Original Research

Adjacent Joint Restriction Differentially Influences Intra- and Inter-rater Reliability and Agreement of Goniometric Measurements

Brendan L. Pinto[1] , Tatjana Stankovic[2], David M. Frost[2], Tyson A. C. Beach[1] [a]

[1] Department of Kinesiology, University of Waterloo, [2] Faculty of Kinesiology and Physical Education, University of Toronto

International Journal of Sports Physical Therapy

Background

Reliability and agreement of goniometric measurements can be altered by variations in measurement technique such as restricting adjacent joints to influence bi-articular muscles. It is unknown if the influence of adjacent joint restriction is consistent across different range of motion (ROM) tests, as this has yet to be assessed within a single study. Additionally, between-study comparisons are challenged by differences between methodology, participants and raters, obscuring the development of a conceptual understanding of the extent to which adjacent joint restriction can influence goniometric ROM measurements.

Purpose

To quantify intra- and inter-rater reliability and levels of agreement of goniometric measurements across five ROM tests, with and without adjacent joint restriction.

Study Design

Descriptive reliability study

Methods

Three trained and experienced raters made two measurements of bilateral ankle dorsiflexion, first metatarsophalangeal dorsiflexion, hip extension, hip flexion, and shoulder flexion, with and without adjacent joint restriction. Intraclass correlation coefficient (ICC), standard error of measurement (SEM), along with participant, measurement/rater and random error variance were estimated.

Results

Eleven females (age 21.4 ±2.3 years) and 19 males (age 22.1 ±2.8 years) participated. Adjacent joint restriction did not influence the reliability and agreement in a consistent way across the five ROM tests. Changes in the inter-rater reliability and agreement were more pronounced compared to the intra-rater reliability and agreement. Assessing variance components (participant, measurement/rater and random error variance) that are used to calculate the ICC and SEM, improved interpretation of ICC and SEM scores.

Conclusion

The effects of adjacent joint restriction on reliability and agreement of goniometric

---

a **Corresponding author:**
Tyson Beach, PhD
Department of Kinesiology
University of Waterloo
200 University Avenue West
Waterloo, ON N2L3G1
CANADA
E-mail: tyson.beach@uwaterloo.ca

measurements depend on the ROM test and should be considered when comparing measurements between multiple raters. Reporting variance components that are used to calculate the ICC and SEM can improve interpretation and may improve between-study comparisons, towards developing a conceptual framework to guide goniometric measurement technique.

Level of Evidence

3b

INTRODUCTION

The universal goniometer is a portable, low cost and easy to use tool which can quantify joint range of motion (ROM). This has favoured its use in clinical settings as well as in research.[1,2] Although it can have a close association with technologically advanced techniques,[3,4] the measurement reliability and agreement of goniometry can vary due to multiple factors. These factors include the rater, the individual being measured, the joint being measured and differences in ROM measurement technique such as test posture, imposed restraints (e.g. adjacent joint(s) restriction) and assistance (e.g. active vs. passive).[5] It is important to understand the extent to which these different factors can influence goniometry to inform best practices in the clinic and research.

Intra-rater reliability and agreement has consistently been shown to be higher than inter-rater reliability and agreement.[5–8] Pain and spasticity have the potential to reduce reliability.[5] Each joint can have different characteristics that can influence alignment and placement of the goniometer to influence ROM measurements.[9] Similarly, different ROM tests measuring different ranges of the same joint (i.e. movement directions such as flexion or extension) can potentially influence reliability and agreement due to different position and goniometer alignment requirements. However, the extent to which different types of techniques can influence goniometry is uncertain. For instance, restricting adjacent joints can be used to assess the flexibility and the contribution of bi-articular and mono-articular muscles to the joint ROM which can inform on function and risk of muscle injury.[10] Yet, it is unknown if restricting adjacent joints influences measurement reliability and agreement in a consistent way across different ROM measurements.

