Intra and Inter-Rater Reliability of Screening for Movement Impairments: Movement Control Tests from The Foundation Matrix

Carolina R. Mischiati [1]✉, **Mark Comerford** [2,3], **Emma Gosford** [1], **Jacqueline Swart** [4], **Sean Ewings** [5], **Nadine Botha** [2,6], **Maria Stokes** [2,6] **and Sarah L. Mottram** [2,3,6]

[1]Department of Sport and Exercise Sciences, University of Chichester, UK; [2] Faculty of Health Sciences, University of Southampton, UK; [3] Movement Performance Solutions, Chichester, UK; [4] Manage Movement Centre, Pretoria, South Afric; [5] Southampton Statistical Sciences Research Institute, University of Southampton, UK; [6] Arthritis Research UK Centre for Sport, Exercise and Osteoarthritis, UK

Abstract

Pre-season screening is well established within the sporting arena, and aims to enhance performance and reduce injury risk. With the increasing need to identify potential injury with greater accuracy, a new risk assessment process has been produced; The Performance Matrix (battery of movement control tests). As with any new method of objective testing, it is fundamental to establish whether the same results can be reproduced between examiners and by the same examiner on consecutive occasions. This study aimed to determine the intra-rater test re-test and inter-rater reliability of tests from a component of The Performance Matrix, The Foundation Matrix. Twenty participants were screened by two experienced musculoskeletal therapists using nine tests to assess the ability to control movement during specific tasks. Movement evaluation criteria for each test were rated as pass or fail. The therapists observed participants real-time and tests were recorded on video to enable repeated ratings four months later to examine intra-rater reliability (videos rated two weeks apart). Overall test percentage agreement was 87% for inter-rater reliability; 98% Rater 1, 94% Rater 2 for test re-test reliability; and 75% for real-time versus video. Intraclass-correlation coefficients (ICCs) were excellent between raters (0.81) and within raters (Rater 1, 0.96; Rater 2, 0.88) but poor for real-time versus video (0.23). Reliability for individual components of each test was more variable: inter-rater, 68-100%; intra-rater, 88-100% Rater 1, 75-100% Rater 2; and real-time versus video 31-100%. Cohen's Kappa values for inter-rater reliability were 0.0-1.0; intra-rater 0.6-1.0 for Rater 1; -0.1-1.0 for Rater 2; and -0.1-1 for real-time versus video. It is concluded that both inter and intra-rater reliability of tests in The Foundation Matrix are acceptable when rated by experienced therapists. Recommendations are made for modifying some of the criteria to improve reliability where excellence was not reached.

Key words: Movement control, movement impairments, screening, reliability.

Introduction

Emphasis on injury prevention is increasing in sport to reduce the health and economic impact of injury and maximise performance (Bahr and Krosshaug, 2005; McBain et al., 2012 Peate et al., 2007; Turbeville et al., 2003; Tyler et al., 2006; Webborn, 2012; Zazulak et al., 2007). Injury results when body tissue is unable to cope with the applied stresses, whether acute or chronic

(McBain et al., 2012). Multiple factors increase the risk of injury and invariably injuries result from a combination of factors including; history of pain, previous injury, acquired hypermobility, aerobic fitness, changes in the control of movement (Bahr and Krosshaug, 2005; McBain et al., 2012; Plisky et al., 2006, Roussel et al 2009; Webborn, 2012; Yeung et al., 2009). Various strategies have been employed to reduce both extrinsic causes of injury (e.g. unavoidable direct contact trauma and intrinsic causes of injury (e.g. non-contact injury related to overuse, or poor movement technique, efficiency or control), but consistent predictors of injury are still lacking (Butler at al., 2010; Garrick, 2004).

Although pre-season screening has been part of the routine in sport for some time, some approaches lack the complexity required to identify movement impairments relevant to everyday activities. In other testing protocols, there has been an emphasis on measuring joint mobility, muscle extensibility, endurance and strength, as well as fitness tests and physiological testing (Butler et al., 2010; Mottram and Comerford, 2008; Myer et al., 2008; Yeung et al., 2009). Such tests have a role in providing benchmarks for rescreening reference during the season and post injury, and provide some indication of limitations that need addressing but many do not predict injury (Butler et al., 2010).

Currently, the strongest predictor of injury is previous injury (Chalmers, 2002; Fulton et al., 2014; Tyler et al., 2006) but this is clearly not desirable as an ongoing predictor. It has been suggested that a change may occur following injury, which could be explained as a change in motor control (Kiesel et al., 2009; MacDonald et al., 2009). Central nervous system mediated motor control is vital to both production and control of movement (Hodges and Smeets 2015), and more recently the focus of assessing movement impairments and developing movement retraining programs has moved towards optimising the control of movement (Cook et al., 2006a; Cook et al., 2006b, Luomajoki et al., 2010, Worsley et al., 2013).

Quality of movement, specifically control of movement, is now being recognised as an important element of assessment of movement efficiency as well as range (Simmonds and Keer, 2007; Roussel et al., 2009). The identification and correction of movement control impairments have been recognised as an important part of assessing and rehabilitating injury (Comerford and Mot-

tram, 2001; Comerford and Mottram 2012; Luomajoki et al., 2007; Sahrmann, 2014) but attention is now focussing on protocols to evaluate uninjured groups to determine non-symptomatic deficits within the kinetic chain of functional movement patterns that might predispose to injury (Cook et al., 2006a; 2006b; Kiesel et al., 2007; Peate et al., 2007; Plisky et al., 2006; Roussel et al., 2009). This new perspective in screening has been used to develop a tool known as The Performance Matrix (TPM). This tool employs generic multiple joint tasks that are functionally relevant, but not necessarily habitual, and have been modified to test the cognitive control of movement. The protocol identifies inefficient control of movement (weak links) within the kinetic chain indicating the presence of uncontrolled movement (UCM; Comerford and Mottram, 2012). The key features of TPM that differ from other movement screens include: 1) detailing the site and direction of UCM to direct specific targeted retraining; 2) evaluation of low and high threshold UCM explained below; 3) testing for active control of movement to benchmark standards (not natural or habitual movement patterns), 4) evaluation of movement control and not pain, which, when present, does not mean a fail in this screen, as some people have good control in the presence of pain; 5) the software produces a risk algorithm that helps establish retraining priorities (although this has yet to be validated). 6) A unique classification of subgroups of movement control impairments as high risk, low risk and assets.

Uncontrolled movement is described for the purpose of this screening process as a lack of ability to cognitively co-ordinate and control motion efficiently to benchmark standards at a particular body segment (Comerford and Mottram, 2012). A loss of the ability to control movement is thought to increase the loads and stresses on the joint, increasing susceptibility to injury (Sterling et al., 2001).

Motor control is key to optimising control of movement and ensuring the coordinated interplay of the various components of the movement system: the articular, myofascial and connective tissue, and neural systems. Changes in motor control result in altered patterns of recruitment, with inhibition in some muscle groups and increased activity in others (Hodges and Richardson, 1996; O'Sullivan, 2005; Sterling et al., 2001). Altered muscle recruitment strategies and motor control impairments may result from previous injury (O'Sullivan, 2005; Sterling et al., 2001) fatigue, stiffness i.e. loss of range of joint of motion or myofascial extensibility (Cook et al., 2006a), or muscle imbalances (Cook et al., 2006a; Sahrmann, 2002). Impaired motor control can also lead to compensatory movements (Roussel et al., 2009; Zazulak et al., 2007).

The Foundation Matrix forms part of The Performance Matrix movement screening system (Movement Performance Solutions Ltd). The Performance Matrix is the generic name for the entire group of screening tests developed by Movement Performance Solutions. The Foundation Matrix, is the most commonly used screening tool in the database, and is the entry level screen, which is designed to identify performance related inefficient

control of movement in the kinetic chain. This entry level matrix was therefore chosen for the present study. Other screens in the database are sport specific, e.g. football and golf, or region specific e.g. low back, or occupation specific e.g. office worker or tactical athlete, such as fire fighter or police. Using a series of multi-joint functionally relevant tests (listed in Table 2), The Foundation Matrix screen evaluates movement control efficiency. The protocol assesses both the site and direction of uncontrolled movement in different joint systems, and evaluates these control impairments under two different, but functionally relevant physiological situations, low (Figure 1 for example of test 1) and high threshold testing (Figure 2 for example of test 9). The screening tool assesses deficits in the control of non-fatiguing alignment and co-ordination skills in what is referred to as 'low threshold' tasks, and assesses deficits in movement control during fatiguing strength and speed challenges in what is referred to as 'high threshold' tasks (Mottram and Comerford 2008). The objective of The Foundation Matrix screening tool is to provide the assessor with details of the site, direction, and threshold of uncontrolled movement, to allow for the development of a specific training programme. When considering the utility of a test, both reliability (intra and inter-rater) and validity must be established.

