



Original Article

The effect of proficiency level on measurement error of range of motion

KAZUNORI AKIZUKI, RPT, PhD^{1)*}, KAZUTO YAMAGUCHI, RPT^{2, 3)}, YOSHIYUKI MORITA, RPT²⁾, YUKARI OHASHI, RPT, PhD⁴⁾

¹⁾ Department of Physical Therapy, Faculty of Health Sciences, Mejiro University: 320 Ukiya, Iwatsuki-ku, Saitama-city, Saitama 339-8501, Japan

²⁾ Department of Rehabilitation, Misato Central General Hospital, Japan

³⁾ Graduate School of Health Sciences, Ibaraki Prefectural University of Health Sciences, Japan

⁴⁾ Department of Physical Therapy, Ibaraki Prefectural University of Health Sciences, Japan

Abstract. [Purpose] The aims of this study were to evaluate the type and extent of error in the measurement of range of motion and to evaluate the effect of evaluators' proficiency level on measurement error. [Subjects and Methods] The participants were 45 university students, in different years of their physical therapy education, and 21 physical therapists, with up to three years of clinical experience in a general hospital. Range of motion of right knee flexion was measured using a universal goniometer. An electrogoniometer attached to the right knee and hidden from the view of the participants was used as the criterion to evaluate error in measurement using the universal goniometer. The type and magnitude of error were evaluated using the Bland-Altman method. [Results] Measurements with the universal goniometer were not influenced by systematic bias. The extent of random error in measurement decreased as the level of proficiency and clinical experience increased. [Conclusion] Measurements of range of motion obtained using a universal goniometer are influenced by random errors, with the extent of error being a factor of proficiency. Therefore, increasing the amount of practice would be an effective strategy for improving the accuracy of range of motion measurements.

Key words: Range of motion measurement, Measurement error, Proficiency level

(This article was submitted Apr. 11, 2016, and was accepted May 31, 2016)

INTRODUCTION

According to a survey conducted by the Japanese Physical Therapy Association of its 59,328 members, between December 20, 2009 to January 31, 2010, hemiplegia (13.6%), pain (13.2%), muscle weakness (13.0%), and limited range of motion (ROM; 10.6%) are the most common impairments treated in physical therapy¹⁾. Although improving ROM is a common therapeutic goal of physical therapy, effective treatment of ROM restrictions is often difficult. Application of evidence-based methods and reliable measurement of ROM are essential components of an effective therapeutic intervention to increase ROM.

The importance of measuring ROM has led to the development of several ROM measurement methods, including visual assessment scales²⁾; the universal goniometer³⁾; the electrogoniometer⁴⁾; digital imaging⁵⁾; three-dimensional motion analysis systems⁶⁾; and, recently, the use of smartphone-based goniometers⁷⁾. Each of these measurement techniques has advantages and drawbacks in particular applications in practice and research. In clinical practice, universal goniometers are most commonly used due to their high practicality. The information obtained from universal goniometers is used to quantify the extent of restriction in joint ROM, to inform intervention selection and application, to assess the effects of an intervention, and to confirm the progress of physical therapy treatment. Therefore, understanding the validity and reliability of measurements of

*Corresponding author. Kazunori Akizuki (E-mail: akiduki@mejiro.ac.jp)

©2016 The Society of Physical Therapy Science. Published by IPEC Inc.

This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial No Derivatives (by-nc-nd) License <<http://creativecommons.org/licenses/by-nc-nd/4.0/>>.

ROM obtained using a universal goniometer is highly relevant to physical therapy practice and has been investigated by a number of researchers.

Currier⁸⁾ defines validity as “the degree to which an instrument measures what it is purported to measure” and, “the extent to which the measurement fulfills its purpose”. In other words, the validity of a measurement instrument refers to the degree to which it can reproduce the ‘true’ value of the variable it is designed to measure. Validity is broadly divided into four types: face validity, content validity, criterion-related validity, and construct validity. Previous studies have supported the face, content and criterion-related validity of ROM measurements using a universal goniometer^{9–11)}. Of these, criterion validity is of particular importance, establishing the measurement obtained against an accepted gold standard. Radiography is considered the gold standard for ROM measurement^{12–14)}. Gogia et al.¹²⁾ evaluated measurements of knee joint positions, between 0° to 120° of 30 subjects, which were made by two physical therapists using a universal goniometer, and compared them with radiographs of the joint position obtained immediately after the measurement. The goniometric measurements correlated with the radiographic angles, with correlation coefficients of 0.97 and 0.98, respectively, for the two therapists. Based on their results, Goglia et al. considered knee ROM measurements to be valid. A similar comparison of knee ROM measures obtained using a universal goniometer and radiographs was completed by Enwemeka¹³⁾. In Enwemeka’s experiment, a universal goniometer was used to position the right knee of 10 participants at angles of 0°, 15°, 30°, 45°, 60°, and 90°, and radiographs of the end-point position were obtained. For knee joint flexion angles between 30° and 90°, the difference between goniometric and radiographic measurements ranged between 0.52° and 3.81°, a difference which was not significant. On the other hand, for knee joint flexion angles between 0° and 15°, the difference between goniometric and radiographic measurements increased to 4.59°, which was significant. Therefore, according to these results, the criterion validity of ROM obtained by universal goniometer varies depending on the ROM, with a high level of agreement between goniometric and radiographic measurements for knee joint angles between 30° and 90°, and low for angles $\leq 15^\circ$.

