Intra- and inter-rater reliability of goniometric finger range of motion using a written protocol

Takuya Nakai[1], Satoru Amano[2,3], Chikako Murao[1], Haruki Taguchi[1], Kayoko Takahashi[2,3]

[1]Department of Occupational Therapy, Kitasato University Hospital, Kanagawa - Japan
[2]School of Allied Health Sciences, Kitasato University, Kanagawa - Japan
[3]Graduate School of Medicine, Kitasato University, Kanagawa - Japan

ABSTRACT

Introduction: Goniometric finger range of motion (ROM) is the most common outcome measure used for functional evaluation of finger joints, but its reliability is not well-evaluated. This study aimed to investigate intra- and inter-rater reliability of goniometric finger ROM using a written protocol for active, passive, and composite movements in healthy adults.
Methods: The design was a single-center, cross-sectional, reliability study. Participants were 20 healthy adults (mean ± standard deviation, 36.4 ± 10.9 years). ROM for active, passive, and composite movements of the fingers was assessed by three occupational therapists with at least 5 years clinical experience in the field of physical disabilities. To standardize the measurement method used, we developed a written protocol, stabilized the wrist position, and trained the evaluators. Intraclass correlation coefficient (ICC) values were used for the reliability analysis. ICC (1,1) was used for intra-rater reliability. ICC (2,1) was used for inter-rater reliability. Hand-shaped heatmaps were used to summarize the reliability data.
Results: Most of the results (88.7%) showed moderate to good intra-rater reliability (ICC ≥ 0.50), while inter-rater reliability showed less (69.0%). Both intra- and inter-rater reliability showed no trends between dominant and non-dominant hands, type of movement, finger, or joint.
Conclusions: Intra-rater reliability was relatively high and using a written protocol was beneficial. Inter-rater reliability tended to be lower, and differences in the physical structure of both raters and participants may have affected inter-rater reliability values.

Keywords: Finger, Range of motion, Reliability, Reproducibility, Standardization

---

What is already known about this topic:

- *Goniometric finger range of motion (ROM) is the most common outcome measure used for functional evaluation of finger joints.*
- *However, the intra- and inter-rater reliability of finger ROM is not well-evaluated.*

What does the study add:

- *Relative intra-rater reliability was relatively high and using a written protocol was beneficial.*
- *Differences in the physical structure of raters and participants may have affected inter-rater reliability values.*
- *The results of ROM cannot be interpreted in terms of absolute reliability at 2-degree and 5-degree increments.*

Introduction

The fingers are indispensable for performance of tasks. These sophisticated body parts have motor (e.g., grasping and releasing) and sensory (e.g., touching and adjusting) functions.

Corresponding author:
Kayoko Takahashi
email: kayo.ot@kitasato-u.ac.jp

Range of motion (ROM) is one measure used for functional evaluation of the finger joints (1). When restrictions occur due to disease or disability, ROM is useful for understanding the patient's joint condition, observing changes over time, and evaluating the outcome of an intervention (2). ROM assessment is also frequently used during post-stroke upper limb rehabilitation (3). There is a consensus that ROM should be used for musculoskeletal injuries (4). Santisteban et al's (3) review found that ROM is not only a traditional tool. It remains a first choice for measurement of outcomes associated with the body function categories of the International Classification of Functioning, Disability, and Health. In addition, due to the current emphasis on evidence-based medicine, the need for objective and reliable measures is increasing rapidly.

There are only a few standardized protocols available for finger ROM measurement (e.g., "Methods for Indication and Measurement of Joint Range of Motion" by the Japanese Orthopaedic Association and the Japanese Society of Rehabilitation Medicine (5), *Measurement of Joint Motion: A Guide to Goniometry*, fifth edition by Norkin and White (6)). However, other than definition of the basic and moving axes, some procedures of measurement are not consistent among references. Therefore, repetition of measurements and limb positions can vary across examiners. In clinical settings, examiner bias can be high because therapists commonly use the goniometer manually. Although several previous studies have been reported on the reliability of finger ROM measurement using goniometers, most of them were limited to the certain fingers/joints (5-9) and movement type (10).

