



Research Article

Volume 7 Issue 5 - March 2020
DOI: 10.19080/JPFMTS.2020.07.555725

J Phy Fit Treatment & Sports

Copyright © All rights are reserved by James McKivigan

Reliability and Validity of a Smartphone-Based Inclinometer Application Measuring Shoulder Internal Rotation



Taylor Lau, Sheng Lin, Tyler True, Wayne Wu and James McKivigan*

Touro University Nevada, School of Physical Therapy, USA

Submission: February 27, 2020; **Published:** March 16, 2020

***Corresponding author:** James McKivigan, Touro University Nevada, School of Physical Therapy, 874 American Pacific Drive, NV 89122, Henderson, USA

Abstract

This study intended to evaluate the reliability and validity of a smartphone-based inclinometer application and compare these results to those of the standard clinical goniometer. The study measured the internal shoulder rotation of 19 men and 20 women. One third-year physical therapy student took all the smartphone-based inclinometer measurements, and another took all the clinical standard goniometer measurements. The subjects were randomly placed into set amounts of internal rotation, and then the two measurements were taken. The study found no significant difference between the smartphone-based inclinometer and the goniometer. The reliability between the app and the goniometers was good to excellent.

Keywords: Smartphone application; Inclinometer; Range-of-motion; Goniometer

Abbreviations: ROM: Range of Motion; RFG: Rate Fast Goniometer; UG: Universal Goniometer; SBI: Smartphone-Based Inclinometer; ICC: Intraclass Correlation

Introduction

The goniometer is an instrument used to measure angles mainly the range of motion (ROM) of joints. The goniometer, the gold standard clinical device to measure ROM, is a quick and inexpensive way to measure joint angles. Although clinicians frequently use it, there can be a measurement error of plus or minus five degrees when using a goniometer. Chapleau, Canet & Rouleau [1] determined that maximal errors of goniometric measurements for elbow ROM ranged from 6.5-10.3 degrees per measurement. The goniometer tool has also been reported to have poor inter-rater reliability [1]. Another tool, the inclinometer, is defined as a device for measuring angles among different body parts, for example, specific bones or joints. It can also be employed to establish the relative motion of these structures during active or passive bending. Current literature has shown that the smartphone-based inclinometer, Goniometer Pro, possesses good to excellent intra-rater and inter-rater reliability and concurrent validity for measurement of wrist ROM. Our study will assess a different application, the Rate Fast Goniometer (RFG), and its ability to reliably and accurately measure shoulder internal rotation through the raising of the arm by a movement at the shoulder. We hypothesize that the RFG, an application very similar to Goniometer Pro, will be significantly more valid and

reliable and will have more intra-rater reliability when compared to the universal goniometer (UG) for measuring shoulder internal rotation.

According to recent literature, the rapid expansion of technology has made smartphones a convenient, cheap, quick way to measure ROM, at the extent of the movement of a joint, measured in degrees of a circle. Previous studies have found that smartphone-based inclinometer (SBI) applications are reliable and valid when measuring the elbow, wrist, knee, and ankle. Barker and his team at Marshall University performed a study where they used an SBI application to measure the ROM of the knee and compared it to a manual goniometer tool (2016). They found that the SBI application had a smaller measurement error and superior reproducibility when compared to the UG [2]. Other articles reported similar findings when measuring the elbow, wrist, and ankle. However, there is minimal research available on the validity and reliability of SBI applications measuring the shoulder. Our study aims to compare the validity, reliability, and intra-rater reliability of the SBI to that of the UG. We chose to compare the SBI to a UG because the UG is considered the clinical gold standard for measuring ROM [3].