Reliability is defined as the consistency of measurements or of an individual’s performance in a test^{15, 16)}. Baumgartner¹⁷⁾ further divided reliability into two types, relative reliability and absolute reliability. Relative reliability is defined as the degree to which individuals maintain their position in a sample with repeated measurements. This type of reliability is often evaluated using Pearson’s product-moment correlation coefficient or the intra-class correlation coefficient. Absolute reliability is defined as the degree to which repeated measurements vary among individuals. This type of reliability is represented in actual units of measurement or as measurement ratios. Measurement error is quantified using repeated measures using absolute reliability indices, such as the Bland-Altman method¹⁸⁾, standard error of measurement, and the minimum detectable change (MDC)^{19, 20)}. The Bland-Altman method is a useful method as it allows visualization and statistical evaluation of the presence or absence of systematic bias in measurements. Standard error of measurement represents the minimum level of change that demonstrates true improvement in a group, while the MDC represents the minimum level of change that represents true improvement for an individual^{21, 22)}. Previous studies evaluating the reliability of ROM measurements have mostly addressed relative reliability, with intra-tester reliability being higher than inter-tester reliability^{23, 24)}. However, very few studies have examined the absolute reliability of ROM measurements. Lenssen et al.²⁵⁾ used the Bland-Altman method and the MDC to evaluate the absolute reliability of knee ROM measurement of 30 hospitalized patients, on day 3 or 4 following total knee arthroplasty, as measured by two physical therapists using a universal goniometer. They reported a fair level of agreement between testers, establishing differences in ROM of $<8^\circ$ as being indistinguishable from measurement error.

In clinical settings, physical therapists routinely use a change in ROM from pre- to post-intervention to assess the effects of the intervention. However, with limited evidence regarding the error in ROM measured using a universal goniometer, it is not known to what degree these measurements actually reflect a change in ROM due to the intervention. Also, although previous studies have examined the effects of factors such as age²⁶⁾, gender²⁷⁾, and body mass index²⁸⁾ on ROM measurement, there has been no investigation of the effects of the proficiency level of the tester on error in measurement of ROM using a universal goniometer. This issue is highly relevant when it is considered that ROM measures obtained by physical therapists with different levels of experience are used to evaluate change related to an intervention in practice. Therefore, the aim of this study was two-fold. The first was to use the Bland-Altman method to evaluate the type and extent of error in the measurement of ROM using a universal goniometer. The second was to evaluate the effect of the proficiency level on the magnitude of the measurement error. Understanding of the type and magnitude of measurement error using a universal goniometer will enable more accurate interpretation of measurements and improve the accuracy of clinical reasoning. Moreover, findings could assist educators in the implementation of strategies to improve students’ proficiency in ROM measurement techniques.

SUBJECTS AND METHODS

The participants were 45 physical therapy students recruited from a university, and 21 physical therapists recruited from a general hospital. Students were recruited from different years of study in a physical therapy program, with the distribution as follows: 15 second-year students, 15 third-year students, and 15 fourth-year students. At the university, the students learn ROM measurement within the first two months of their second year (April or May). Therefore, the lapse of time from learning the basic techniques of ROM measurement to participation in our study was approximately 9 months for second-year students, 21 months for third-year students and 33 months for fourth-year students. The student group was further differentiated by their clinical experience, with only fourth year students having completed all their clinical education. Among the 21 physical

therapist recruited from the hospital, 7 were in their first year of clinical experience, 8 in their second year and 6 in their third year. Their training institution, university or vocational school, was not taken into account. Overall in our study group, the lapse of time from learning of basic techniques of ROM measurement ranged between 1 year, for second year students, to 6 years, for physical therapists in their third year of clinical practice. All participants performed ROM measurements on the same volunteer, a healthy young male with no past history of orthopedic disorders of his right knee.

The methods and procedures for our study were approved by the Mejiro University Institutional Review Board (approval number: 15-013). All participants, and the volunteer, received a full explanation of the study and provided their informed consent.