Although some movement control tests have been evaluated for reliability and validity (Luomajoki et al., 2007; Roussel et al., 2009; Teyhen et al., 2012), the reliability of the battery of movement control tests in The Foundation Matrix has not been examined. Therefore, the aim of the present study was to establish both the intra and inter-rater reliability of experienced therapists in rating performance of nine of the 10 tests from The Foundation Matrix. These include five low threshold tests of alignment and coordination control and four high threshold tests of strength and speed control. The reason for excluding one of the high threshold tests is explained below.

Methods

Participants
Twenty university sports students (11 females; 9 males; aged = 21 ± 3) participated in the study. Participants were asymptomatic and were excluded if they had a present pathology, injury, pain, surgery, a musculoskeletal injury within the past six months, or were pregnant. Prior to screening, all participants gave written, informed consent, and the Research Ethics Committee of the University approved the study.

Raters
Two experienced musculoskeletal therapists, who were specialists in the field of movement control, assessed the efficiency of movement control during the performance of tests. Therapist 1 had 23 years' experience in musculoskeletal physiotherapy and 14 years' experience in movement control assessment and re-training, and Therapist 2 had 16 years' experience in musculoskeletal physiotherapy and 7 years' experience in movement control assessment and re-training.

Protocol

Both therapists assessed participants using the battery screening protocol comprising nine movement control tests (Table 2). Each participant was scheduled to a 45 minute session. Both therapists completed the screening process at the same time independently, without conferring, during which time testing was video recorded for retrospective analysis of intra-rater reliability of assessing test performance on another occasion (Butler et al., 2012; Fersum et al., 2009; Luomajoki et al., 2007). Six digital cameras (Casio exfh20) were used to record participants performing the tests, and were set up to give anterior, posterior and lateral views. Tripods and angle adjustment allowed for variation in positioning of the tests. All participants wore black lycra shorts and females wore a sports top that allowed observation of movement and bony landmarks.

Table 1. Order of performance of tests.

The tests are reported by the name of the test in The Foundation Matrix (TFM).

NB Test 6 was not included in the present study (see Methods section of text)

Standing Tests

1- Double Knee Swing (Low threshold TFM Test 1)
2- Single leg ¼ squat+hip turn (Low threshold TFM Test 2)
3- Controlled shoulder internal rotation (Low threshold TFM Test 4)
4- Split squat+fast feet change (High threshold TFM Test 9)
5- Lateral stair hop + rotational landing control (High threshold TFM Test 10)

Floor Tests

6- Bridge + straight leg lift & lower (Low threshold TFM Test 3)
7- 4 point - arm reach forward and back (Low threshold TFM Test 5)
8- Plank + lateral twist (High threshold TFM Test 7)

Wall test

• One arm wall push (Low threshold TFM Test 8)

Movement control tests

Nine of the 10 movement control tests of The Foundation Matrix were used (Table 1). Each of the 10 tests in the Foundation Matrix has five criteria posed as questions (n = 50) which require an observational judgement regarding the person's ability to adequately control movement to a pass or fail benchmark standard. Not all movement evaluation criteria on movement control faults could be evaluated. Some movement evaluation criteria involving the ability to control movement could not be included due to: 1) appropriate views not being possible to obtain clearly on video; 2) passive tests of movement restrictions not being part of the present study of observational testing; 3) tests requiring the use of a pressure biofeedback unit which gave objective measures of control; and 4) criteria assessing repositioning ability (proprioception). Therefore, 40 of the 50 movement evaluation criteria, and 9 of the 10 tests were used in the study. Test 1, the Double Knee Swing (Figure 1) has been described and illustrated previously (McNeill 2014). Each test was rated by a series of movement evaluation criteria (Table 2) posed as ques-

tions, aimed at identifying observational markers that indicate uncontrolled movement. Each criterion is given a pass or fail response. The final report identifies both performance assets and weak links (movement control impairments), which are the priority risk factors (Comerford, 2006; Mottram and Comerford, 2008).
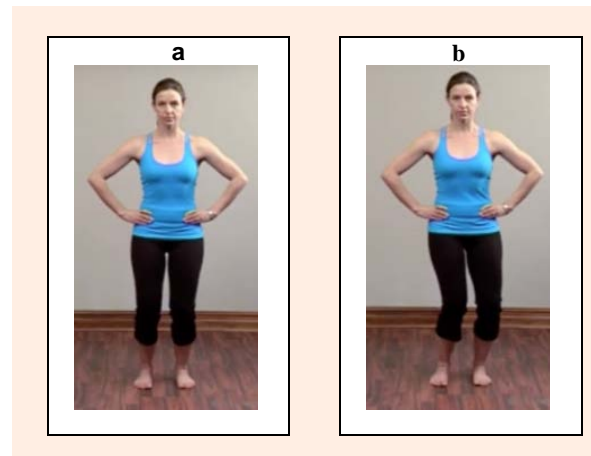
Figure 1. Test 1, **a)** start position, **b)** end position double knee swing.
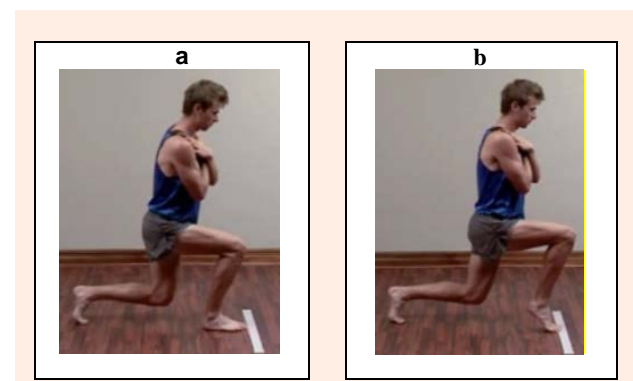
Figure 2. Test 9, **a)** start position, **b)** end position split squat and fast feet change

Prior to testing, the study co-ordinator and therapists reviewed the assessment criteria to ensure consistency. The participant was taught each test following standardised instruction (summarised in Table 2) regarding how to perform the movement task correctly using visual, audio and kinaesthetic techniques. Before the screening process commenced, each participant viewed a video of the correct performance of the movements and Therapist 1 verbally explained each test in detail, and the main objectives. Because these tests evaluate the performance of an unfamiliar skill of movement control, and not a natural functional movement, a period of familiarisation is necessary so as not to skew the results for the wrong reason. It is important that a person is judged to fail a test because of poor active cognitive control of movement, not because they were unsure of what the control task required.

Experience has shown that four to six practice attempts with feedback and cueing was sufficient for people with good control abilities to learn and pass the test (Luomajoki et al., 2007; Worsley et al., 2013).

Table 2. Test details and scoring system for movement efficiency criteria.