Comparing results between studies is challenged by differences between methodology, participants and the raters, as these all have the potential to influence the intraclass correlation coefficient (ICC) and the standard error of measurements (SEM) which have been recommended as the preferred estimate of reliability and agreement.[11–13] The ICC is calculated as the ratio of between-participant variance to total variance, where the total variance is the sum of the between participant variance and error variance.[14] When there are an equal number of repeated measurements for each participant (at least two per participant), a two-way model can be used to partition the error into systematic error and random error,[13,14] such that the absolute agreement, two-way random effects, single measurement ICC (i.e. ICC[2,1]) can be calculated as:

$$ICC\,[2,1] = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_r^2 + \sigma_e^2}$$

Where σ_p^2 is the variance between participants, σ_r^2 is the systematic error variance due to systematic differences between the measurements or raters being compared, and σ_e^2 is the random error variance.[11,13,15] There are other formulas for calculating the ICC that reduce the influence of σ_r^2 and σ_e^2 which results in an increased ICC score (e.g. ICC[3,1] and ICC[2,k]).[13] However, the use and reporting of ICCs between studies is inconsistent and often lacks sufficient detail.[2,16,17] Additionally, the SEM can be calculated as SEM= SD_{pooled} x $\sqrt{1 - ICC}$.[14] Hence, both the ICC and SEM are influenced by any changes that could influence the components of variance within the data. For this reason, differences between the methodology, participants and raters, challenges the comparison of ICC and SEM scores between studies. Further, few studies usually report the components of variance which limit the consolidation of different or conflicting results.[11]

There is a paucity of research investigating the effect of technique, such as adjacent joint restriction, across more than one joint or ROM test within a single study to overcome the difficulties in comparing results between different studies.[5] Additionally, differences between restricting and not restricting the adjacent joint could influence the variance between participants (e.g. due to differences in bi-articular muscle flexibility leading to changes in ROM), variance between measurements/raters and random error in different ways. Thus, understanding if the effect is consistent across different ROM tests and investigating why any changes occur can contribute to the conceptual framework used to guide goniometry in research and practice.

The purpose of this study was to quantify intra- and inter-rater reliability and levels of agreement of goniometric measurements across five ROM tests, with and without adjacent joint restriction. Using the same participants and raters as well as reporting the components of variance may improve understanding of whether adjacent joint restriction consistently influences goniometer measurement reliability and agreement across different ROM tests. This knowledge may contribute to creating a conceptual framework to guide effective use of goniometers.

MATERIALS AND METHODS

A convenience sample of 30 healthy participants provided written informed consent. All procedures were approved by the University of Toronto's Office of Research Ethics. Two Certified Athletic Therapists and one Registered Physiotherapist rated each participant in a random order. Raters had similar field experience and worked interchangeably

within the same clinic. Raters underwent specific training for this study, which included standard instructions for each ROM test (Appendix A) and a practice session involving three participants who were not part of the study. To control for differences in repositioning participants between measurements, a trained research assistant positioned all participants. A second research assistant documented the measured values.

Participants performed a warm up session upon arrival that consisted of five minutes of light stationary cycling and 10 repetitions of bodyweight squats, lunges, calf-raises and arm circles. The ROM tests included ankle dorsiflexion, first metatarsophalangeal (MTPJ1) dorsiflexion, hip extension, hip flexion, and shoulder flexion. Each ROM test was performed twice bilaterally, with and without restraining the adjacent joint position. The left side was measured first, followed by the right side, and then both sides were repeated to obtain two measurements per side. This repositioning between each measurement avoided residual effects from holding the testing position (i.e. stretching). Each rater performed all measurements in the same order but rated each participant in a random order with approximately two minutes of rest between each rater. This allowed for equal time between each rater's evaluation of each ROM test.

Details of the testing positions and goniometer placements can be found in Appendix A. The following provides a brief description:

ANKLE DORSIFLEXION

Measured in a lunge/split stance. Unrestricted ROM was measured on the front foot, while the rear foot was used to measure restricted ROM by keeping the rear knee extended to influence the gastrocnemius muscles (Figure 1A).

MTPJ1 DORSIFLEXION

Measured while participants sat on a table with their feet hanging freely above the floor. Unrestricted ROM was measured with the ankle supported in its natural hanging position, whereas the restricted ROM was measured in maximal ankle dorsiflexion to influence flexor hallucis longus (Figure 1B).

HIP EXTENSION

Measured with the participant lying supine with their hips on the edge of a table and a rolled towel beneath their lower back. The test leg hung off of the edge while the non-test leg was supported in hip and knee flexion by the research assistant. Unrestricted hip extension was measured by pushing the hanging leg down until the pelvis started to rotate anteriorly, while controlling for leg rotation. Restricted hip extension was performed in the same manner but while holding the hanging leg in 90° of knee flexion to influence rectus femoris (Figure 1C).

