

# **Inter-examiner and intra-examiner reliability of the seated flexion test**

**Paul Michael O'Keefe BSc**

**Gary Fryer, BAppSc (Osteo)**

School of Health Sciences, Victoria University, Melbourne, Australia

**Address Correspondence to:**

Gary Fryer, School of Health Sciences, Victoria University, PO Box 14428 MCMC,

VIC 8001, Australia

Gary.fryer@vu.edu.au

## ABSTRACT

**Background:** Motion palpation is widely used in the field of manual medicine despite a lack of research demonstrating its reliability. The seated flexion test is a motion test that has been advocated for the detection of sacroiliac joint dysfunction and has not been examined for reliability.

**Objective:** The aim of this study was to determine the inter- and intra-examiner reliability of the seated flexion test, and examine the influence of examiner training on reliability.

**Methods:** Ten final year osteopathic students were recruited as examiners. Five of the examiners participated in two training sessions to standardise the testing protocol whilst the remaining five examiners did not participate. The ten examiners performed the test on ten asymptomatic women, three times each.

**Results:** The mean inter-examiner reliability coefficient ( $k$ ) was 0.105, indicating “slight” agreement, whereas the mean intra-examiner reliability ( $k$ ) was 0.213, indicating “fair” agreement. Reliability of the trained group was slightly higher for both intra-examiner reliability ( $k=0.41$ ) and inter-examiner reliability ( $k=0.14$ ).

**Conclusion:** Examiner training appeared to produce a slight improvement in the inter-examiner and intra-examiner reliability of the seated flexion test. Neither group, however, achieved acceptable reliability for the seated flexion test to be recommended as a useful clinical test.

## Introduction

Low-back pain (LBP) is a significant health problem and is a major expense to the healthcare system.<sup>1</sup> Most population-based surveys of back pain report a point prevalence of 15%-30%, a one-year prevalence of 50%, and a lifetime prevalence of approximately 80%.<sup>2</sup> Present research suggests the aetiology of low back pain is unknown in 80-90% of cases indicating that commonly employed physical tests may not be helpful.<sup>3</sup> The sacroiliac joint (SIJ) has long been implicated as a source of LBP.<sup>4,5</sup> Bogduk *et al.*<sup>6</sup> used controlled, diagnostic, anaesthetising SIJ blocks to 43 patients with chronic low back pain. 30% of patients obtained gratifying relief of their pain, coincidentally nine of these also exhibited tears of their ventral capsule with radiography, with the authors concluding that this possible causative factor requires further study.

Lewit and Liebenson<sup>7</sup> concluded that in order to accurately assess the SIJ, a multitest regimen must be applied. Given the fact that there are at least twelve tests commonly used for the SIJ, it is logical that we should only use those tests that have been shown individually to have a degree of reliability. A number of studies have investigated the reliability of specific tests for the SIJ. To date, many studies indicate that reliability for motion palpation procedures of the SIJ, without pain provocation, is poor.<sup>8-11</sup>

Potter & Rothstein<sup>11</sup> examined the intertester reliability of 13 tests for sacroiliac joint (SIJ) dysfunction, of which the seated flexion test was one. Eight therapists examined 17 patients in two clinical settings but only two examiners tested each subject per test. Reliability was poor; 11 of the 13 tests resulted in less than 70% agreement. The fact that only two examiners were used per test, and no attempt was made to correct for

“agreement by chance”, this indicates that further research is required into the reliability of these tests.

Reliability of an examination procedure is the accuracy, consistency, stability and reproducibility of the technique.<sup>12</sup> In order to accurately diagnose and therefore prescribe a certain treatment regimen, it is paramount that reliable forms of assessment are implemented. Validity, on the other hand, refers to the ability of a procedure to test the structure(s) that it claims to do so. Tests must display both reliability and validity before they can be considered effective examination tools.<sup>12</sup>

Motion palpation has long been used in manual medicine as an integral part of the clinical examination.<sup>7</sup> It is believed that motion palpation allows the therapist to detect alterations in segmental motion and thus provides important information as to the focus of manual intervention and the choice of treatment modality.<sup>13</sup> However motion palpation has yet to be proven as a reliable method of determining spinal joint dysfunction.<sup>8</sup> Liebenson & Lewit state that instead of abandoning palpation entirely, we must investigate the possible reasons for this lack of reliability, including that individual tests may have not been properly investigated.<sup>8</sup> This is a common theme in research papers that have examined the reliability of palpation testing. Gibbons & Tehan<sup>14</sup>, Boline *et al*<sup>15</sup> and Moir *et al*.<sup>16</sup> have all stated that further research must be undertaken to substantiate motion palpation as an accurate form of diagnosis.