Sato et al (11) examined intra- and inter-rater reliability of finger ROM at 2- versus 5-degree intervals. They found that the error was smaller for the 2-degree interval measurement than for the 5-degree interval measurement. This result suggested that smaller angle changes can be captured using a goniometer with smaller measurement intervals. Therefore, it is necessary to verify intra- and inter-rater reliability for all fingers, joints, and types of movement (active, passive, and composite). Thus, the purpose of this study was to investigate the intra- and inter-rater reliability of goniometric finger ROM using a written protocol for active, passive, and composite movements in healthy adults.

Methods

Research design

We used an observational, descriptive study design to examine the intra- and inter-rater reliability of a new protocol for goniometric measurement of finger motions. The risk of bias of the present study was assessed using the COSMIN checklist (Reliability: relative measures) in the supplementary tables. The Kitasato University School of Medicine and Hospital Ethics Committee (2020-027) approved this study.

Participants

The participants were recruited from among the staff members of the hospital where the first author was employed. The exclusion criteria were as follows: (1) history of musculoskeletal condition, such as arthritis, orthopedic conditions involving the upper limbs, (2) neurological, (3) psychiatric conditions, and (4) an unstable general condition due to other complications.

Evaluator

Finger ROM was assessed by three occupational therapists (TN, CM, HT) with ≥5 years of clinical experience in the field of physical disabilities (Rater A/B/C, mean years of experience: 8.3 years).

Procedure

We developed a measurement protocol manual that was based on "Joint Range of Motion Indication and Measurement Methods" by the Japanese Orthopaedic Association and Japanese Society of Rehabilitation Medicine (5) and *Measurement Evaluation for PT/OT: ROM Measurement*, Second Edition (2). To ensure uniformity of the measurement method used, raters received a 15-minute course on the contents of the written protocol and trained for 15 minutes individually using the measurement manual.

Each participant was seated in a chair facing the table with the assessed side of the arm placed on the table. The forearm position was 0-degree rotation with a 20-degree wrist dorsiflexion. A sheet of paper with a diagram of the basic fixed axis was placed under the arm as a guide (Fig. 1A). The goniometer was placed from the dorsal side of the hand with the long handle (with fixed axis) on the basic axis and the short handle (with meter printed) on the moving axis (Fig. 1B). The thumb was measured first, followed by the index, middle, ring, and little fingers. Measurement of each finger followed the order of metacarpophalangeal (MP), proximal interphalangeal (PIP), and distal interphalangeal (DIP) joints.

First, active (voluntary) movement was measured with the accompanying verbal instruction, "Please bend XX joint of your XX finger utmost, without moving your wrist." If other fingers were flexed at the same time, the raters instructed the participant to "try to move only your XX (targeted) finger." Second, passive ROM was measured in the same order, with the instructions, "Please relax and let me bend your XX finger's XX joint to the maximum." While measuring the MP joint, extreme flexion of the interphalangeal (IP) joint was avoided, and it kept its natural orientation. The MP and DIP joints were straightened (0-degree flexion/extension) during PIP joint measurement. When the DIP joint was measured, the MP joint was straightened (0-degree flexion/extension) with the PIP joint flexed at 70-90 degrees.

Last, active composite movements of all finger flexion positions were performed following the same orders. The instructions were, "Please bend all fingers utmost without moving your wrist." The thumb was placed closely over the basal phalange of the index finger to avoid interfering with ROM of the other fingers. If the goniometer could not fully contact the joint, we allowed measurement on a line parallel to the basic axis and axis of movement.