HIP FLEXION

Tested in a supine position. The test hip was flexed with the test side knee either flexed or extended to influence the hamstring muscle group, for the unrestricted and restricted
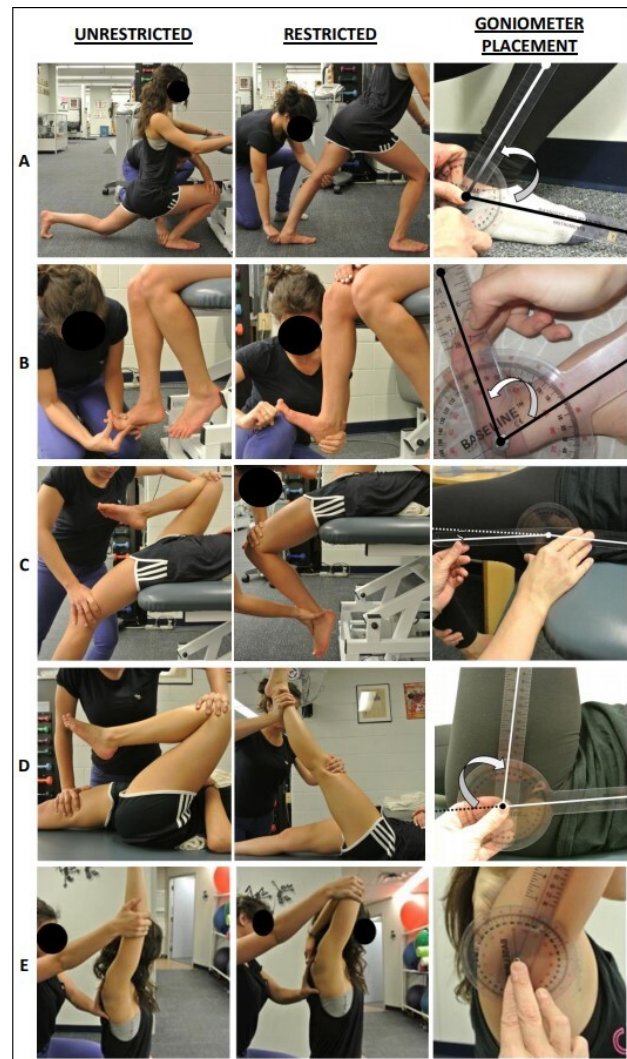

Figure 1. Participant position and goniometer placement for unrestricted and restricted (A) ankle dorsiflexion (B) MTPJ1 dorsiflexion, (C) hip extension, (D) hip flexion and (E) shoulder flexion.

measurements respectively (Figure 1D).

SHOULDER FLEXION

Measured while participants sat with their feet firmly on the ground. The research assistant supported the scapula while raising the participants' arm with the elbow either extended or flexed to influence the triceps brachii, for the unrestricted and restricted measurements respectively (Figure 1E).

To determine the effects of adjacent joint restriction on intra- and inter-rater reliability and agreement, changes in ICC and SEM scores were assessed between the unrestricted and restricted conditions across the five ROM tests. An absolute agreement, two-way random effects, single measurement ICC (i.e. ICC [2,1]) was used to measure intra- and inter-rater reliability (psych package, RGui Version 4.0.2, The R Foundation, Vienna, Austria). Inter-rater ICC was calculated using the mean of each rater's two measurements to account for the deviation within their respective mea-

Table 1. Participant characteristics.

Participants	N	Height (cm)		Mass (kg)		Age (yrs)	
		Mean	SD	Mean	SD	Mean	SD
Female	11	167.1	6.4	62.3	9.9	21.4	2.3
Male	19	179.1	5.8	78.5	7.7	22.1	2.8

surements. ICC scores were interpreted as <0.40 is poor, 0.41-0.59 is fair, 0.60-0.74 is good, and 0.75-1.0 is excellent.[18] SEM was calculated as SEM= SD_{pooled} x $\sqrt{1 - ICC}$. The mean squares from the ICC calculation were used to estimate σ^2_p, σ^2_r and σ^2_e.[15]

RESULTS

Participant characteristics are provided in Table 1.