Prior to conducting the full study, we completed a pilot study to confirm the accuracy of knee ROM measurements obtained using an electrogoniometer (SG150, Biometrics Ltd., UK), which was used as the criterion value for the evaluation of the measurement error of the universal goniometer. A standardized procedure was used to position the electrogoniometer for the recording of knee joint ROM positions. The electrogoniometer was attached to the right leg using double-sided tape, with the knee flexed to 60°. The proximal point of attachment was localized 1 cm from the lateral epicondyle of the femur, along a line connecting the greater trochanter and the lateral epicondyle of the femur. The distal point of attachment was localized 1 cm from the head of the fibula, along a line connecting the head of the fibula and the lateral malleolus. The bony landmarks used as a reference for the localization of the electrogoniometer were identified by palpation and marked. A thin string was attached with tape between both pairs of landmarks (i.e., greater trochanter and lateral epicondyle of the femur; head of the fibula and lateral malleolus). The thin string was used as a guide to attach the electrogoniometer. Once the electrogoniometer was placed and secured, an elastic wrap was used to prevent slipping of the electrogoniometer during measurements, as well as to hide its position from the participants performing measurements using a universal goniometer. Reflective markers were then placed over the bony landmarks, namely the greater trochanter, lateral epicondyle of the femur, head of the fibula, and lateral malleolus to visualize their positions on digital images captured obtained by a two-camera system (B-CAM, DKH, Japan). Passive knee ROM was measured by one experimenter, from full extension to maximum flexion, with the experimenter taking care to maintain the knee motion in the sagittal plane, avoiding adduction/abduction and internal/external rotation motions of the hip as much as possible. The measurement was recorded by the two-camera system with the signals between the cameras and the electrogoniometer synchronized using a trigger switch. The knee joint position data obtained from the electrogoniometer and the imaging data captured by the cameras were analyzed using a motion analysis system (Frame-DIAS V, DKH, Japan). For the imaging data, the knee joint angle was defined as the angle formed by the line connecting the greater trochanter and the lateral epicondyle and the line connecting the head of the fibula and the lateral malleolus. Agreement between the imaging and electrogoniometer data was evaluated by simple linear regression, using the imaging data as the dependent variable and the electrogoniometer data as the independent variable. The level of agreement between the electrogoniometer and imaging data was high at knee flexion angles of 20° to 90° ($y=0.960x$, $R^2=0.998$, $p<0.001$). Therefore, we used this range of knee motion for the main experiment. Statistical analyses were performed with IBM SPSS Statistics software (ver. 23).

For the main experiment, each participant performed two consecutive measurements of passive knee flexion ROM, and the limb was returned to its original position between measurements. For all measurements, the electrogoniometer was attached by one experimenter, using the standardized procedures, with its position hidden from participants by the elastic wrap. The electrogoniometer was connected to a data logger (PH-7010, DKH, Japan), providing a real time display of measurements to the experimenter, with signals saved at a sampling frequency of 100 Hz for off-line analysis. Audio feedback of threshold knee positions of 60° and 75° was provided to the volunteer through stereo earphones. An audio signal was provided when knee ROM exceeded 60° and disappeared when the ROM exceeded 75°. The volunteer used the audio feedback to simulate knee flexion restrictions. On the first measurement trial, upon hearing the audio feedback (i.e., 60° of knee flexion), the volunteer would say that ‘the knee will not bend any further’ and resist the force being applied to flex the knee by the participant. In the second measurement trial, the volunteer would say the same and resist motion at the cessation of the audio feedback (i.e., 75° of knee flexion). These scenarios were used to simulate a pre- to post-intervention change in ROM, as might be found in practice.

All measurements were performed using a metal universal goniometer. This goniometer had a handle 30 cm in length and angle marks with 1° increments. The measurement landmarks were the greater trochanter, the lateral epicondyle of the femur, the head of the fibula, and the lateral malleolus. The angle formed by the line connecting the greater trochanter and the lateral epicondyle of the femur and the line connecting the head of the fibula and the lateral malleolus was measured by the participants in 1° increments. Measurements were performed using participants’ *usual* technique, with no instructions provided regarding confirmation of localization of landmarks by palpation, the technique for passively flexing the knee, or the method used to position the goniometer on the leg. As soon as the participants completed their measurement, they were instructed to announce “done” and to report the measurement to the experimenter. The experimenter marked the end of the trial using an external ‘event’ trigger (IS3, DKH, Japan) input into the data logger. This event marker was used to synchronize the signal from the electrogoniometer with the end-position measured using the universal goniometer. The angle provided by the electrogoniometer was used as the criterion reference value, with the measurement error of the universal goniometer quantified as the difference between the measure stated by the participant and the reference value recorded using the electrogoniometer. The time required to complete the measurement was the time from the start of the trial, including palpation of