All three raters measured all participants twice with at least 24-hour interval to test intra-rater reliability. For inter-rater reliability, the dates of assessment were distributed so a participant was not assessed by more than one rater on the same day. Before each assessment, it was confirmed with the participants that there had been no injury or change in hand function since the last assessment. Assessment was conducted individually in a separate room to ensure the other raters were blinded, and discussion or comparison between rates was strictly prohibited.

Data analysis

Intraclass correlation coefficients (ICCs) were used for the relative reliability analysis (ICC (1,1) for intraclass reliability, ICC (2,1) for inter-rater reliability) (12). R (version 4.0.2) was used for the statistical analysis. We used heatmaps to summarize the reliability data because the study included a large number of values, based on 366 ICC calculations. Heatmaps were
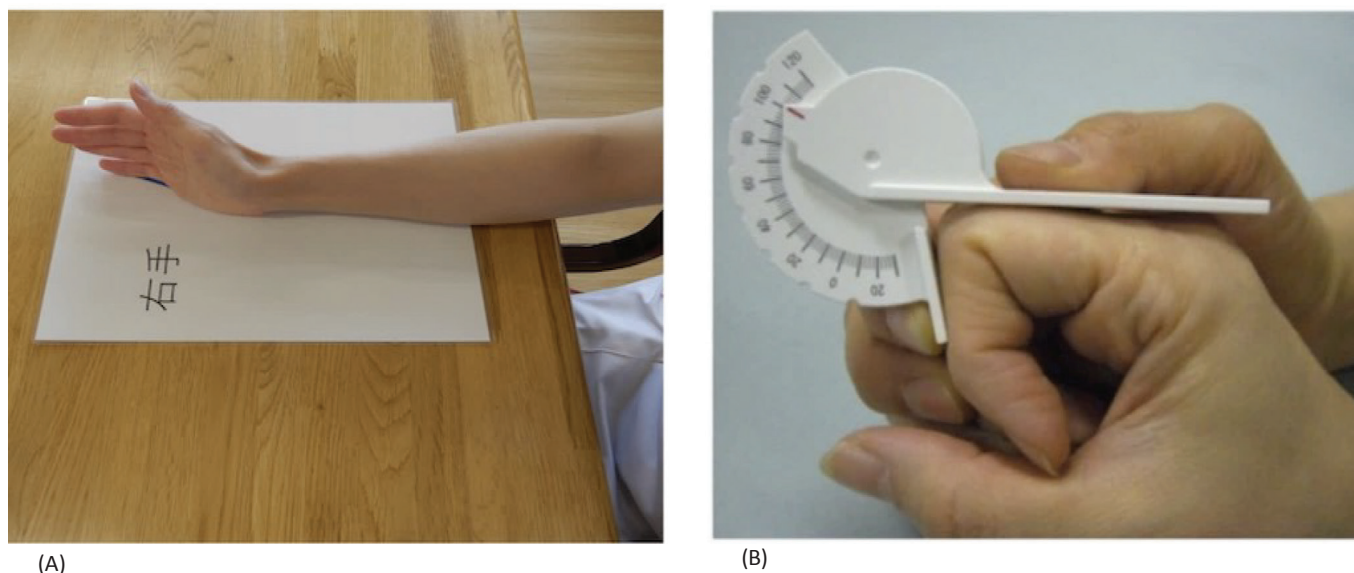
(A)

(B)

FIGURE 1 - Goniometric measurement of finger range of motion. A) Alignment of wrist during finger measurement. The axis of movement and basic axis of wrist dorsiflexion are shown on a sheet placed on a desk, so that the 20-degree dorsiflexion fixation is not displaced during measurement. The paper with both the fixed and moving axes was placed under the arm to stabilize the 20-degree dorsiflexion of the measured arm. B) Placement of goniometer on finger. The goniometer was placed from the dorsal side of the hand with the long handle (with fixed axis) on the basic axis and the short handle (with meter printed) on the moving axis. Note: Numerical values are measured to the first digit in 2-degree increments.

also of great value for presentations based on the shape of the hand. However, because heatmaps alone did not include all necessary information, we provide ICC precision data for more in-depth interpretation (Supplementary Table). In addition, minimal detectable change (MDC) was calculated for absolute reliability. The standard error of measurement (SEM) was used to calculate an MDC value with the following formula: $MDC_{95} = 1.96*\sqrt{(2)}*(SEM)$. A SEM value was calculated as $\sqrt{\sigma^2_{error}}$ (square root of the error variance) (13).