ICC and SEM scores changed with the restricted technique in different ways across ROM tests (Tables 2 and 3). Specifically, the mean intra-rater ICC and SEM scores across raters respectively decreased and increased (left and right ankle and MTPJ1 dorsiflexion), increased and decreased (left and right shoulder flexion and left hip extension), or both increased (left and right hip flexion and right hip extension) with adjacent joint restriction. However, each rater had different changes between unrestricted and restricted techniques. For example, rater 2 had a decrease in ICC and increase in SEM for shoulder flexion (Table 2).

Changes in intra-rater ICC and SEM scores between the unrestricted and restricted techniques matched changes in variance between participants (σ^2_p) and random error variance (σ^2_e) respectively. Specifically, σ^2_p and σ^2_e respectively decreased and increased (left and right ankle and MTPJ1 dorsiflexion), increased and decreased (left and right shoulder flexion and left hip extension) or both increased (left and right hip flexion and right hip extension) with adjacent joint restriction. Although adjacent joint restriction modified ICC scores, intra-rater ICCs were still good to excellent across the ROM tests even with the associated decreases with or without adjacent joint restriction (Table 2).

Inter-rater ICC and SEM scores showed similar changes to intra-rater scores with adjacent joint restriction for some tests (left and right ankle dorsiflexion, right hip extension, left hip flexion and right shoulder flexion) but not others (increased and decreased with restriction for left and right MTPJ1, decreased and increased with restriction for right side hip flexion and both ICC and SEM increased for left hip extension and left shoulder flexion, Table 3).

Inter-rater ICCs ranged from poor to good and had broader 95% CIs than the intra-rater comparison. Similarly, SEM was higher between raters compared to SEM within each rater's measurements (Tables 2 and 3). The effects of adjacent joint restriction were more pronounced in the inter-rater comparison. The restricted technique reduced the inter-rater ICC score from fair to poor for left and right ankle dorsiflexion. Conversely, the restricted technique increased ICC scores from poor to fair for left and right hip extension and left shoulder flexion, and improved the ICC score from fair to good for left MTPJ1 flexion. SEM also decreased by approximately 3° with the restricted technique for right MTPJ1 flexion (Table 3).

The systematic variance between measurements/raters (σ^2_r) and the random error variance (σ^2_e) were higher for the inter-rater comparison than any of the intra-rater comparisons, but the changes in the ICC and SEM scores due to adjacent joint restriction varied due to differing changes across each component of variance (i.e. σ^2_p, σ^2_r and σ^2_e). The lower inter-rater ICC for the restricted ankle dorsiflexion was due to a decrease in σ^2_p and increase in σ^2_r and σ^2_e compared to unrestricted ankle dorsiflexion. The notable increases in ICC for restricted hip extension and left side shoulder flexion were also due to σ^2_p, as it increased but σ^2_r and σ^2_e remained similar compared to the unrestricted techniques. Conversely, the improved ICC and SEM scores for restricted MTPJ1 were not due to changes in σ^2_p, but due to decreases in σ^2_r and σ^2_e compared to the unrestricted technique (Table 3).

DISCUSSION

No known study has evaluated intra- and inter-rater goniometer measurement reliability and agreement across multiple ROM tests, with and without adjacent joint restriction. Reliability and agreement were differentially influenced by adjacent joint restriction across the five ROM tests. Joint restriction does not appear to substantially influence intra-rater reliability and agreement. Conversely, inter-rater reliability and agreement were more substantially influenced by adjacent joint restriction, but this was due to different changes across the components of variance used to calculate the ICC and SEM.

Table 2. Intra-rater ICC with 95% CI, SEM, variance between participants (σ^2_p), systematic error variance (σ^2_r), and random error variance (σ^2_e).