Test Details	Marking Criteria
Double Knee Swing	
In standing, bend the knees into a ¼ squat position Swing both legs simultaneously to the left, then right to 20° of hip rotation The pelvis should not rotate or laterally shift to follow the knees Keep the 1st metatarsal head fully weight bearing on the floor.	Can you prevent the pelvis and back rotating to follow the legs? Can you prevent side bending of the trunk and lateral movement of the shoulders? Can you keep the trunk upright and prevent further forward bending at the hips? Can you prevent the foot from turning out as the knee swings out to 20°? Can you prevent the big toe from lifting as the knee swings out to 20°?
Single Leg ¼ Squat + Hip Turn	
Stand on one foot keeping pelvis and shoulders level, and arms across the chest Take a small knee bend 30°, and hold this position for 5 seconds Then moving the trunk and pelvis together, turn 30° away from the standing foot Hold this position for 3 seconds Turning back to the front straighten the knee Repeat the movement, standing on the other leg	Can you keep the pelvis facing straight ahead as you lower into the small knee bend and hold the position for 5 seconds? Can you prevent side bending of the low back and trunk in the small knee bend position or during the rotation? Can you prevent the trunk from leaning further forward in the small knee bend position? Can you prevent the (WB) knee turning in across the foot to follow the pelvis as you turn the pelvis away from the standing foot? Can you prevent the (WB) arch from rolling down or toes clawing?
Bridge + Heel Lift + Single Straight Leg Raise & Lower	
Lying in crook lying position, lumbopelvic neutral position, arms folded across the chest Maintaining position, lift the pelvis just clear of the floor (about 2 cm) Lift heels into full plantar flexion Maintaining position, slowly take weight off one foot and straighten that knee keeping thighs level. Then slowly raise the straight leg, moving the thigh up towards the vertical position, then slowly lower the straight leg (extend the hip) to horizontal Return to crook lying and repeat on the opposite side	Can you prevent low back flexion as the straight leg raises? Can you prevent low back extension as the leg lowers? Can you prevent pelvic rotation against asymmetrical single leg load?
Controlled Shoulder Internal Rot	
Stand tall with the scapular in neutral position, shoulder abducted to 90°, 15-30° forward of the body in scapular plane, elbow flexed to 90° Ensure humeral head and shoulder blade, are in neutral position Maintaining upper arm and scapular position, rotate the arm to lower the hand down towards the floor. Monitor the scapular at the coracoid with one finger and the front of humeral head W with another finger during medial rotation There should be 60° of independent medial rotation of the shoulder joint	Can you prevent the upper back and chest from dropping forward as you rotate the arm? Can you prevent the upper back and chest from turning as you rotate the arm? Can you prevent the coracoid rolling or tilting forward? Can you prevent forward protrusion of the humeral head?
4 Point - Arm Reach Forward And Back	
Start on all fours, knees under the hips and hands under the shoulders Position the spine, scapulae and head in neutral mid position Maintaining neutral position, shift body weight onto one hand, slowly lift the other arm off the floor to reach behind you to 15° shoulder extension. Then move to lift and reach the arm in front to ear level. Repeat to other side	Can you prevent either shoulder blade hitching? Can you prevent either shoulder blade dropping or tilting forward? Can you prevent winging of the weight-bearing shoulder blade? Can you prevent forward protrusion of the head of the shoulder joint as the non weight-bearing arm extends?
Plank + Lateral Twist	
Lie face down supported on elbows, positioned under shoulders and forearms across the body, side by side. Maintaining the knees and feet together, bend the knees to 90°, and push the body away from the floor taking the weight through the arms into a ¾ plank, keeping a straight line with legs, trunk and head. Maintaining lumbopelvic neutral position shift the upper body weight onto one elbow, during the weight shift the body should move laterally (approx 5-10cm). Turn the whole body 90° from the (WB) shoulder to a ¾ side plank, the trunk, pelvis and legs should turn together and remain in a straight line. Return to starting position again maintaining position. Repeat the movement to the other side	Can you prevent the weight-bearing shoulder blade dropping? Can you prevent the weight-bearing shoulder blade winging or retracting? Can you prevent forward protrusion of the humeral head of the weight-bearing shoulder joint as you turn onto one arm? Can you prevent the low back from arching? Can you prevent the pelvis from leading the twist as you turn from the front plank position towards the side plank position?

Table 2. Continued.

Test Details	Marking Criteria
One Arm Wall Push	
Stand tall in front of a wall, hold the arm at 90° flexion, hand placed on the wall, scapular in neutral, move the feet one foot length further back away from the wall, lean forward and take body weight on the hand	Can you prevent the upper back from flexing or rounding out as the arm pushes away from the wall?
Keeping the shoulder blade, trunk and pelvis in neutral, slowly bend the elbow to lower the forearm down to the wall	Can you prevent the upper back from rotating?
Lower the elbow so the forearm is vertical and fully weight-bearing against the wall, then push the body slowly away from the wall to fully straighten the elbow	Can you prevent the weight-bearing shoulder blade from hitching or retracting?
Do not allow the trunk and pelvis to rotate or arch towards the wall.	Can you prevent forward tilt or winging of the weight-bearing shoulder blade?
Repeat with the other arm	
Split Squat + Fast Feet Change	
Step out with one foot (4 foot length), feet facing forwards and arms folded across chest	Can you prevent side bending of the trunk?
Keeping the trunk upright, drop down into a lunge, rapidly switch feet in a split squat movement, control the landing	Can you keep the trunk upright and prevent the trunk leaning forward at the hips towards the front foot?
Then lift the heel of the front foot to full plantarflexion and hold this heel lift in the deep lunge for 5 seconds, then lower the heel and without straightening up, rapidly switch feet in a split squat movement, control the landing	Can you prevent the front knee moving in across the line of the foot?
After the landing, again lift the heel of the front foot to full plantarflexion and hold this heel lift in the deep lunge for 5 seconds	Can you prevent the foot from turning out or the heel pulling in as you land?
Repeat the heel lift twice with each leg in the forward position	Can you prevent the heel of the front foot from rolling out during the heel lift?
Lateral Stair Hop + Rotational Landing Control	
Stand side on to a box/step (approx 15 cm) with the feet together, and arms by your side	Can you prevent the trunk or pelvis from rotating?
Keeping the back straight bend the knees into a 'small knee bend' position, lift the outside leg off the floor to balance on the inside leg	Can you prevent side-bending of the trunk as you land on the hop down?
Hop laterally up onto the box/step / keeping the back upright and controlling the landing into the 'small knee bend' position	Can you prevent the body from leaning forwards at the hip as you land?
Hold this position for 5 seconds	Can you prevent the landing knee turning in across the foot as you hop down?
Then hop back down off the box to rotate through 90° to land on the same leg turning to face away from the box/step	Can you prevent the arch from rolling down or toes clawing as you hop down?
Repeat with the other leg	

Once the participant indicated that they clearly understood how to perform the test, the test procedure was commenced and scored independently by the assessors. Each test was repeated up to three times, and the two therapists recorded their scores of performance. The order of the tests was standardised (Table 1) to ensure all participants were assessed the same way as recommended by Luomajoki et al., (2007). The test sessions took approximately 40 minutes.

Scoring system

As the participant carried out the movement task, Therapists 1 and 2 recorded their observation on the efficacy the participant's ability to control movement by a number of criteria, which involved scoring a pass or fail to a set of criteria (Table 2). Since the intention of The Foundation Matrix is to measure impairment, a low score indicates less impairment and a high score indicates greater impairment. Therefore, fail is rated as 1 and pass is rated as 0. After the therapist had recorded their observations, the participant was taught the next test and the procedure repeated until all nine tests had been recorded.

Inter-rater reliability was determined from the scoring conducted real-time, reflecting routine practice. Conversely, intra-rater reliability was established from the video recordings of the screening protocol as this ensured consistency of performance of the movement control tests. Videos were downloaded onto a hard drive, where each participant was identified by number only. Videos were then edited using Final cut pro (2001).

The video recordings were distributed to Therapists 1 & 2 for intra-rater reliability evaluation four months following the real-time assessment of the movement control tests. This time lapse was due to logistical reasons, including the need to edit and compile the videos ready for assessment, which was conducted two weeks apart, as previously described by Luomajoki et al., (2007). The video recordings were viewed on a laptop with a maximum of three views per test based on Ekegren et al., (2009). As before, the Therapists' observations on the control of movement were recorded. All scoring sheets were transcribed to an Excel spread sheet for analysis.

Statistical analysis

Percentage agreement is presented for each site and direction of uncontrolled movement (identified as a fail) as a whole score (e.g. left and right sides combined). The overall score for an individual is the sum of the scores across all individual site and direction failures of each test. The reliability of this overall score was assessed

using intraclass correlation coefficient (ICC) and classi-fied according to Fleiss (2007). Real-time versus video- and intra-rater reliability were assessed with the ICC(1,1) model (with subjects as the only effect), and inter-rater reliability was assessed with the ICC(2,1) model (with subjects and judges both considered random effects).

Agreement was examined using the Kappa test. Cohen's Kappa (κ) is commonly used to assess agreement between different judges when ratings are on a nominal scale (Cohen, 1960). Kappa describes agreement beyond chance relative to perfect agreement beyond chance (as opposed to percentage agreement, which describes agreement relative to perfect agreement). Kappa was used to assess inter- and intra-rater agreement, across each individual criterion of each test; Kappa was also used to evaluate agreement between real-time and video assess-ment by one rater. All Kappa values are presented with 95% confidence intervals. Where movement evaluation criteria of a test are applicable to two sides of the body (e.g. left leg and right leg), Kappa is presented for each separate side in order to preserve the assumption of inde-pendence between observations.

Results

The analysis was based on nine tests, comprising varying numbers of components, i.e. criteria about movement control (between 3 and 5 components in each test, total-ling 40 criteria). These criteria may be further broken down into left- and right-hand side for assessment using κ, referred to here as sub-components (35 of the criteria may be broken down into two sub-components, making 75 criteria in total).

Table 4. Reliability of overall scores: intraclass correlation coefficients (ICC).

