**South African Journal of Physiotherapy**
ISSN: (Online) 2410-8219, (Print) 0379-6175



# Intra-rater and inter-rater reliability of six musculoskeletal preparticipatory screening tests

**Authors:**
Nosipho Zumana[1]
Benita Olivier[1]
Lonwabo Godlwana[1]
Candice Martin[1]

**Affiliations:**
[1]Department of Physiotherapy, University of the Witwatersrand, Johannesburg, South Africa

**Corresponding author:**
Benita Olivier,
benita.olivier@wits.ac.za

**Dates:**
Received: 13 June 2018
Accepted: 04 Feb. 2019
Published: 24 Apr. 2019

**How to cite this article:**
Zumana, N., Olivier, B., Godlwana, L. & Martin, C., 2019, 'Intra-rater and inter-rater reliability of six musculoskeletal preparticipatory screening tests', *South African Journal of Physiotherapy* 75(1), a469. https://doi.org/10.4102/sajp.v75i1.469

**Copyright:**
© 2019. The Authors. Licensee: AOSIS. This work is licensed under the Creative Commons Attribution License.

**Read online:**
Scan this QR code with your smart phone or mobile device to read online.

**Background:** High injury prevalence rates call for effective sports injury prevention strategies, which include the development and application of practical and reliable pre-participatory screening tools.

**Objectives:** The aim of this study was to investigate the intra-rater and inter-rater reliability of the one-legged hyperextension test (1LHET), the empty can (EC) and full can (FC) tests, the standing stork test (SST), the bridge-hold test (BHT) and the 747 balance test (747BT).

**Method:** Thirty-five healthy, injury-free male athletes (cricket and soccer players), aged 16–24 years, were evaluated by two physiotherapists. For each of the tests, the participants were evaluated twice (on two consecutive days) by each physiotherapist. Both the intra- and inter-rater reliability were determined. Cohen's kappa ($k$) was calculated for the 1LHET, the EC and FC tests and the SST. The intraclass correlation coefficient (ICC) was used for the BHT and the 747BT. A confidence level of 95% ($p \leq 0.05$) was applied as the criterion for determining the statistical significance of the results.

**Results:** The SST presented with the lowest level of intra-rater agreement (ICC = –0.20 to 0.10). On the other hand, the EC test was the only test where one rater achieved an excellent intersessional agreement ($k$ = 0.80; 95% confidence interval [CI] 0.40–1.20). Substantial to excellent results for the inter-rater agreement for both sessions were recorded for the 1LHET ($k$ = 0.70–0.90) and the BHT (ICC = 0.70–0.90).

**Conclusion:** Reliability values need to be considered when making clinical decisions based on screening tests. A more refined description of the testing procedures and criteria for interpretation might be necessary before including the six screening tests investigated in this study in formal screening protocols.

**Clinical implication:** Confirmed reliability of screening tests would enable sports professionals to make informed decisions when designing preparticipatory musculoskeletal screening tools and when dealing with the management of injury risks in athletes.

**Keywords**: musculoskeletal screening; injury risk management; intra-rater reliability; inter-rater reliability; soccer; cricket.

## Introduction

In South Africa, soccer and cricket remain popular sports. Injury prevalence studies highlight that musculoskeletal injuries are inevitably a component in the career of the professional soccer (Naidoo 2007) and cricket (Stretch 2001:336) player. Naidoo (2007) reported that over a competitive season, the majority (57%) of soccer players in a professional South African team were found to have sustained injuries. Lower limb injuries were most prevalent among defenders and midfielders, while goalkeepers and forwards were more prone to injuries of the trunk (Naidoo 2007). Cricket injury prevalence rates pose an equal challenge.

Stretch (2001) conducted a 3-year longitudinal study and concluded that cricketers tend to be more prone to lower limb injuries (49.50%), followed by injuries to the upper limbs (23.30%), back and trunk (22.80%). Bowling accounts for more injuries (41.3%) than fielding, including wicketkeeping (28.6%) and batting (17.1%).

These high injury rates call for effective sports injury prevention strategies, which include the development and application of preparticipatory screening tools (Madsen, Drezner & Salerno 2014:142). The ultimate goals of musculoskeletal screening are to identify the modifiable and non-modifiable risks to injury, to facilitate optimal musculoskeletal health and to optimise performance (Cook, Burton & Hoogenboom 2006:62; Ekstrom, Donatelli & Carp 2007:754; Lehr et al. 2013:225).

Tests included in preparticipation screening tools should be practical and reliable. These tests should enable health professionals, including physiotherapists, to determine the athlete's musculoskeletal condition and risk of injury. A screening test is considered to be reliable if there is an error-free consistency, whereby the test measurements can be reproduced by two different raters (inter-rater reliability) and repeatedly by the same rater (intra-rater reliability) (Portney & Watkins 2000:768). Agreement between ratings ensures that results are comparable and that accurate conclusions can therefore be drawn from the results.

The tests included in this study, namely the one-legged hyperextension test (1LHET), the empty can (EC) and full can (FC) tests, the standing stork test (SST), the bridge-hold test (BHT) and the 747 balance test (747BT), attempt to identify intrinsic, person-related risk factors. These tests have been included in the screening protocols of the regulatory bodies of different professional sporting teams, including those of the South African National Cricket (Gray 2015) and Rugby teams (Gray & Naylor 2009), as well as that of the International Football Federation's Medical and Research Centre (Dvorak & Junge 2009). To demonstrate the need for an investigation into the reliability, a brief overview of the literature on each of these tests will follow.

## One-legged hyperextension test

Sporting activities that require repetitive lumbar extension and rotation such as cricket pace bowling predispose athletes to lumbar spondylosis (Masci et al. 2006:940; Wiesel 2018). Moderate sensitivity (50% – 75%) and low specificity (12% – 32%) have been reported in the 1LHET and serve as a means to diagnose spondylolysis (Gregg, Dean & Schneiders 2009:121; Masci et al. 2006). Although results from these validity studies present reasons for conducting further investigations, this test is still included in the preparticipatory screening and diagnostic procedures in sports such as cricket (Gray 2015).

It is important to note, however, that only limited research has been conducted in terms of the reliability of the 1LHET.

## Empty can and full can tests

The subacromial space accommodates, among others, the tendon of the supraspinatus muscle, which is responsible for glenohumeral joint compression, abduction and, to a lesser degree, external rotation.

Supraspinatus activity increases with resisted scapular plane motions (Hughes & Na 1996:75). The EC test (Beaudreuil et al. 2009:15) and the FC test (Kelly, Kadrmas & Speer 1996:581) were designed to identify a supraspinatus tendon pathology that might lead to the encroachment of the subacromial space during activation. Humeral internal rotation, a component of the EC test (Cools, Cambier & Witvrouw 2008:628), blocks greater tuberosity movement, preventing the humerus from giving way under the acromion during its elevation, thus leading to further subacromial space encroachment (Hughes & Na 1996:75). For this reason, the FC test might be favoured above the EC test (Hughes & Na 1996:75). Results from several studies propose that the FC and EC tests demonstrate acceptable diagnostic accuracy, that is sensitivity, specificity and likelihood ratios, for full or partial thickness in supraspinatus tendon ruptures (Itoi et al. 1999:65; Kim et al. 2006:223; Lasbleiz et al. 2014:228; Somerville et al. 2014:1911).