ROM Test	Side	Rater 1					Rater 2					Rater 3					Mean				
		ICC [95% CI]	SEM	σ^2_p	σ^2_r	σ^2_e	ICC [95% CI]	SEM	σ^2_p	σ^2_r	σ^2_e	ICC [95% CI]	SEM	σ^2_p	σ^2_r	σ^2_e	ICC	SEM	σ^2_p	σ^2_r	σ^2_e
Ankle Dorsiflexion Unrestricted	Left	0.92 [0.85, 0.96]	2.2	57.7	0.0	5.1	0.93 [0.87, 0.96]	1.8	43.5	0.0	3.3	0.94 [0.90, 0.97]	2.0	61.8	0.0	3.6	0.93	2.0	54.3	0.0	4.0
	Right	0.83 [0.71, 0.91]	3.0	45.2	0.0	9.1	0.93 [0.87, 0.96]	2.0	50.7	0.0	4.0	0.9 [0.83, 0.95]	2.4	50.3	0.1	5.3	0.89	2.5	48.7	0.0	6.1
Ankle Dorsiflexion Restricted	Left	0.93 [0.87, 0.96]	1.8	43.1	0.0	3.3	0.92 [0.86, 0.96]	2.0	47.2	0.0	4.1	0.87 [0.78, 0.93]	2.8	51.5	0.0	7.6	0.91	2.2	47.3	0.0	5.0
	Right	0.89 [0.81, 0.94]	2.0	31.1	0.1	3.7	0.91 [0.85, 0.95]	2.1	43.9	0.0	4.2	0.81 [0.67, 0.89]	3.0	37.9	0.0	9.0	0.87	2.3	37.6	0.0	5.6
MTPJ1 Dorsiflexion Unrestricted	Left	0.91 [0.84, 0.95]	3.6	131.9	0.0	12.9	0.83 [0.70, 0.90]	4.6	101.2	0.0	21.2	0.94 [0.90, 0.97]	3.7	215.4	0.0	12.8	0.89	4.0	149.5	0.0	15.6
	Right	0.92 [0.86, 0.96]	4.5	235.3	0.0	20.7	0.9 [0.82, 0.94]	4.2	157.7	0.0	17.5	0.96 [0.93, 0.98]	3.5	296.7	1.0	10.5	0.93	4.1	229.9	0.3	16.2
MTPJ1 Dorsiflexion Restricted	Left	0.89 [0.80, 0.94]	4.0	131.4	0.0	16.9	0.8 [0.66, 0.89]	5.4	117.3	0.0	28.9	0.8 [0.65, 0.89]	5.9	141.8	1.5	34.5	0.83	5.1	130.2	0.5	26.8
	Right	0.83 [0.70, 0.90]	4.0	79.4	0.0	16.5	0.76 [0.60, 0.86]	6.5	133.5	1.5	40.5	0.88 [0.78, 0.93]	4.1	124.2	0.0	17.5	0.82	4.9	112.4	0.5	24.8
Hip Extension Unrestricted	Left	0.86 [0.75, 0.92]	2.2	29.5	0.5	4.4	0.74 [0.57, 0.85]	3.3	31.8	0.0	10.9	0.79 [0.65, 0.88]	3.9	57.8	0.0	15.0	0.80	3.1	39.7	0.2	10.1
	Right	0.94 [0.89, 0.97]	1.4	31.4	0.0	2.1	0.8 [0.66, 0.89]	3.1	39.6	0.2	9.6	0.9 [0.81, 0.94]	2.6	59.0	0.5	6.4	0.88	2.4	43.3	0.2	6.0
Hip Extension Restricted	Left	0.92 [0.85, 0.95]	2.1	48.9	0.0	4.4	0.9 [0.83, 0.95]	3.2	91.8	0.0	9.8	0.92 [0.86, 0.96]	2.7	86.1	0.2	7.3	0.91	2.7	75.6	0.1	7.2
	Right	0.92 [0.83, 0.96]	2.0	48.3	0.8	3.4	0.94 [0.89, 0.97]	2.6	104.1	0.0	6.8	0.85 [0.74, 0.92]	3.5	67.6	0.0	11.7	0.90	2.7	73.3	0.3	7.3
Hip Flexion Unrestricted	Left	0.96 [0.93,	2.0	98.9	0.0	4.2	0.88 [0.79,	3.8	106.8	0.0	14.3	0.88 [0.78,	4.2	128.1	1.8	15.4	0.91	3.3	111.3	0.6	11.3