	Scenario	ICC	(95% CI)
ICC(1,1)	Intra-rater (1)	.96	(.93, .99)
	Intra-rater (2)	.88	(.78, .98)
	Real-time versus video	.23	(.00, .65)
	Intra-rater (1)	.96	(.93, .99)
ICC(2,1)	Inter-rater agreement	.81	(.50, .93)

Composite analysis

Overall percentage test agreement (as a combination of each test's criteria) were: 86.5% for inter-rater reliability; 97.5% for test re-test reliability in Rater 1 and 93.9% in Rater 2; and 74.5% for real-time versus video. Table 3

shows the percentage agreement for each of the nine tests for the four scenarios examined.

The ICCs for overall agreement (combination of all tests) for each scenario were excellent between raters (0.81) and within raters (Rater 1, 0.96; Rater 2, 0.88) but poor for real-time versus video (0.23) (Table 4).

Criteria analysis within movement tests

More detailed analysis of the movement evaluation crite-ria in each test revealed variability within each of the reliability scenarios, indicating which criteria were more reliable than others.

Inter-rater agreement

The percentage agreement for criteria ranged from 67.5 to 100% (mean overall agreement 86.5%), with only two of the 40 criteria in Table 5 with less than 70% agreement. Each test as a whole had agreement above 70% as shown in Table 3. Agreement varied between tests, with some having ratings for all criteria agreeing >80% (Test 3, 4, 7) and for one test >90% (Test 8). Tests 5 and 9 were less reliable, with criteria below 75% (Test 5 D 67.5%; Test 9 D 72.5 and E 67.5%). The percentage agreements for the inter-rater reliability are presented in Figure 3, which highlights the level of agreement for each test.

Cohen's Kappa with 95% confidence intervals, and criteria percentage agreement results are presented in Table 5 for inter-rater reliability of real-time observations. Kappa values ranged from 0-1, with κ>0.6 for 36 of the 75 criteria, and 12 of these with κ>0.8. A further six crite-ria had κ=N/A and hence 100% agreement. Kappa values were generally similar between the right and left sides, although there were some exceptions. The range of values for criteria within each test varied, although some tests could be identified as having particularly good reliability or otherwise. For example, Tests 4 and 8 had generally high Kappa and percentage agreement values, whereas Test 9 had some low Kappa scores that were reflected by lower percentage agreement than other tests.

Intra-rater agreement

Intra-rater agreement for repeated ratings from videos is presented for Rater 1 in Table 6 and Rater 2 in Table 7. Reliability for Rater 1 was good, with percentage agree-ment values ranged from 87.5 to 100% (mean 97.5%). Kappa values ranging from 0.64 to 1.0, with only 5/75 criteria (right and left shown) with κ<0.70 (10/75 were κ=N/A) and all nine tests containing values of 1.0.

Table 3. Overall test agreement (as a combination of each test's criteria) and overall scenario agreement (as a combination of all tests).

Test	Test agreement (%)			
	Inter-rater	Intra-rater (1)	Intra-rater (2)	Real-time / video
1	85.0	99.5	93.0	76.9
2	83.5	99.5	90.5	77.5
3	87.2	95.0	100.0	67.9
4	95.0	97.1	93.6	74.6
5	81.9	96.7	94.1	67.8
7	91.7	98.3	90.0	82.2
8	95.0	94.3	96.4	87.1
9	77.5	97.5	95.5	77.3
10	87.0	97.0	96.0	58.3
Overall	86.5	97.5	93.9	74.5

Table 5. Inter-rater agreement for real-time assessment: Cohen's kappa with 95% confidence intervals and percentage agreement for criteria (n = 20 except * n = 18)

Test	Criteria	Left-hand side Kappa (95% CI)		Right-hand side Kappa (95% CI)		Criteria agreement (%)
1	A	.86	(.59-1.00)	.88	(.64-1.00)	95.0
	B	.55	(.19-.90)	.38	(-.02-.78)	75.0
	C	.74	(.40-1.00)	.74	(.40-1.00)	90.0
	D	.58	(.19-.98)	.70	(.39-1.00)	85.0
	E	.66	(.31-1.00)	.43	(.01-.85)	80.0
2	A	.22	(-.33-.76)	.00	N/A	82.5
	B	.57	(.14-1.00)	.32	(-.27-.90)	85.0
	C	.12	(-.28-0.52)	.27	(-.17-.71)	75.0
	D	.88	(.64-1.00)	.76	(.45-1.00)	92.5
	E	.76	(.45-1.00)	.38	(-.07-.82)	82.5
3	A	.77 (.35-1.00)				95.0
	B	.80 (.54-1.00)				90.0
	C	.00	N/A	.34 *	(-.17-0.85)	84.2
4	A	1.00 (1.00-1.00)				100.0
	B	N/A	N/A	N/A	N/A	100.0
	C	1.00	(1.00-1.00)	.64	(.01-1.00)	97.5
	D	.69	(.38-1.00)	.68	(.35-1.00)	85.0
5	A	N/A	N/A	.00	N/A	97.5
	B	.63	(.24-1.00)	.74	(.40-1.00)	87.5
	C	.38	(-.07-.82)	.39	(-.03-.81)	75.0
	D	.24	(-.17-.65)	.13	(-.33-.58)	67.5
7	A	.00	N/A	.00	N/A	95.0
	B	.77	(.35-1.00)	.44	(-.20-1.00)	92.5
	C	.77	(.35-1.00)	.86	(.59-1.00)	95.0
	D	.66 (.308-1.00)				85.0
	E	.89	(.67-1.00)	.58	(.22-.95)	87.5
8	A	N/A N/A				100.0
	B	.80	(.54-1.00)	.89	(.67-1.00)	92.5
	C	N/A	N/A	N/A	N/A	100.0
	D	.69	(.29-1.00)	.78	(.50-1.00)	90.0
9	A	.29	(-.14-.73)	.63	(.24-1.00)	77.5
	B	.68	(.35-1.00)	.60	(.26-.94)	82.5
	C	.55	(.19-.90)	.89	(.67-1.00)	87.5
	D	.38	(-.12-.87)	.15	(-.28-.57)	72.5
	E	.35	(-.07-.76)	.30	(-.06-.66)	67.5
10	A	.44	(-.20-1.00)	.000	N/A	92.5
	B	.39	(-.09-.86)	.615	(.15-1.00)	85.0
	C	.50	(.12-.88)	.600	(.25-.95)	77.5
	D	.64	(.01-1.00)	1.000	(1.00-1.00)	97.5
	E	.60	(.24-.95)	.634	.28-.99)	82.5

NB. Test 6 from the Performance Matrix was not included in the study, for reasons explained in the methodology section of the text

For Rater 2, Kappa values ranged from -0.1-1.0, with 35/75 criteria with κ>0.70 and seven of the nine tests containing values of 1.0. Overall percentage values ranged from 0.75 to 100% (mean 93.9%). Figure 3 illustrates these percentage intra-rater results clearly, with Rater 1 values close to or on the outer edge of the radar graph and those of Rater 2 slightly more towards the centre, while those for inter-rater and then video versus real-time becoming more central, indicating less reliability.

Real-time versus video agreement

Results for agreement between real-time and video ratings are presented in Table 8. Percentage agreement ranged from 30.8 to 100% (74.5%). Figure 3 highlights the low agreement of specific test criteria, for example 10D, which showed percentage agreement of 30.8% with correspondingly low Kappa coefficients on both the right (0.04) and left (0.04) sides (Table 8). Kappa values

ranged from -0.1 to 1, with only 12 of the 75 criteria with κ>0.60 and only four κ>0.70.

Discussion

The results of the present study indicate that reliability of The Foundation Matrix movement assessment system was generally acceptable. The present findings show variable reliability, with intra-rater reliability for Rater 1 being highest, then intra-rater reliability for Rater 2, then inter-rater reliability, with comparison between video and real-time assessments showing less robust reliability for some criteria.