Liu et al. (2016:147) reported sensitivity levels of 84.30% and 78.90% and specificity levels of 74.50% and 80.90% for the EC and FC tests, respectively (Liu et al. 2016:147). Michener et al. (2009:1898) investigated the inter-rater reliability of the EC test and reported a kappa value of 0.45 to 0.67. However, unlike in our study, the inter-rater reliability test was based only on evidence of weakness and disregarded pain as a component (Kelly et al. 1996). No literature specifically reporting on the inter- and intra-rater reliability of the FC and EC tests among physiotherapists, who are often responsible for the preseason screening of players in a team setting, could be found.

## Standing stork test

The optimal function of the lumbo–pelvic–hip complex allows for the effective generation and transfer of forces during athletic activity (Kibler, Press & Sciascia 2006:189). The SST assesses the ability of the pelvis to remain stable as load is transferred between the spine and the limbs (Hungerford et al. 2007:879). Hungerford et al. (2007) investigated the ability of physiotherapists to evaluate intrapelvic movement using the SST and found good inter-rater reliability ($k = 0.67$). Conversely, Tong et al. (2006:464) found poor inter-rater reliability for the SST. However, the sample size was small ($n = 24$) and consisted only of females with lower back pain, which limits the generalisation of findings to other populations.

## Bridge-hold test

The BHT assesses gluteal strength and endurance, as well as the static stability of the trunk and pelvis (Dennis et al. 2008:25). The stability of the core allows for improved balance and for the motion of the trunk over the pelvis (Andrade et al. 2012:268). Andrade et al. (2012) investigated the intra- and inter-rater reliability of the BHT using a two-dimensional motion analysis and reported kappa values of 0.32–0.58 and 0.80, respectively. In the light of the costs and logistics related to two-dimensional motion analysis, there is a need to determine the reliability of the BHT without the application of movement analysis software, which is also often the case in clinical practice.

## 747 Balance test

The 747BT (also known as the 'Romanian deadlift') assesses general balance, coordination and stability in a single-leg body position (Strauts & Tate 2015:43) and is, therefore,

considered to be applicable to sporting activities that require a combination of strength, flexibility and speed (Gamble 2013). It is important to note, however, that limited research related to the validity and reliability of the 747BT is currently available.

From the literature, it is clear that research related to the reliability of these six screening tests is limited. The intra- and inter-rater reliability of the aforementioned six tests were therefore investigated in order to provide guidance as to the inclusion of these tests in the official musculoskeletal screening protocols of professional sporting teams.

# Materials and methods

This reliability study was conducted at the sports fields of the cricket and soccer clubs of a tertiary institution.

Thirty-five healthy, injury-free male players aged between 16 and 24 years from the university's respective soccer and cricket clubs were randomly selected for the study. Players with a history of spinal or lower limb surgery were excluded. The sample size was based on the findings and suggestions by Sim and Wright (2005:257). Effect sizes (ES) were calculated using Cohen's *d*-test, where ES values of 0.20, 0.50 and 0.80 were respectively interpreted as small, medium and large. An a priori power analysis, using G-power relating to the medium ES category (ES = 0.5) was used in the calculation to determine sample size. A power analysis for estimating the size of the sample that would yield a power of 80% was conducted prior to the data collection phase.

## Procedures

Three participants (±10% of the main sample size), other than those included in the main study, were included in the pilot study, which used the same inclusion and exclusion criteria specified for the main study.

The pilot study aimed to familiarise the raters with the testing procedures, to ensure that the testing instructions and procedures were standardised and to establish the time required for the completion of each test. The data collected from the pilot study were not included for the analysis of the main study results as changes to the standardised testing instructions and conditions (i.e. time of day: before, during or after training) had been made to the study procedure subsequent to the pilot study.

The main study was conducted over 2 weeks. To minimise the effect of physiological and biomechanical changes and to allow the symptoms that might have been provoked by the tests to subside, the first and second testing sessions for the individual participants occurred on two consecutive days. The second session for a specific participant occurred under the same conditions (i.e. before, during or after training) as those for the first. The screening tests were conducted by two qualified physiotherapists (Rater 1 and Rater 2), each with more than 5 years of clinical experience. Video recordings were made of each test for digital storage purposes and were in turn managed by a research assistant.

The screening tests were conducted according to a standard set of instructions and procedures (Figure 1) and performed in the following order: 1LHET, BHT, 747BT, ECTFCT, SST, without any period of rest between tests. Each rater assessed each participant. The FC and EC tests and the 1LHET and SST required a 'hands-on' assessment by the respective raters and were conducted and rated separately by each of them. Being observational tests, the BHT and the 747BT were rated simultaneously by the raters. During the simultaneous ratings, no communication was allowed between the raters, who were blinded to each other's findings.

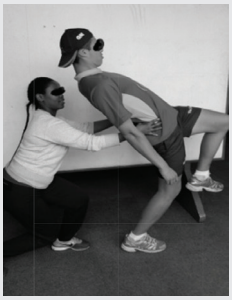
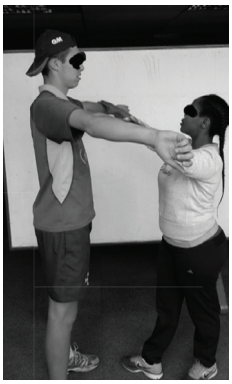
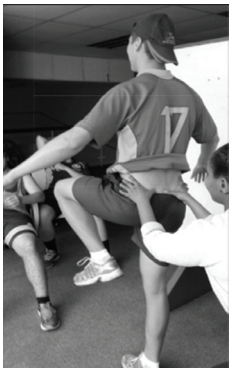
## Data analysis

Data were recorded on specifically designed data collection sheets and later captured by the first author on an Excel spreadsheet. Statistical analyses were accomplished using SPSS Version 23 (IBM Corporation, Armonk, NY, USA). Descriptive analysis was used to describe the basic features of the data.

Agreement in the test results by two different raters (inter-rater reliability) and repeatedly by the same rater (intra-rater reliability) was determined. The inter-rater reliability was determined by comparing the per-session ratings of Rater 1 as opposed to those of Rater 2. Between-day intra-rater reliability was tested by comparing the ratings of a rater for Session 1 with those of the same rater for Session 2. To determine both the inter- and intra-rater reliability, Cohen's kappa (*k*) was used for the 1LHET, EC and FC tests, and the SST because the outcomes (yes or no) for these tests were nominal (Cohen 1960; Sim & Wright 2005). The intraclass correlation coefficient ($ICC_{3,2}$) was used for the BHT and the 747BT, the data for which were continuous. The ICC was measured through a two-way random effect for inter-rater reliability, and a mixed random effect for intra-rater reliability was used because each participant from this random sample was assessed more than once (Shrout & Fleiss 1979:420). A confidence interval of 95% ($p < 0.05$) was used to determine the statistical significance of the data. The *k* and ICC values were interpreted according to the guidelines as set out by Landis and Koch (1977:159) (Table 1).