		0.98]					0.93]					0.94]									
	Right	0.95 [0.92, 0.98]	2.3	97.7	0.0	4.6	0.91 [0.84, 0.95]	3.2	103.9	0.0	10.2	0.92 [0.86, 0.96]	3.4	134.2	0.0	11.3	0.93	3.0	111.9	0.0	8.7
Hip Flexion Restricted	Left	0.98 [0.96, 0.99]	2.2	244.1	0.1	5.8	0.92 [0.85, 0.95]	3.3	121.7	0.2	11.0	0.88 [0.78, 0.93]	5.6	231.5	0.0	32.3	0.93	3.7	199.1	0.1	16.4
	Right	0.95 [0.91, 0.97]	3.7	263.7	0.4	13.6	0.85 [0.73, 0.92]	4.8	131.0	2.6	20.0	0.93 [0.88, 0.96]	4.4	261.2	0.4	18.2	0.91	4.3	218.6	1.2	17.3
Shoulder Flexion Unrestricted	Left	0.92 [0.86, 0.96]	3.1	112.6	0.0	9.6	0.92 [0.86, 0.96]	3.2	114.4	0.0	10.0	0.78 [0.64, 0.88]	4.8	82.4	0.0	22.6	0.87	3.7	103.1	0.0	14.1
	Right	0.94 [0.89, 0.97]	3.1	153.2	0.0	10.0	0.94 [0.89, 0.97]	2.9	132.3	0.0	8.7	0.41 [0.13, 0.63]	13.4	125.1	0.0	177.2	0.76	6.5	136.9	0.0	65.3
Shoulder Flexion Restricted	Left	0.93 [0.86, 0.96]	3.4	155.1	1.1	11.2	0.87 [0.78, 0.93]	4.1	110.4	0.0	16.0	0.91 [0.84, 0.95]	3.6	132.6	0.1	12.7	0.90	3.7	132.7	0.4	13.3
	Right	0.96 [0.93, 0.98]	2.5	153.4	0.0	6.3	0.89 [0.80, 0.94]	3.8	115.5	0.4	14.0	0.88 [0.79, 0.93]	4.6	156.4	0.2	20.7	0.91	3.6	141.8	0.2	13.7

Table 3. Inter-rater ICC with 95% CI, SEM, variance between participants (σ^2_p), systematic error variance (σ^2_r), and random error variance (σ^2_e).

ROM Test	Side	ICC [95% CI]	SEM	σ^2_p	σ^2_r	σ^2_e
Ankle Dorsiflexion Unrestricted	Left	0.43 [0.18, 0.63]	5.7	47.2	17.6	24.9
	Right	0.52 [0.26, 0.71]	5.0	49.9	12.4	18.2
Ankle Dorsiflexion Restricted	Left	0.32 [0.10, 0.53]	5.8	34.0	22.1	27.1
	Right	0.28 [0.08, 0.48]	5.4	23.9	17.2	24.5
MTPJ1 Dorsiflexion Unrestricted	Left	0.55 [0.33, 0.72]	8.4	148.2	23.2	58.5
	Right	0.56 [0.38, 0.72]	10.2	214.8	17.4	94.5
MTPJ1 Dorsiflexion Restricted	Left	0.66 [0.50, 0.78]	7.0	146.5	5.1	45.9
	Right	0.57 [0.38, 0.73]	7.3	117.2	12.5	46.7
Hip Extension Unrestricted	Left	0.34 [0.10, 0.56]	5.4	34.3	22.7	21.9
	Right	0.34 [0.14, 0.54]	5.5	29.8	12.0	26.5
Hip Extension Restricted	Left	0.49 [0.24, 0.69]	6.4	73.5	20.0	30.2
	Right	0.5 [0.27, 0.69]	6.2	70.5	16.5	29.9
Hip Flexion Unrestricted	Left	0.56 [0.34, 0.73]	7.2	113.2	17.8	41.4
	Right	0.67 [0.52, 0.79]	6.2	120.8	4.1	35.8
Hip Flexion Restricted	Left	0.59 [0.38, 0.75]	9.2	206.4	25.8	69.7
	Right	0.62 [0.44, 0.77]	9.3	230.1	19.3	73.8
Shoulder Flexion Unrestricted	Left	0.37 [0.14, 0.58]	8.3	82.9	38.9	54.9
	Right	0.48 [0.26, 0.66]	9.4	144.9	31.9	72.9
Shoulder Flexion Restricted	Left	0.47 [0.18, 0.68]	8.6	134.9	51.4	49.4
	Right	0.56 [0.32, 0.73]	8.1	149.2	27.3	51.3

The differing changes in intra-rater reliability and agreement with adjacent joint restriction were primarily influenced by σ^2_p and σ^2_e. Although changes in σ^2_p and σ^2_e each have the potential to influence the ICC and SEM, it appears as though changes in ICC were more influenced by changes in σ^2_p and changes in SEM were more influenced by changes in σ^2_e (Table 2). However, it may not be appropriate to generalize these associations. Regardless, intra-rater reliability and agreement were not substantially influenced by adjacent joint restriction as the ICC scores were still excellent and the differences between the SEM scores were not larger than approximately 1°, with the exception of right-side unrestricted shoulder flexion (Table 2). Upon examining the data, the lower ICC and higher SEM for right side shoulder flexion was likely caused by an outlier trial from rater 3, which measured 270° of shoulder flexion. It appears as though this may be unrealistic but the trial was not removed from the data to maintain equal number of trials across raters.