The three statistical analyses (Kappa, ICC, percentage agreement) and ways of managing data (group and individual criteria) produced different levels of agreement but had been used in similar studies of movement tests (see section on comparison with other movement screening reliability studies), enabling comparison
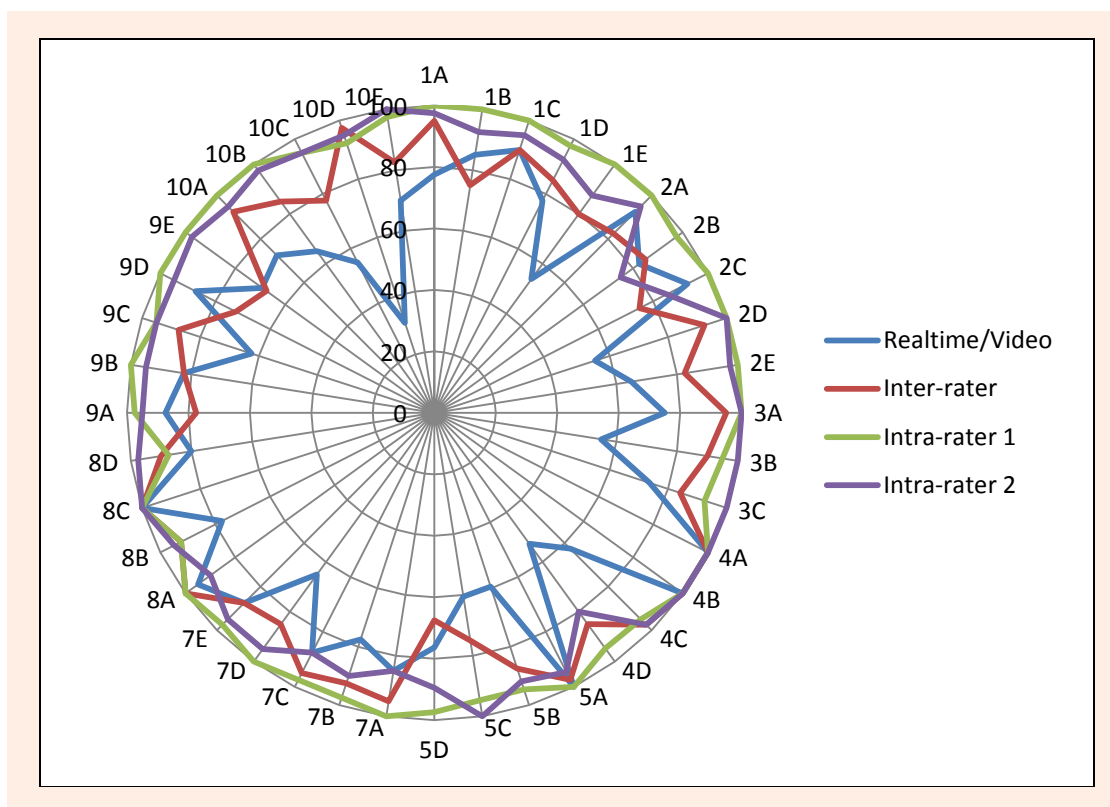
Figure 3. Radar graph of percentage agreement, indicating intra and inter-rater reliability and agreement between real-time and video ratings.

with earlier studies. The ICC values for each test overall, gave the strongest reliability results, followed by percentage agreement and then Kappa, the suitability of which is questioned for the type of data being examined. As discussed below, overall analyses of tests do not allow weaknesses in specific test criteria to be revealed, as illustrated by the present study, which provides the opportunity to modify the assessment tool to make it more robust.

Inter-rater reliability

Agreement between the two raters real-time was generally good, although the strength of agreement depended on which analysis method was used, i.e. overall analysis was excellent with ICC = 0.81 (Table 4) and 86.5% for percentage agreement, whereas analysis of individual criteria produced lower percentage scores (although only 2/40 criteria had <70% agreement) and Kappa values that indicated the majority had moderate to poor agreement, with approximately 50% of the values κ<0.60 (Table 5).

The more detailed analyses enable the criteria to be scrutinized to determine aspects contributing to relative weakness in the overall scores in Table 3. For example, Test 9, the split squat with fast feet change (Figure 2) showed the lowest percentage agreement (78%). When assessing an individual performing a high speed manoeuvre, such as the split squat with fast feet change, it becomes more challenging to observe deviations from the standardised criteria. A reliability study of another screening tool also found that more complex movements are more difficult to judge consistently (Reid et al., 2014). Indeed, the more experienced rater (Therapist 1) showed slightly better intra-rater reliability on this test than Ther-

apist 2 (98% and 96% agreement respectively; Tables 6 & 7). Some of the criteria for this test focused on the control of the alignment of the lower limb; "Can the foot be prevented from turning out or the heel from pulling in as they land?" and "Can the heel of the front foot be prevented from rolling out (inversion) during the heel lift?" Such observations may be difficult to detect, especially as some movement would occur while the limb is stabilising.

The study used three processes to help increase consistency in evaluating these more difficult movement control tasks. Firstly, the testers had undergone the same training in the delivery and scoring of the movement control tests. Secondly, each movement control task was taught to the participant the same way and the same familiarisation process was followed. Thirdly, each movement control task is a specific movement pattern to be performed the same way by all participants. The participants were instructed to perform the task to a standard benchmark level. The test is not an evaluation of their natural or habitual movement pattern or strategy. Although the marking criteria regarding such observations are very precise, it may be necessary to modify the instructions to the therapist and state a time point by which stabilisation should have occurred.

Intra-rater test re-test reliability

Intra-rater agreement was excellent for overall analyses for both raters (ICC 0.96 Rater 1; 0.88 Rater 2; percentage agreement 97.5% Rater 1; 93.9% Rater 2). As for inter-rater reliability, strength of agreement was lower for the analysis of individual criteria, although this was still good (Tables 6 and 7).

Table 6. Rater 1 intra-rater agreement: Cohen's kappa with 95% confidence intervals and criteria percentage agreement (n = 20 except *n = 19, ** n= 18).

Test	Criteria	Left-hand side Kappa	(95% CI)	Right-hand side Kappa	(95% CI)	Criteria agreement (%)
1	A	1.00	(1.00-1.00)	1.00	(1.00-1.00)	100.0
	B	1.00	(1.00-1.00)	1.00	(1.00-1.00)	100.0
	C	1.00	(1.00-1.00)	1.00	(1.00-1.00)	100.0
	D	1.00	(1.00-1.00)	.83	(.50-1.00)	97.5
	E	1.00	(1.00-1.00)	1.00 *	(1.00-1.00)	100.0
2	A	N/A	N/A	N/A	N/A	100.0
	B	1.00	(1.00-1.00)	.83	(.50-1.00)	97.5
	C	N/A	N/A	N/A	N/A	100.0
	D	1.00	(1.00-1.00)	1.00	(1.00-1.00)	100.0
	E	1.00	(1.00-1.00)	1.00	(1.00-1.00)	100.0
3	A	1.00 (1.00-1.00)				100.0
	B	.88 (.64-1.00)				95.0
	C	.64	(.01-1.00)	.64	(.01-1.00)	.64
4	A	1.00 (1.00-1.00)				100.0
	B	N/A	N/A	N/A	N/A	N/A
	C	.90	(.70-1.00)	.90	(.70-1.00)	.90
	D	.83 *	(.50-1.00)	.83 *	(.50-1.00)	.83 *
5	A	N/A	N/A	N/A	N/A	100.0
	B	.89 *	(.69-1.00)	.88 **	(.65-1.00)	94.6
	C	.83 *	(.50-1.00)	.87 *	(.63-1.00)	94.7
	D	1.00*	(1.00-1.00)	.00 *	N/A	97.4
7	A	1.00	(1.00-1.00)	1.00	(1.00-1.00)	100.0
	B	1.00	(1.00-1.00)	0.83	(.50-1.00)	97.5
	C	1.00	(1.00-1.00)	0.64	(.01-1.00)	97.5
	D	1.00 (1.00-1.00)				100.0
	E	1.00	(1.00-1.00)	0.90	(.71-1.00)	97.5
8	A	1.00 (1.00-1.00)				100.0
	B	.79	(.52-1.00)	.79	(.52-1.00)	.79
	C	N/A	N/A	N/A	N/A	N/A
	D	.61	(.11-1.00)	.61	(.11-1.00)	.61
9	A	1.00	(1.00-1.00)	.89	(.69-1.00)	97.5
	B	1.00	(1.000-1.00)	1.00	(1.00-1.00)	100.0
	C	.90	(.70-1.00)	.89	(.69-1.00)	95.0
	D	1.000	(1.00-1.00)	1.00 *	(1.00-1.00)	100
	E	1.000	(1.00-1.00)	1.00 *	(1.00-1.00)	100
10	A	1.00	(1.00-1.000)	1.00	(1.00-1.00)	100.0
	B	1.00	(1.00-1.000)	1.00	(1.00-1.00)	100.0
	C	.89	(.69-1.000)	.90	(.70-1.00)	95.0
	D	.88	(.64-1.000)	.76 *	(.45-1.00)	92.3
	E	1.00	(1.00-1.000)	.90	(.70-1.00)	97.5

When examining reliability of each of the criteria, intra-rater reliability was more robust than that between the two raters, as evidenced in (Figure 3). It is clear that some assessment criteria could be revised, to ensure greater consistency in terms of agreement. Perhaps some criteria need greater clarification regarding what constitutes a pass or fail are determined, i.e. there may need to be additional benchmarks for some of the scoring guidelines. For example, in Test 4 that assesses the control of shoulder movement in the 4 point arm reach, palpation would aid evaluation of winging of the scapula and control of the humeral head. Observation of these may also be difficult if there is additional recruitment of the anterior musculature (pectorals). With the high threshold tests, for example the split squat Test 9, one criterion is to determine if side bending of the trunk can be prevented. However, with such a dynamic test it may take a few seconds to stabilise, so loss of control would be determined by an inability to stabilise and then control side bending. This should also be considered when reflecting on the lateral stair hop (Test 10) and the ability to prevent the body from leaning forward on landing. However, a small degree of flexion would be expected but the key is the ability to correct this to then return to an upright position once landed. It may therefore be useful to have a brief period, perhaps 2-3 seconds, to allow for stabilisation before grading, e.g. in the plank and lateral twist (Test 7), and then this would ensure the scoring is made at that time point.

