J. Phys. Ther. Sci. 29: 1869–1873, 2017

The Journal of Physical Therapy Science

Original Article

Reliability and validity of angle measurements using radiograph and smartphone applications: experimental research on protractor

Takenori Awatani, RPT, MS[1, 2)*], Taisuke Enoki, PhD[3)], Ikuhiro Morikita, MD, PhD[2, 4)]

[1)] *Faculty of Sports Science, Kyushu Kyoritsu University: 1-8 Jiyuugaoka, Yahatanishi-ku, Kitakyushu, Fukuoka 807-8585, Japan*
[2)] *Graduate School of Sport and Exercise Sciences, Osaka University of Health and Sport Sciences, Japan*
[3)] *Faculty of Education, Osaka Kyoiku University, Japan*
[4)] *Faculty of Physical Education, Osaka University of Health and Sport Science, Japan*

Abstract. [Purpose] The present study aimed to demonstrate the following by using measurements for the definite angles provided by the digital protractor: inter-rater reliability and validity in radiograph measurements and smartphone application measurements. [Subjects and Methods] The subject angles were 26 angles between 15° and 180° that were selected randomly using a computer. Three examiners measured the angles using the radiograph and smartphone application. The radiograph was obtained at a position 250 cm from the chest shooting cassette holder. The smartphone photograph was obtained at positions 50, 100, 150, 200, and 250 cm from the holder. [Results] Under all conditions, intra-class correlation coefficients showed 0.999. The correlation coefficient was 0.999 for all conditions. The mean absolute difference to the protractor was ≤0.28° for all conditions. [Conclusion] In comparison with the protractor, radiograph measurements and smartphone application measurements, the results of the present study showed high inter-rater reliability, validity, and small error. The results indicated that radiograph and smartphone application measurements could be used as criteria of validity in angle measurements. It supported the legitimacy of high-quality previous studies that used radiograph measurements as a criterion for validity.

Key words: Radiograph measurement, Smartphone application measurement, Measurement error

(This article was submitted May 23, 2017, and was accepted Jul. 24, 2017)

INTRODUCTION

In various clinical practices, angle measurements have been used for assessing physical alignment and articular range of motion. In particular, for clinical diagnosis, angle measurements have been utilized as information for evaluating physical characteristics, determining the effects of rehabilitation, and preventing injury induced by physical activity. For a number of clinical practices, a goniometer is commonly used to measure the joint angles. However, clinical angle measurements using the goniometer have lower reliability and validity[1, 2)]. Radiograph measurements (RMs) are strongly recommended for reliable measurements of several joint angles to improve the accuracy[3)]. However, RM also has disadvantages such as radiation exposure, higher medical costs, and a specialized medical implementation site. Therefore, a non-invasive, simple, and low cost measurement method that has high correlation to RM would be beneficial for various clinical practices.

In recent years, angle measurements using smartphone applications (apps)[4–6)] or digital photographs[7–10)] have been clinically researched as new angle measurement methods. These innovative methods have shown high reliability[4–10)]. In addition, previous studies have demonstrated that these new measurement methods had high correlation with RM, which was used as a criterion for validity[6, 8–11)].

*Corresponding author. Takenori Awatani (E-mail:awtn9831@gmail.com)

In 2013, Ege et al.[11] used smartphone application measurement (AM) methods to analyze the radiograph image. However, the RM images analyzed by the AM method were captured using the radiographic image from the computer display, not directly captured from the actual joint. Thus, regarding validation, the study of Ege et al.[11] does not have sufficient clarity to ensure consistency between RM and digital photographic measurement. In addition, previous studies showed correlation and reliability coefficients within the measurements. Thus, the measurement error was not sufficiently demonstrated[8, 9, 11].

Therefore, the present study aimed to demonstrate the following three research aspects using the measurement of definite angles provided from the digital protractor: (1) inter-rater reliability and validity in RM, (2) inter-rater reliability and validity in AM, and (3) measurement error in RM and AM.