## Ethical considerations

Ethical clearance (reference number: M150626) was obtained from the University of the Witwatersrand's Human Research Ethics Committee (Medical). Each participant received an information leaflet presenting the goals and procedures of the study and was requested to voluntarily provide consent to participate in the study and to permit a video recording of their performance in the respective tests.

| Screening test | Procedure and instructions | Instructions to participants |
|---|---|---|
| One-legged hyperextension test  | Facing away from the tester, the athlete is asked to stand on the left leg and raise the right leg with the right hip and knee slightly flexed to ±80°. The athlete is then asked to perform lumbar extension. The rater asks the athlete if any pain is felt (yes or no). The test is then repeated while standing on the right leg. The test is considered positive if lumbar pain is produced. | 'Stand on the left leg and raise the right leg with the hip and knee slightly bent. Then bend your back backwards and tell me if you feel any pain.' |
| Empty can test  | Facing the examiner, the participant elevates both shoulders to 90° in the scapular plane and internally rotates the shoulders so that both thumbs point to the floor. Further elevation is resisted by the examiner applying pressure to the mid-forearm. The test is considered positive if the athlete experiences pain. For the purposes of this study only the dominant arm was considered. | 'Stand with both arms stretched out to 90° in the scapular plane and both thumbs facing down; maintain this position and do not let me push you down. Tell me if you feel any pain in your dominant shoulder while maintaining this position.' |
| Full can test | Facing the examiner, the participant elevates both shoulders to 90° in the scapular plane and externally rotates both shoulders so that both thumbs point upward. The test is considered positive if the athlete experiences pain in the shoulder. For the purposes of this study only the dominant arm was considered. | 'Stand with both arms stretched out to 90° in the scapular plane and thumbs both facing up; maintain this position and do not let me push you down. Tell me if you feel any pain in your dominant shoulder while maintaining this position.' |
| Standing stork test  | The participant stands facing away from the examiner. The examiner places his or her left thumb on the PSIS while the right thumb palpates the midline of the sacrum at the same level as the PSIS. The participant is asked to flex the left hip to 90°. Sacro-iliac joint movement is considered normal when the thumb on the PSIS moves upward with flexion of the hip and abnormal when the thumb does not move or moves downward. The test is repeated on the right. | 'Stand on one leg. Raise the other so that your hip is flexed to 90°.' |
| Bridge-hold test | The participant is positioned supine on the floor with knees flexed to 90°. The hips and pelvis are then raised so that the shoulder, hip and knee joints are aligned. With the ASIS aligned and no arching of the back, the participant is asked to extend the right knee and maintain the position for as long as possible. Duration (seconds) is timed from full knee extension until termination of the test as a result of fatigue, pain or cramping of the low back, gluteal or hamstring muscles or because of the athlete's inability to maintain the ASIS level or shoulder, hip and knee aligned. The test is repeated on the opposite side. Time (s) and reason for termination of the test is recorded. | 'Lie on your back on the floor with both knees bent about 90°, then lift the buttocks from the floor until your shoulders, hips and knees are in a straight line. Straighten your right knee so that your right ankle is now also in line with your right shoulder, hip and knee. Also maintain the horizontal level of your pelvis in this position. Maintain this position for as long as you can.' |
| 747 balance test | The participant stands on one leg with arms abducted to the side. He or she then leans forward, extending the opposite hip backward until the shoulder, hip and knee are horizontally aligned ('flying position'). This position is maintained until failure because of pain or cramping of the low back, gluteal or hamstring muscles or the athlete cannot maintain ASIS or shoulder–hip–knee alignment. Duration (s) is recorded from acquisition of flying position until failure. Time (s) and reason for termination is recorded. | 'Stand on one leg with the non-weight-bearing leg straight behind you and the arms stretched out to the sides, as if in the flying position. Lift your leg behind you until your shoulder, hip, knee and ankle are in one horizontal line. Maintain this position for as long as you can.' |

*Source*: Photos courtesy of the authors taken during their study

PSIS, posterior superior iliac spine; ASIS, left and right anterior superior iliac spines.

**FIGURE 1:** Procedures and standard instructions for the one-legged hyperextension test, full can test, empty can test, standing stork test, bridge-hold test and 747 balance test.

## Results

Of the 35 selected participants, four (11%) could not return for the second assessment because of unexpected time conflicts with training and study-related responsibilities. Therefore, data from 31 participants (89%) were eligible for analysis. Table 2 summarises the demographic (age) and anthropometric data of the 31 participants included in the main study.

The intra-rater reliability results are summarised in Table 3. Only Rater 2's assessment of the EC test showed substantial intra-rater reliability ($k$ = 0.80; 95% confidence interval [CI] 0.40–1.20), while the intra-rater reliability levels for the SST for both raters were poor or slight.

**TABLE 1:** Guidelines for interpretation of kappa and intraclass correlation coefficient values.

| Value of $k$†or ICC | Strength of agreement |
|---|---|
| < 0.00 | No agreement |
| 0.00–0.20 | Slight |
| 0.21–0.40 | Fair |
| 0.41–0.60 | Moderate |
| 0.61–0.80 | Substantial |
| 0.81–1.00 | Excellent |

*Source*: Landis and Koch 1977:159

†, $k$ = kappa; ICC, intraclass correlation coefficient.

**TABLE 2:** Demographic and anthropometric data of participants ($n$ = 31).

| Variable | Mean ± SD | Minimum | Maximum |
|---|---|---|---|
| Age (yr) | 18.60 ± 1.50 | 16.00 | 24.00 |
| Weight (kg) | 74.90 ± 9.40 | 60.00 | 100.00 |
| Height (m) | 1.80 ± 0.10 | 1.50 | 1.90 |

SD, standard deviation.

The inter-rater reliability levels for each of the screening tests included are shown in Table 4. Notably, with the exception of the SST, the left BHT and the left 747BT, the inter-rater agreement always tended to be higher during Session 2, and the agreement between the results for this session for the EC test ($k$ = 0.80; 95% CI 0.40–1.20), the FC test ($k$ = 0.80; 95% CI 0.50–1.10) and the right BHT (ICC = 0.80; 95% CI 0.60–0.90) was substantial. Only the 1LHET (bilaterally) revealed substantial to excellent agreement for both sessions. A poor agreement between the raters was noted for the EC test for Session 1 ($k$ = –0.05; 95% CI –0.10 to 0.01) and for the SST (right) for Session 2 ($k$ = -0.06; 95% CI –0.20 to 0.10).

## Discussion

Sporting teams often include preparticipatory screening tools as part of their injury prevention strategies (van Mechelen,

**TABLE 3c:** Intra-rater reliability of the screening tests in this study.

| Rater | Test | Session 1 outcomes ($n$ = 31) | Session 2 outcomes ($n$ = 31) | $ICC_{3,2}$ (95%CI) | SEM |
|---|---|---|---|---|---|
| | | Seconds held (mean ± SD) | Seconds held (mean ± SD) | | |
| | **BHT** | | | | |
| 1 | Left | 26.30 ± 10.80 | 17.30 ± 12.50 | 0.30 (-0.40–0.60) | 15.70 |
| | Right | 21.00 ± 8.70 | 18.10 ± 10.40 | 0.50 (0.20–0.70) | 11.50 |
| 2 | Left | 22.90 ± 12.40 | 16.50 ± 12.50 | 0.40 (0.10– 0.70) | 15.80 |
| | Right | 18.20 ± 8.50 | 16.30 ± 10.10 | 0.40 (0.10–0.70) | 11.90 |
| | **747 BT** | | | | |
| 1 | Left | 17.80 ± 9.10 | 15.80 ± 10.9 | 0.60 (0.30–0.80) | 11.70 |
| | Right | 16.50 ± 9.70 | 15.20 ± 10.80 | 0.30 (-0.10–0.60) | 13.90 |
| 2 | Left | 15.70 ± 10.10 | 15.20 ± 8.40 | 0.40 (0.1–0.7) | 11.90 |
| | Right | 15.30 ± 7.50 | 16.40 ± 9.30 | 0.20 (-0.4–0.7) | 11.60 |

BHT, bridge-hold test; 747BT, 747 balance test; ICC, intraclass correlation coefficient; SEM, standard error measurement; CI, confidence interval.