A UG was compared to an SBI to assess the validity, reliability, and minimal detectable change. It was found that the SBI had a higher intraclass correlation (ICC) and a smaller standard error of measurement when compared to the UG. The minimal detectable change was the same for both tools, which suggests there is a smaller change required to infer a noticeable change in measurement for each instrument [2]. In another study, the ROM of 60 healthy volunteers was measured using an SBI and a UG. The study found that the SBI had higher reliability and validity in comparison to the UG. The subjects were examined on elbow flexion, pronation, and supination, where the reliability was highest in pronation and lowest in flexion. The SBI was also easier to access and use when used in a clinic [4]. The cost of the goniometer app ranges from zero to five dollars. However, clinicians need to own a phone with a gyro-sensor to utilize the application, which may increase the overall cost of using it. Kolber, Fuller, Marshall, Wright, and Hanney suggest that the difference in measurement of shoulder ROM between a UG and SBI can be expected to range from 2-20 degrees. Previous studies have compared goniometers to SBIs but have not compared those to measurements set by a licensed physical therapist.

The purpose of this study is to evaluate the SBI application and the UG in a clinical setting against a degree set by a licensed physical therapist. Recent research has found the SBI to be cheaper and more convenient, valid, and reliable compared to a goniometer in the clinical setting [3]. If we find that the SBI is more reliable and valid compared to a UG, clinicians will be able to use the SBI instead of the UG in the clinic. We hypothesize that the SBI will have better validity, reliability, and intra-rater reliability compared to the UG when measuring internal shoulder rotation.

Methods

Subjects

Using a convenience sample, 39 healthy adult volunteer subjects were recruited to participate in the study. All volunteers were students at Touro University Nevada. The subjects were recruited via written communication through email and by word of mouth. The inclusion criteria for subject participation included age between 21 and 40, a normal ROM, and the ability to stand for 45 minutes without discomfort. Subjects were screened for exclusion criteria, which included shoulder pathologies, shoulder pain, abnormal ROM, previous shoulder surgical procedures or injections, and an allergy to Sharpie Permanent Markers. The dominant arm of each subject was used in the study, identified by which arm the subject used to throw a ball. If the subject did not throw, the dominant arm was identified by what hand they used for writing. We recruited 19 men and 20 women, of which 2 were right-arm dominant, and 37 were left-arm dominant.

Measures/materials

Shoulder internal rotation was measured for each subject using two methods: the clinical standard goniometer and the RFG

smartphone-based application on the iPhone. The application was downloaded from the Apple App Store. Clinician number one performed all the standard goniometer measurements, and clinician number two performed all the RFG measurements (Alchemy Logic Systems Inc, 2014). The clinician using the RFG used an Apple iPhone 7s + with a three-axis gyro and accelerometer. Each clinician served as an expert in using their respective measurement tool. Both clinicians had the same training and experience and were third year Doctor of Physical Therapy Students at Touro University Nevada. The clinicians measured the subject's shoulder internal rotation, the degree of which was set and maintained by the licensed physical therapist. The degree of shoulder internal rotation was randomly selected and set by a third clinician. Randomization was used to ensure that both clinicians performing the measurements, using the clinical standard goniometer and the RFG, were blind to the set value and that no advantages were given to either clinician.

Procedure

Subjects were recruited via convenience sampling at Touro University Nevada, via email and through word of mouth. There were no incentives given to participating subjects. Data collection was completed in 2018 at the Touro University Physical Therapy Research Lab. Upon entering the research facility, the subjects were greeted by all four clinicians. The subjects were thanked for their participation in the study and were then verbally informed about the procedure and the data collection. The subjects were informed that there would only be one 20-minute session of data collection. They were then read the informed consent document by one of the clinicians and were asked to sign and initial it if they agreed to participate in the study and understood the informed consent. The subjects were also informed of the risks and rewards of this study. Risks included possible slight epidermal irritation from the Sharpie marking, and the benefits included assisting in the advancement of implementing modern technology into the clinical setting. The subjects had the option to decline to sign the photography and videography form if they were not comfortable with being photographed or videoed. If the subject had hearing deficits, they could read the informed consent and media consent on their own and sign and initial when completed. The subjects were informed that their data would remain anonymous and that a numerical system would be used to log the data; their name would not appear anywhere within the study. After obtaining the participants' consent, we had the subjects answer a questionnaire asking about past shoulder pathologies, pain, surgeries, injections, and any other factors that might exclude them from the study. An active ROM quick screen was completed to ensure the subjects had a normal ROM: 80 to 100 degrees (ACSM, 2013). All measurement tools had been calibrated before the subjects' arrival. The RFG was calibrated to 0 degrees horizontal and vertical using a level. Once the patients were cleared for participation, the two expert clinicians who were measuring the internal shoulder rotation exited the room. The third clinician used a random number

