

Rafael P. Ribeiro*, MSc, Filipe G. Guerrero, DO, Eduardo N. Camargo, Luiza R. Pivotto, MSc, Mateus A. Aimi, Jefferson F. Loss, PhD and Cláudia T. Candotti, PhD

Construct validity and reliability of tests for sacroiliac dysfunction: standing flexion test (STFT) and sitting flexion test (SIFT)

<https://doi.org/10.1515/jom-2021-0025>

Received February 23, 2021; accepted June 29, 2021;

published online September 22, 2021

Abstract

Context: Sacroiliac dysfunction is characterized by a hypomobility of the range of motion of the joint, followed by a positional change regarding the relationship between the sacrum and the iliac. In general, the clinical tests that evaluate the sacroiliac joint (SIJ) and its dysfunctions lack validity and reliability values.

Objectives: This article aims to evaluate the construct validity and intra- and inter-rater reliability of the standing flexion test (STFT) and sitting flexion test (SIFT).

Methods: In this prospective study, the sample consisted of 30 individuals of both sexes, and the evaluation team was composed of five researchers. The evaluations took place on two different days: first day, inter-rater reliability and

construct validity; and second day, intra-rater reliability. The reference standard for the construct validity was 3-dimensional measurements obtained utilizing the BTS SMART-DX system. For statistical analysis, the percentage (%) agreement and the kappa statistic (K) were utilized.

Results: The construct validity was determined for STFT (70% agreement; $K=0.49$; $p<0.01$) and SIFT (56.7% agreement; $K=0.29$; $p<0.05$). The intra-rater reliability was determined for STFT (66.3% agreement; $K=0.43$; $p<0.01$) and SIFT (56.7% agreement; $K=0.38$; $p<0.01$). The inter-rater reliability was determined for STFT (10% agreement; $K=-0.02$; $p=0.825$) and SIFT (13.3% agreement; $K=0.01$; $p=0.836$).

Conclusions: The STFT confirmed the construct validity and was reliable when applied by the same rater to healthy people, even if the rater had no experience. It was not possible to achieve minimum scores using the SIFT either for construct validity or reliability. We suggest that further studies be conducted to investigate the measurement properties of palpatory clinical tests for SIJ mobility, especially in symptomatic patients.

Keywords: articular; musculoskeletal manipulations; range of motion; sacroiliac joint; sitting flexion test (SIFT); standing flexion test (STFT).

Sacroiliac dysfunction is characterized by hypomobility of the range of motion of the sacroiliac joint (SIJ), followed by a positional change regarding the relationship between the sacrum and the iliac [1]. This dysfunction can be a source of pain in the lower back and buttocks [2–6] and can also radiate to the posterior part of the thighs and anterior part of the inguinal region [7].

Of the range of SIJ assessment possibilities [7], the standing flexion test (STFT) and the sitting flexion test (SIFT) are among the options for assessing the mobility of this joint [8]. The STFT is utilized mainly to investigate the iliac, while the SIFT investigates the sacrum, and both tests are among those most utilized by osteopathic physicians [8–10]. However, based on a recent systematic review [11], there are no studies in the scientific literature reporting on the diagnostic

*Corresponding author: **Rafael P. Ribeiro**, MSc, School of Physical Education, Physiotherapy and Dance (ESEFID), Brazilian Institute of Osteopathy (IBO), Federal University of Rio Grande do Sul (UFRGS), 500 São Vicente St, Porto Alegre, Rio Grande do Sul, 90630-180, Brazil, E-mail: rpaivaribeiro2@gmail.com. <https://orcid.org/0000-0002-2696-9026>

Filipe G. Guerrero, DO, Brazilian Institute of Osteopathy (IBO), Porto Alegre, Rio Grande do Sul, Brazil. <https://orcid.org/0000-0003-1243-3850>

Eduardo N. Camargo, School of Physical Education, Physiotherapy and Dance (ESEFID), Brazilian Institute of Osteopathy (IBO), Federal University of Rio Grande do Sul (UFRGS), Porto Alegre, Rio Grande do Sul, Brazil; and Brazilian Institute of Osteopathy (IBO), Porto Alegre, Rio Grande do Sul, Brazil. <https://orcid.org/0000-0002-1562-6431>

Luiza R. Pivotto, MSc, **Mateus A. Aimi**, **Jefferson F. Loss**, PhD and **Cláudia T. Candotti**, PhD, School of Physical Education, Physiotherapy and Dance (ESEFID), Federal University of Rio Grande do Sul (UFRGS), Porto Alegre, Rio Grande do Sul, Brazil. <https://orcid.org/0000-0001-9746-9996> (L.R. Pivotto). <https://orcid.org/0000-0003-1667-8409> (M.A. Aimi). <https://orcid.org/0000-0001-5948-6357> (J.F. Loss). <https://orcid.org/0000-0002-8676-9157> (C.T. Candotti)

validity of these tests. This is a worrying scenario, because diagnostic validity refers to how well the test truly assesses the characteristic it is intended to evaluate as judged by external criteria (i.e., gold standard) [12]. There is no widely accepted reference standard for diagnosing SIJ mobility. Thus, we speculate that the lack of diagnostic validity identified in the literature is related to the lack of a gold standard for these tests. Alternatively, the STFT and SIFT tests could be compared with other tests that purport to measure the same characteristic, a procedure called construct validity [13].