Table 1. Means and standard deviations of measurement values and measurement times of the electrogoniometer and the universal goniometer by proficiency level

Proficiency level	First measurement			Second measurement		
	EG (deg)	UG (deg)	MT (s)	EG (deg)	UG (deg)	MT (s)
Students						
1 (n=15)	64.5 ± 5.1	67.1 ± 6.9	51.7 ± 22.7	85.3 ± 4.6	87.5 ± 5.9	34.3 ± 9.3
2 (n=15)	61.2 ± 6.7	62.1 ± 11.0	57.3 ± 24.7	88.3 ± 5.4	86.6 ± 6.7	42.8 ± 14.6
3 (n=15)	63.7 ± 8.2	65.0 ± 9.8	54.1 ± 22.9	84.5 ± 6.3	83.9 ± 6.7	41.5 ± 16.9
Physical therapists						
4 (n=7)	63.1 ± 4.9	62.7 ± 3.9	51.6 ± 25.8	82.7 ± 4.5	80.9 ± 3.0	40.6 ± 15.7
5 (n=8)	64.1 ± 4.9	64.8 ± 5.2	43.7 ± 12.8	78.7 ± 2.6	77.4 ± 3.9	33.6 ± 13.5
6 (n=6)	65.3 ± 4.2	64.5 ± 5.5	48.3 ± 16.8	81.0 ± 5.1	81.3 ± 4.9	37.0 ± 22.3

Values are reported as the mean ± SD; EG: electrogoniometer; UG: universal goniometer; MT: measurement time

Table 2. Mean, standard deviation and 95% CI of the absolute measurement error (deg)

Proficiency level	First measurement		Second measurement		Average	
	Mean ± SD	95% CI	Mean ± SD	95% CI	Mean ± SD	95% CI
Students						
1 (n=15)	6.8 ± 4.3	4.4–9.2	6.7 ± 3.5	4.7–8.7	6.8 ± 3.9	5.5–8.0
2 (n=15)	5.6 ± 3.4	3.7–7.5	4.9 ± 4.1	2.6–7.1	5.3 ± 3.8	4.0–6.5
3 (n=15)	4.2 ± 2.3	2.9–5.5	4.1 ± 2.0	3.0–5.2	4.2 ± 2.1	2.9–5.4
Physical therapists						
4 (n=7)	4.3 ± 1.1	3.3–5.3	3.3 ± 2.1	1.4–5.3	3.8 ± 2.4	2.1–5.6
5 (n=8)	2.0 ± 1.5	0.7–3.2	2.6 ± 1.7	1.2–3.9	2.3 ± 1.0	0.6–3.9
6 (n=6)	2.2 ± 1.5	0.6–3.7	1.1 ± 0.7	0.3–1.8	1.6 ± 1.3	0.0–3.5

95% CI: 95% confidence interval

bony landmarks, to the time when the participant announced “done”.

The following variables were used in the analysis: level of proficiency, defined as the number of years since receiving basic instruction in goniometry measurement; measurement error; and measurement time. The means and standard deviations (SD) of the measurement error and measurement time were calculated for the proficiency level (1–6 years). Reproducibility of the end-point of knee ROM by the volunteer was evaluated for the first (60°) and second (75°) trials by calculating the mean and SD of the values recorded by the electrogoniometer. The absolute measurement error was calculated as the difference between the ROM measure of the electrogoniometer and the measure reported by the participants. The mean, SD and 95% confidence interval (CI) of the absolute error was calculated for each proficiency level. Split-plot analysis of variance (ANOVA) was used to evaluate the effect of proficiency level on the absolute measurement error and measurement time as dependent variables. A Bland-Altman analysis was performed to evaluate the presence or absence of systematic bias in the ROM measurements obtained using a universal goniometer. The differences between the measurements obtained with the electrogoniometer and the universal goniometer were plotted along the *y*-axis, with the mean of both measurements plotted along the *x*-axis. The mean difference (md), SD of the md (SD_{dif}), and 95% CI of the md were calculated to identify systematic (fixed) errors. A simple linear regression analysis was performed to identify possible proportional bias. The magnitude of the random error in measurement was evaluated by calculating the MDC at the 95% CI level (MDC_{95}) as follows: $MDC_{95}=1.96 SD_{dif}$. All statistical analyses were performed using IBM SPSS Statistics software (ver. 23).