**TABLE 3a:** Intra-rater reliability of the screening tests in this study.

| Rater | Test | Session 1 outcomes ($n$ = 31) | | Session 2 outcomes ($n$ = 31) | | $k$ (95% CI) | SEM |
|---|---|---|---|---|---|---|---|
| | | Yes $n$ (%) | No $n$ (%) | Yes $n$ (%) | No $n$ (%) | | |
| | **1LHET** | | | | | | |
| 1 | Left | 10.00 (32.30) | 21.00 (67.00) | 7.00 (22.60) | 24.00 (77.40) | 0.60 (0.30–0.90) | 0.20 |
| | Right | 11.00 (35.50) | 20.00 (64.50) | 5.00 (16.10) | 26.00 (83.90) | 0.50 (0.20–0.80) | 0.20 |
| 2 | Left | 13.00 (11.90) | 18.00 (58.10) | 6.00 (19.40) | 25.00 (80.60) | 0.40 (0.10–0.70) | 0.20 |
| | Right | 11.00 (35.50) | 20.00 (64.50) | 4.00 (12.90) | 27.00 (87.10) | 0.40 (0.10–0.70) | 0.20 |
| | **EC** | | | | | | |
| 1 | Dominant arm | 2.00 (6.50) | 29.00 (93.50) | 1.00 (3.20) | 30.00 (96.80) | 0.60 (0.20–1.1) | 0.20 |
| 2 | Dominant arm | 0.00 (0.00) | 31.00 (100.00) | 1.00 (3.20) | 30.00 (96.80) | 0.80 (0.40–1.20) | 0.20 |
| | **FC** | | | | | | |
| 1 | Dominant arm | 1.00 (3.20) | 30.00 (96.80) | 1.00 (3.20) | 30.00 (96.80) | 0.10 (0.10–0.20) | 0.10 |
| 2 | Dominant arm | 1.00 (3.20) | 30.00 (96.80) | 1.00(3.20) | 30.00 (96.80) | 0.60 (0.20–1.10) | 0.20 |

1LHET, one-legged hyperextension test; EC, empty can test; FC, full can test; $k$, kappa; SEM, standard error measurement; CI, confidence interval.

**TABLE 3b:** Intra-rater reliability of the screening tests in this study.

| Rater | Test | Session 1 outcomes ($n$ = 31) | | | Session 2 outcomes ($n$ = 31) | | | $k$ (95% CI) | SEM |
|---|---|---|---|---|---|---|---|---|---|
| | | Motion detected by rater | | | Motion detected by rater | | | | |
| | | Up | Down | None | Up | Down | None | | |
| | **SST** | | | | | | | | |
| 1 | Left | 16.00 (51.60) | 4.00 (12.90) | 11.00 (35.50) | 14.00 (45.20) | 13.00 (41.90) | 3.00 (9.70) | -0.06 (-0.27–0.10) | 0.10 |
| | Right | 17.00 (54.80) | 6.00 (19.40) | 8.00 (25.80) | 11.00 (35.50) | 18.00 (58.10) | 2.00 (6.50) | -0.15 (-0.40–0.40) | 0.10 |
| 2 | Left | 0.00 (0.00) | 16.00 (51.60) | 15.00 (48.40) | 0.00 (0.00) | 22.00 (70.90) | 9.00 (29.00) | 0.10 (-0.20-0.40) | 0.20 |
| | Right | 0.00 (0.00) | 24.00 (77.40) | 7.00 (22.60) | 0.00 (0.00) | 27.00 (87.10) | 4.00 (12.90) | -0.20 (-0.30–0.10) | 0.10 |

$k$, kappa; SEM, standard error measurement; CI, confidence interval.

**TABLE 4a:** Inter-rater reliability of the screening tests in this study.

| Session | Test | Rater 1 outcomes ($n$ = 31) | | Rater 2 outcomes ($n$ = 31) | | $k$ (95% CI) | SEM |
|---|---|---|---|---|---|---|---|
| | | Yes $n$ (%) | No $n$ (%) | Yes $n$ (%) | No $n$ (%) | | |
| | **1LHET** | | | | | | |
| 1 | Left | 10.00 (32.30) | 21.00 (67.70) | 13.00 (41.90) | 18.00 (58.10) | 0.80 (0.60–1.00) | 0.1 |
| | Right | 11.00 (35.50) | 20.00 (64.50) | 11.00 (35.50) | 20.00 (64.50) | 0.80 (0.60–1.00) | 0.1 |
| 2 | Left | 7.00 (22.60) | 24.00 (77.40) | 6.00 (19.40) | 25.00 (80.60) | 0.90 (0.70–1.10) | 0.1 |
| | Right | 5.00 (16.10) | 26.00 (83.90) | 27.00 (87.10) | 0.90 (0.60–1.10) | 0.90 (0.60–1.10) | 0.1 |
| | **EC** | | | | | | |
| 1 | Dominant arm | 2.00 (6.50) | 29.00 (93.50) | 0.00 (0.00) | 31.00 (100.00) | -0.05 (-0.10–0.01) | 0.003 |
| 2 | Dominant arm | 1.00 (3.20) | 30.00 (96.80) | 1.00 (3.20) | 30.00 (96.80) | 0.80 (0.40–1.20) | 0.20 |
| | **FC** | | | | | | |
| 1 | Dominant arm | 1.00 (3.20) | 30.00 (96.80) | 1.00 (3.20) | 30.00 (96.80) | 0.50 (0.03–0.90) | 0.20 |
| 2 | Dominant arm | 1.00 (3.20) | 30.00 (96.80) | 1.00 (3.20) | 30.00 (96.80) | 0.80 (0.50–1.10) | 0.10 |

1LHET, one-legged hyperextension test; EC, empty can test; FC, full can test; *k*, kappa; SEM, standard error measurement; CI, confidence interval.

**TABLE 4b:** Inter-rater reliability of the screening tests in this study.

| Session | Test | Rater 1 outcomes ($n$ = 31) | | | Rater 2 outcomes ($n$ = 31) | | | $ICC_{3,2}$ (95% CI) | SEM |
|---|---|---|---|---|---|---|---|---|---|
| | | Motion detected by rater | | | Motion detected by rater | | | | |
| | | Up | Down | None | Up | Down | None | | |
| | **SST** | | | | | | | | |
| 1 | Left | 16.00 (51.60) | 4.00 (12.90) | 11.00 (35.50) | 0.00 (0.00) | 16.00 (51.60) | 15.00 (48.40) | 0.30 (0.10–0.40) | 0.10 |
| | Right | 17.00 (54.80) | 6.00 (19.40) | 8.00 (25.80) | 0.00 (0.00) | 24.00 (77.40) | 7.00 (22.60) | 0.10 (-1.10–1.30) | 0.60 |
| 2 | Left | 14.00 (45.20) | 13.00 (41.90) | 3.00 (9.70) | 0.00 (0.00) | 22.00 (70.90) | 9.00 (29.00) | 0.30 (0.20–0.40) | 0.10 |
| | Right | 11.00 (35.50) | 18.00 (58.10) | 2.00 (6.50) | 0.00 (0.00) | 27.00 (87.10) | 4.00 (12.90) | -0.06 (-0.20–0.10) | 0.10 |

SST, standing stork test; ICC, intraclass correlation coefficient; SEM, standard error measurement; CI, confidence interval.