generator (random.org) to determine the ROM to be set by the licensed physical therapist. Then the third clinician positioned the subject, and the licensed physical therapist set the random joint angle.

All measurements in the study were standardized, and the standard anatomical landmarks for measuring internal shoulder rotation were used. These landmarks included the stationary arm being aligned with the olecranon of the humerus and the moving arm being aligned with the ulnar styloid process [5]. Each clinician measured shoulder internal rotation while standing. The licensed physical therapist set a random angle. The first clinician coming into the room used the RFG application on the iPhone. They placed the iPhone in alignment with the olecranon and ulnar styloid process and then pressed the "Start" button at the top of the screen. The screen showed the angle of measurement, and the third clinician wrote it down. The first clinician then exited the room. Next, the second clinician entered the room with a UG. This person aligned the stationary arm to vertical, which was the 0 degrees set point. They then moved the moving arm of the goniometer in alignment with the ulnar styloid process and read the measurement to the third clinician to note. All measurements were recorded by the third clinician on an Excel sheet.

Statistical Analysis

Statistical analysis was performed using data from the goniometer and RFG measurements. Descriptive statistics (mean, standard deviation) using customary procedures were calculated for descriptive and anthropometric variables. Intra-rater reliability was examined using intraclass correlation coefficients (ICC 2,1) and p-value. The levels of reliability were either excellent ($ICC > 0.80$), good ($0.80 > ICC > 0.60$), moderate ($0.60 > ICC > 0.40$), or poor ($ICC < 0.40$). A very high correlation was represented by a p-value higher than 0.7, while coefficients between 0.7 and 0.5 showed moderate correlation. Values between 0.5 and 0.3 were considered poor correlation. For criterion validity, p-values and Pearson correlation coefficients were calculated to examine the association between smartphone photographic and inertial measurements.

Accuracy was determined by taking the goniometric and RFG data and calculating the percent error compared to the angle set by the licensed physical therapist. The standard error of measurement (SEM) was used in the formula: $SEM = SD \sqrt{1-r}$. The SEM was measured to estimate the repeated measures of the testers to find their true score. By finding the SEM, we calculated the reliability of the clinical standard goniometer and the RFG application. The smaller the SEM score, the increased reliability, with the opposite also being true. An analysis was done using JASP version 0.8.2.0 for Windows and Mac. The accuracy, validity, and reliability measurements for each measuring device were then compared using a paired t-test. This test was chosen because there was only one variable being compared the type of measurement tool. The null hypothesis was that there were no differences

between the data gathered by the clinical standard goniometer and that collected by the RFG smartphone application.

Literature Review

A study by Chapleau et al. [1] found that the maximal error of goniometric measurement was 10.3 degrees 95% of the time. This study revealed that the UG had a high range of error, and it pushed us to research a new method of measuring ROM. Our study analyzes the accuracy, reliability, and validity of a clinical standard goniometer and compares those factors to those of a smartphone-based goniometer. Our study measures the shoulder internal rotation ROM of 35 subjects, both male and female. Subjects were required to meet several inclusion criteria, including an age between 18 and 40, having a full, healthy shoulder ROM, and being able to stand for 45 minutes. Werner et al. [3] used similar inclusion criteria in their study, in which they measured the accuracy and reliability of a smartphone inclinometer application and compared these results to those from a gold standard goniometer. Subjects were required to answer a series of screening questions to screen for our exclusion criteria, which included shoulder pain, limited ROM, any prior shoulder surgeries or injections, or any known shoulder pathologies. By screening subjects for these inclusion and exclusion criteria, Werner and colleagues were able to minimize the amount of subject dropout and ensure accurate, reliable, valid results.