RESULTS

The mean ± SD measurement values and measurement times, as a function of proficiency level, are reported in Table 1. The mean ± SD of the first and second reference positions measured by the electrogoniometer were $63.4 \pm 6.1^\circ$ and $84.3 \pm 5.8^\circ$, respectively, indicative of a good reproducibility of the reference angles across trials. The mean, SD and 95%CI of the difference between the criterion value of the electrogoniometer and the measurement using the universal goniometers are reported, as a function of proficiency level, in Table 2. Overall, both absolute error and variability in measurement decreased as the proficiency level increased. Split-plot ANOVA, with absolute measurement error used as the dependent variable and proficiency level (1–6) and measurement order (first or second measurement) as factors, identified a main

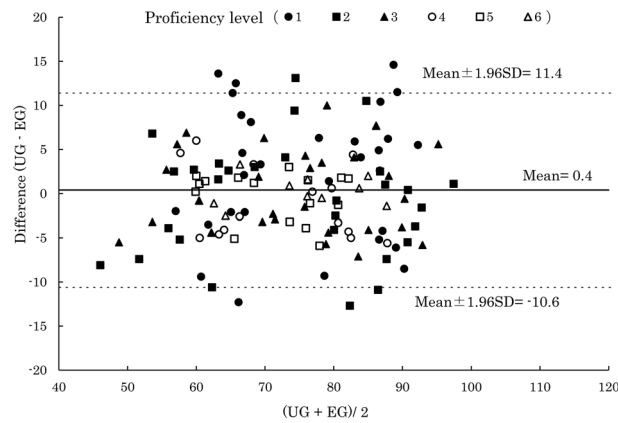


Fig. 1. Bland-Altman plot of the measurement values of the electrogoniometer (EG) and the universal goniometer (UG)

The solid line shows the mean difference between the 2 measurement methods and the 2 dotted lines indicate the ± 1.96 standard deviation; x-axis, mean angle of both measurement instruments (UG and EG); y-axis, the difference between the angle as measured by UG and as measured by EG.

Table 3. Bland-Altman analysis of measurement error

Proficiency level	md \pm SD _{dif} (deg)	Fixed error (deg)	Proportional bias	MDC ₉₅ (deg)
		95%CI	regression coefficient	
Students				
1 (n=15)	2.4 \pm 7.5	-0.4-5.2	0.037	14.7
2 (n=15)	-0.4 \pm 6.5	-2.8-2.1	0.046	12.8
3 (n=15)	0.3 \pm 4.7	-1.5-2.1	-0.020	9.3
Physical therapists				
4 (n=7)	-1.1 \pm 4.2	-3.5-1.3	-0.28	8.3
5 (n=8)	-0.3 \pm 2.8	-1.7-1.1	-0.20	5.5
6 (n=6)	-0.2 \pm 2.1	-1.5-1.1	0.37	4.1

md: mean difference; SD_{dif}: standard deviation of the mean difference; 95% CI: 95% confidence interval; MDC₉₅: minimal detectable change at the 95% confidence level

effect of proficiency level ($F_{5,60}=6.517$, $p<0.001$, $\eta_p^2=0.352$) on measurement error, with no significant effect of order of measurement ($F_{1,60}=0.756$, $p=0.388$, $\eta_p^2=0.012$) or proficiency level \times measurement order interaction ($F_{5,60}=0.256$, $p=0.935$, $\eta_p^2=0.021$). Multiple comparisons for proficiency levels, using the Bonferroni method, identified significant differences in ROM measurements between the following proficiency levels: 1 and 5 years ($p=0.001$); 1 and 6 years ($p<0.001$); and 2 and 6 years ($p=0.033$).

Split-plot ANOVA, with measurement time used as the dependent variable and proficiency level (1–6) and measurement order (first or second measurement) as factors, identified a significant main effect of measurement order ($F_{1,60}=29.735$, $p<0.001$, $\eta_p^2=0.331$) on measurement time, with no significant effect of proficiency level ($F_{5,60}=0.651$, $p=0.662$, $\eta_p^2=0.051$) or proficiency level \times measurement order interaction ($F_{5,60}=0.267$, $p=0.929$, $\eta_p^2=0.022$).

The Bland-Altman plot used to evaluate systematic bias is shown in Fig. 1. The mean absolute difference in measurement was low at 0.4° (95% CI, 1.7° to -1.0°) and, therefore, there was no evidence of fixed error in ROM measurements. Simple linear regression analysis also showed there was no evidence of proportional bias in measurement, with a regression coefficient of nearly zero, and the regression line was not significant: $y=0.018x - 0.94$; $r=0.037$; $p=0.67$.

The Bland-Altman analysis performed as a function of proficiency level confirmed the absence of systematic bias on measurements obtained using the universal goniometer, regardless of proficiency level (Table 3). The MDC₉₅ values, however, did decrease as proficiency level increased, indicating a lowering in the random error of measurement as the proficiency level increased.