Inter-rater comparisons provided a more complex scenario due to the higher σ^2_r, which was almost nonexistent in the intra-rater comparison. The decreased inter-rater ICC and increased SEM can be attributed to the systematic and random error that occurred between raters, as σ^2_r and σ^2_e

were higher while $\sigma^2{}_p$ was similar to the intra-rater values. Adjacent joint restriction differentially influenced these components of variance and the resulting ICC and SEM. Improvements in reliability and agreement observed with hip extension and shoulder flexion were due to changes $\sigma^2{}_p$, which suggests that raters were not necessarily better at making the measurements even though ICC and SEM scores improved with adjacent joint restriction. Conversely, the improvements in MTPJ1 with joint restriction, can be attributed to fewer differences between raters and less random error as decreases in $\sigma^2{}_r$ and $\sigma^2{}_e$ caused the observed improvements in ICC and SEM. When multiple raters are involved in the measurement process, it may be beneficial to determine which technique (unrestricted/restricted) provides higher reliability and agreement and standardize it across raters. Although this may also be important with intra-rater comparisons, it may not be as important, because the intra-rater comparisons scores were excellent even with changes due to adjacent joint restriction.

The inter-rater comparison demonstrates the utility of reporting the components of variance used to calculate the ICC and SEM, as it improved interpretation of the scores. Without investigating the components of variance, it is difficult to determine why the changes in ICC and SEM occurred which can lead to misleading conclusions. Reporting the components of variance has the potential to also improve between-study comparisons that use different participants, raters and methodologies, towards enhancing the consolidation of knowledge to guide goniometry. For example, standardizing the measurement protocol and training raters has been investigated as a means to improve reliability and agreement.[5] Evaluating the variance components within the data can elucidate how different protocols may differentially influence the systematic and random error, which can be used to inform on how protocols can be tailored to target improvements in reliability and agreement, beyond the constraints of a specific study. Interpreting the data in this way can contribute to developing a conceptual framework that can inform future research and decision making in practice.

Considering the limitations associated with the current investigation, maximal ROM may not have changed between measurements, but it is possible that the repositioning between raters could have contributed to $\sigma^2{}_r$ and $\sigma^2{}_e$. Although a single researcher positioned participants to control these differences, a criterion measure to confirm repositioning consistency and accuracy was not performed. The study design only provides rater reliability within a single session or day. Although apparently healthy participants were investigated, it is certainly possible that they could have had limitations in bi-articular muscle flexibility which would have influenced their ROM with adjacent joint restriction and the variance between participants ($\sigma^2{}_p$). Thus, the results from this study could represent the reliability and agreement when assessing heterogeneous groups such as sport teams but would not necessarily be well representative of other homogenous groups such as those with a specific condition that would influence their muscle flexibility. Due to challenges in making comparisons between studies and the scarcity of research on the specific ROM tests and techniques investigated here, it is difficult to make meaningful comparisons to other studies.

CONCLUSION

Restriction of adjacent joints influenced measurement reliability and agreement differentially across the five ROM tests. The changes due to adjacent joint restriction were more pronounced in the inter-rater reliability and agreement, whereas intra-rater reliability and agreement were not substantially influenced. Thus, the effects of adjacent joint restriction on reliability and agreement depends on the ROM test and is of higher importance when multiple raters are involved. Estimating the components of data variance improved the interpretation of the ICC and SEM scores, which demonstrates the utility in reporting variance components in future work, to potentially improve between-study comparisons and towards developing a conceptual framework to guide goniometry.