For any testing instrument to be considered useful, it must be both a valid and reliable measure of the variable it is designed to assess [14]. Reliability refers to the consistency of the test in repeated trials. In addition, the same systematic review pointed out that good agreement for intra-rater reliability was only found for SIFT [11]. Intra-rater reliability, which is the agreement between the assessments of the same rater when applying the test at different times [15], needs to be confirmed, so that it is possible, for example, to be sure that the changes detected between tests are due to an intervention. It also pointed out there was no information in the literature about inter-rater reliability (i.e., the agreement between different raters, assessing the same subject) [16], which is necessary to interchange information between professionals.

Thus, the objective of this study was to evaluate the construct validity and determine the intra- and inter-rater reliabilities of the STFT and SIFT.

Methods

This study was registered in the Brazilian Clinical Trials Registry (ReBec) under approval number RBR-9kb7km9. The date range was between July 2019 and November 2020, and it was written according to the Guidelines for Reporting Reliability and Agreement Studies (GRRAS) [15] for the analyses of reliability and validity.

Study design and participants

This was a prospective study with a sample composed of individuals of both sexes. The research participants were volunteers who were identified by invitation on social networks. The evaluations took place in the biomechanics sector of a university research laboratory. For the reliability and validity analyses, the sample was defined based on a two-tailed test, adopting a 90% power, assuming a null hypothesis, a kappa value (K) of 0.00, and a detectable kappa (K) of 0.70, with a proportion of positive ratings of 0.50, resulting in 22 individuals [16]. Predicting 30% of sample loss, a target number of 30 individuals was adopted.

The inclusion criteria were individuals who were: (1) between 18 and 60 years of age; (2) non-obese (BMI < 30 kg/m²) [17]; and (3) without surgery of the lumbar area, pelvis, or hip. The exclusion criteria were: (1) low back or hip pain at the time of data collection; (2) inability to

carry out the protocol tests, that is, not being able to flex the trunk and not being able to remain seated without a back support; and (3) more than 2 cm difference in length between the lower limbs (MsIs) [18].

Evaluation protocol

The evaluation team was composed of three raters (A, B, C) — one osteopath DO (A) and two third-year osteopathy students (B and C) — in addition to two researchers (D, E), both of whom were physiotherapists. Rater A was the most experienced, with 11 years of clinical practice in applying the protocol tests. Raters B and C had 2 years of experience in applying the tests. Researcher D was responsible for registering the history of the individuals and randomizing the evaluations, and researcher E was responsible for data collection using the BTS SMART-DX system (BTS Bioengineering, Milan, Italy). Raters A, B, and C underwent a 20-h training course, in which the way to carry out the tests, their commands, and the complete evaluation protocol were tested.

The study was approved by the university ethics committee (CAAE: 15499219.4.0000.5347), and the participants signed an informed consent form (ICF). The evaluations took place on two different days: first day, inter-rater reliability and construct validity; and second day (after 24–72 h) intra-rater reliability. The standard reference utilized for construct validity was the measurements based on 3-dimensional (3D) kinematics using the BTS SMART-DX system (BTS Engineering, Milan, Italy), which is a motion-capture system.

First day: inter-rater reliability and construct validity

The history taking was obtained by researcher D, and familiarization was carried out by rater A, which consisted of two to three executions of the sequence of test movements. When necessary, corrections were made, such as flexing the column correctly, carrying out the movement more slowly, not flexing the knees, and placing the feet closer or farther apart. After familiarization, the participants underwent three consecutive evaluations of the STFT and SIFT. The order of the tests and of the three raters (A, B, C) were randomized using the simple randomization method, with envelopes administered by researcher D. During the evaluation by rater A, 3D kinematic measurements were carried out simultaneously.

At the end of the first day, the participant received a reminder sheet showing the date and time of the second evaluation day and guidelines for not carrying out activities that involved physical effort or therapeutic treatments, such as physiotherapy, osteopathy, and chiropractic.