In order to accurately assess the reliability of a test it is accepted that the examiners must be proficient in conducting the test. In a recent study, Gerwin *et al*<sup>17</sup> demonstrated the importance of examiner training and the standardisation of

procedures. Initially, a group of examiners were examined for agreement in detecting myofascial trigger points, and were found to have poor reliability. The examiners then undertook a training session to discuss discrepancies among their findings and what constituted a positive test. This produced a notable increase in inter-examiner reliability for palpation of myofascial trigger points.

The SIJ has been established as a cause of LBP, but the nature of the lesion is contentious. Osteopaths and other manual therapists claim to detect functional disorders of the joint using clinical methods such as motion testing, but few of these have been investigated for reliability, and of those that have been examined, few have been determined to be reliable.

Despite the perceived simplicity of tests such as the seated flexion test, examination of the SIJ continues to be one of the most contentious areas for manual therapists.<sup>18</sup> The SIJ is a synovial joint and as such may be susceptible to the same biomechanical dysfunction that affects other synovial joints.<sup>19</sup> Its deep location, limited movement and irregular anatomy add to the difficulty in assessing the joint.

Recently Peace & Fryer<sup>20</sup> surveyed the Australian osteopathic profession to determine what clinical tests osteopaths use to detect SIJ dysfunction. 44% of respondents routinely used the seated flexion test on patients with potential SIJ dysfunction, thus indicating that the test is commonly utilised in osteopathic clinical practice.

The seated flexion test involves the seated patient bending forward while the practitioner palpates the posterior superior iliac spines (PSIS).<sup>21</sup> A normal

(asymptomatic) SIJ will allow for the sacral aspect of the joint to flex forward while the ilial sides of the joint (the palpated PSIS) remain relatively symmetrical. Any dysfunction in the SIJ will cause the flexing sacrum to drag one side of the ilial aspect forward, which will be detected as greater cephalid excursion of the PSIS on the involved side.

The standing flexion test is commonly used in conjunction with the seated flexion test. In a recent study, Vincent-Smith and Gibbons<sup>10</sup> attempted to demonstrate inter-examiner and intra-examiner reliability of the standing flexion test. The authors used nine senior osteopathic students as examiners and nine subjects were recruited as patients. Statistical methodology employed was percentage agreement and kappa coefficient. A realistic attempt to blind examiners was made by disguising individual subjects. However, only a very small reference was made regarding the procedures followed to standardise the testing protocol. As a result, it is unclear as to the contents of this training session and the subsequent competencies of the examiners. Intra-examiner reliability was “moderate” ( $k=0.46$ ), and inter-examiner reliability was only slight” ( $k=0.05$ ). Despite these poor results, the design of this research is in concordance with the guidelines suggested by Haas,<sup>12</sup> and therefore represents a sufficient method for determining the reliability of the standing flexion test.

The seated flexion test is a test commonly advocated by authors of osteopathic texts, but the reliability or validity of this procedure has not been investigated. This study aimed to investigate the intra- and inter-examiner reliability of the seated flexion test, and examine the influence of a standardising training session on this reliability.

## **Method**

### Examiners

Ten senior post-graduate osteopathic students (mean age 23.9, range 22-27) comprised the examiner group of the study. All examiners had 3-4 years experience using the seated flexion test. Five of the examiners underwent two one-hour training sessions to discuss and enhance the testing procedure. At each training session examiners conducted the seated flexion test on five volunteers. At the completion of each test examiners discussed their findings so they were clear what constituted a positive test. This was undertaken to ensure a standardised protocol was implemented.

### Subjects

Ten asymptomatic subjects were recruited from Victoria University. These participants were aged 21-40 and were approximately the same height and weight, to reduce the possibility of recognition.

Participants were screened for the study to rule out any visible identifiable characteristics (eg birthmarks, body piercings skin lesions etc). Subjects who were not able to bend forward, whilst seated, were excluded from the study, as were potential participants who were unable to sit for the required one hour. No subjects were excluded as a result of the above criteria.