To determine the gold standard for clinical ROM measurement, [1] studied the validity of goniometric measurements of ROM for the elbow when compared to the radiographic method. In their study, they established the goniometer as a clinical gold standard. This study is related to our study because we wanted to compare the accuracy, reliability, and validity of a smartphone-based goniometer to the clinical gold standard. We wished to determine which method of measurement was best by comparing the measurements of the smartphone-based goniometer to the measurements made by a licensed physical therapist. To find a reasonable new method of measuring ROM, we looked at Barker et al. [2], who designed a study to compare a UG to a smartphone goniometer application, to test reliability, validity, and minimal detectable change. The testers in that study found that when measuring knee ROM, the ICC of the smartphone application used was 0.97 and 0.94 for knee flexion and extension, respectively, compared to that for the UG, which was 0.95 and 0.87. The standard error of the measure for flexion and extension was 2.72 & 1.18 degrees, respectively, for the smartphone application compared to that for the UG, which was 3.41 and 1.62 degrees. This information was relevant to our study because we wanted to determine how a smartphone application compared to a UG. Several other studies have also analyzed the results of an application-based goniometer or inclinometer. For example, [4] tested the reliability of the UG versus a digital inclinometer in measuring elbow ROM. They concluded that the smartphone application had a higher ICC than the UG at 0.95, 0.98, and 0.98 for elbow flexion, pronation, and

supination, respectively, compared to the goniometer at 0.77, 0.79, and 0.91. The reason for testing the two tools is that smartphone applications are becoming more available to practitioners. They are readily available and user-friendly, and if testing by multiple practitioners shows them consistently simple and user-friendly, they could be possible substitutions or replacements for the UG. This action is relevant to the study because having a smartphone application that reduces error and is easier to use can give clinicians a better, more reliable alternative to the UG. Kolber et al. [6] studied how digital inclinometer applications of a smartphone application are more reliable and valid than UGs for measuring the shoulder in different ranges of motion. The reason these researchers tested the two tools was that UGs require both hands to use and can exhibit a higher risk of human error while measuring. On the other side of the argument, digital inclinometers are more portable and lightweight. They are more consistent in that they can establish a zero point of measurement in the application, so there is a reduced error in measurement. This article is relevant to our study because it tested the validity and reliability of a smartphone digital inclinometer application compared to a UG with different users taking measurements.

Johnson et al. [7] studied the inter-rater and intra-rater reliability of UGs versus smartphone application digital inclinometers. These authors used a UG as the base measurement and had three physical therapists who were experienced with a UG measure shoulder abduction. They covered the angles on the UG and took those measurements and compared them to the digital inclinometer of the smartphone application to test for inter-rater and intra-rater reliability. The testers found that both tools were reliable in repeated measurements, with an average concordance correlation coefficient of 0.997 and a standard deviation of ± 4 degrees. This study is relevant to ours because we wanted to determine the inter-rater and intra-rater reliability of both tools when measuring the shoulder. We took the results of this study and compared the inter-rater and intra-rater reliability of both those tools to the angle set by the licensed physical therapist. We adopted aspects of Werner et al.'s [3] procedure into our study. This study identified the patient's dominant arm by which arm they threw with, and if the subject did not throw, the researchers determined hand dominance using writing-handedness. We used these same parameters to define the dominant arm. The study also established a set of questions to screen patients for shoulder pathologies. The patients needed to be free of pain, have full ROM, and have no prior shoulder injections or surgeries. We adopted these parameters to identify subjects for our study. Werner et al. [3] also demonstrated that SBI obtained measurements that agreed jointly with their clinical gold standard: the goniometer. Their procedure also showed a good correlation among different providers' skill levels. Russo et al. [8] also studied how the experience affected the measurement of joint ROM of the shoulder, elbow, hip, and knee. In this study, three investigators (orthopedic

surgeons, physical therapists, and residents) with different types and levels of training were determined to make accurate and precise measurements. The precision level was similar for all shoulder motions in this study. For this reason, our third year DPT clinician needed to be able to obtain accurate measurements in our study.