Second day: intra-rater reliability

On the second evaluation day, the participant was asked about any pain he/she had at that moment and if the reminder sheet—given at the end of the first day—had been utilized. STFT and SIFT were applied by rater B. The order of the tests was also randomized using the simple randomization method, with envelopes administered by researcher D. The interval between the first and second days (24–72 h) was stipulated in order to minimize any possible changes in the anatomical characteristics of the participants between the evaluation days.

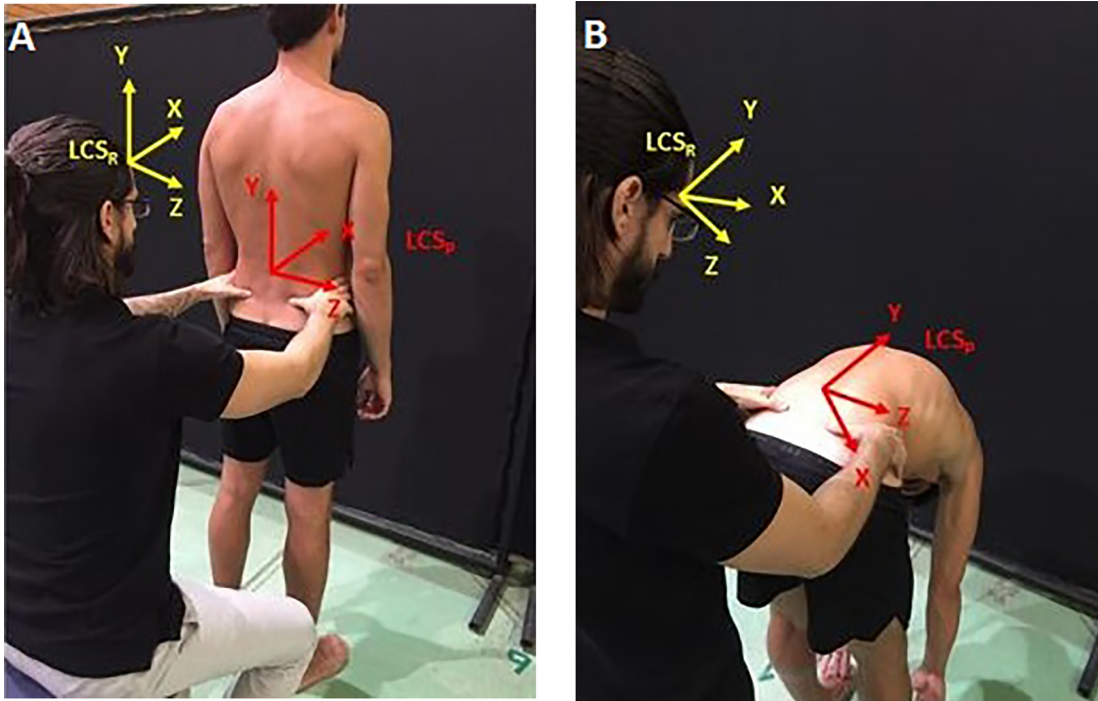


Figure 1: The standing flexion test (STFT) with the rater's local coordinate system (LCS_R) is indicated in yellow, and the participant's local coordinate system (LCS_p) is indicated in red. The Y axes of each system were kept parallel in each position: (A) Initial test position, and (B) final test position. The rater and participant were both male and 26 years old.

History taking

The clinical and demographic information was obtained by researcher *D* on the first evaluation day to identify whether the participant met the eligibility criteria. It also made it possible to collect data such as the date of birth, height, and body mass, based on the self-report made by the subject.

Standing flexion test (STFT)

The participants were instructed to remain in a standing position, with the upper limbs beside the body and the feet in line with the hips [19], positioned in parallel with no angle of rotation. The rater positioned himself behind the participant, placed his hands laterally on the iliac crests, and moved his thumbs to find the posterior superior iliac spines (PSISs). The pads of the thumb tips were positioned on the lower obliquity of the PSISs (Figure 1A). The participants were then instructed to slowly carry out maximum back flexion, starting the movement in the cervical region and keeping the knees extended (Figure 1B). The test was considered negative if the movement of the PSISs was symmetrical or positive if one side moved more than the other in the cephalic and/or ventral directions [5, 9, 20, 21]. Three results were possible: negative test, positive on the right (*R*), and positive on the left (*L*) [5, 9, 20, 21]. Raters *A*, *B*, and *C* noted the test results on a spreadsheet, grouped in three blocks: (1) initial position (cephalic *R* PSIS, cephalic *L* PSIS, or symmetrical PSISs); (2) final

position (cephalic *R* PSIS, cephalic *L* PSIS, or symmetrical PSISs); and (3) test conclusion (negative, positive for *R*, or positive for *L*). For the initial and final positions, the raters compared the *R* PSIS with the *L* one.

