 16.0°). The CVs ranged from 4.0 (Rater 5 PIP Flexion [A] and Rater 3 PIP Flexion [P]) to 38.0 (Rater 6 DIP Flexion [P]). Although DIP Flexion (P) did not show significant differences between raters, the large CVs for each rater indicate that this was because of the large intrindividual variability for this measurement.

A summary of means and SDs for ROM × Type of Movement × Trial for all raters is shown in Table 2. The only measurement that showed significant differences across trials was active DIP Flexion (A; Table 3). Table 3 summarizes main effects (rater and trial) and interaction effects for all movements. All means were within 2° across trials. No Rater × Trial interaction effects were observed. The ICC values for each rater for his or her three trials for each measurement are reported in Table 4 (ICC range for raters across all measurements: .43–.99). Raters 6 and 7 consistently had the highest ICC values across measurements and the smallest ME and CVME. Raters 3 and 4 had the lowest ICC values across measurements and the largest ME and CVME. Table 5 summarizes ICC values across trials for all raters (ICC ranges: .24–.88). DIP Flexion (A) had the highest ICC values, whereas DIP Flexion (P) had the lowest ICC values. The ICC values for all raters across all trials by measurement are also reported in Table 5 (ICC range: .35–.85).

Table 3. Analysis of Variance Summary for Trial and Rater Effects by Joint Range of Motion

<table>
<thead>
<tr>
<th>Joint Measured</th>
<th>Main Effect Rater</th>
<th>Main Effect Trial</th>
<th>Interaction Effect Rater × Trial</th>
</tr>
</thead>
<tbody>
<tr>
<td>MCP Flexion (A)</td>
<td>.001</td>
<td>.32</td>
<td>.18</td>
</tr>
<tr>
<td>PIP Flexion (A)</td>
<td>.001</td>
<td>.14</td>
<td>.43</td>
</tr>
<tr>
<td>DIP Flexion (A)</td>
<td>.002</td>
<td>.02</td>
<td>.69</td>
</tr>
<tr>
<td>MCP Flexion (P)</td>
<td>&lt;.000</td>
<td>.94</td>
<td>.80</td>
</tr>
<tr>
<td>PIP Flexion (P)</td>
<td>&lt;.000</td>
<td>.68</td>
<td>.15</td>
</tr>
<tr>
<td>DIP Flexion (P)</td>
<td>.12</td>
<td>.45</td>
<td>.14</td>
</tr>
</tbody>
</table>

Note. A = active; P = passive.

goniometer would have been made. The equations for these measurements, as described in Portney and Watkins (2009), are as follows:

\[
\text{ME} = \frac{SD \text{ of Trial to Trial Difference Score}}{\sqrt{2}}
\]

\[
\text{CV}_{ME} = \frac{2 \text{ME}}{\text{Mean of Trial 2} + \text{Mean of Trial 3}}
\]

Results were deemed significant for two-tailed tests if \( p < .05 \).

Discussion

ROM intrarater reliability was high using the Rolyan finger goniometer. Differences in ROM >4° between trials were present <5% of the time, even with passive movements, a result that is similar to previously reported intrarater reliability measures (Ellis & Bruton, 2002). Table 5 demonstrates that active flexion had higher reliability measurements than passive flexion. This finding could be expected because with full strength and full ROM, a participant can actively flex to his or her end range with each repetition. When measuring active ROM, the therapist only has to manipulate the goniometer. When measuring passive ROM, however, the therapist must manipulate the joint passively while manipulating the goniometer. In addition, with passive ROM, the therapist is moving the joint, and pressure applied is likely to be more variable.

Table 4. Reliability Measures by Rater Across Trials

<table>
<thead>
<tr>
<th>Rater</th>
<th>MCP Flexion (A)</th>
<th>PIP Flexion (A)</th>
<th>DIP Flexion (A)</th>
<th>MCP Flexion (P)</th>
<th>PIP Flexion (P)</th>
<th>DIP Flexion (P)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ICC ME CVME</td>
<td>ICC ME CVME</td>
<td>ICC ME CVME</td>
<td>ICC ME CVME</td>
<td>ICC ME CVME</td>
<td>ICC ME CVME</td>
</tr>
<tr>
<td>Rater 1</td>
<td>0.89 1.96 2.28</td>
<td>0.92 1.87 2.18</td>
<td>0.94 2.12 2.96</td>
<td>0.84 1.87 2.04</td>
<td>0.93 2.04 1.89</td>
<td>0.99 2.12 2.62</td>
</tr>
<tr>
<td>Rater 3</td>
<td>0.68 3.47 3.73</td>
<td>0.76 2.24 2.38</td>
<td>0.84 3.84 2.55</td>
<td>0.63 3.93 3.87</td>
<td>0.43 3.73 3.44</td>
<td>0.72 5.54 6.69</td>
</tr>
<tr>
<td>Rater 4</td>
<td>0.64 3.05 3.55</td>
<td>0.68 4.32 4.38</td>
<td>0.78 7.13 4.90</td>
<td>0.76 2.50 2.70</td>
<td>0.62 5.24 4.97</td>
<td>0.86 5.07 6.76</td>
</tr>
<tr>
<td>Rater 5</td>
<td>0.79 2.13 2.37</td>
<td>0.73 2.28 2.54</td>
<td>0.96 1.89 2.57</td>
<td>0.82 2.66 2.80</td>
<td>0.79 2.33 2.21</td>
<td>0.91 3.11 3.96</td>
</tr>
<tr>
<td>Rater 6</td>
<td>0.93 1.23 1.40</td>
<td>0.91 1.38 1.34</td>
<td>0.99 1.02 1.16</td>
<td>0.84 1.65 1.75</td>
<td>0.91 1.35 1.23</td>
<td>0.99 2.98 3.34</td>
</tr>
<tr>
<td>Rater 7</td>
<td>0.93 1.23 1.41</td>
<td>0.94 1.26 1.44</td>
<td>0.99 0.71 0.98</td>
<td>0.73 1.81 1.91</td>
<td>0.90 1.71 1.62</td>
<td>0.99 0.87 1.08</td>
</tr>
<tr>
<td>Rater 8</td>
<td>0.74 2.79 3.28</td>
<td>0.84 2.06 2.12</td>
<td>0.91 3.23 4.42</td>
<td>0.57 2.37 3.60</td>
<td>0.75 2.34 2.32</td>
<td>0.78 3.65 4.66</td>
</tr>
</tbody>
</table>