Real-time versus video agreement

Agreement between real-time and video assessment was lower than that for the other scenarios (ICC 0.23; overall percentage agreement 74.5%), with criteria showing lower Kappa scores and percentage agreement (Table 8). There were seven criteria that returned κ=N/A, which corresponded to 100% agreement for those criteria.

The purpose of the video recordings was to standardise the experimental conditions for examining intra-rater reliability on different occasions. However,

Table 7. Rater 2 intra-rater agreement: Cohen's Kappa with 95% confidence intervals and overall criteria percentage agreement (n = 20 except *n = 19, **n = 18).

Test	Criteria	Left-hand side Kappa (95% CI)		Right-hand side Kappa (95% CI)		Criteria agreement (%)
1	A	.83	(.50-1.00)	1.00	(1.00-1.00)	97.5
	B	.89	(.69-1.00)	.76	(.45-1.00)	92.5
	C	.90	(.71-1.00)	.90	(.71-1.00)	95.0
	D	.79	(.52-1.00)	.89	(.69-1.00)	92.5
	E	.90	(.70-1.00)	.60	(.26-.94)	87.5
2	A	.86	(.59-1.00)	.83	(.50-1.00)	95.0
	B	.69	(.38-1.00)	.31	(-.08-.71)	75.0
	C	.70	(.39-1.00)	.68	(.35-1.00)	85.0
	D	N/A	N/A	N/A	N/A	100.0
	E	N/A	N/A	.00	N/A	97.5
3	A	1.00 (1.00-1.00)				100.0
	B	N/A N/A				100.0
	C	N/A	N/A	N/A	N/A	N/A
4	A	1.00 (1.00-1.00)				100.0
	B	N/A	N/A	N/A	N/A	N/A
	C	1.00	(1.00-1.00)	1.00	(1.00-1.00)	1.00
	D	.60	(.28-.92)	.60	(.28-.92)	.60
5	A	.64 *	(.00-1.00)	.77 *	(.35-1.00)	94.7
	B	.83 *	(.50-1.00)	.44 **	(-.21-1.00)	91.9
	C	N/A	N/A	N/A	N/A	100.0
	D	-.08 *	(-.23-.08)	.83*	(.50-1.00)	89.5
7	A	.69	(.29-1.00)	.53	(.14-0.92)	85.0
	B	.00	N/A	.00	N/A	90.0
	C	.79	(.53-1.00)	.67	(.34-1.00)	87.5
	D	0.77 (0.35-1.00)				95.0
	E	1.00	(1.00-1.00)	.69	(.31-1.00)	95.0
8	A	-0.053 (-0.18-0.07)				90.0
	B	1.00	(1.00-1.00)	1.00	(1.00-1.00)	1.00
	C	N/A	N/A	N/A	N/A	N/A
	D	.00	N/A	.00	N/A	.00
9	A	.89	(.67-1.00)	.88	(.64-1.00)	95.0
	B	.88	(.64-1.00)	.88	(.64-1.00)	95.0
	C	.77	(.35-1.00)	.77	(.35-1.00)	95.0
	D	.83	(.50-1.00)	.83	(.50-1.00)	95.0
	E	1.00	(1.00-1.00)	.64	(.01-1.00)	97.5
10	A	.89 *	(.69-1.00)	.90	(.71-1.00)	94.9
	B	.64	(.01-1.00)	N/A	N/A	97.5
	C	.64	(.01-1.00)	.64	(.01-1.00)	95.0
	D	.64	(.01-1.00)	.00	N/A	95.0
	E	1.00	(1.00-1.00)	N/A	N/A	100.0

However, visibility of some test movements on the video recordings may not have been as clear compared to the real time scoring. The video cameras were set up to provide views from all sides but still this does not enable the freedom to observe from all angles and distances. Percentage agreement for Criterion 1e was 54% for real-time / video agreement and 80% for inter-rater and 100% for intra-rater (Therapist 1) and 88% (Therapist 2). The criterion scored was: "Can you prevent the big toe from lifting as the knee swings 20 degrees (inversion)?" Interpretation of this may have been different in real time, as therapists were able to move around the participant to gain the best view to observe inversion and this may have been limited on the videos.

Comparison with other movement screening reliability studies

Some examples of reliability studies using various movement screening tools are discussed here briefly. Reliability of the Functional Movement Screen (FMS) has been examined between and within raters in various stud-

ies, using the 21 point screen (e.g. Gribble et al., 2013; Onate et al., 2012; Parenteau-G et al., 2014; Schneiders et al., 2011; Smith et al., 2013; Teyhen et al., 2012). The literature generally reports substantial to excellent reliability for the FMS between raters, using videotaped or real time scoring methods. However, direct comparison between studies is limited due to differences in study designs, test viewing methods and statistical analyses used, and no study has compared real-time with video scores. For example, Schneiders et al., (2011) recorded excellent inter-rater reliability for the composite FMS testing (ICC 0.971), while the inter-rater reliability for the individual test components of the FMS tests demonstrated substantial to excellent agreement with Kappa ranging from 0.70-1.0. Onate et al. (2012) found excellent inter-rater reliability (ICC 0.98) for real-time scoring by one certified FMS specialist and a FMS novice. Teyhen et al., (2012) recorded an ICC of 0.76 for inter-rater reliability between four FMS trained physical therapy doctoral students. Smith et al., (2013) used two sessions where FMS was scored live and the inter-rater reliability was good for

Table 8. Real-time versus video agreement: Cohen's Kappa with 95% confidence intervals and criteria percentage agreement; (n = 20, except * n = 19, ** n = 18).

Test	Criteria	Left-hand side Kappa	(95% CI)	Right-hand side Kappa	(95% CI)	Criteria agreement (%)
1	A	.55	(.19-.90)	.43	(.01-.85)	77.5
	B	.67	(.34-1.00)	.69	(.38-1.00)	85.0
	C	.69	(.29-1.00)	.69	(.29-1.00)	90.0
	D	.46	(-.14-1.00)	.34	(.04-.64)	77.5
	E	.13	(-.16-.46)	.23 *	(-.11-.57)	53.8
2	A	.00	N/A	N/A	N/A	92.5
	B	.38	(-.12-.87)	.318	(-.27-.90)	82.5
	C	.00	N/A	.000	N/A	92.5
	D	.14	(-.16-.43)	.259	(-.08-.60)	55.0
	E	.00	(-.40-.40)	.529	(.14-.92)	65.0
3	A	.31 (-.14-.75)				75.0
	B	.10 (-.28-.48)				55.0
	C	-.08	(-.24-.08)	-.08	(-.24-.08)	-.08
4	A	1.00 (1.00-1.00)				100.0
	B	N/A	N/A	N/A	N/A	N/A
	C	.26	(-.12-.63)	.26	(-.12-.63)	.26
	D	.24 *	(-.02-.50)	.24 *	(-.02-.50)	.24 *
5	A	N/A*	N/A	N/A	N/A	100.0
	B	.18 *	(-.16-.53)	.11 **	(-.33-.55)	59.5
	C	-.02 *	(-.46-.42)	.03 *	(-.43-.48)	60.5
	D	-.09 *	(-.25-.08)	.22 *	(-.15-.58)	76.3
7	A	.00	N/A	.000	N/A	85.0
	B	-.11	(-.29-.06)	-.14	(-.33-.06)	77.5
	C	.64	(.01-1.00)	.27	(-.17-.71)	87.5
	D	.38 (.08-.67)				65.0
	E	.78	(.50-1.00)	.69	(.38-1.00)	87.5
8	A	0.00 N/A				95.0
	B	.58	(.22-.95)	.58	(.22-.95)	.58
	C	N/A	N/A	N/A	N/A	N/A
	D	.83	(.50-1.00)	.83	(.50-1.00)	.83
9	A	.73	(.39-1.00)	.66	(.31-1.00)	87.5
	B	.58	(.22-.95)	.70	(.38-1.00)	82.5
	C	.49	(.11-.88)	.07	(-.28-.43)	62.5
	D	.35	(-.17-.86)	.46 *	(-.14-1.00)	87.2
	E	.42	(.06-.78)	.38 *	(.01-.76)	69.2
10	A	.50	(.05-.95)	.00	N/A	72.5
	B	-.07	(-.48-.35)	.21	(-.24-.67)	65.0
	C	.03	(-.37-.43)	.20	(-.22-.62)	55.0
	D	.03	(-.04-.11)	.04 *	(-.04-.12)	30.8
	E	.38	(-.02-.78)	.38	(-.02-.78)	70.0