Sitting flexion test (SIFT)

The SIFT test is similar to the STFT, but the individuals start from a sitting position. The participants were instructed to sit in a height-adjustable seat, with the back erect and the feet placed on a flat surface in parallel with no angle of rotation, the knees and hips at shoulder width, and approximately 90° of flexion. The position of the rater was the same as for the STFT for palpation of the PSISs. The participants were then instructed to place their hands behind their heads, bring their elbows together, and slowly carry out maximum back flexion, starting the movement in the cervical region [5, 20–22]. The possible test results were the same as those for the STFT.

3D kinematic measurements

For the construct validity of the tests, the 3D measurements (3D kinematics) were carried out with 10 infrared cameras (4 MPixels) with a sampling rate of 100 Hz and assisted by the BTS Smart Capture software. Spherical, 15-mm-diameter reflective markers were fixed on the thumbs of rater *A* using double-sided tape (Figure 2). Prior to collection, the BTS System was calibrated according to the

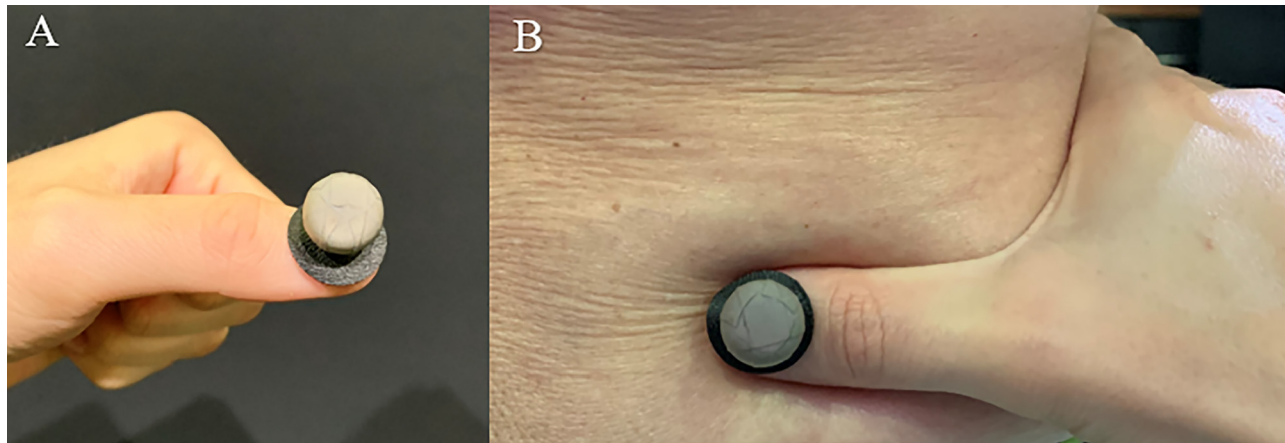


Figure 2: Reflective marker on the thumb of the rater A (A) and palpation of the PSIS (B).

manufacturer's instructions (movement of known-distance points within the calibration volume) reaching an error of less than 0.3 mm.

The location of the PSISs in space was obtained from a local coordinate system (LCS_R). The construction of the LCS consisted of a cluster with four points fixed to a band on the forehead of rater A. A second coordinate system (LCS_P) was placed in the lumbar region of the participant, based on the representative points of the PSISs (thumbs of rater A) and another marker placed on L3 (Figure 1). During the execution of the tests, rater A was instructed to follow the participant's movement with his head, so that the LCS_R and LCS_P remained at a similar angle to each other. The kinematic data were smoothed using a fourth-order low-pass Butterworth filter with a cutoff frequency of 1 Hz.

Two static measurements were made for each participant, each lasting 10 s: (1) initial position, standing (STFT) or sitting (SIFT); and (2) final position, in maximum standing (STFT) or sitting (SIFT) back flexion. The position of the points was defined from the average of the central 8 s of each static measurement. Any asymmetry between the PSISs was obtained by varying the position of the points (right PSIS and left PSIS) on the Y axis (participant's caudal-cranial region), taking into account the final and initial positions. The conclusion of the test was obtained by analyzing the displacement between the initial and final positions of the individual.

For analytical purposes, the result of the asymmetry between the PSISs was determined using a 3-mm cutoff point. The choice of this value was based on the experience of the raters in carrying out the tests and according to the systematic review by Goode et al. [23]. The authors reported a translation mobility of the SIJ of 4–5 mm with a standard error of 1.3 mm, during the bilateral hip flexion movement. The asymmetries were then classified as: (1) negative, positive to R, and positive to L, for the conclusion of the test; and (2) cephalic R PSIS, cephalic L PSIS, and symmetrical, for the initial and final positions.