DISCUSSION

In this study, the Bland-Altman method was used to quantitatively determine the type and magnitude of error in ROM measurements made using a universal goniometer and to evaluate the effects of learning on identified measurement error. While systematic error was not found in the measurements, the magnitude of the random error did decrease with increasing level of proficiency.

The validity of our methodology was robust. The accuracy of the electrogoniometer measurements used as criterion values was first established, and it showed that electrogoniometer measurements were reliable for the ROM of knee flexion between 20° and 90°. Piriyaarasarth et al.²⁹⁾ previously reported the standard error of measurement with an electrogoniometer, placed by the same tester across repeated trials, to be <1.7°. Kettelkamp et al.³⁰⁾ proposed that errors in electrogoniometer measurements resulted from misalignment of the electrogoniometer with the anatomical axis of the knee joint and/or due to slippage during operation. Therefore, in the preset study, all measurements were performed on a single healthy young adult, with the electrogoniometer positioned and secured by a single experimenter. In order to minimize any misalignment of the electrogoniometer with the anatomical axis of the knee joint, reference lines connecting the greater trochanter and the lateral epicondyle of the femur was used, as well as the head of the fibula and the lateral malleolus, to guide the experimenter to reproducibly attach the electrogoniometer to the right lower limb. In order to prevent slippage of the electrogoniometer during flexion motions of the knee, the electrogoniometer was attached to the skin with double-sided tape, and then fixed with surgical tape and an elastic wrap. Any errors due to positioning and slippage would have manifested as systematic bias in measurement. No evidence of such systematic bias was found (Fig. 1, Table 3). Therefore, we are confident that differences between measurements obtained using the universal goniometer and the criterion reference, obtained using an electrogoniometer, were 'true' errors.

ANOVA identified a significant main effect of proficiency level on absolute measurement error, with error decreasing as a function of increased proficiency. The proficiency level was defined the number of years since receiving basic instruction in goniometry measurement. Therefore, participants with a higher proficiency level were inferred to have performed goniometric measures a greater number of times. From the results, it can be inferred that practice significantly improves the accuracy of ROM measurements. Previous studies of skill acquisition have demonstrated that the amount of practice is an important factor in skill acquisition, both for accuracy and consistency in performance^{31, 32)}. Although decrease in performance time is also an important indicator of skill acquisition, an effect of proficiency level on measurement time was not found. We postulate that this finding can be explained within the previously described phenomenon of a speed-accuracy trade-off³³⁾. In the present study, 'speed' would be the measurement time and 'accuracy' would be the measurement error. Participants in the present study were not given any instructions regarding the measurement time, therefore, it is likely that participants focused on being as accurate as possible. We recognize that the physical therapists recruited from the general hospital for this study only had a maximum of three years of clinical experience. It is possible that with more years of clinical experience, physical therapists may become better at measuring ROM accurately, without sacrificing time.

Our Bland-Altman analysis did not identify evidence of systematic bias in measurements using the universal goniometer, with no fixed or proportional error detected. Therefore, measures of ROM using a universal goniometer were only influenced by random errors. MDC₉₅ was calculated as an index of absolute reliability, which represents the minimum change that demonstrates a 'true' improvement. Change in ROM less than the MDC₉₅ cannot be differentiated from random error in measurement. Therefore, in practice, it is impossible to determine whether pre- to post-physical therapy intervention changes in ROM that are within the MDC₉₅ range represent a 'true' improvement in ROM. MDC₉₅ was found to be influenced by the level of proficiency, with the MDC₉₅ value of 14.7° at 1 year decreasing to 4.1° after three years of clinical experience. At an MDC₉₅ level of 14.7°, changes in ROM <15° cannot be differentiated from measurement error, thus making it impossible to identify small, but possibly relevant, changes in ROM. In contrast, the MDC₉₅ value of 4.1° of the more experienced therapists indicates that changes ≥5° can be interpreted as 'true' changes in ROM, which would increase the sensitivity of ROM measurements to the effects of physical therapy intervention.

The effect of random error in ROM measures obtained by a novice therapist can be reduced by using the mean value of multiple measurements, an approach that has been reported in previous studies to improve the reliability of ROM measurements^{34, 35)}. For the measurement of knee and elbow ROM, Rothstein et al.³⁴⁾ reported that inter-tester reliability obtained from the mean of two measurements was slightly better than inter-tester reliability obtained from a single measurement. Thus, if the only type of error involved in ROM measurements is random in nature, then averaging the means of multiple measurements can cancel out random error and consequently yield more accurate values. This is a strategy that might benefit novice therapists and learners and improve the reliability of the interpretations of their measurements in practice, with the number of measurements reducing as the magnitude of random error decreases. Moreover, increasing the amount of practice (i.e., the number of measurements performed) could be an effective teaching strategy for improving the accuracy of ROM measurements using a universal goniometer, with high repetition enhancing motor learning and skill acquisition^{36, 37)}.