**TABLE 4c:** Inter-rater reliability of the screening tests in this study.

| Session | Test | Rater 1 outcomes ($n$ = 31) | Rater 2 outcomes ($n$ = 31) | ICC3,2 (95% CI) | SEM |
|---|---|---|---|---|---|
| | | Seconds held (mean ± SD) | Seconds held (mean ± SD) | | |
| | **BHT** | | | | |
| 1 | Left | 26.30 (±10.80) | 22.90 (±12.40) | 0.80 (0.70–0.90) | 9.10 |
| | Right | 21.00 (±8.70) | 18.20 (±8.50) | 0.70 (0.40–0.80) | 9.20 |
| 2 | Left | 17.30 (±12.50) | 16.50 (±12.50) | 0.90 (0.80–0.90) | 6.90 |
| | Right | 18.10 (±10.40) | 16.30 (±10.10) | 0.80 (0.60–0.90) | 9.40 |
| | **747 BT** | | | | |
| 1 | Left | 17.80 ± 9.10 | 15.70 ± 10.10 | 0.70 (0.50–0.90) | 9.30 |
| | Right | 16.50 ± 9.70 | 15.30 ± 7.50 | 0.20 (0.00–0.70) | 11.30 |
| 2 | Left | 15.80 ± 10.90 | 15.20 ± 8.40 | 0.40 (0.10–0.70) | 11.90 |
| | Right | 15.20 ± 10.80 | 16.40 ± 9.30 | 0.20 (-0.20–0.50) | 11.60 |

BHT, bridge-hold test; 747BT, 747 balance test; ICC, intraclass correlation coefficient; SEM, standard error measurement; CI, confidence interval.

Hlobil & Kemper 1992:82). Reliable, cost- and time-effective screening tools might allow medical and fitness professionals to make informed decisions regarding the management of an athlete's injury risk. The purpose of this study was therefore to investigate the reliability of six screening tests often included in the screening protocols of various sporting disciplines.

Among other factors, body composition and specific physical attributes have been related to elite and sub-elite level cricketers (Koley 2011:427; Stuelcken, Pyne & Sinclair 2007:1587) and soccer players (Hencken & White 2006:205). Considering the mean age and level of participation, the weight and height measurements of the participants were similar to those of the cricketers (21.03 ± 1.72 years; 61.83 ± 9.6 kg; 171.00 ± 7.1 cm) (Koley 2011) and soccer players (66.60–78.00 kg; 171.2–178.1 cm) (Rebelo et al. 2012:312) investigated in other studies. Although body composition and specific physical characteristics have been associated with advanced performance in general athletic and sport-specific skills (Rodriguez, DiMarco & Langley 2009), these specifics do not fall within the scope of this study. A summation of the intra- and inter-rater reliability results of the screening tests investigated in this study are presented in Table 5.

## One-legged hyperextension test

While the 1LHET was the only test presenting with substantial to excellent inter-rater agreement in this study, the intra-rater agreement was moderate (Rater 1) to fair (Rater 2). This was also the only bilateral test (i.e. performed on the left and right sides) in which both raters achieved the same level of intersessional agreement for the left and right sides. This might indicate that the test was performed in a uniformly bilateral manner by each rater during Session 1 and Session 2 but that the level of pain experienced by the participants during the respective sessions differed.

Another explanation could be related to the lack of specification in terms of the lumbar extension range according to which the test was performed. The designers of the 1LHET hypothesised that in the presence of spondylolysis, compressive forces on the pars interarticularis, associated with lumbar extension, would exacerbate the pain (Jackson et al. 1981:304). A specific lumbar spine extension range was not described, however, and was therefore apparently left to the discernment of the examiner. During the execution of the test, a manipulation of the lumbar extension range by the participant from one assessment session to the next, as well as the resultant change in compression of the pars

**TABLE 5:** Summation of strength of intra- and inter-rater agreement for the screening tests included in this study.

| Test | Side | Intra-rater reliability | | Inter-rater reliability | |
|---|---|---|---|---|---|
| | | Rater 1 | Rater 2 | Session 1 | Session 2 |
| 1LHET | Left | Moderate | Fair | Substantial | Excellent |
| | Right | Moderate | Fair | Substantial | Excellent |
| EC | Dominant | Moderate | Substantial | No agreement | Substantial |
| FC | Dominant | Slight | Moderate | Moderate | Substantial |
| SST | Left | No agreement | Slight | Fair | Fair |
| | Right | No agreement | No agreement | Slight | No agreement |
| BHT | Left | Fair | Fair | Substantial | Excellent |
| | Right | Moderate | Fair | Substantial | Substantial |
| 747BT | Left | Moderate | Fair | Substantial | Fair |
| | Right | Fair | Slight | Slight | Slight |

1LHET, one-legged hyperextension test; EC, empty can; FC, full can; SST, standing stork test; BHT, bridge-hold test; 747BT, 747 balance test.

interarticularis, might account for different levels, if any, of pain.

Despite the substantial to excellent inter-rater reliability measured in this study, the less-than-substantial intra-rater reliability and conclusions from studies investigating the validity of the 1LHET (Alqarni et al. 2015:268; Masci et al. 2006:940) place doubt on its usefulness as the first-line pathognomonic test for spondylosis.

## Empty can test

In this study, the intra-rater reliability of the EC test proved to be moderate to substantial, with a small standard error measurement (SEM) (0.20), which indicates a higher level of rater agreement compared to that for the 1LHET, specifically in respect of Rater 2. Limited research related to the intra-rater reliability of the EC test has been conducted. As such, a comparison of the results in this study proved to be difficult. However, other studies investigating the diagnostic accuracy of the EC test have reported moderate ($k$ = 0.4–0.43 [0.13–0.67]) inter-rater reliability (Magee, Sueki & Chepeha 2011; Michener et al. 2009:1898). Our study, however, found no agreement between the ratings of Raters 1 and 2 for Session 1 but substantial agreement between their respective ratings for Session 2. However, the range for the 95% confidence level for both sessions was broad and the inter-rater kappa values should therefore be interpreted with caution. The limited homogeneity of the rater outcomes for a screening test might highlight the defects of the screening tools or suggest that the raters require additional training in the use of the tool (Martin &Altman 1986:307).

In another study investigating the inter-rater reliability of, among others, the EC test, the outcomes of a research nurse (with no formal musculoskeletal training) and a specialist consultant (a rheumatologist with a special interest in shoulders), as well as the outcomes of the same research nurse and specialist rheumatology registrar, reported fair inter-rater agreement ($k$ = 0.38–0.46) (Ostor 2004:1288). These results might indicate that regardless of the expertise of the examiner (expert vs. expert or novice vs. expert), the inter-rater agreement for the EC test was at most moderate. In our study, however, regardless of similar examiner qualifications and experience, the difference in the level of rater agreement between the two sessions was noteworthy (no agreement for Session 1 vs. substantial agreement for Session 2). One might therefore infer that additional training in the execution of the EC test and in the interpretation of the test results might be warranted.