session one and two with ICCs of 0.89 and 0.87 respectively, whilst intra-rater reliability was good for each rater with an ICC from 0.81-0.91. Gribble et al., (2013) scored three video-taped subjects on two separate testing sessions one week apart and found excellent intra-rater reliability (ICC 0.96) for trained athletic trainers (ATC) with greater than six months FMS experience (n=7), an intra-rater reliability (ICC 0.77) for trainers with less than six months FMS experience (n=15), and poor intra-rater reliability (ICC 0.37) for ATC students (n=16). However, comparing scores across all participants, without classification of experience, an intra-rater ICC of 0.75 was produced (Gribble et al., 2013) but the number of subjects in that study was small (n=3). The overall ICC of 0.81 for inter-rater reliability in the present was lower than some FMS studies and higher than others. The present intra-rater reliability ICCs for the two raters (0.96 and 0.88) were comparable with the FMS findings for experienced raters and higher than less experienced raters.

Frohm et al., (2012) evaluated the inter- and intra-rater reliability of the Nine-Test Screening Battery to screen athletic movement patterns and found no significant difference (p = 0.31) between eight physiotherapists assessing 26 participants on two test occasions, with ICCs of 0.80 and 0.81, respectively. Using the same Screening Battery, Rafnsson et al., (2014) found excellent intra-rater reliability (ICC 0.95) between test-retest sessions. The retest had significantly higher total score than the first test (p = 0.041), indicating some learning effect by the participant.

The Single-limb Mini Squat (SLMS) is used to distinguish between those with knee-over-foot and knee-medial to-foot positioning. Ageberg et al., (2010) found no difference between examiners (p = 0.317) who assessed 25 participants, with a Kappa value of 0.92 (95% CI, 0.75 to 1.08) and 96% agreement between examiners.

Other considerations

Sample sizes for reliability testing vary widely between studies. A sample of 20 is recommended as the minimum for reliability testing (Atkinson and Nevill, 2001; Walter et al., 1998), so the present sample size was adequate.

It has been acknowledged previously that achieving good reliability from visual information can be difficult (Luomajoki et al., 2007), although it has also been suggested that with sufficient training on each test, such judgments can be enhanced. It is worth considering that the more experienced therapists in the present study and that by Gribble et al., (2013) demonstrated higher intra-rater reliability than the less experienced therapists, which is an observation consistent with other reliability studies (Luomajoki et al., 2007).

There are situations where it is not possible to assess agreement using Kappa (κ), and the phenomenon of high agreement but low Kappa scores was discussed by Cicchetti and Feinstein, (1990). In the present study, for example, if both therapists score each participant "0", κ is undefined; this is because there are no data relating to "1", and so agreement cannot be assessed across all outcomes (0 and 1). In such cases, κ was presented as "N/A" (although it is worth noting these cases always had 100% agreement on one level of the outcome, and no information regarding agreement on the other level of the outcome). There were also situations where κ was zero or negative, suggesting that agreement was no better or worse than agreement expected by chance alone, respectively. We believe this is not necessarily a true reflection of the agreement of the test criterion itself, but rather a reflection of the limited sample size – again, these situations arise when there is very little data relating to one of the outcomes (0 or 1). For example, Table 7 shows the results for criterion 8a in the intra-rater assessment for Rater 2. We observe κ = -0.05 (95% CI [-0.18, 0.07]) for this criterion, i.e. agreement is worse than that by chance, or at best very slightly better than chance, but also observe 90% agreement. The reason for low Kappa is obvious when we consider the cases where the rater scores 0: on both first and second viewings, the rater scores a single 0, but these do not coincide; therefore, we have total disagreement whenever the rater scores 0. Further investigation is needed to establish agreement adequately in such a situation.

Limitations of the study

Factors related to the research design may have impacted on the scoring of the tests. For example, participants were asked to repeat the test up to three times to allow for the two therapists to move around freely to score their ratings and to ensure the test had been recorded. This repetition may have allowed improvement in performance but there was no observed change after the initial practice sessions as described in the protocol above. When examining inter-rater reliability, it is important that each criterion (Table 2) is marked in the same order and therefore at the same time point, especially on those tests that have a greater number of marking criteria. For future research, it is recommended that: a) the number of repeat performances is standardised; and b) both therapists observe the same area of the body, and score at the same time point. Only two experienced observers (physiotherapists) were

involved in the present study. For the findings to be generalizable, more observers with different levels of experience and from a range of professions need to be studied. The lack of agreement between the real-time and video scores indicates that video evaluation of movement tasks may have increased the ability to judge movement control impairments more consistently. This may have been because evaluation was made from four standardised views i.e. the four video cameras as opposed to the therapist moving around the participant whilst performing the test. For future studies comparing real-time with video evaluation, the criteria should be scored by observing from the same angle of view for the two types of assessment.

For the purposes of large epidemiological studies that require retrospective analysis to save time during testing, the most robust tests and criteria would need to be determined. The videos provided a valid means of presenting consistent performance of tests by the participants to assess intra-rater reliability of the observers.

Future research

Acceptable inter and intra-rater reliability has now been established for the tests selected from The Foundation Matrix and has indicated specific criteria that require modification to the scoring benchmarks and pass or fail guidelines (Figure 3) to improve the robustness of the tool. Reliability of other tests in The Performance Matrix needs to be established in experienced and novice therapists. The validity of the screening tool against objective markers also needs to be established, as well as sensitivity to change. An example of validity and responsiveness was demonstrated for a shoulder movement control test (from the database of tests available in The Performance Matrix series of movement screens) that was examined objectively using 3-dimensional motion analysis. That study found abnormal scapular movements in people with shoulder impingement, which improved (kinematics and recruitment) following a motor control retraining programme (Worsley et al., 2013). More comprehensive validity and reliability testing of The Performance Matrix screening system is warranted.

Conclusion

The Foundation Matrix has demonstrated good inter-rater reliability and excellent intra-rater reliability for two experienced therapists. Agreement for real-time versus video assessment was generally moderate to poor. The findings indicate the clinical utility of this screening tool in identifying performance assets and uncontrolled movements. Recommendations have been made for refining the criteria and number of repetitions of some tests to improve reliability.

References

Ageberg, E., Bennell, K.L., Hunt, M.A., Simic, M., Roos, E.M. and Creaby, M.W. (2010) Validity and inter-rater reliability of me-dio-lateral knee motion observed during a single-limb mini squat. *BMC Musculoskeletal Disorders* **11,** 265-265.

Atkinson, G., and Nevill, A. (2001) Selected issues in the design and analysis of sport performance research. *Journal of Sports Science* **19,** 811-827.

Bahr, R. and Krosshaug, T. (2005) Understanding injury mechanisms: a key component of preventing injuries in sport. *British Journal of Sports Medicine* **39,** 324-329.

Butler, R.B., Plisky, P., Southers, C., Scoma, C. and Kiesel, B.K. (2010) Biomechanical analysis of the different classification of the functional movement screen deep squat test. *Sports Biomechanics* **9(4),** 270-279.

Butler, R.B., Plisky, P. and Kiesel, B.K. (2012) Interrater reliability of videotaped performance of the functional movement screen us-ing the 100 – point scoring scale. *Athletic Training & Sports Health Care* **4(3),** 103-109.

Chalmers, D.J. (2002) Injury prevention in sport: not yet part of the game? *Injury Prevention* **8,** 22-25.

Cicchetti, D.V. and Feinstein, A.R. (1990) High agreement but low kappa: II. Resolving the paradoxes. *Journal of Clinical Epidemiology* **43,** 551-558.

Cohen, J. (1960) A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* **20,** 37-46.

Comerford, M. (2006) Screening to identify injury and performance risk: movement control testing- the missing piece of the puzzle. *Sport Exercise Medicine* **29,** 21-26.

Comerford, M.J. and Mottram, S.L. (2001) Movement and stability dysfunction – contemporary developments. *Manual Therapy* **6(1),** 15-26.