Results

Participant demographic characteristics

Twenty healthy adults were included in this study; no participants met the exclusion criteria and no data were missing. The mean ± standard deviation age of the participants was 36.4 ± 10.9 years (33.8 ± 8.3 years) for males and 40.3 ± 13.1 years for females, 40% were female, and 90% were right-handed (Tab. 1).

TABLE 1 - Demographic characteristics of participants (N = 20)

Characteristics	N (%)
Gender, N (%)	
Male	12 (60)
Female	8 (40)
Age, mean (SD)	36.4 (10.9)
Dominant hand, N (%)	
Right	18 (90)
Left	2 (10)

SD = standard deviation.

Relative intra- and inter-rater reliability

Figure 2-5 presents the results for the heatmap of intra-rater reliability of each rater and inter-rater reliability among the three raters. A darker red color indicated a higher ICC value; a lighter color indicated a lower ICC value. Both intra- and inter-rater reliability values showed no trends between dominant and non-dominant hand, type of movement, finger, or joint. Rater C's heatmap tended to be lighter than that of rater A or B. Reliability results varied among the different raters. Compared with intra-rater reliability (Figure 2-4), ICC values for inter-rater reliability were generally low (Figure 5). Detailed ICC information, including precision data, is presented in the supplementary tables.

Absolute intra- and inter-rater reliability

Both intra- and inter-rater reliability values showed no clear trends between dominant and non-dominant hand, type of movement, finger, or joint. Absolute reliability varied depending on the different evaluators, but in many cases MDC fitted between 10 and 15. Compared with intra-rater reliability, MDC values for inter-rater reliability were generally high. Detailed MDC and SEM information, including precision data, is presented in the supplementary tables.