SUBJECTS AND METHODS

The sample size of the study was calculated as 10 using effect size of 0.9, statistical power of 0.99, and significance level of 1% with reference to the results (r=0.96) of Nix et al.[9] G*Power 3.1.9.2 statistical software was used for calculating the sample size.

The subject angles were 30 angles between 0° and 180° that were selected randomly using a computer. Angles were selected in units of 0.01°. Imaging subjects were assessed using a digital protractor (protractor) (DUP-360; Niigata seiki Co., Ltd., Japan). The protractor (the blade; 150mm) was set at the defined angle, and the image was acquired by both a smartphone camera (iPhone5, Apple Inc., USA) and radiography (70 kV, 200 mA, 0.25 s; Toshiba Medical Systems Co., Ltd., Japan). After measuring the 30 angles, four angles between 0° and 15° could not be measured because they did not fit the measurement axis in the radiographic image editor software (Dyna V500, Medical Standard Co., Ltd., Japan).

One operation coordinator captured all the photographs. Three examiners measured the angles using the radiograph and app. The examiners did not have knowledge of the subject angles measured by the coordinator. The examiners had no previous experience in angle measurement using the radiograph or app. Based on the Declaration of Helsinki, all examiners were explained about this investigation and confirmed the consent before the investigation.

The images of the protractor were taken at the same time using the radiograph and smartphone. The radiograph was obtained at a position 250 cm from the chest shooting cassette holder (holder). The smartphone photograph was obtained at positions 50, 100, 150, 200, and 250 cm from the holder (Fig 1). The projected center of the radiograph was aligned to the center of the protractor. The smartphone was set perpendicular to the floor and protractor, and the camera lens was aligned to the center of the protractor.

The angle in the radiograph was measured so that the measurement axis was parallel to the blade of the protractor. The angle was measured in intervals of 0.1°. The RM and AM measurements were separated by a 1-month interval to prevent examiner bias.

The angle measurement using the app (Measurement of ROM Angle −use 4 markers, Apple Application store) and devise (Apple Inc., iPad mini, USA) was performed using five condition photographs that were obtained at positions 50 cm (AM50), 100 cm (AM100), 150 cm (AM150), 200 cm (AM200), and 250 cm (AM250) from the holder (Fig. 1). Images were edited and standardized to the size of AM50. Measurements performed for each condition were separated by a 1-day interval and were performed once for each image. The measurement order was randomized using a computer system. The angle in the app was measured so that the measurement axis was parallel to the blade of the protractor, after the image was enlarged in the app. The measurement was performed using a touch pen for smartphones (Jot Pro, Adonit Corp., USA). The angle was measured in intervals of 0.01°.

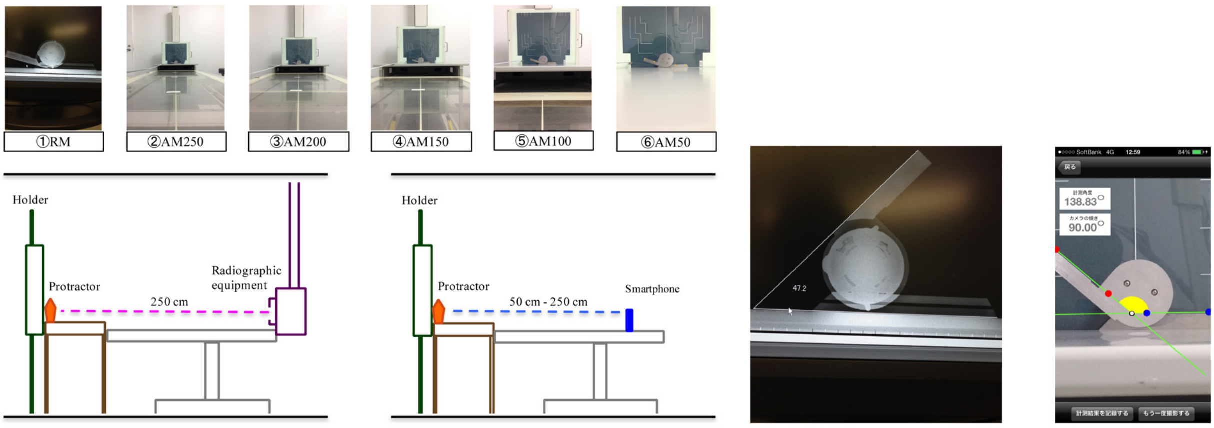
Descriptive data were calculated for each measurement, including means and standard deviations (SD). Inter-rater reliability was determined using intra-class correlation coefficients ($ICC_{2,1}$) and associated 95% confidence intervals (95% CIs). The ICC values were assessed according to the criteria of Landis et al.[12]: 0.00 to 0.20, slight; 0.21 to 0.40, fair; 0.41 to 0.60, moderate; 0.61 to 0.80, substantial; and 0.81 to 1.00, almost perfect. The inter-rater measurement error value was calculated using the standard error of measurement (SEM) using the SD of the differences ($SD_{difference}$) between the two measurements[13], 95% CI of the minimal detectable change (MDC)[14]. The validity was determined using Pearson's correlation coefficient. Furthermore, the measurement error was determined using the 95% limits of agreement (LOA)[15] and the absolute difference (AD)[9] to protractor. The measurement error was also determined for RM. Statistical analyses were performed with R 2.8.1 and Excel for Windows 2010 (Microsoft Japan Co., Ltd.), p values less than 0.05 were considered statistically significant.