## Full can test

Prior to this study, research investigating the reliability of the FC test had not been documented (Gray 2015), making the comparison of results challenging. However, the validity of the FC test in the diagnosis of supraspinatus pathology has been confirmed by several studies (Itoi et al. 1999:65; Kelly et al. 1996:581). In our study, intra-rater reliability was found to be slight and moderate for Raters 1 and 2, respectively. On the other hand, inter-rater agreement proved to be moderate to substantial. One explanation for the differences in agreement between the respective sessions, as well as between the raters, might be related to differences in the symptoms experienced by the participants. Another might be on account of a variation in the amount of resistance applied by the raters, which in turn elicits varying levels of isometric muscle activity and possible symptoms.

## Standing stork test

No intra-rater agreement was found for Session 1, while Rater 2 found only slight agreement for right-sided sacro-iliac joint (SIJ) dysfunction in Session 2. Inter-rater agreement was at most fair. Reasons for this less-than-optimal reliability level may include the observational and palpatory nature of this test. Compared to pain provocation test results, palpatory SIJ test results show moderate inter-rater agreement ($k$ = –0.60) (Robinson et al. 2007:72). This is not unique to SIJ-related testing as similar difficulties have been reported for Craig's test, which requires the palpation of the greater trochanter for the measurement of femoral anteversion (Choi & Kang 2015:1141).

Like in our study, Hungerford et al. (2007:879) investigated the ability of three physiotherapists to assess SIJ movement using the SST. The authors found that when bone motion (movement of the innominate bone on the sacrum) was recorded on the basis of a two-point scale (occurrence or non-occurrence of bone motion), the agreement between the therapists on intrapelvic motion, which occurs during load transfer, proved to be substantial ($k$ = 0.67–0.77) (Hungerford et al. 2007).

However, the use of a three-point scale that is innominate – remains neutral, moves up or moves down – brought moderate reliability ($k$ = 0.59) for both the left and the right sides to light (Hungerford et al. 2007:879). The difference in rater agreement using a three-point scale, as was the case for both this and the last-mentioned study, might be a result of the number of physiotherapists assessed. This means that the use of more

examiners might result in higher inter-rater reliability levels. Research confirming the association between the level of inter-rater reliability and the number of examiners assessed is yet to be conducted. Tong et al. (2006:464) reported fair inter-rater agreement ($k$ = 0.27) between two physiotherapists with regards to the bone motion of the SIJ during testing.

Considering our results and those of the studies mentioned, the reliability of the SST seems dependent on the outcome measure (a two- or a three-point scale) used. Currently, the lack of uniformity in the SST outcome measures and the low measure of reliability of the SST do not justify the inclusion of this test in formal screening procedures.

## Bridge-hold test

The intra-rater reliability for the BHT was found to be fair to moderate, which is similar to the results obtained by Dennis, Elliott and Farhart (2008:25) and Andrade et al. (2012:268), who reported an intra-rater reliability of ICC = 0.56 (95% CI: 0.00, 0.83) and Kw = 0.32–0.58, respectively. The SEM (11.50–15.80) related to the intra-rater reliability in our study points to a large number of errors that might have occurred during testing. This is not surprising considering the observational nature of the test and the number of reasons for terminating it.

Andrade et al. (2012:268) attempted to minimise the subjective component of observational tests to some extent by using two-dimensional motion analyses requiring participants to maintain the unilateral bridge position for a fixed time (10 s) and limiting the test outcomes to the participants. The intra-rater agreement on the ability of the participants to maintain the horizontal alignment of the anterior superior iliac spine for termination still brought only moderate agreement. Numerous studies investigating the reliability of observational musculoskeletal tests that require the assessment of more than one component have been found to have low intra-rater reliability levels (Monnier et al. 2012:1471; Moreland et al. 1997:200; Whatman et al. 2015:210). The BHT also assesses numerous physical fitness aspects such as motor control, endurance, strength and so on, which could be influenced by several factors including training type and intensity and nutritional intake, in a 24-h window period.

The inter-rater reliability of the BHT in this study was substantial to excellent. Andrade et al. (2012:268) reported substantial reliability (Kw = 0.80), while Dennis et al. (2008:25) (ICC = 0.56) reported only moderate inter-rater reliability. The examiners in the latter study assessed the video-recorded performances of the participants in the BHT in separate cubicles as opposed to collectively and simultaneously in one particular facility. This could possibly be the reason for the difference in the inter-rater reliability between the Dennis et al. (2008:25) study and our findings. Our results might indicate that although there is a strong case for inter-rater reliability, the technicalities behind the BHT might require more refined criteria to be applied in the termination phase of the test.

## 747 Balance test

Moderate or less-than-moderate intra-rater reliability was recorded. The inter-rater reliability of the 747BT varied from slight to substantial. Substantial agreement was related only to Session 1's screening of the left side. To the authors' knowledge, this was the first study to investigate the reliability of the 747BT. Therefore, it was not possible to compare these results with those of other studies.

Noteworthy, however, are the large SEM values associated with the inter- and intra-rater reliability ICC values.

Like the BHT, the 747BT has numerous test termination criteria and challenges numerous physical fitness components, which could explain the lower level of intra-rater reliability and the large SEM values. Moreover, this is an observational test that was done in real time – similar to what happens in clinical practice – without using video footage or two-dimensional motion analysis, which perhaps allow for greater human error and lower agreement in the sessional observations.

Studies assessing the reliability of real-time observational data have reported poor intra- and/or inter-rater reliability in respect of the various musculoskeletal screening tests (DiMattia et al. 2005:108; Nilstad et al. 2014:358; Örtqvist et al. 2011:2060). Because the two raters evaluated the 747BT simultaneously, it should be kept in mind that their visual vantage points were different, as they could not stand in the exact same spot, which could influence their observations of movement.

We used the recommendations for interpretation of reliability results by Landis and Koch (1977:159) (Table 1). These cut-off values are arbitrary, as no absolute descriptions are possible; however a test with a moderate rating (0.41–0.60) is generally not considered accurate, and results from all screening tests should always be interpreted together with other findings that form part of the holistic assessment of the athlete.

More research is needed in terms of the reliability of clinical tests before they are included in formal screening protocols. Considering our findings, as well as those of other referenced authors, clear instructions in terms of testing procedures and positive test criteria might improve the reliability of the tests. Whatman et al. (2015:210) noted the importance of accurate observational skills in the clinicians responsible for the musculoskeletal evaluations because they allow for instantaneous results in terms of an athlete's physical condition and performance.

Future research should therefore focus on investigating the effect of more refined testing procedures on the reliability of the screening tests. The fact that our study involved only physiotherapists might make for its limited practical value because the athletes were not also assessed by other medical and fitness professionals.

## Conclusion

The intra-rater reliability of the EC test proved to be moderate to substantial, while the respective values for all of the other tests showed moderate intra-rater reliability to no agreement. The inter-rater reliability of the 1LHET and the BHT, respectively, proved to be substantial to excellent, whereas the other tests performed less satisfactorily in terms of this criterion. Results from the BHT and the 747BT suggest that in order to be reproduced optimally, observational tests should be based on simplified but clearly defined test termination criteria.