Statistical analysis

The data was first organized using Microsoft Office Excel 2016 software, and the statistical analysis carried out using the Statistical Package for the Social Sciences (IBM SPSS Statistics 21) software. The significance level adopted was <0.05 .

The validity and reliability data were analyzed using the percentage (%) agreement and the unweighted K value, considering all disagreements equally. For K values, the classification adopted was according to Altman [24], where K values < 0.20 = poor, 0.21 – 0.40 = light, 0.41 – 0.60 = moderate, 0.61 – 0.80 = good, and 0.81 – 1 = very good. For the % agreement values, the classification adopted was according to Janse et al. [25], in which % agreement < 0.30 = poor, 0.31 – 0.50 = weak, 0.51 – 0.70 = moderate, 0.71 – 0.90 = good, and 0.91 – 1.00 = excellent. In this study, the minimum criterion adopted to consider the tests valid and reliable was a moderate agreement classification (% agreement > 0.50 and $K > 0.40$).

Results

The sample consisted of 30 individuals, 17 males and 13 females, with an average age of 30 years (standard deviation [SD], ± 8 ; range, 22–60 years), an average height of 171 cm (SD, ± 9 ; range, 153–196 cm), an average weight of 71 kg (SD, ± 12 ; range, 56–105 kg), and mean body mass index (BMI) of 24 kg/m^2 (SD, ± 2 ; range, 21–27 kg/m^2). There was no sample loss, and no individuals were excluded by the eligibility criteria.

Regarding the conclusion of the STFT, the comparison of the results of rater A with the 3D kinematic measurements (construct validity) presented good % agreement (70%) and a moderate K value (0.49). For the SIFT, the test conclusion did not show sufficient results for the minimum criteria adopted, indicating that the validity of this test was not confirmed (Table 1).

In addition, for the STFT, the K value was moderate and the % agreement good, both for the initial (0.57; 80%) and final (0.60; 80%) positions. On the other hand, for the SIFT, although the K value was moderate (0.56) and the % agreement was good (76.7%) for the initial position, The K

Table 1: Construct validity (rater A \times 3D kinematic measurements) of the STFT and SIFT.

	Standing flexion test (STFT)			Sitting flexion test (SIFT)			
	n (% agreement)	K (p-Value)	CI 95%	n (% agreement)	K (p-Value)	CI 95%	
Construct validity	Initial position	24 (80%)	0.57 (<0.01)*	0.281–0.861	23 (76.7%)	0.56 (<0.01)*	0.291–0.835
	Final position	24 (80%)	0.60 (<0.01)*	0.316–0.884	20 (66.7%)	0.40 (<0.01)*	0.126–0.666
	Conclusion of the test	21 (70%)	0.49 (<0.01)*	0.198–0.774	17 (56.7%)	0.29 (<0.05)*	0.005–0.581

p: significance level; *: $p < 0.05$; n: absolute agreement values; K: Kappa. The minimum criterion adopted to consider the tests valid and reliable was a moderate agreement classification (% agreement > 0.50 and $K > 0.40$).

Table 2: Intra-rater ($B \times B$) and inter-rater ($A \times B \times C$) reliability of the STFT and SIFT.

	Standing flexion test (STFT)			Sitting flexion test (SIFT)			
	n (% agreement)	K (p-Value)	CI 95%	n (% agreement)	K (p-Value)	CI 95%	
Intra-rater	Initial position	20 (66.7%)	0.38 (<0.01)*	0.131–0.625	19 (66.3%)	0.39 (<0.01)*	0.142–0.636
	Final position	19 (66.3%)	0.43 (<0.01)*	0.160–0.700	16 (53.3%)	0.31 (<0.01)*	0.056–0.566
	Conclusion of the test	19 (66.3%)	0.43 (<0.01)*	0.160–0.700	17 (56.7%)	0.38 (<0.01)*	0.154–0.612
Inter-rater	Initial position	8 (26.7%)	0.05 (0.56)	–0.116–0.215	9 (30%)	0.07 (0.45)	–0.110–0.249
	Final position	4 (13.3%)	0.02 (0.77)	–0.127–0.173	4 (13.3%)	0.04 (0.63)	–0.116–0.192
	Conclusion of the test	3 (10%)	–0.02 (0.83)	–0.166–0.132	4 (13.3%)	0.01 (0.84)	–0.132–0.163

p: significance level; *: $p < 0.05$; n: absolute agreement values; K: kappa. The minimum criterion adopted to consider the tests valid and reliable was a moderate agreement classification (% agreement > 0.50 and $K > 0.40$).

was light (0.40) and the % agreement moderate (66.7%) for the final position (Table 1).

Considering the test conclusions, the intra-rater reliability ($B \times B$) was only confirmed for the STFT, with moderate 66.3% % agreement and a K value of 0.43, whereas the inter-rater reliability ($A \times B \times C$) the results showed poor (10%) agreement and a K value of -0.02 (Table 2).