Discussion

This study examined the intra- and inter-rater reliability of goniometric finger ROM measurements with ICC using a written protocol for various type of movements in healthy adults. Koo and Li (12) define moderate reliability (ICC 0.5-0.75), good reliability (ICC 0.75-0.90), and excellent reliability
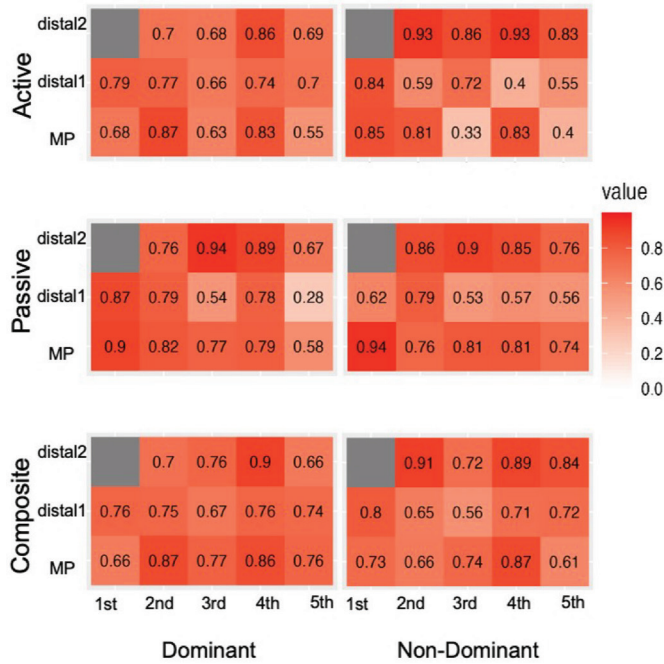
FIGURE 2 - Intraclass correlation coefficient (ICC) values for intra-rater A reliability. Darker red color indicates higher ICC, lighter color indicates lower ICC. The number represents the type of finger. Detailed ICC information and standard error of the measurement (SEM), including precision data, are presented in the supplementary tables.
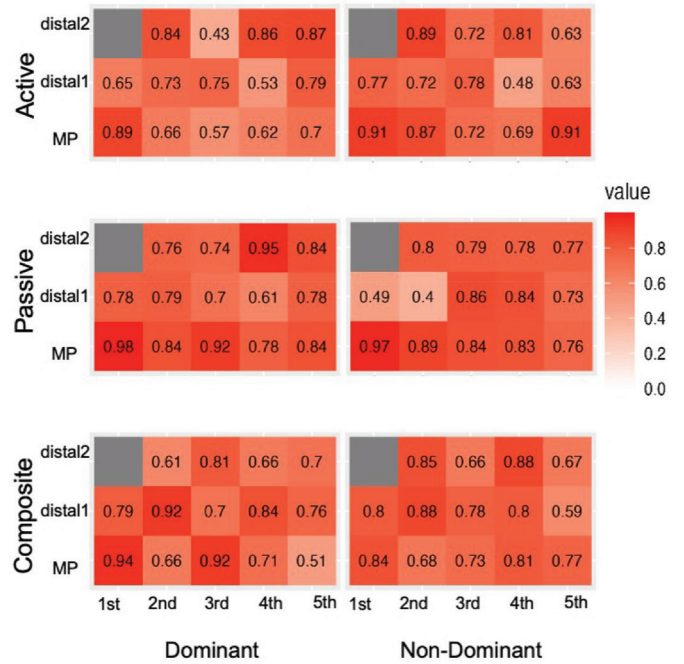

FIGURE 3 - Intraclass correlation coefficient (ICC) values for intra-rater B reliability. Darker red color indicates higher ICC, lighter color indicates lower ICC. The number represents the type of finger. Detailed ICC information and standard error of the measurement (SEM), including precision data, are presented in the supplementary tables.
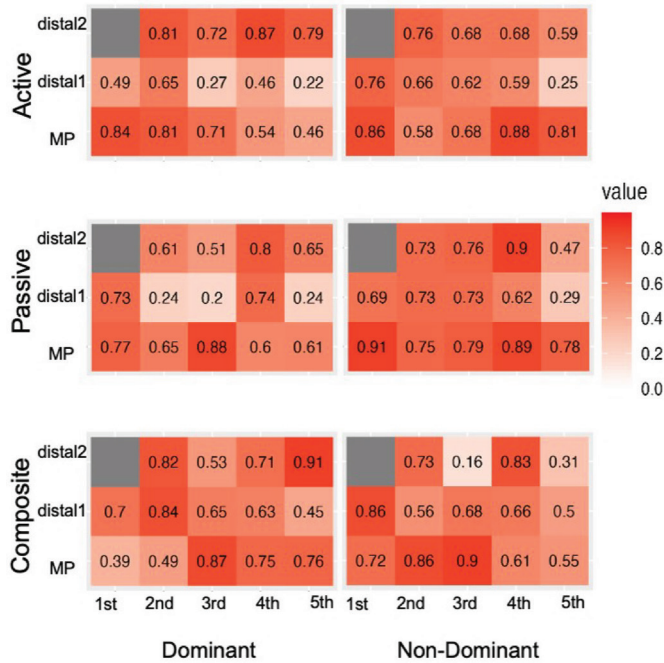

FIGURE 4 - Intraclass correlation coefficient (ICC) values for intra-rater C reliability. Darker red color indicates higher ICC, lighter color indicates lower ICC. The number represents the type of finger. Detailed ICC information and standard error of the measurement (SEM), including precision data, are presented in the supplementary tables.