## Acknowledgements

The authors would like to thank the University of the Witwatersrand, Faculty of Health Sciences Research Committee and the South African Society of Physiotherapy for the financial support for this study.

### Competing interests

The authors declare that they have no financial or personal relationships that may have inappropriately influenced them in writing this article.

### Authors' contributions

N.Z. developed the proposal, collected the data and wrote up the article. B.O. conceptualised the study, collected the data and reviewed all drafts. L.G. performed general supervision of project and revision of all drafts. C.M. wrote up the article and revised all drafts.

### Funding

Funding was received from the University of the Witwatersrand, Faculty of Health Sciences Research Committee and the South African Society of Physiotherapy.

## References

Alqarni, A.M., Schneiders, A., Cook, C.E. & Hendrick, P.A., 2015, 'Clinical tests to diagnose lumbar spondylolysis and spondylolisthesis: A systematic review', *Physical Therapy in Sport* 16(3), 268–275. https://doi.org/10.1016/j.ptsp.2014.12.005

Andrade, J.A., Figueiredo, L.C., Santos, S.T., Paula, A.C.V, Bittencourt, N. & Fonseca, S.T., 2012, 'Reliability of transverse plane pelvic alignment measurement during the bridge test with unilateral knee extension', *Revista Brasileira De Fisioterapia (Sao Carlos (Sao Paulo, Brazil))* 16(4), 268–274. https://doi.org/10.1590/S1413-35552012000400007

Beaudreuil, J., Nizard, R., Thomas, T., Payre, M., Loitard, J.P., Boileau, P. et al., 2009, 'Contribution of clinical tests to the diagnosis of rotator cuff disease: A systematic literature review', *Joint Bone Spine* 76(1), 15–19. https://doi.org/10.1016/j.jbspin.2008.04.015

Choi, B. & Kang, S., 2015, 'Intra- and inter-examiner reliability of goniometer and inclinometer use in Craig's test', *Journal of Physical Therapy Science* 27(4), 1141–1144. https://doi.org/10.1589/jpts.27.1141

Cohen, J., 1960, 'A coefficient of agreement for nominal scales', *Educational and Psychological Measurement* 20(1), 37–46. https://doi.org/10.1177/001316446002000104

Cook, G., Burton, L. & Hoogenboom, B., 2006, 'Pre-participation screening: The use of fundamental movements as an assessment of function – Part 1', *North American Journal of Sport Physical Therapy* 1(2), 62–72.

Cools, A.M., Cambier, D. & Witvrouw, E.E., 2008, 'Screening the athlete's shoulder for impingement symptoms: A clinical reasoning algorithm for early detection of shoulder pathology', *British Journal of Sports Medicine* 42(8),628–635. https://doi.org/10.1136/bjsm.2008.048074

Dennis, R.J., Elliott, B.C. & Farhart, P.J., 2008, 'The reliability of musculoskeletal screening tests used in cricket', *Physical Therapy in Sport* 9(1), 25–33. https://doi.org/10.1016/j.ptsp.2007.09.004

Dennis, R.J., Finch, C.F., McLntosh, A.S. & Elliott, B.C., 2008, 'Use of field-based tests to identify risk factors for injury to fast bowlers in cricket', *British Journal of Sports Medicine* 42(6), 477–482. https://doi.org/10.1136/bjsm.2008.046698

DiMattia, M.A., Livengood, A.L., Uhl T.L., Mattacola, C.G. & Malone, T.R., 2005, 'What are the validity of the single-leg-squat test and its relationship to hip-abduction strength?', *Journal of Sport Rehabilitation* 14(2), 108–123. https://doi.org/10.1123/jsr.14.2.108

Dvorak, J. & Junge, A., 2009, 'FIFA Pre-Competition Medical Assessment (PCMA)', Fédération Internationale de Football Association, Zurich, Switzerland.

Ekstrom, R.A., Donatelli, R.A. & Carp, K.C., 2007, 'Electromyographic analysis of core trunk, hip, and thigh muscles during 9 rehabilitation exercises', *Journal of Orthopaedic & Sports Physical Therapy* 37(12), 754–762. https://doi.org/10.2519/jospt.2007.2471

Gamble, P., 2013, *Strength and conditioning for team sports: Sport-specific physical preparation for high performance*, 2nd edn., Routledge, New York.

Gray, J., 2015, *Cricket South Africa musculoskeletal screening document*, Cricket South Africa, Cape Town.

Gray, J. & Naylor, R., 2009, *BokSmart musculoskeletal assessment form*, BokSmart, Cape Town, South Africa.

Gregg, C.D., Dean, S. & Schneiders, A.G., 2009, 'Variables associated with active spondylolysis', *Physical Therapy in Sport* 10(4), 121–124. https://doi.org/10.1016/j.ptsp.2009.08.001

Hencken, C. & White, C., 2006, 'Anthropometric assessment of premiership soccer players in relation to playing position', *European Journal of Sport Science* 6(4), 205–211. https://doi.org/10.1080/17461390601012553

Hughes, R.E. & Na, K.N., 1996, 'Force analysis of rotator cuff muscles', *Clinical Orthopaedics and Related Research* 330, 75–83. https://doi.org/10.1097/00003086-199609000-00010

Hungerford, B.A., Gilleard, W., Moran, M. & Emmerson, C., 2007, 'Evaluation of the ability of physical therapists to palpate intrapelvic motion with the stork test on the support side', *Physical Therapy* 87(7), 879–887. https://doi.org/10.2522/ptj.20060014

Itoi, E., Kido, T., Sano, A. & Sato, K., 1999, 'Which is more useful, the "Full Can Test" or the "Empty Can Test," in detecting the torn supraspinatus tendon?', *The American Journal of Sports Medicine* 27(1), 65–68. https://doi.org/10.1177/03635465990270011901

Jackson, D., Wiltse L.L., Dingeman, R.D. & Hayer M., 1981, 'Stress reactions involving the pars interarticularis in young athletes', *The American Journal of Sports Medicine* 9(5), 304–312. https://doi.org/10.1177/036354658100900504

Kelly, B.T., Kadrmas, W.R. & Speer, K.P., 1996, 'The manual muscle examination for rotator cuff strength: An electromyographic investigation', *The American Journal of Sports Medicine* 24(5), 581–588. https://doi.org/10.1177/036354659602400504

Kibler, W.B., Press, J. & Sciascia, A., 2006, 'The role of core stability in athletic function', *Sports Medicine* 36(3), 189–198. https://doi.org/10.2165/00007256-200636030-00001

Kim, E., Jeong, E., Won, L.K. & Song, J.S., 2006, 'Interpreting positive signs of the supraspinatus test in screening for torn rotator cuff', *Acta Medica Okayama* 60(4), 223–228. https://doi.org/10.18926/AMO/30715

Koley, S., 2011, 'A study of anthropometric profile of Indian inter-university male cricketers', *Journal of Human Sport and Exercise* 6(2), 427–435. https://doi.org/10.4100/jhse.2011.62.23

Landis, J.R. & Koch, G.G., 1977, 'The measurement of observer agreement for categorical data', *Biometrics* 33(1), 159–174. https://doi.org/10.2307/2529310

Lasbleiz, S., Quintero, N., Ea., Petrover, D., Aout, M. Laredo, J.D., et al., 2014, 'Diagnostic value of clinical tests for degenerative rotator cuff disease in medical practice', *Annals of Physical and Rehabilitation Medicine* 57(4), 228–243. https://doi.org/10.1016/j.rehab.2014.04.001