Cools et al. [5] also provided a protocol to measure internal rotation using a goniometer and an inclinometer. Their protocol demonstrated good to excellent reliability in measuring internal ROM (ICC, 0.85-0.99). For this reason, we incorporated their testing position and procedure used to obtain measurements for goniometer and hand-held inclinometer. An inadequacy of this study was that patient position and equipment possibly influenced the results. We kept this in mind and standardized the participants' position as well as the equipment to minimize errors. Lastly, Cuesta-Vargas & Roldán-Jiménez [9] studied the reliability and validity of a picture-based application for measuring shoulder abduction. The smartphone-based application showed an ICC for intra-reliability and inter-reliability higher than 0.956. The results from the present study are in line with those results, showing higher levels of reliability and validity for shoulder abduction when measuring ROM using the application. Measurements for the picture-based measurement and inclinometer-based measurements were more highly correlated (Pearson $r=0.963$). This article was relevant to our study because it could demonstrate an alternative method of measuring ROM to explore in the future, due to its similar reliability and validity. Researchers can compare the accuracy and reliability of these two smartphone-based apps to determine the better smartphone-based application.

Data Collection/Analysis

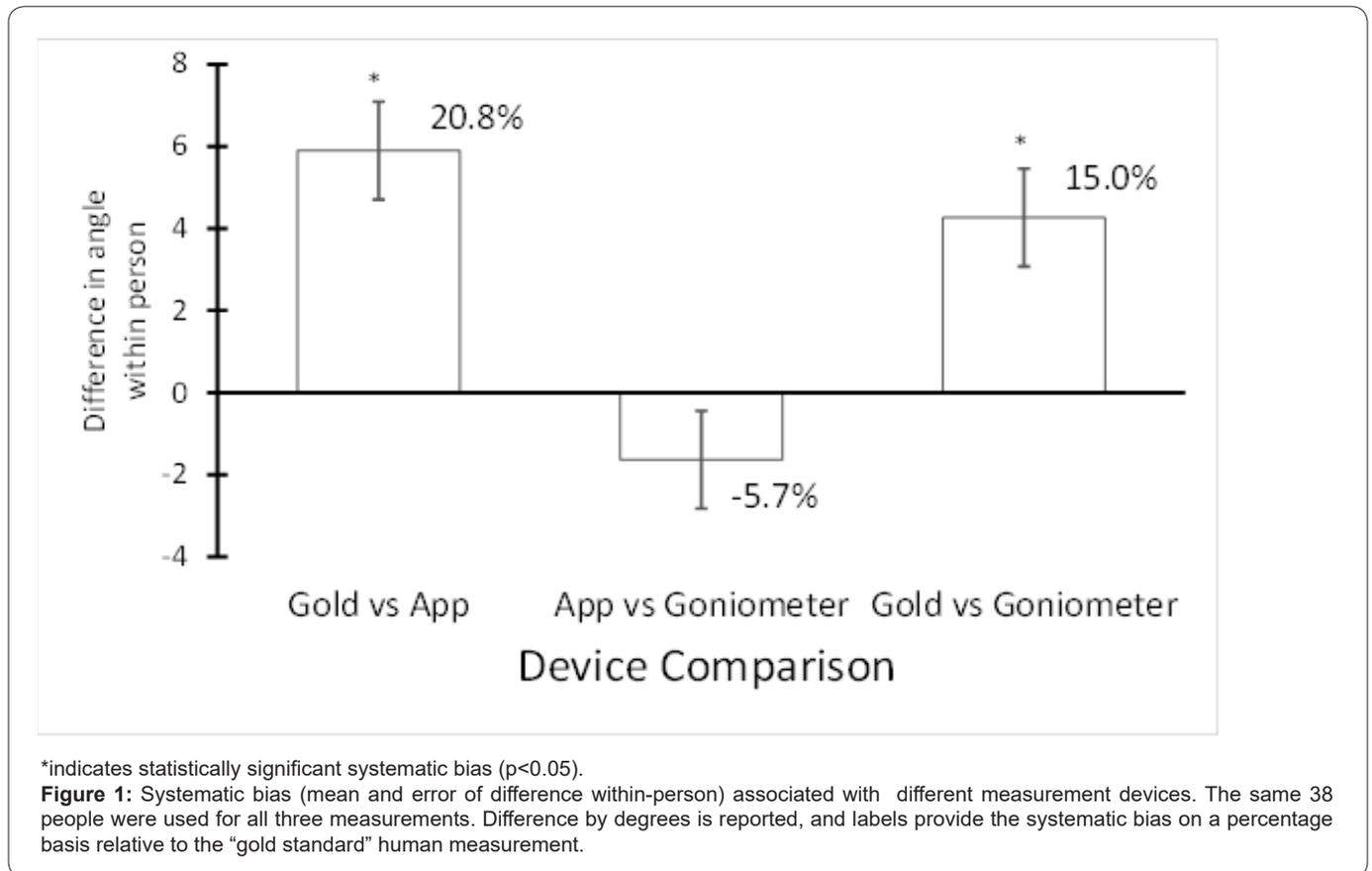
Analysis methods. Systematic bias was estimated by looking at the effect of the measurement device, using a mixed-effects model ANOVA with the subject as a random effect, followed by a pairwise Tukey test between methods (the fixed effect). The random error was assessed using the intraclass correlation coefficient (ICC; type (3, k) in Shrout & Fleiss [10,11], based on a single rating, looking for consistency, with a 2-way mixed-effects model. The analysis assumed that subject by device interactions were not present. The 95% confidence intervals for the ICC were based on the F distribution. Analyses were done in Rv3.5.0 [12-14].