FIGURE 5 - Intraclass correlation coefficient (ICC) values for inter-rater reliability. Note: Darker red color indicates higher ICC, lighter color indicates lower ICC. The number represents the type of finger. Detailed ICC information and standard error of the measurement (SEM), including precision data, are presented in the supplementary tables.
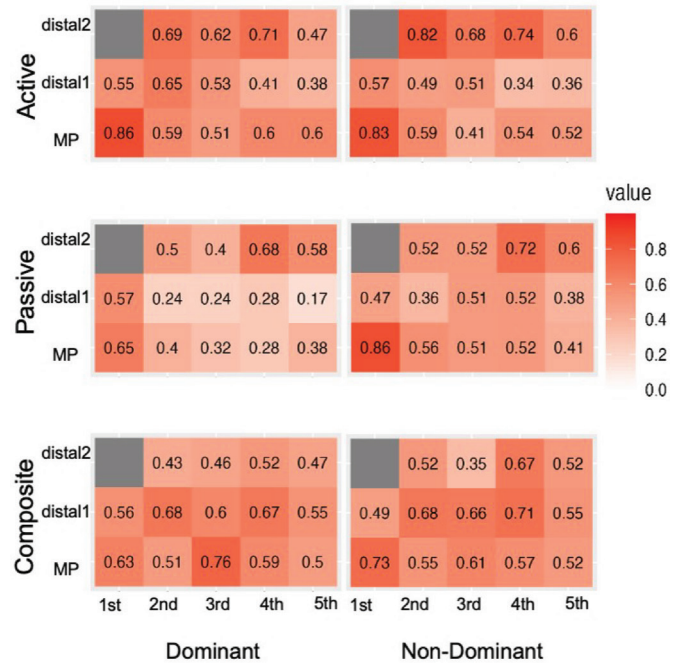
(ICC ≥0.90). This study had a certain degree of reliability in intra-rater reliability. Whereas the ICC tended to have lower inter-rater reliability than intra-rater reliability, the results supported previous studies.

Relative intra-rater reliability

Heatmap analysis revealed a constant dark red color that indicated the presence of a relatively high intra-rater reliability. There were only a few differences in reliability, depending on the type of movement (active, passive/composite), dominant or non-dominant hand, and each finger and each joint. Lewis et al (10) examined intra-rater reliability of the MP, PIP, and DIP joints of the middle finger of the dominant hand in 20 healthy adults. The raters were 10 therapists using Rolyan goniometers to measure both active and passive movement. The ICC values ranged from 0.43 to 0.99. The rater with the highest reliability had ICC values of 0.84-0.99; the rater with the lowest reliability had ICC values of 0.43-0.84. In this study, rater A had the highest reliability (ICC 0.66-0.90 for active composite movement). Thus, the results of this study had acceptable reliability. A certain degree of intra-rater reliability was achieved because we developed a measurement protocol and used raters who were trained to ensure good reproducibility.

Relative inter-rater reliability

For inter-rater reliability, heatmap analysis revealed lighter red color than intra-rater reliability that indicated inter-rater reliability was relatively low compared with the ICC values of intra-rater reliability. Similar results for low inter-rater reliability for finger ROM measurements, compared with intra-rater reliability, have been published (9,10,14). Lewis et al (10) found that inter-rater reliability is lower than intra-rater reliability with ICC values in the range of 0.35-0.85. They also found that errors in ROM angle were due to biarticular muscles and short DIP joints. Ellis et al (14) found that inter-examiner measurements are less reliable than intra-examiner measurements for the comparative reliability of finger ROM measurements using goniometry and wire tracing. They included the amount of force applied to the goniometer, the accuracy of alignment during goniometer application, and identification of anatomical landmarks as reasons for inconsistent measurement outcomes with respect to errors in goniometer measurements. Short et al (15) mentioned that the size of the rater's body (height difference) may affect the interpretation of goniometer readings. In our study, the maximum palm lengths of each rater varied from 19.5, 17.5, and 16.3 cm (average 17.8 cm), and the hand size of each participant also varied. Handling difficulties due to differences in the body structure of both raters and participants may have affected measurement consistency.