Also, according to the intra-rater reliability of the STFT, although the % agreement was moderate (66.3%) with a K value of (0.43) for the final position, the % agreement was also moderate (66.7%), but the K was light (0.38) for the initial position (Table 2).

For the intra-rater reliability of the SIFT, the results for the initial and final positions were similar to those of the test conclusion, with a moderate % agreement (53.3–66.3%) and light K value (0.31–0.39) (Table 2).

For both STFT and SIFT, the inter-rater reliability results for both the initial and final positions showed poor % agreement (13.3–30%) and poor K values (0.02–0.07) (Table 2).

Discussion

The results only allowed for the validation of the STFT, because the conclusions of this test presented moderate percentage agreement and a moderate K value. However, it

was not possible to validate the SIFT, since the values for the percentage agreement and K of the test conclusions did not reach the minimum criterion adopted (% agreement > 0.50 and $K > 0.40$). No evidence was found in the literature on the performance of the construct validity analysis for STFT and SIFT. Some studies were carried out investigating SIJ mobility but did not utilize a clinical test. Bussey et al. [26] utilized the computed tomography exam and also a magnetic tracking device, digitizing the anatomical references and calculating the measurements with 3D coordinates, during the abduction movement and external rotation of the hip in the prone position. Stuesson et al. [4] and Kibsgård et al. [27] utilized radiostereometric analysis (RSA), and SIJ mobility was also calculated using 3D coordinates with the implantation of markers in the joint. Stuesson et al. [4] evaluated SIJ mobility during movements from the supine to sitting positions and from the supine to standing positions, and also hyperextension of the hip in the prone position. Thus, we consider our study to be pioneer, utilizing a 3D system that has historically been utilized for the analysis of biomechanical motion to study clinical tests.

Considering the lack of evidence regarding the validation of these tests, we sought to expand the forms of analysis, subdividing the results obtained in the initial and final positions and the conclusions. For the STFT, the establishment of a difference of at least 3 mm seems to have been sufficient for an identification by the human eye of

asymmetries between the PSISs, in agreement with clinical practice, in which small asymmetries are important for the clinician. However, it is important to highlight that these motion palpation tests do not provide a definitive diagnostic, and according to Nejati et al. [28], it is advisable to utilize a combination of such tests in conjunction with provocation tests and other data sources, including the patient's history and imaging exams, to accurately diagnose SIJ dysfunction. In addition, when choosing the test, it is also important to consider its reliability.

Regarding reliability (Table 2) in the present study, only the STFT was reliable when applied by the same rater. These results suggest that evaluations made using the STFT, applied by the same osteopath, may be reliable for monitoring the evolution of the treatment and to assess interventional changes, when asymptomatic patients are evaluated and the present results were corroborated by other studies that also assessed healthy patients [9, 29]. However, caution is advised when using this test with multi-professional teams, considering the poor agreement and poor K values for inter-rater reliability.

Previous studies on the intra-rater reliability of STFT corroborated the present results, showing agreement ranging from moderate to good, with the percentage agreement from 68 to 87% [9, 29] and K values from 0.46 to 0.70 [9, 29, 30]. Concerning the STFT inter-rater reliability, previous studies showed agreement ranging from poor to moderate, with the percentage agreement between 42.7 and 59% [5, 9, 20, 29, 31], and K values for most studies between 0.05 and 0.32 [9, 20, 29, 31], in agreement with the present results. Only one study had a moderate K value (0.51) [30].

In previous studies, the SIFT intra-rater reliability agreement ranged from mild to good. However, the results of these studies were heterogeneous, with K values ranging from 0.29 to 0.73 [22, 30, 32], and only one study analyzed the % agreement and obtained a value of 58.1% [22]. The present results were lower than the pre-established threshold of $K > 0.40$, demonstrating the difficulty of the same rater to repeat the results of this test. The difficulty of this test to be repeated is illustrated by the results of a recent systematic review [11].

In previous studies, the SIFT inter-rater reliability agreement ranged from poor to good, with the percentage agreement ranging from 34.4 to 71% [5, 22, 32]. The K values ranged from 0.06 to 0.14 [22, 32, 33], with only one study showing a good K [30]. According to Fryer et al. [32], there are many factors that can contribute to inter-rater inconsistency, such as expectations and clinical diagnostic skills, fatigue, distraction, degree of asymmetry, movements of the subject, fat composition, and tissue thickness.