Result

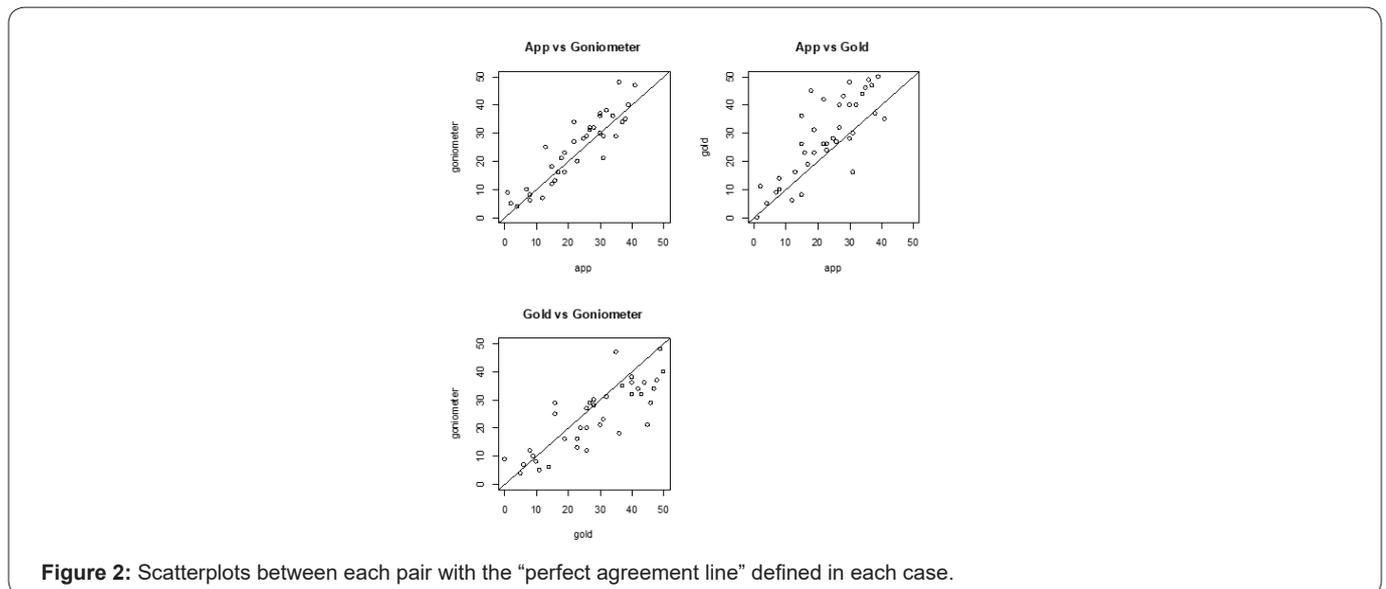
The human-measured gold standard measurements were significantly larger than measurements from the two devices ($F=13.1$, $df=2$, 74 , $P<0.001$; Tukey test p -value <0.05 ; Figure 1). The app and the goniometer were not significantly different ($p>0.05$). Based on the ratings from Koo and Li (2016) [15-18], reliability between the human gold standard and the app or goniometer was moderate to good. The reliability between the app and the goniometers was good to excellent (Table 1).