Absolute reliability

Measurement error was considered as absolute reliability. Even if the interpretation of relative reliability was acceptable, the results of absolute reliability may not be clinically acceptable. However, rather than clearly judging it to be "clinically unusable," we would like to recommend that medical professionals leave it to the "system" for interpreting ROM. The Mayo Wrist Score (16) is a good example of a practice that takes this approach. In section 3 of the assessment (regarding ROM), the assessment is based on an ordinal scale in increments of approximately 25%, with emphasis on % normal. Even if the ROM is clinically acceptable in terms of relative reliability, medical clinicians should pay attention to the results of this study, which show that the results cannot be interpreted in terms of absolute reliability at 2-degree and 5-degree increments.

Strength of this study

The strength of our study is that we verified the reliability of all active, passive, and composite movements of all joints in all fingers of the participants' dominant and non-dominant hands. In previous studies (7,8), the validation was limited to certain fingers, joints, and types of movement, and this study was the first to compare and validate the results by all joints, fingers, and types of movement. In the clinical setting, ROM should be measured at all affected joints and fingers, and ROM of different types of movement would help define the problem and plan the intervention. Therefore, the results of this study contributed to the field of hand therapy by validating all fingers, all joints, and various types of movements. The results also indicated that a certain degree of intra-rater reliability was obtained.

As with other assessments (17,18), the creation of a manual to reduce variation in measurement methods among raters may have contributed to a certain reliability. In our study, ROM was measured using a written protocol, and multiple trainings were conducted among raters. These components could have helped standardize the measurement methods and improved reliability. ROM angle is significantly affected by the position of the proximal joint. Thus, our manual, with its concrete description of wrist position, could have minimized rater bias and error.

Limitations and direction for future research

One of the limitations of this study is the sample size. According to Borg et al (19), a sample size estimation for a reliability study with three raters requires an ICC planning value of 0.8, a minimum acceptable reliability of 0.6, a power of 80%, and an alpha equal to 0.05, with a necessary sample size of 33 patients. However, the small variability observed in ROM scores in this study may have mitigated the impact of the small sample size on the reliability results. For these reasons, future studies with larger sample sizes are warranted to confirm our findings, particularly in cases involving diseases or pathologies that result in limitations in hand ROM.

The ROM measurement procedure was designed to measure all types of movement of both the dominant and non-dominant hands and to measure all fingers and joints twice; 30 to 40 minutes were required to measure ROM for each participant. This time constraint could have negatively affected rater concentration and the ability to accurately interpret the goniometer scale. Future research should be modified to better reflect actual clinical settings.

This study was a single-center study, and future validation in multicenter studies are recommended. It is also possible that differences in the physical structure of both the raters and participants affected inter-rater reliability. Future validation studies should consider the effects of different body structures of both raters and participants.

Conclusions

This study examined the intra- and inter-rater reliabilities of finger ROM in healthy adults using a finger goniometer. The results indicated that relative intra-rater reliability was relatively acceptable and that inter-rater reliability tended to be lower than intra-rater reliability. In clinical practice, having the same rater is recommended to achieve a certain degree of reliability, regardless of the type of movement or joint, and to capture finger ROM changes over time.

Acknowledgments

Disclosures

Conflict of interest: There is no conflict of interest for all authors.

Authors' contributions: Conceptualization: NT, SA, KT; Data Curation: NT, CM, HT; Formal Analysis: NT, SA; Funding Acquisition: KT; Investigation: NT, SA, KT; Methodology, NT, SA, KT; Project Administration: NT, CM, HT, SA, KT; Resources: KT, NT; Supervision: KT; Validation: SA; Visualization: SA, KT; Writing – Original Draft: NT, KT; Writing – Review and Editing: NT, CM, HT, SA, KT.