REFERENCES

1. Medeiros DM, Martini TF. Chronic effect of different types of stretching on ankle dorsiflexion range of motion: Systematic review and meta-analysis. *The Foot*. 2018;34:28-35. doi:10.1016/j.foot.2017.09.006

2. van Trijffel E, van de Pol RJ, Oostendorp RAB, Lucas C. Inter-rater reliability for measurement of passive physiological movements in lower extremity joints is generally low: a systematic review. *J Physiother*. 2010;56(4):223-235. doi:10.1016/s1836-9553(10)70005-9

3. Moreside JM, McGill SM. Quantifying normal 3D hip ROM in healthy young adult males with clinical and laboratory tools: Hip mobility restrictions appear to be plane-specific. *Clin Biomech*. 2011;26(8):824-829. doi:10.1016/j.clinbiomech.2011.03.015

4. Chapleau J, Canet F, Petit Y, Laflamme GY, Rouleau DM. Validity of goniometric elbow measurements: Comparative study with a radiographic method. *Clin Orthop*. 2011;469(11):3134-3140. doi:10.1007/s11999-011-1986-8

5. Mohsin F, McGarry A, Bowers RJ. Factors influencing the reliability of the universal goniometer in measurement of lower-limb range of motion: A literature review. *J Prosthet Orthot*. 2015;27(4):140-148. doi:10.1097/jpo.0000000000000074

6. Jonson LSR, Gross MT. Intraexaminer reliability, interexaminer reliability, and mean values for nine lower extremity skeletal measures in healthy naval midshipmen. *J Orthop Sports Phys Ther*. 1997;25(4):253-263. doi:10.2519/jospt.1997.25.4.253

7. Kim PJ, Peace R, Mieras J, Thoms T, Freeman D, Page J. Interrater and intrarater reliability in the measurement of ankle joint dorsiflexion is independent of examiner experience and technique used. *J Am Podiatr Med Assoc*. 2011;101(5):407-414. doi:10.7547/1010407

8. Prather H, Harris-Hayes M, Hunt DM, Steger-May K, Mathew V, Clohisy JC. Reliability and agreement of hip range of motion and provocative physical examination tests in asymptomatic volunteers. *PM&R*. 2010;2(10):888-895. doi:10.1016/j.pmrj.2010.05.005

9. Keogh JWL, Cox A, Anderson S, et al. Reliability and validity of clinically accessible smartphone applications to measure joint range of motion: A systematic review. Müller J, ed. *PLOS ONE*. 2019;14(5):e0215806. doi:10.1371/journal.pone.0215806

10. Cejudo A, Ayala F, De Baranda PS, Santonja F. Reliability of two methods of clinical examination of the flexibility of the hip adductor muscles. *Int J Sports Phys Ther*. 2015;10(7):976-983.

11. de Vet HCW, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. *J Clin Epidemiol*. 2006;59(10):1033-1039. doi:10.1016/j.jclinepi.2005.10.015

12. Koo TK, Li MY. A Guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*. 2016;15(2):155-163. doi:10.1016/j.jcm.2016.02.012

13. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychol Methods*. 1996;1(1):30-46. doi:10.1037/1082-989x.1.1.30

14. Weir JP. Quantifying Test-retest reliability using the intraclass correlation coefficient and the SEM. *J Strength Cond Res*. 2005;19(1):231. doi:10.1519/15184.1

15. Liljequist D, Elfving B, Skavberg Roaldsen K. Intraclass correlation – A discussion and demonstration of basic features. Chiacchio F, ed. *PLOS ONE*. 2019;14(7):e0219854. doi:10.1371/journal.pone.0219854

16. Rankin G, Stokes M. Reliability of assessment tools in rehabilitation: an illustration of appropriate statistical analyses. *Clin Rehabil*. 1998;12(3):187-199. doi:10.1191/026921598672178340

17. van Kooij YE, Fink A, Nijhuis-van der Sanden MW, Speksnijder CM. The reliability and measurement error of protractor-based goniometry of the fingers: A systematic review. *J Hand Ther*. 2017;30(4):457-467. doi:10.1016/j.jht.2017.02.012

18. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol Assess*. 1994;6(4):284-290. doi:10.1037/1040-3590.6.4.284

SUPPLEMENTARY MATERIALS

Appendix A

Download: https://ijspt.scholasticahq.com/article/30998-adjacent-joint-restriction-differentially-influences-intra-and-inter-rater-reliability-and-agreement-of-goniometric-measurements/attachment/78678.pdf?auth_token=Q5RVIcgi-9-cJe7Gm0To