Table 1: Intraclass correlation coefficients and 95% confidence intervals for estimating consistency between pairs of devices for measuring joint flexion.

	ICC	95% Confidence Interval	
Gold and App	0.691	0.483	0.826
App and Goniometer	0.895	0.808	0.944
Gold and Goniometer	0.771	0.603	0.874



Discussion



The standard goniometer demonstrated a smaller percent difference relative to the gold standard (15.0% vs. 20.8%). However, the standard goniometer was not significantly more accurate than the RFG. One explanation for this variance may be the volunteers' muscle fatigue. The test position required that a volunteer hold the set angle at 90 degrees of abduction and 90 degrees of elbow flexion, a position that can be quickly fatiguing. In our study, the clinician measuring with the RFG always measured second, which can account for the percent difference. To improve this study for future use, we need to address threats to the study, such as muscle fatigue and instrumentation, which are further discussed in the limitations section. Reliability between the human gold standard and both the app and goniometer was moderate to good. However, our study failed to produce the good to excellent reliability found in Barker et al. [2]. This situation was likely due to error introduced during the data collection, including muscle fatigue and instrumentation error. These errors were likely significant enough to interfere with the data analysis as well, negating the findings of this study.

Limitations

There are several limitations to the validity and reproducibility of this study. The first limitation we encountered was an error introduced during patient positioning before measurement by the two clinicians. Even though we standardized our approach to measuring by adapting the protocol used by Cools et al. [5], we did not account for shoulder fatigue setting in so quickly. During the first day of data collection, our clinicians noticed that the volunteer would occasionally drift out of position after being set by the licensed clinician. This drift could have increased the percent error of both measurements when compared to the intended degree. The second clinician to measure would be more exposed to error due to such fatigue. To rectify this error, the gold standard clinician would reset the arm position after the first clinician took the standard goniometer measurement. In future studies, a change to the protocol to measure volunteers in a supine position will likely increase the accuracy and validity of the study. This position requires less muscle activation to hold and will likely decrease errors in data collection. Another threat to the internal validity of the study involves instrumentation. As the clinicians gained experience taking measurements using the RFG, likely, their accuracy improved as well. Through this logic, we can expect data measurements to become more accurate and reliable with practice. To standardize this for future protocols, extensive training before data collection is recommended for smartphone goniometers such as RFG.