Feedback of results is an important factor in motor learning. However, during practice of ROM measurement, there is no effective feedback provided to the learning regarding the accuracy of measurement. We speculate that development of a feedback device based on electrogoniometry might promote better acquisition of ROM measurement skill. Therefore, going

forward, we would like to focus on developing such feedback devices and evaluate their usefulness in learning and practice.

The limitations of the present study need to be acknowledged in the interpretation of results. Foremost, the results are only valid within the context of ROM measurement of right knee flexion using a metal universal goniometer. Although evaluation was limited to the joint that is most commonly evaluated in practice, it is important to note that previous studies have reported the reliability of ROM measurements to be influenced by the joint being measured^{23,24)} and the device used^{38,39)}. In addition, the participants in the present study were limited to students from a single university and physical therapists working at a single general hospital. Therefore, caution should be exercised when generalizing the results to other groups of students and therapists. In order to overcome these limitations and enable more accurate interpretation of measurements, a variety of joints must be examined and participants in future studies must include students from other universities and physical therapists working at other institutions.

REFERENCES

- 1) Japanese Physical Therapy Association: Physical Therapy White Paper. Tokyo: Shinohara publication, 2010.
- 2) Rowe CR: Joint measurement in disability evaluation. *Clin Orthop Relat Res*, 1964, 32: 43–53. [[Medline](#)]
- 3) Moore ML: The measurement of joint motion; the technic of goniometry. *Phys Ther Rev*, 1949, 29: 256–264. [[Medline](#)]
- 4) Karpovich PV, Herden EL Jr, Asa MM: Electrogoniometric study of joints. *U S Armed Forces Med J*, 1960, 11: 424–450. [[Medline](#)]
- 5) Blonna D, Zarkadas PC, Fitzsimmons JS, et al.: Validation of a photography-based goniometry method for measuring joint range of motion. *J Shoulder Elbow Surg*, 2012, 21: 29–35. [[Medline](#)] [[CrossRef](#)]
- 6) Tojima M, Ogata N, Yozu A, et al.: Novel 3-dimensional motion analysis method for measuring the lumbar spine range of motion: repeatability and reliability compared with an electrogoniometer. *Spine*, 2013, 38: E1327–E1333. [[Medline](#)] [[CrossRef](#)]
- 7) 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](#)]
- 8) Currier DP: *Elements of Research in Physical Therapy*, 3rd ed. Baltimore: Williams & Wilkins, 1990, p 171.
- 9) Portney LG, Watkins MP: *Foundations of Clinical Research: Applications to Practice*, 2nd ed. Upper Saddle River: Prentice-Hall, 2000.
- 10) Gajdosik RL, Bohannon RW: Clinical measurement of range of motion. Review of goniometry emphasizing reliability and validity. *Phys Ther*, 1987, 67: 1867–1872. [[Medline](#)]
- 11) Sim J, Arnell P: Measurement validity in physical therapy research. *Phys Ther*, 1993, 73: 102–110, discussion 110–115. [[Medline](#)]
- 12) Gogia PP, Braatz JH, Rose SJ, et al.: Reliability and validity of goniometric measurements at the knee. *Phys Ther*, 1987, 67: 192–195. [[Medline](#)]
- 13) Enwemeka CS: Radiographic verification of knee goniometry. *Scand J Rehabil Med*, 1986, 18: 47–49. [[Medline](#)]
- 14) Kato M, Echigo A, Ohta H, et al.: The accuracy of goniometric measurements of proximal interphalangeal joints in fresh cadavers: comparison between methods of measurement, types of goniometers, and fingers. *J Hand Ther*, 2007, 20: 12–18, quiz 19. [[Medline](#)] [[CrossRef](#)]
- 15) Safrit MJ: An overview of measurement. In: Safrit MJ, Wood TM (eds.), *Measurement concepts in physical education and exercise science*. Champaign: Human Kinetics, 1989, pp 3–20.
- 16) Atkinson G, Nevill AM: Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Med*, 1998, 26: 217–238. [[Medline](#)] [[CrossRef](#)]
- 17) Baumgartner TA: Norm-referenced measurement: reliability. In: Safrit MJ, Wood TM (eds.), *Measurement concepts in physical education and exercise science*. Champaign: Human Kinetics, 1989, pp 45–72.
- 18) Bland JM, Altman DG: Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*, 1986, 1: 307–310. [[Medline](#)] [[CrossRef](#)]
- 19) Lexell JE, Downham DY: How to assess the reliability of measurements in rehabilitation. *Am J Phys Med Rehabil*, 2005, 84: 719–723. [[Medline](#)] [[CrossRef](#)]
- 20) Hopkins WG: Measures of reliability in sports medicine and science. *Sports Med*, 2000, 30: 1–15. [[Medline](#)] [[CrossRef](#)]
- 21) Beckerman H, Roebroek ME, Lankhorst GJ, et al.: Smallest real difference, a link between reproducibility and responsiveness. *Qual Life Res*, 2001, 10: 571–578. [[Medline](#)] [[CrossRef](#)]
- 22) Schuck P, Zwingmann C: The ‘smallest real difference’ as a measure of sensitivity to change: a critical analysis. *Int J Rehabil Res*, 2003, 26: 85–91. [[Medline](#)] [[CrossRef](#)]
- 23) Boone DC, Azen SP, Lin CM, et al.: Reliability of goniometric measurements. *Phys Ther*, 1978, 58: 1355–1360. [[Medline](#)]
- 24) Pandya S, Florence JM, King WM, et al.: Reliability of goniometric measurements in patients with Duchenne muscular dystrophy. *Phys Ther*, 1985, 65: 1339–1342. [[Medline](#)]
- 25) Lenssen AF, van Dam EM, Crijns YH, et al.: Reproducibility of goniometric measurement of the knee in the in-hospital phase following total knee arthroplasty. *BMC Musculoskelet Disord*, 2007, 8: 83. [[Medline](#)] [[CrossRef](#)]
- 26) Roach KE, Miles TP: Normal hip and knee active range of motion: the relationship to age. *Phys Ther*, 1991, 71: 656–665. [[Medline](#)]
- 27) James B, Parker AW: Active and passive mobility of lower limb joints in elderly men and women. *Am J Phys Med Rehabil*, 1989, 68: 162–167. [[Medline](#)] [[CrossRef](#)]
- 28) Lichtenstein MJ, Dhanda R, Cornell JE, et al.: Modeling impairment: using the disablement process as a framework to evaluate determinants of hip and knee flexion. *Aging (Milano)*, 2000, 12: 208–220. [[Medline](#)]
- 29) Piriyaaparth P, Morris ME, Winter A, et al.: The reliability of knee joint position testing using electrogoniometry. *BMC Musculoskelet Disord*, 2008, 9: 6. [[Medline](#)] [[CrossRef](#)]
- 30) Kettelkamp DB, Johnson RJ, Smidt GL, et al.: An electrogoniometric study of knee motion in normal gait. *J Bone Joint Surg Am*, 1970, 52: 775–790. [[Medline](#)]
- 31) Crossman E: A theory of the acquisition of speed-skill. *Ergonomics*, 1959, 2: 153–166. [[CrossRef](#)]