## Procedures

Only participants who gave informed written consent were included and the procedures of this study were approved by the Victoria University Human Research Ethics Committee.

Data collection took place in one 45 minute session in a practical laboratory at Victoria University. Subjects were assigned a bench on the periphery of the room where they remained seated for the duration of the session. The ten examiners were instructed to simultaneously complete the seated flexion. This test involves the seated patient bending forward while the examiner palpates the PSIS's. At the completion of each test the examiners moved around the room in a clockwise direction. Three circulations of the room were completed thus creating data for intra-examiner reliability, as well as inter-examiner reliability. Each individual result was immediately recorded on separate data sheets. Three results were possible: positive left (+L) indicating the left PSIS was palpated greater than 5mm above the level of the right PSIS; positive right (+R) indicating the right PSIS was palpated greater than 5mm above the level of the left PSIS; or negative (-) indicating that PSIS's on both sides were level (up to 5mm).

**Insert Figure 1**



## Statistical Analysis

Analysis of reliability was determined by percentage agreement and Cohen's Kappa score, as recommended by Haas.<sup>12</sup> Cohen's Kappa score is utilised because it takes into account the possibility of agreement by chance. It provides a numerical value between -1 and +1. Although there is no formal protocol for the interpretation of the reliability coefficient, we have employed the guidelines suggested by Landis and Koch.<sup>22</sup> They have proposed the following interpretation for the Cohen k statistic, as displayed in Table 1.

### **Insert Table 1**

Kappa is calculated using proportion of observed agreement ( $P_o$ ), and proportion of expected agreement ( $P_e$ ) (refer to Tables 3 & 4). In doing so the k coefficient effectively discounts the proportion of agreements which is expected by chance:  $K = (P_o - P_e) / (1 - P_e)$ . These calculations were made using Windows Microsoft Excel.

## Results

Based on the guidelines suggested by Landis and Koch,<sup>22</sup> intra-examiner reliability for the total group was only fair ( $k=0.21$ ), and inter-examiner agreement was only slight ( $k=0.11$ ). Intra-examiner reliability for the trained group was moderate ( $k=0.41$ ), compared with slight agreement ( $k=0.02$ ) for the untrained group. Inter-examiner reliability for both the trained group ( $k=0.14$ ) and untrained group ( $k=0.07$ ) were only slight.

The findings of all examiners for each test are presented in Table 2.

Inter- and intra-examiner agreement have been calculated and presented in Tables 3 and 4.

**Insert Table 2**

**Insert Table 3**

**Insert Table 4**

## Discussion

This study examined the reliability of the seated flexion test, commonly advocated by osteopathic authors for the detection of SIJ dysfunction, and used widely in the profession.<sup>20</sup> Intra-examiner reliability was higher than inter-examiner reliability, and those in the trained group had higher intra- and inter-examiner reliability than the group which did not undertake any training to achieve standardisation. The training sessions appeared to have increased the reliability of the test, most notably intra-examiner reliability. These results are consistent with previous research that has demonstrated that intra-examiner reliability is stronger than inter-examiner reliability,<sup>10</sup> and that standardisation of testing protocols increases reliability.<sup>17</sup>

Bogduk<sup>6</sup> has suggested that poor reliability of physical tests in manual medicine may stem from insufficient education of practitioners. The fact that there was a discrepancy between trained and untrained in the present study is encouraging however K scores for both categories did not demonstrate adequate reliability.

There are several limitations of this study. All subjects used in the study were asymptomatic. Prior knowledge of this by examiners may have lead to observer bias, because seventy-five percent of the test findings were negative. As a result of this, high agreement existed. Kappa scores, however, were quite low, particularly for intra-examiner reliability. For example, examiner 7 had an observed agreement of Po=93%, yet kappa was -0.03. This is explained by the fact that most agreement fell into the negative category, and kappa becomes unstable when there is limited variability.<sup>12</sup> Future studies may wish to include symptomatic subjects or blind

examiners as to the physical status of subjects to reduce over-representation in one category. Haas<sup>12</sup> has recommended that studies use a representative sample of subjects seen in practice, because symptomatic subjects should provide a more even distribution of negative-positive findings, and maintain the stability of the kappa statistic. However, many osteopathic authors claim that SIJ dysfunction commonly exists in asymptomatic individuals (creating altered mobility and predisposing, but not directly causing strain and subsequent pain), and so it was reasonable to examine this sample group.<sup>10</sup>