$$LOA=\bar{x} \pm 1.96 \times SD$$
$$SEM=SD_{difference} / \sqrt{2}$$
$$MDC=SEM \times \sqrt{2} \times 1.96$$

RESULTS

In all conditions, $ICC_{2,1}$ showed 0.999 (0.999 to 1.00, almost perfect) (Table 1). For reliability, SEM and MDC showed ≤0.36° and 0.99°, respectively (Table 1).

The correlation coefficient was 0.999 (0.999 to 1.00) under all conditions, and significant correlation was observed



a. b. c.

Fig. 1. Setup and measurements
a; Setup: The images of the protractor were taken at the same time using the radiograph and smartphone. The radiograph was obtained at a position 250 cm from the chest shooting cassette holder (holder). The smartphone photograph was obtained at positions 50, 100, 150, 200, and 250 cm from the holder. The projected center of the radiograph was aligned to the center of the protractor. The smartphone was set perpendicular to the floor and protractor, and the camera lens was aligned to the center of the protractor.
b; Radiographic measurement (RM): The angle in the radiograph was measured so that the measurement axis was parallel to the blade of the protractor.
c; Smartphone application measurement (AM): Images were edited and standardized to the size of AM50. The angle in the app was measured so that the measurement axis was parallel to the blade of the protractor, after the image was enlarged in the app.

(Table 2). The mean difference (LOA) to the protractor and mean difference (LOA) to RM showed ≤0.22° (−0.30° to 0.74°) and 0.16° (−0.49° to 0.81°), respectively (Table 3). Mean AD (95%CI) to the protractor and mean AD (95%CI) to RM showed ≤0.28° (0.20° to 0.36°) and 0.28° (0.17° to 0.39°), respectively (Table 3).

DISCUSSION

The AM method was developed to provide high inter-rater reliability and validity. Moreover, AM is performed using only simple digitizing and does not require complicated techniques. Many app measurements are performed by moving three markers, but our measurements were performed in a detailed setting by moving four markers. However, the basic function of measuring the image is similar to other applications, and this study is not designed to discuss the difference between applications.

The irradiation angle of X-rays is spread in cases with short distances between the X-ray generator and the subject. Then, the magnification is changed in the subject, and RM may cause an error in the measured angle. Thus, the condition of RM was to place 250 cm long distances between the X-ray generator and the subject. In addition, the radiograph editing software could not measure angles close to 0°. Therefore, four angles were excluded, and the subject of this study included 26 angles. Celebić et al. reported to the effect of camera-object distance on accuracy of the dimension[16]. The photographs using a simple digital camera should be taken from a distance from 70 cm to 1 m [16]. In addition, 35 cm was unacceptable[16]. Thus, the conditions of AM were to place 50 cm to 250 cm distances between the smartphone and the subject.