- 32) Ericsson KA, Krampe RT, Tesch-Römer C: The role of deliberate practice in the acquisition of expert performance. *Psychol Rev*, 1993, 100.3: 363. [[CrossRef](#)]
- 33) Fitts PM: The information capacity of the human motor system in controlling the amplitude of movement. *psychology. Exp Psychol*, 1954, 47: 381. [[CrossRef](#)]
- 34) Rothstein JM, Miller PJ, Roettger RF: Goniometric reliability in a clinical setting. Elbow and knee measurements. *Phys Ther*, 1983, 63: 1611–1615. [[Medline](#)]
- 35) Low JL: The reliability of joint measurement. *Physiotherapy*, 1976, 62: 227–229. [[Medline](#)]
- 36) Guadagnoli MA, Lee TD: Challenge point: a framework for conceptualizing the effects of various practice conditions in motor learning. *J Mot Behav*, 2004, 36: 212–224. [[Medline](#)] [[CrossRef](#)]
- 37) Akizuki K, Ohashi Y: Measurement of functional task difficulty during motor learning: what level of difficulty corresponds to the optimal challenge point? *Hum Mov Sci*, 2015, 43: 107–117. [[Medline](#)] [[CrossRef](#)]
- 38) Goodwin J, Clark C, Deakes J, et al.: Clinical methods of goniometry: a comparative study. *Disabil Rehabil*, 1992, 14: 10–15. [[Medline](#)] [[CrossRef](#)]
- 39) Rome K, Cowieson F: A reliability study of the universal goniometer, fluid goniometer, and electrogoniometer for the measurement of ankle dorsiflexion. *Foot Ankle Int*, 1996, 17: 28–32. [[Medline](#)] [[CrossRef](#)]