Data availability statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to the nature of this research.

References

1. Shiraishi H. What is the evaluation of hand function from various evaluations on the hand. SOBIM Japan. 2010;34(4):291-296. CrossRef

2. Fukuda O. Measurement evaluation for PT/OT: ROM measurement. 2nd ed. MIWA SHOTEN; 2010.

3. Santisteban L, Térémetz M, Bleton J-P, Baron JC, Maier MA, Lindberg PG. Upper limb outcome measures used in stroke rehabilitation studies: a systematic literature review. PLoS One. 2016;11(5):e0154792. CrossRef PubMed

4. Crasto JA, Sayari AJ, Gray RR-L, Askari M. Comparative analysis of photograph-based clinical goniometry to standard techniques. Hand (N Y). 2015;10(2):248-253. CrossRef PubMed

5. Yonemoto K, Isigami S, Kondo T. Joint range of motion and measurement methods. Jpn J Rehabil Med. 1995;32(4):207-217. CrossRef

6. Norkin CC, White DJ. Measurement of joint motion: a guide to goniometry. 5th ed. F A Davis Co; 2016.

7. Hamilton GF, Lachenbruch PA. Reliability of goniometers in assessing finger joint angle. Phys Ther. 1969;49(5):465-469. CrossRef PubMed

8. Burr N, Pratt AL, Stott D. Inter-rater and intra-rater reliability when measuring interphalangeal joints: comparison between three hand-held goniometers. Physiotherapy. 2003;89(11):641-652. CrossRef

9. Ellis B, Bruton A. A study to compare the reliability of composite finger flexion with goniometry for measurement of range of motion in the hand. Clin Rehabil. 2002;16(5):562-570. CrossRef PubMed

10. Lewis E, Fors L, Tharion WJ. Interrater and intrarater reliability of finger goniometric measurements. Am J Occup Ther. 2010;64(4):555-561. CrossRef PubMed

11. Sato A, Oi H, Abe Y. Reliability of hand joint angle measurements using the manual for hand joint range of motion measurements. Jpn Soc Surg Hand. 2015;31(4):512-516.

12. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med. 2016;15(2):155-163. CrossRef PubMed

13. Stratford PW, Goldsmith CH. Use of the standard error as a reliability index of interest: an applied example using elbow flexor strength data. Phys Ther. 1997;77(7):745-750. CrossRef PubMed

14. Ellis B, Bruton A, Goddard JR. Joint angle measurement: a comparative study of the reliability of goniometry and wire tracing for the hand. Clin Rehabil. 1997;11(4):314-320. CrossRef PubMed

15. Short N, Almonreoder T, Mays M, et al. Interrater reliability of a novel goniometric technique to measure scapular protraction and retraction. Am J Occup Ther. 2022;76(1):7601205010. CrossRef PubMed

16. Cooney WP, Bussey R, Dobyns JH, Linscheid RL. Difficult wrist fractures. Perilunate fracture-dislocations of the wrist. Clin Orthop Relat Res. 1987;(214):136-147. PubMed

17. Platz T, Pinkowski C, van Wijck F, Kim I-H, di Bella P, Johnson G. Reliability and validity of arm function assessment with standardized guidelines for the Fugl-Meyer Test, Action Research Arm Test and Box and Block Test: a multicentre study. Clin Rehabil. 2005;19(4):404-411. CrossRef PubMed

18. See J, Dodakian L, Chou C, et al. A standardized approach to the Fugl-Meyer assessment and its implications for clinical trials. Neurorehabil Neural Repair. 2013;27(8):732-741. CrossRef PubMed

19. Borg DN, Bach AJE, O'Brien JL, Sainani KL. Calculating sample size for reliability studies. PM R. 2022;14(8):1018-1025. CrossRef PubMed