In a previous study of inter-rater reliability, Ferriero et al.[4,5] reported a high ICC value for the app at the elbow (ICC=0.998; 95%CI=0.996 to 0.999) and knee (ICC=0.994; 95%CI=0.989 to 0.997). Nix et al.[9] reported ICC, SEM, and MDC_{90} for radiographs ($ICC_{2,1}$=0.99; 95% CI=0.98 to 1.00; SEM=1°; MDC_{90}=2.4°) and photographs in hallux ($ICC_{2,1}$=0.96; 95% CI=0.93 to 0.98; SEM=2°; MDC_{90}=4.7°). The $ICC_{2,1}$ value of the present study was 0.999 (0.999 to 1.00, almost perfect) in all conditions (Table 1), SEM for RM and AM250 showed 0.15° and 0.36° respectively (Table 1). The RM and AM showed high inter-rater reliability that was consistent with previous studies. The inter-rater reliability was also shown to have exceptional results within 1.00°. Therefore, RM and AM were defined as measurement methods with high precision.

In previous studies of validity, Nix et al.[9] reported very high correlations of 0.98 to 0.96 obtained by comparing RM and photographic measurement. Moreover, Ege et al.[11] reported very high correlation of 0.99 by comparing RM and AM. The results of the present study showed that the significant correlation coefficient represents 0.999 (0.999 to 1.00) in all conditions. Moreover, RM and AM were shown to have high correlation that was consistent with previous studies. Therefore, RM and AM were defined as measurement methods with high accuracy.

The measurement error was examined using LOA and AD. Naylor et al.[8] reported the comparison of RM and photographic measurement, showing that the mean difference was within 1.6° on all occasions and the LOA was within 5° to 10° in flexion of knee. Nix et al.[9] reported that the measurement error showed mean AD (1.9°), mean difference (−0.7°), and LOA

Table 1. Inter-rater reliability and error for angle measurements by RM and AM

	Mean ± SD (°)	Difference of examiner	SEM (°)	MDC (°)	$ICC_{2,1}$
		Mean ± 2SD (LOA) (°)			Mean (95%CI)
Protractor	115.33 ± 42.83				
RM	115.27 ± 42.92	0.03 ± 0.41 (−0.38 to 0.44)	0.15	0.41	0.999 (0.999 to 1.00)
AM250	115.17 ± 42.19	0.07 ± 0.72 (−0.65 to 0.79)	0.26	0.72	0.999 (0.999 to 1.00)
AM200	115.21 ± 42.92	0.09 ± 0.99 (−0.90 to 1.08)	0.36	0.99	0.999 (0.999 to 1.00)
AM150	115.26 ± 42.81	0.09 ± 0.93 (−0.84 to 1.02)	0.34	0.93	0.999 (0.999 to 1.00)
AM100	115.12 ± 42.87	0.10 ± 0.75 (−0.65 to 0.85)	0.27	0.75	0.999 (0.999 to 1.00)
AM50	115.27 ± 42.87	0.09 ± 0.76 (−0.67 to 0.85)	0.27	0.76	0.999 (0.999 to 1.00)

RM: radiographic measurement; AM: smartphone application measurement; SD: standard deviation; LOA: 95% limits of agreement; SEM: the standard error of measurement; MDC: 95% confidence interval of minimal detectable change; ICC: interclass correlation coefficient; CI: confidence interval

Table 2. Validity for angle measurements by RM and AM

	RM	AM250	AM200	AM150	AM100	AM50
Protractor	0.999 * (0.999 to 1.00)	0.999 * (0.999 to 1.00)	0.999 * (0.999 to 1.00)	0.999 * (0.999 to 1.00)	0.999 * (0.999 to 1.00)	0.999 * (0.999 to 1.00)
RM		0.999 * (0.999 to 1.00)	0.999 * (0.999 to 1.00)	0.999 * (0.999 to 1.00)	0.999 * (0.999 to 1.00)	0.999 * (0.999 to 1.00)
AM250			0.999 * (0.999 to 1.00)	0.999 * (0.999 to 1.00)	0.999 * (0.999 to 1.00)	0.999 * (0.999 to 1.00)
AM200				0.999 * (0.999 to 1.00)	0.999 * (0.999 to 1.00)	0.999 * (0.999 to 1.00)
AM150					0.999 * (0.999 to 1.00)	0.999 * (0.999 to 1.00)
AM100						0.999 * (0.999 to 1.00)