In the study by Fryer et al. [32], the raters were divided into two groups: trained and untrained. Both groups carried out the STFT and the PSIS palpations, but both groups obtained inter-rater reliability results with K values < 0.20 . For the SIFT intra-rater reliability, the trained group obtained a K value of 0.41 and the untrained group achieved a K value of 0.02. For PSIS palpation, the K value was similar for the two groups (0.54 and 0.49, respectively). These results suggest that the rater's experience contributes to the repetition of the results achieved by the same rater, but experience is not sufficient when intending to share information obtained from different raters.

For both the STFT and SIFT, the percentage agreement and K values were higher when the initial and final positions were analyzed, as compared to the conclusion of the test. This difference indicates that the rater was more capable of identifying the symmetry of the PSISs in static situations, at the beginning and at the end of the test. For the 3D kinematic measurements, the conclusion of the test is just a difference of positions in space, and the rater needs to decide if there has been sufficient displacement of the anatomical references. This rater's interpretation carries a degree of subjectivity that seems to have an effect on the agreement of the tests. The authors believe that the fact that the magnitude of the differences was to the order of a few millimeters was a determining factor in the difficulty of agreement between the rater's interpretations and the data of the 3D kinematic measurements.

A possible interference in the results of the SIFT was the position of the head of rater A when applying the test. Because both the rater and the participant were seated, this may have made it difficult to angle the local coordinate systems (LCS_R and LCS_P). If this hypothesis is true, we assume it was a limitation of the evaluation protocol for this study. Furthermore, it is important to highlight the possibility of the SIJ mobility changing both during the three successive assessments on the same day and from one day to the next. In this sense, the reliability values, both intra- and inter-rater reliability, may have been impacted by a possible change in the condition of the participant. Also, the fact that the most experienced rater did not take part in the intra-rater reliability can also be considered a limitation. On the other hand, if a non-experienced rater reached a reliable index, one can speculate that the experienced rater would also reach a good index of reliability.

Another limitation is that the interpretation of the STFT and SIFT varies according to the reference, positive tests can be found on the side of the PSIS that moved the most or the side that moved last, and in the current study, the first form was utilized. Finally, it is important to emphasize that the use of the K score has limitations and that its interpretation

is not so straightforward. There are other factors that can influence the magnitude of this coefficient, such as prevalence, bias, and non-independence of the ratings [13], which are factors not addressed in the present study.

Conclusions

The construct validity of the STFT was confirmed, and it is reliable when applied by the same rater to healthy people, even if the rater has no experience. On the other hand, under the same conditions, minimum scores were not obtained in the SIFT for either construct validity or reliability. Thus, osteopaths can utilize the STFT as a scientifically based test to carry out clinical practice on asymptomatic patients. We suggest that further studies be conducted to investigate the measurement properties of palpatory clinical tests for SIJ mobility, especially in symptomatic patients.

Research funding: None reported.

Author contributions: All authors provided substantial contributions to conception and design, acquisition of data, or analysis and interpretation of data; R.P.R., E.N.C., J.F.L., and C.T.C. drafted the article or revised it critically for important intellectual content; R.P.R., E.N.C., J.F.L., and C.T.C. gave final approval of the version of the article to be published; all authors agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Competing interests: None reported.

Informed consent: All participants in this study provided written informed consent prior (paper format) to participation.

Ethical approval: This study was reviewed and approved (number 15499219.4.0000.5347) by the research ethics committee of the Federal University of Rio Grande do Sul. This study was registered in the Brazilian Clinical Trials Registry (ReBec) under approval number RBR-9kb7km9.