Note. A = active; P = passive; ICC = intraclass coefficient; ME = method error of Trials 2 and 3; CVME = coefficient of variation of method error for Trials 2 and 3; MCP = metacarpophalangeal; PIP = proximal interphalangeal; DIP = distal interphalangeal.
Anecdotaly, participants reported that different therapists applied different amounts of pressure during passive ROM measurements, and one rater told the researchers that she has a tendency to apply lighter pressure when measuring a normal joint than when measuring a contracted joint that needs to be stretched.

Differences in interrater reliability measures were greater than intrarater differences, in agreement with previous research (Bovens, van Baak, Vrencken, Wijnen, & Verstappen, 1990; Ellis & Bruton, 2002; Ellis et al., 1997). Active DIP flexion showed the smallest range (lowest to highest mean scores) of deviations in mean rater differences of $4.7^\circ$ $(76.0^\circ$–$71.3^\circ$; Raters 3 and 8), whereas with passive DIP flexion there was a mean range difference of $16.1^\circ$ $(90.6^\circ$–$74.6^\circ$; Raters 6 and 4). Differences between raters in the various types of measurements exceeded $5^\circ >80\%$ of the time.

An examination of raters who had the highest and lowest levels of intrarater reliability showed Raters 6 and 7 were the highest, whereas Raters 3 and 4 consistently had the lowest reliabilities. However, whereas Raters 6 and 7 had good internal consistency, their agreement with each other did not always match. MCP Flexion (A), DIP Flexion (A), and MCP Flexion (P) were within $1.5^\circ$ of each other on their mean flexion measurement. However, for DIP Flexion (P), there was in excess of a $10^\circ$ difference in measurements. Hellebrandt et al. (1949) and Bovens et al. (1990) reported that therapists with more experience were more reliable. There were only seven raters, precluding a systematic analysis of reliability of measurements by experience, but it is likely that experience and time in the clinic were related to increased reliability.

These results are similar to those of Hellebrandt et al. (1949), who suggested that it may be more difficult to obtain reliable measures of the MCP joint because of the biaxial nature of the joint. The DIP joints may have had less reliability because they are shorter levers and more difficult to measure, as reported by Ellis et al. (1997). In addition, the goniometer used in this study has particularly short lever arms. The moving arm is particularly short, making it difficult to assess the DIP joints where the distal phalanx is already a short lever and difficult to manipulate. One reason this is the instrument of choice in this clinic is because the length of the stationary arm is shorter than the more traditional metal finger goniometer. The stationary arm tends to approximate the wrist while measuring MCP ROM.

Although effort was taken to standardize the procedures, one possible source of error could be that some of the therapists measured DIP passive flexion in an intrinsic minus position. Although this position is easy to obtain in a healthy population, it may not be practical in a patient population in which the PIP is less likely to have maximum ROM. Other therapists may have measured DIP passive flexion with the PIP in slight flexion. Anatomically, the oblique retinacular ligament causes the DIP joint to extend when the PIP is in extension (Tubiana, Thomine, & Mackin, 1996). Similarly, when the DIP joint is flexed, the PIP is also under tension to flex (Tubiana et al., 1996). If this indeed occurred, an important implication is that the position the finger is in when measured will partially determine the ROM actually measured. Therefore, proper training procedures and consistency in technique are important for therapists to have when assessing finger ROM. This study was limited to measuring the middle finger and used healthy, uninjured participants. Because this goniometer for these measurements was not a reliable tool, especially for intrarater assessments of finger ROM, measuring ROM on injured people would likely only produce less reliable measures.

### Conclusions

These results provide clinicians with evidence for using a specific goniometer (Rolyan finger goniometer) and issues associated with various therapists assessing finger ROM. Intrarater reliability shows that assessment of ROM of a healthy population over time by the same therapist should not be problematic but that having different therapists measure the healthy participant could
lead to intertherapist measurement error. Further studies are needed to apply this information to a patient population where less control over variables may exist. Experience and proper consistent training with this specific goniometer across the staff of therapists who may work together are keys to reducing this source of error.

Acknowledgments

We thank Pella Anderson, Rachel Blakeslee, Rebecca D’Amour, Sheela Mudholkar Zerilli, and Nancy Tabor for their hard work and dedication to this research study. We also thank all of the hand therapists at UMass Memorial Health Care and participants for their active involvement in this research study. This research was funded in part by a grant from the American Hand Therapy Foundation and the University of Massachusetts Lowell School of Health and Environment. No part of this work has been presented previously.

Disclaimer

This project was reviewed by the University of Massachusetts Lowell Office of Institutional Compliance and the University of Massachusetts Medical School Office of Institutional Compliance and classified as expedited.

References