Correlation coefficient and p values for Pearson's correlation coefficient.
Mean (95% confidence interval)
*p<0.001
RM: radiographic measurement; AM: smartphone application measurement

Table 3. Measurement error for angle measurements by RM and AM

	Difference	AD	SEM	MDC
	Mean ± 2SD (LOA)	Mean (95%CI)		
Protractor vs. RM	0.06 ± 0.28 (−0.22 to 0.34)	0.12 (0.08 to 0.16)	0.10	0.28
Protractor vs. AM250	0.16 ± 0.35 (−0.19 to 0.51)	0.19 (0.13 to 0.25)	0.13	0.35
Protractor vs. AM200	0.12 ± 0.40 (−0.28 to 0.52)	0.20 (0.15 to 0.25)	0.14	0.40
Protractor vs. AM150	0.17 ± 0.40 (−0.23 to 0.57)	0.21 (0.15 to 0.27)	0.14	0.40
Protractor vs. AM100	0.22 ± 0.52 (−0.30 to 0.74)	0.28 (0.20 to 0.36)	0.19	0.52
Protractor vs. AM50	0.06 ± 0.77 (−0.71 to 0.83)	0.27 (0.16 to 0.38)	0.28	0.77
RM vs. AM250	0.10 ± 0.40 (−0.30 to 0.50)	0.17 (0.11 to 0.23)	0.14	0.40
RM vs. AM200	0.06 ± 0.47 (−0.41 to 0.53)	0.21 (0.16 to 0.26)	0.17	0.47
RM vs. AM150	0.12 ± 0.52 (−0.40 to 0.64)	0.21 (0.14 to 0.28)	0.19	0.52
RM vs. AM100	0.16 ± 0.65 (−0.49 to 0.81)	0.27 (0.18 to 0.36)	0.23	0.65
RM vs. AM50	0.01 ± 0.81 (−0.80 to 0.82)	0.28 (0.17 to 0.39)	0.29	0.81

Unit: degree (°)
RM: radiographic measurement; AM: smartphone application measurement; SD: standard deviation; AD: absolute difference; CI: confidence interval; SEM: the standard error of measurement; MDC: 95% confidence interval of minimal detectable change

(−4.9° to 3.5°) in the comparison of RM and photographic measurement. Ege et al.[11] reported that the measurement error was 1.25° ± 1.02° when the RM and app measurements were compared. The results for RM and AM in the present study had a 0.10° mean difference (LOA= −0.30° to 0.50°) in AM250, which was a small error compared to previous reports. In the comparison with the definite angle, RM was shown to have a 0.06° mean difference (LOA= −0.22° to 0.34°) and 0.12° mean AD (95% CI=0.08° to 0.16°), while AM250 was shown to have a 0.16° mean difference (LOA= −0.2° to 0.51°) and 0.19° mean AD (95% CI=0.13° to 0.25°). Both RM and AM showed high accuracy and small error similar to previous studies. The present study used identical images. Therefore, this result indicated the error of digitization.

Both RM and AM were found to have high validity, inter-rater reliability, and small error in the present study. Therefore, angle measurements using images taken in a vertical position and a certain distance from the subject were accurate for the definite angle. The results showed what angle measurements using images could use as criterion for validity, and it supported the legitimacy of high-quality previous studies that used RM as a criterion for validity. On the other hand, AM has high clinical benefits, such as being a simple, non-invasive technique that is cost effective. Thus, AM can provide high-quality data in clinical studies and can be used as a criterion of validity. The AM is recommended placing approximately 2 m distance between the smartphone and the subject. This study found that, with appropriate and consistent shooting conditions, AM measurements can be performed with high reliability and validity.