Lehr, M.E., Plinsky P.J., Butler, R.J., Fink, M.L., Kiesel, K.B., Underwood, F.B. et al., 2013, 'Field-expedient screening and injury risk algorithm categories as predictors of noncontact lower extremity injury: Field screens predict lower extremity injury', *Scandinavian Journal of Medicine & Science in Sports* 23(4), 225–232. https://doi.org/10.1111/sms.12062

Liu, Y., Ao, Y., Yan, H., Cui, G., 2016, 'The hug-up test: A new, sensitive diagnostic test for supraspinatus tears', *Chinese Medical Journal* 129(2), 147. https://doi.org/10.4103/0366-6999.173461

Madsen, N.L., Drezner, J.A. & Salerno, J.C., 2014, 'The preparticipation physical evaluation: An analysis of clinical practice', *Clinical Journal of Sport Medicine* 24(2), 142–149. https://doi.org/10.1097/JSM.0000000000000008

Magee, D.J., Sueki, D. & Chepeha, J., 2011, *Orthopedic physical assessment atlas and video: Selected special tests and movements*, Saunders (Musculoskeletal rehabilitation series), Philadelphia, PA.

Martin, B.J. & Altman, D., 1986, 'Statistical methods for assessing agreement between two methods of clinical measurement', *The Lancet* 327(8476), 307–310. https://doi.org/10.1016/S0140-6736(86)90837-8

Masci, L., Pike, J., Malara, F., Phillips, B., Bennell, K., Brukner, P., 2006, 'Use of the one-legged hyperextension test and magnetic resonance imaging in the diagnosis of active spondylolysis', *British Journal of Sports Medicine* 40(11), 940–946. https://doi.org/10.1136/bjsm.2006.030023

Michener, L.A., Walsworth, M., Doukas, W.C. & Murphy, K.P., 2009, 'Reliability and diagnostic accuracy of 5 physical examination tests and combination of tests for subacromial impingement', *Archives of Physical Medicine and Rehabilitation* 90(11), 1898–1903. https://doi.org/10.1016/j.apmr.2009.05.015

Monnier, A., Heuer, J., Norman, K. & Ang, B.O., 2012, 'Inter- and intra-observer reliability of clinical movement-control tests for marines', *BMC Musculoskeletal Disorders* 13(1). https://doi.org/10.1186/1471-2474-13-263

Moreland, J., Finch, E., Stratford, P., Balsor, B. & Gill, C., 1997, 'Interrater reliability of six tests of trunk muscle function and endurance', *Journal of Orthopaedic & Sports Physical Therapy* 26(4), 200–208. https://doi.org/10.2519/jospt.1997.26.4.200

Naidoo, M.A., 2007, *The epidemiology of soccer injuries sustained in a season of a professional soccer team in South Africa*, University of the Western Cape, viewed 18 May 2018, from http://etd.uwc.ac.za/bitstream/handle/11394/3786/Naidoo_MSc_2007.pdf?sequence=1

Nilstad, A., Andersen, T.E., Kristianslund, E., Bahr, R., Myklebust, G., Steffen, K., et al., 2014, 'Physiotherapists can identify female football players with high knee valgus angles during vertical drop jumps using real-time observational screening', *Journal of Orthopaedic & Sports Physical Therapy* 44(5), 358–365. https://doi.org/10.2519/jospt.2014.4969

Örtqvist, M., Mostrom, E.B., Roos, M., Lundell, P. Janarv, P.M., Werner, S. et al., 2011, 'Reliability and reference values of two clinical measurements of dynamic and static knee position in healthy children', *Knee Surgery, Sports Traumatology Arthroscopy* 19(12), 2060–2066. https://doi.org/10.1007/s00167-011-1542-9

Ostor, A.J.K., 2004, 'Interrater reproducibility of clinical tests for rotator cuff lesions', *Annals of the Rheumatic Diseases* 63(10), 1288–1292. https://doi.org/10.1136/ard.2003.014712

Portney, L.G. & Watkins, M.P., 2000, *Foundations of clinical research: Applications to practice*, 2nd edn., Prentice Hall, Upper Saddle River, NJ.

Rebelo, A., Brito, J., Maia, J., Coelho-e-Silva, M.A. Figueiredo, A.J., Bangsbo, J., et al., 2012, 'Anthropometric characteristics, physical fitness and technical performance of under-19 soccer players by competitive level and field position', *International Journal of Sports Medicine* 34(04), 312–317. https://doi.org/10.1055/s-0032-1323729

Robinson, H.S., Brox, J.I., Bjelland, E., Solem, S. Telje, T., 2007, 'The reliability of selected motion- and pain provocation tests for the sacroiliac joint', *Manual Therapy* 12(1), 72–79. https://doi.org/10.1016/j.math.2005.09.004

Rodriguez, N., DiMarco, N. & Langley, S., 2009, 'Nutrition and athletic performance', *Medscape*, viewed 01 March, from https://www.medscape.com/viewarticle/717046

Shrout, P.E. & Fleiss, J.L., 1979, 'Intraclass correlations: Uses in assessing rater reliability', *Psychological Bulletin* 86(2), 420–428. https://doi.org/10.1037/0033-2909.86.2.420

Sim, J. & Wright, C.C., 2005, 'The kappa statistic in reliability studies: Use, interpretation, and sample size requirements', *Physical Therapy* 85(3), 257–268.

Somerville, L.E., Wilits, K., Johnson, A.M., Litchfield, R. LeBel, M.E., Moro, J., et al., 2014, 'Clinical assessment of physical examination maneuvers for Rotator Cuff Lesions', *The American Journal of Sports Medicine* 42(8), 1911–1919. https://doi.org/10.1177/0363546514538390

Strauts, J. & Tate, K., 2015, 'Exercise highlight: Single leg Romanian deadlift and variations', *Journal of Australian Strength Conditioning* 23(3), 43–48.

Stretch, R.A., 2001, 'Incidence and nature of epidemiological injuries to elite South African cricket players', *South African Medical Journal* 91(4),336–339.

Stuelcken, M., Pyne, D. & Sinclair, P., 2007, 'Anthropometric characteristics of elite cricket fast bowlers', *Journal of Sports Sciences* 25(14), 1587–1597. https://doi.org/10.1080/02640410701275185

Tong, H.C., Heyman, O.G., Lado D.A. & Isser, M.M., 2006, 'Interexaminer reliability of three methods of combining test results to determine side of sacral restriction, sacral base position, and innominate bone position', *The Journal of the American Osteopathic Association* 106(8),464–468.

van Mechelen, W., Hlobil, H. & Kemper, H.C., 1992, 'Incidence, severity, aetiology and prevention of sports injuries. A review of concepts', *Sports Medicine* 14(2), 82–99. https://doi.org/10.2165/00007256-199214020-00002

Whatman, C., Hume, P. & Hing, W., 2015, 'The reliability and validity of visual rating of dynamic alignment during lower extremity functional screening tests: A review of the literature', *Physical Therapy Reviews* 20(3), 210–224. https://doi.org/10.1179/1743288X15Y.0000000006

Wiesel, S., 2018, 'The dilemma of spondylolysis; inadequate evidence, poorly documented outcomes', *The Back Letter* 17(4), 37–45.