The examiner group comprised senior osteopathic students. As a result we can assume that this particular group may be less likely to produce more reliable test findings than their experienced counter-parts. The average results obtained in this study may well be a reflection on the competency level of the examiner sample group. Future research into this test and other forms of motion palpation may wish to utilise examiners with a greater level competency. This will ensure that the variable of examiner competency is of a high standard, therefore offering greater likelihood that reliability can be tested more accurately. In addition, this study didn't take into account the possibility that there may have been a discrepancy in competency of members of either the trained or untrained groups. As a result of this, the results investigating the effect of examiner training must be viewed with caution. A better research design in future should initially test a group of examiners for reliability, and then retest the same examiners after undergoing the training sessions.

Based on the low reliability of individual tests for the SIJ, Cibulka<sup>18</sup> claimed that a multi-test regimen, the so-called multi-test score, is required to assess the SIJ. In

other words, no one definitive test may be able to diagnose the SIJ. Haas<sup>12</sup> has suggested the utilisation of a multi-test score will allow for two observers to be in total disagreement about the exact number of positive tests, yet be in complete agreement with the final conclusion. However it is essential that only those tests that have shown a relative degree of reliability should be used in such a protocol.<sup>7</sup>

Therefore more research is required to test the reliability of individual tests. Once all individual tests are scientifically tested it is at this point manual therapists can choose which tests are used in a standardised examination procedure for the SIJ. Future research may be directed at testing the reliability of different combinations of multi-test regimens, utilising individual tests that have shown the highest levels of reliability. However, at this point, the seated flexion test is typical of all other motion palpation tests for the SIJ<sup>9-11</sup> in that it has yet to show a satisfactory degree of reliability for it to be considered in a multi-test regimen and possibly for its continued use in clinical practice.

## **Conclusion**

The seated flexion test, when performed by senior osteopathic students and with the addition of examiner training sessions produced only 'fair' intra-examiner reliability and 'slight' inter-examiner reliability. This suggests that the reliability of this procedure is questionable and makes this test unsuitable for clinical use.

## References

1. Li L, Maetzel A. The economic burden of low back pain: a review of studies published between 1996 and 2001. *Best Practice and Research Clinical Rheumatology*. 2002;16(1):23-30.
2. Grant WD, Muller R, Walker FW. Low Back Pain in Australian Adults: Prevalence and Associated Disability. *Journal of Manipulative and Physiological Therapeutics*, 2004; 27(4):238-244.
3. E M Gass In: K Refshauge and E Gass. Editors. Musculoskeletal and sports injuries: Musculoskeletal Physiotherapy. 1st Edn. Butterworth & Heinemann, Oxford (1995):206-217.
4. Norman GF and May A. Sacroiliac conditions simulating intervertebral disc syndrome. *Western Journal of Surgery, Obstetrics and Gynaecology*. 1956;64:461-462.
5. Shaw JL. The role of the sacroiliac joint as a cause of low back pain and dysfunction. *First Interdisciplinary World Congress on Low Back Pain and its Relation to the Sacroiliac Joint*, ECO, Rotterdam. 1992.
6. Bogduk N, Schwarzer AC & Aprill CN. The sacroiliac joint in chronic low back pain. *Spine*. 1995;20(1):31-37.

7. Lewit K, Liebenson C. Palpation's reliability: a question of science vs. art? *Journal of Bodywork and Movement Therapies*. 2003;7(1):46-48.
8. Deursen van LLJM, Patijn J, Ockhuysen AL et al. The value of some clinical tests of the sacroiliac joint. *Journal Manual Medicine*. 1990;5:96-99.
9. Carmichael JP. Inter- and intra-examiner reliability of palpation for sacroiliac joint dysfunction. *Journal of Manipulative and Physiological Therapeutics*. 1987;10(4):164-171
10. Vincent-Smith B, Gibbons P. Inter-examiner and intra-examiner reliability of the standing flexion test. *Manual Therapy*. 1999; 4:87-93.
11. Potter NA, Rothstein JM. Intertester reliability for selected clinical tests of the sacroiliac joint. *Physical Therapy*. 1985;11:1671-1675.
12. M Haas, Statistical methodology for reliability. *Journal of Manipulative and Physiological Therapeutics*. 1991;14:119-132.
13. Kirkaldy-Willis WH, Hill RJ. A more precise diagnosis for low-back pain. *Spine*. 1979;4(2):102-9.
14. Gibbons P, Tehan P. Manipulation of the Spine, Thorax and Pelvis: An Osteopathic Perspective. Churchill Livingstone, Melbourne; 2000: 43-48.