This study uses a protractor as the subject. Therefore, future studies of angle measurement using the app are necessary to evaluate the reliability and validity of human body measurements. In addition, this study has verified the reliability and validity of smartphone-subject distances 50 to 250 cm, but less than 50 cm is still unclear.

In comparison with the protractor, RM, and AM, the results of the present study showed high inter-rater reliability, validity, and small error. The results indicated that RM and AM could be used as valid criteria in angle measurements. Moreover, AM is a highly beneficial clinical measurement method.

Conflict of interest

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

ACKNOWLEDGEMENTS

We thank Mr. Shogo Uota, Mr. Yusuke Ataka and Mr. Kuniyasu Shinomiya for contribution to the study. We would like to thank Editage (www.editage.jp) for English language editing.

REFERENCES

1) van de Pol RJ, van Trijffel E, Lucas C: Inter-rater reliability for measurement of passive physiological range of motion of upper extremity joints is better if instruments are used: a systematic review. J Physiother, 2010, 56: 7–17. [Medline] [CrossRef]
2) van Trijffel E, van de Pol RJ, Oostendorp RA, et al.: Inter-rater reliability for measurement of passive physiological movements in lower extremity joints is generally low: a systematic review. J Physiother, 2010, 56: 223–235. [Medline] [CrossRef]
3) Chapleau J, Canet F, Petit Y, et al.: Validity of goniometric elbow measurements: comparative study with a radiographic method. Clin Orthop Relat Res, 2011, 469: 3134–3140. [Medline] [CrossRef]
4) Ferriero G, Sartorio F, Foti C, et al.: Reliability of a new application for smartphones (DrGoniometer) for elbow angle measurement. PM R, 2011, 3: 1153–1154. [Medline] [CrossRef]
5) Ferriero G, Vercelli S, Sartorio F, et al.: Reliability of a smartphone-based goniometer for knee joint goniometry. Int J Rehabil Res, 2013, 36: 146–151. [Medline] [CrossRef]
6) Walter R, Kosy JD, Cove R: Inter- and intra-observer reliability of a smartphone application for measuring hallux valgus angles. Foot Ankle Surg, 2013, 19: 18–21. [Medline] [CrossRef]
7) Bennett D, Hanratty B, Thompson N, et al.: Measurement of knee joint motion using digital imaging. Int Orthop, 2009, 33: 1627–1631. [Medline] [CrossRef]
8) Naylor JM, Ko V, Adie S, et al.: Validity and reliability of using photography for measuring knee range of motion: a methodological study. BMC Musculoskelet Disord, 2011, 12: 77. [Medline] [CrossRef]
9) Nix S, Russell T, Vicenzino B, et al.: Validity and reliability of hallux valgus angle measured on digital photographs. J Orthop Sports Phys Ther, 2012, 42: 642–648. [Medline] [CrossRef]
10) Ha SM, Kwon OY, Weon JH, et al.: Reliability and validity of goniometric and photographic measurements of clavicular tilt angle. Man Ther, 2013, 18: 367–371. [Medline] [CrossRef]
11) Ege T, Kose O, Koca K, et al.: Use of the iPhone for radiographic evaluation of hallux valgus. Skeletal Radiol, 2013, 42: 269–273. [Medline] [CrossRef]
12) Landis JR, Koch GG: The measurement of observer agreement for categorical data. Biometrics, 1977, 33: 159–174. [Medline] [CrossRef]
13) Ramanathan AK, Kiran P, Arnold GP, et al.: Repeatability of the Pedar-X in-shoe pressure measuring system. Foot Ankle Surg, 2010, 16: 70–73. [Medline] [CrossRef]
14) Haley SM, Fragala-Pinkham MA: Interpreting change scores of tests and measures used in physical therapy. Phys Ther, 2006, 86: 735–743. [Medline]
15) Ludbrook J: Statistical techniques for comparing measurers and methods of measurement: a critical review. Clin Exp Pharmacol Physiol, 2002, 29: 527–536. [Medline] [CrossRef]
16) Celebić A, Stipetić J, Nola P, et al.: Use of digital photographs for artificial tooth selection. Coll Antropol, 2004, 28: 857–863. [Medline]