References

- Dreyfuss P, Dryer S, Griffin J, Hoffman J, Walsh N. Positive sacroiliac screening tests in asymptomatic adults. *Spine* 1994;19:1138–43.
- Maigne J, Aivaliklis A, Pfefer F. Results of sacroiliac joint double block and value of sacroiliac pain provocation tests in 54 patients with low back pain. *Spine* 1996;3:175–89.
- Shaw JL. The role of the sacroiliac joint as a cause of low back pain and dysfunction. In: Vleeming A, Mooney V, Snijders C, Dorman T, eds. *Proceedings from the first Interdisciplinary world congress on low back pain and its relation to the sacroiliac joint*. San Diego, CA; 1992: 67–80.
- Sturesson B, Selvik G, Uden A. Movements of the sacroiliac joints. A roentgen stereophotogrammetric analysis. *Spine* 1989;14:162–5.
- Potter NA, Rothstein JM. Intertester reliability for selected clinical tests of the sacroiliac joint. *Phys Ther* 1985;65:1671–5.
- Slipman CW, Whyte WS, Chow DW, Chou L, Lenrow D, Ellen M. Sacroiliac joint syndrome. *Pain Physician* 2001;4:143–52.
- Prather H. Pelvis and sacral dysfunction in sports and exercise. *Phys Med Rehabil Clin* 2001;11:805–36.
- Fryer G, Morse CM, Johnson JC. Spinal and sacroiliac assessment and treatment techniques used by osteopathic physicians in the United States. *Osteopath Med Prim Care* 2009;1:4.
- Vincent-Smith B, Gibbons P. Inter-examiner and intra-examiner reliability of the standing flexion test. *Man Ther* 1999;4:87–93.
- Potter NA, Rothstein JM. Intertester reliability for selected clinical tests of the sacroiliac joint. *J Women's Heal Phys Ther* 2006;30: 21–5.
- Ribeiro RP, Guerrero FG, Camargo EN, Beraldo LM, Candotti CT. Validity and reliability of palpatory clinical tests of sacroiliac joint mobility: a systematic review and meta-analysis. *J Manip Physiol Ther* 2021;44:307–18.
- Thomas RL, Zidan MA, Slovis TL. What you need to know about statistics part I: validity of diagnostic and screening tests. *Pediatr Radiol* 2015;45:146–52.
- Karros DJ. Statistical methodology: II. Reliability and validity assessment in study design, part B. *Acad Emerg Med* 1997;4:144–7.
- Karras DJ. Statistical methodology: II. Reliability and variability assessment in study design, part A. *Acad Emerg Med Off J Soc Acad Emerg Med* 1997;4:64–71.
- Kottner J, Gajewski BJ, Streiner DL. Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. *Int J Nurs Stud* 2011;48:661–71.
- Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Phys Ther* 2005; 85:257–68.
- World Health Organization. *Obesity: preventing and managing the global epidemic*. World Health Organization; 2000.
- Gurney B. Leg length discrepancy. *Gait Posture* 2002;15:195–206.
- Greenman PE. *Principles of manual medicine*. Lippincott Williams & Wilkins; 2003.
- Riddle DL, Freburger JK, North American Orthopaedic Rehabilitation Research Network. Evaluation of the presence of sacroiliac joint region dysfunction using a combination of tests: a multicenter intertester reliability study. *Phys Ther* 2002;82:772–81.
- Cibulka MT, Koldehoff R. Clinical usefulness of a cluster of sacroiliac joint tests in patients with and without low back pain. *J Orthop Sport Phys Ther* 1999;29:83–92.
- Paydar D, Thiel H, Gemmell H. Intra- and interexaminer reliability of certain pelvic palpatory procedures and the sitting flexion test for sacroiliac joint mobility and dysfunction. *J Neuromusculoskel Syst* 1994;2:65–9.
- Goode A, Hegedus EJ, Sizer P, Brismee JM, Linberg A, Cook CE. Three-dimensional movements of the sacroiliac joint: a systematic review of the literature and assessment of clinical utility. *J Man Manip Ther* 2008;16:25–38.
- Altman D. *Practical statistics for medical research*. London: CRC Press; 1990.
- Janse AJ, Gemke RJB, Uiterwaal CSPM, Van Der Tweel I, Kimpen JLL, Sinnema G. Quality of life: patients and doctors don't always agree: a meta-analysis. *J Clin Epidemiol* 2004;57:653–61.

26. Bussey MD, Yanai T, Milburn P. A non-invasive technique for assessing innominate bone motion. *Clin Biomech* 2004;19: 85–90.
27. Kibsgård TJ, Røise O, Stuge B, Röhrh SM. Precision and accuracy measurement of radiostereometric analysis applied to movement of the sacroiliac joint. *Clin Orthop Relat Res* 2012;470:3187–94.
28. Nejati P, Sartaj E, Imani F, Moeineddin R, Nejati L, Safavi M. Accuracy of the diagnostic tests of sacroiliac joint dysfunction. *J Chiropr Med* 2020;19:28–37.
29. Åström M, Gummesson C. Assessment of asymmetry in pelvic motion—an inter- and intra-examiner reliability study. *Eur J Physiother* 2014;16:76–81.
30. Arab AM, Abdollahi I, Joghataei MT, Golafshani ZKA. Inter- and intra-examiner reliability of single and composites of selected motion palpation and pain provocation tests for sacroiliac joint. *Man Ther* 2009;14:213–21.
31. Bowman C, Gribble R. The value of the forward flexion test and three tests of leg length changes in the clinical assessment of movement of the sacroiliac joint. *J Orthop Med* 1995;17: 66–7.
32. Fryer G, McPherson HC, O’Keefe P. The effect of training on the inter-examiner and intra-examiner reliability of the seated flexion test and assessment of pelvic anatomical landmarks with palpation. *Int J Osteopath Med* 2005;8:131–8.
33. Tong HC, Heyman OG, Lado DA, Isser MM. Interexaminer reliability of three methods of combining test results to determine side of sacral restriction, sacral base position, and innominate bone position. *J Am Osteopath Assoc* 2006;106:464–8.