15. Boline PD, Keating JC, Brist J, Denver G. Interexaminer reliability of palpatory evaluation of the lumbar spine. *American Journal of Chiropractic Medicine*. 1988;1: 5-11.
16. Mior SA, McGregor M, Schut B. The role of experience in clinical accuracy. *J Manipulative and Physiological Therapies*. 1990;13(2):68-71.
17. Gerwin DR, Shannon S, Hong C-Z, Hubbard D, Gevirtz R. Interrater reliability in myofascial trigger point examination. *Pain*. 1997;69:65-73.
18. Cibulka MT. Understanding sacroiliac joint movement as a guide to the management of a patient with unilateral low back pain. *Manual Therapy* 2002;7(4):215-221.
19. Bernard TN, Cassidy JD. The sacroiliac joint syndrome. In Frymoyer JW (ed): *The Adult Spine: Principles and Practice*. New York, Raven Press, 1991:2111-2112.
20. Peace S, Fryer G. Methods used by members of the Australian osteopathic profession to assess the sacroiliac joint. *Journal of Osteopathic Medicine*, 2004; 7(1) 25-32.
21. Greenman PE. *Principles of Manual Medicine*, Williams and Wilkins, Baltimore; 1989:230-231.



22. Landis JR, Koch G. The measurement of observer agreement for categorical data.

*Biometrics*. 1977;33:159-174.

(c) 2005  
Victoria University

## Appendix

**Figure 1.** Demonstration of the seated flexion test



**Table 1.** Interpretation of Kappa (k) scores

<b>K&lt;0</b>	Less than chance
<b>K=0</b>	Chance agreement
<b>K=0.0-0.20</b>	Slight
<b>K=0.21-0.40</b>	Fair
<b>K=0.41-0.60</b>	Moderate
<b>K=0.61-0.80</b>	Substantial
<b>K=0.81-1.0</b>	Almost perfect

**Table 2.** Findings for all examiners on all subjects for the seated flexion test

Examiners	Subjects									
	A	B	C	D	E	F	G	H	I	J
1	LLL	NNL	NNN	NNN	NNL	RLR	NNN	NNN	NNN	NNN
2	NLR	NNN	NNN	NNN	RNN	NRN	NRN	NNN	RRR	NNN
3	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNL	NNN	NNN
4	NLL	NNN	NNN	NNN	LNR	LRN	LLN	LNN	NNN	NNN
5	NNN	NNN	NNN	NNN	LRR	NNN	NNN	NNN	NNN	NNN
6	RNR	RNN	NNR	RNN	NNN	NNR	RNL	NNN	RRR	NNN
7	NNN	NNN	NNN	NNN	NNN	LNN	NNN	NNN	NNN	NNN
8	NNN	NNN	NNN	RNN	NNN	NNN	NNN	NNN	NNN	NNN

9	RLL	NNL	NRL	RNN	RRR	NNN	NNN	NRN	RNN	NNN
10	LRN	RNN	RNN	NNN	NRR	NLR	RNN	LRR	RNL	RNN

L = positive left, R = positive right, N = negative

**Table 3.** Intra-examiner reliability of the seated flexion test

Examiner no.	Po	Pe	K
1	80	58	0.52
2	70	59	0.26
3	93	76	0.72
4	60	55	0.11
5	87	76	0.44
6	57	51	0.11
7	93	94	-0.03
8	93	94	-0.03
9	57	46	0.2
10	30	40	0.16
Mean	72	64.9	0.213
Mean Trained	78	64.8	0.41
Mean Untrained	66	65	0.016

Examiners 1-5 trained; Examiners 6-10 untrained; Po=Percentage of observed agreement; Pe=percentage chance agreement; K=kappa coefficient (agreement beyond chance)

**Table 4.** Inter-examiner reliability of the seated flexion test

Examiner	Po	Pe	K
Trained	71	66	0.14
Untrained	55	52	0.07
Mean	63	59	0.105

Po=Percentage of observed agreement; Pe=percentage of chance agreement; K=kappa coefficient (agreement beyond chance)