References

1. Chapleau J, Canet F, Petit Y, Laflamme GY, Rouleau DM (2011) Validity of goniometric elbow measurements: Comparative study with a radiographic method. *Clinical Orthopaedics and Related Research* 469(11): 3134-3140.
2. Barker K, Bowman B, Galloway H, Oliashirazi N, Oliashirazi A, et al. (2016) Reliability, concurrent validity, and minimal detectable change for iPhone goniometer app in assessing knee range of motion. *The Journal of Knee Surgery* 30(6): 577-584.
3. Werner BC, Holzgrefe RE, Griffin JW, Lyons ML, Cosgrove, et al. (2014) Validation of an innovative method of shoulder range-of-motion measurement using a smartphone clinometer application. *Journal of Shoulder and Elbow Surgery* 23(11): e275-e282.
4. Behnouth B, Tavakoli N, Bazmi E, Nateghi Fard F, Pourgharib Shahi M, et al. (2016) Smartphone and universal goniometer for measurement of elbow joint motions: A comparative study. *Asian Journal of Sports Medicine* 7(2): e30668.
5. Cools A, De Wilde L, Van Tongel A, Ceysens C, Ryckewaert R, et al. (2014) Measuring shoulder external and internal rotation strength and range of motion: Comprehensive intra-rater and inter-rater reliability study of several testing protocols. *Journal of Shoulder and Elbow Surgery* 23(10): 1454-1461.
6. Kolber M, Fuller C, Marshall J, Wright A, Hanney W (2011) The reliability and concurrent validity of scapular plane shoulder elevation measurements using a digital inclinometer and goniometer. *Physiotherapy Theory and Practice* 28(2): 161-168.
7. Johnson L, Sumner S, Duong T, Yan P, Bajcsy R, et al. (2015) Validity and reliability of smartphone magnetometer-based goniometer evaluation of shoulder abduction: A pilot study. *Manual Therapy* 20(6): 777-782.
8. Russo RR, Burn MB, Ismaili SK, Brayden GB, Han S, et al. (2018) How does level and type of experience affect measurement of joint range of motion? *Journal of Surgical Education* 75(3): 739-749.
9. Cuesta-Vargas A, Roldán-Jiménez C (2016) Validity and reliability of arm abduction angle measured on smartphone: A cross-sectional study. *BMC Musculoskeletal Disorders* 17(1).
10. Shrout PE, Fleiss JL (1979) Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin* 86(2): 420-428.
11. McGraw KO, Wong SP (1996) Forming inferences about some intraclass correlation coefficients. *Psychological Methods* 1(4): 30-46.
12. Core Team R (2018) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
13. Bates D, Maechler M, Bolker B, Walker S (2015) Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*. 67(1): 1-48.
14. Martire RL (2017) Package "rel:" Reliability Coefficients.
15. Koo TK, Li MY (2016) A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine* 15(2): 155-163.
16. Ager A, Roy J, Roos M, Belley A, Cools A, Hébert L (2017) Shoulder proprioception: How is it measured and is it reliable? A systematic review. *Journal of Hand Therapy* 30(2): 221-231.
17. Dougherty J, Walmsley S, Osmotherly P (2015) Passive range of movement of the shoulder: A standardized method for measurement and assessment of intrarater reliability. *Journal of Manipulative and Physiological Therapeutics* 38(3): 218-224.
18. Shin S, Ro D, Lee O, Oh J, Kim S (2012) Within-day reliability of shoulder range of motion measurement with a smartphone. *Manual Therapy* 17(4): 298-304.



This work is licensed under Creative Commons Attribution 4.0 License
DOI: [10.19080/JPFMTS.2020.07.555725](https://doi.org/10.19080/JPFMTS.2020.07.555725)

**Your next submission with Juniper Publishers
will reach you the below assets**

- Quality Editorial service
- Swift Peer Review
- Reprints availability
- E-prints Service
- Manuscript Podcast for convenient understanding
- Global attainment for your research
- Manuscript accessibility in different formats
(Pdf, E-pub, Full Text, Audio)
- Unceasing customer service

Track the below URL for one-step submission
<https://juniperpublishers.com/online-submission.php>