Comerford, M. and Mottram, S. (2012) *Kinetic Control. The Management of uncontrolled movement.* Elsevier, Sydney.

Cook, G., Burton, L. and Hoogenboom, B. (2006a) Pre- participation screening: the use of fundamental movements as an assessment of function-part 1. *North American Journal of Sports Physical Therapy* **1,** 62-72.

Cook, G., Burton, L. and Hoogenboom, B. (2006b) Pre- participation screening: the use of fundamental movements as an assessment of function- part 2. *North American Journal of Sports Physical Therapy* **1,** 132-139.

Ekegren, C.L., Miller, W.C., Celebrini, R.G., Eng, J.J. and Macintyre, D.L. (2009). Reliability and validity of observational risk screening in evaluating dynamic knee valgus. *Journal of Orthopaedic & Sports Physical Therapy* **39(9),** 665-674.

Fersum, K.V., O'Sullivan, P.B., Kvale, A. and Skouen, J.S. (2009) Inter-examiner reliability of a classification system for patients with non-specific low back pain. *Manual Therapy* **14,** 555-561.

Fleiss, J (2007). Reliability of measurement in the Design and Analysis of Clinical Experiments. New York: Wiley and Sons pp 1-33.

Frohm, A., Heijne, A., Kowalski, J, Svensson, P. and Myklebust, G. (2012) A nine-test screening battery for athletes: a reliability study. *Scandinavian Journal of Medicine & Science in Sports* **22(3),** 306-315.

Fulton, J.,Wright, K., Kelly, M., Zebrosky, B., Zanis, M., Drvol, C., Butler, R. (2014) Injury risk is altered by previous injury: a sys-tematic review of the literature and presentation of causative neuromuscular factors. *International Journal of Sports Physical Therapy* **9(5),** 583-595

Garrick, J.G. (2004) Preparticipation orthopaedic screening evaluation. *Clinical Journal Sports Medicine* **14(3),** 123-126.

Gribble, P.A., Brigle, J., Pietrosimone, B.G., Pfile, K.R. and Webster, K.A. (2013) Intrarater reliability of the functional movement screen. *Journal of Strength and Conditioning Research* **27(4),** 978-981.

Hodges, P.W. and Smeets, R.J. (2015) Interaction between pain, move-ment and physical activity: Short-term benefits, long-term con-sequences, and targets for treatment *The Clinical Journal of Pain* 31(2), 97-107.

Hodges, P.W. and Richardson, C.A. (1996) Inefficient muscular stabili-sation of the lumbar spine associated with low back pain. *Spine* **21,** 2640-2650.

Kiesel, K., Plisky, P. and Voight, M.L. (2007) Can serious injury in professional football be predicted by a preseason functional movement screen? *North American Journal of Sports Physical Therapy* **2(3),** 132-139.

Kiesel, K., Plisky, P. and Butler, R. (2009) Functional movement test scores improve following a standardised off-season intervention program in professional football players. *Scandinavian Journal of Medicine & Science in Sports* **29,** 287-292.

Luomajoki, H., Kool, J., Bruin, E.D. and Airaksinen, O. (2007) Reliabil-ity of movement control tests in the lumbar spine. *BMC Musculoskeletal Disorders* **8,** 90.

Luomajoki, H., Kool, J., Bruin, E.D. and Airaksinen, O. (2010) Im-provement in low back movement control, decreased pain and disability, resulting from specific exercise intervention. *Sports Medicine, Arthroscopy, Rehabilitation, Therapy and Technology* **23(2),**11

MacDonald, D., Moseley, G.L. and Hodges, P.W. (2009) Why do some patients keep hurting their back? Evidence of on going back muscle dysfunction during remission from recurrent back pain. *Pain* **142(3),** 183-188.

McBain, K., Shrie, I., Shultz, R., Meeuwisse, W.H., Klügl, M., Garza, D. and Matheson, G. (2012) Prevention of sports injury I: a sys-tematic review of applied biomechanics and physiology out-comes research. *British Journal of Sports Medicine* **46,** 169-173.

McNeill, W. (2014) The double knee swing test - a practical example of The Performance Matrix Movement Screen. *Journal Bodywork & Movement Therapies* **18(3),** 477-481.

Mottram, S. and Comerford, M. (2008) A new perspective on risk assessment. *Physical Therapy in Sport* **9,** 40-51.

Myer, G.D., Ford, K.Y., Paterno, M.V., Nick, T.G. and Hewett, T.E. (2008) The effects of generalised joint laxity on risk of anterior cruciate ligament injury in young female athltetes. *American Journal of Sports Medicine* **36(6),** 1073-1080.

Onate, J.A., Dewey, T., Kollock, R.O., Thomas, K.S., Van Lunen, B.L., De Maio, M. and Ringleb, S.I. (2012) Real-time intersession and interrater reliability of the functional movement screen. *Journal of Strength & Conditioning Research* **26(2),** 408-415.

O'Sullivan. (2005) Diagnosis and classification of chronic low back pain disorders: Maladaptive movement and motor control impair-ments as underlying mechanism. *Manual Therapy* **10,** 242-255.

Parenteau-G, E., Gaudreaulta, N., Chambersa, S., Boisverta, C., Greniera, A., Gagnéa, G. and Balgb, F. (2014) Functional movement screen test: A reliable screening test for young elite ice hockey players. *Physical Therapy in Sport* **15,** 169-175

Peate, W.F., Bates, G., Lunda, K., Francis, S. and Bellamy, K. (2007) Core strength: A new model for injury prediction and preven-tion. *Journal of Occupational Medicine and Toxicology* **2,** 3.

Plisky, P.J., Rauh, M.J, Kaminski, T.W. and Underwood F.B. (2006) Star Excursion Balance Test as a predictor of lower extremity injury in high school basketball players. *Journal of Orthopae-dic Sports Physical Therapy* **36(12),** 911-919.

Rafnsson, E., Frohm, A., Myklebust, G., Bahr, R., Valdimarsson, Ö. and Árnason, Á. (2014) Nine Test Screening Battery- Intra-rater Reliability and Screening on Icelandic Male Handball Players. *British Journal of Sports Medicine* **48(7),** 674.

Reid, D., Vanweerd, R., Larmer, P. and Kingstone, R. (2014) The inter and intra rater reliability of the Netball Movement Screening Tool. *Journal of Science and Medicine in Sport* (In press).

Roussel, N.A., Nijs, J., Mottram, S., Moorsel, A.V., Truijen, S. and Stassijns, G. (2009) Altered lumbopelvic movement control but not generalised joint hypermobility is associated with increased injury in dancers. A prospective study. *Manual Therapy* **14(6),** 630-635.

Sahrmann, S.A. (2002) *Diagnosis and treatment of movement impair-ment syndromes.* Mosby, St Louis.

Sahrmann, S. (2014) The Human Movement System: Our Professional Identity. *Physical Therapy* **94,** 1034-1042.

Schneiders, G., Davidsson, A., Hörman, E. and Sullivan, S.J. (2011) Functional movement screen normative values in a young, ac-tive population. *International Journal of Sports & Physical Therapy* **6(2),** 75-82.

Smith, C.A., Chimera, N.J., Wright, N.J. and Warren, M. (2013) Inter-rater and intrarater reliability of the functional movement screen. *Journal of Strength & Conditioning Research* **27(4),** 982-987.

Simmonds, J.V. and Keer, R.J. (2007). Hypermobility and hypermobili-ty syndrome. *Manual Therapy* **12,** 298-309.

Sterling, M., Jull. G. and Wright. A. (2001) The effect of musculoskeletal pain on motor activity and control. *The Journal of Pain* **2(3),** 135-145.

Teyhen, D.S., Shaffer, S.W., Lorenson, C.L., Halfpap, J.P., Donofry, D.F., Walker, M.J., Dugan, J.L. and Childs, J.D. (2012) The functional movement screen: A reliability study. *Journal of Orthopaedic & Sports Physical Therapy* **42(6),** 530-537.

Turbeville, S., Cowan, L. and Owen, W. (2003) Risk factors for injury in high school football players. *American Journal of Sports Medicine* **31(6),** 974-980.

Tyler, T.F., McHugh, M,P., Mirabella, M.R., Mullaney, M.J. and Nicholas, S.J. (2006) Risk factors for noncontact ankle sprains in high school football players: the role of previous ankle sprains and body mass index. *American Journal of Sports Medicine* **34(3),** 471-475.

Walter, S.D, Eliasziw, M. and Donner A. (1998) Sample size and optimal design for reliability studies. *Statistics in Medicine* **17,** 101-110.

Webborn, N. (2012) Lifetime injury prevention: the sport profile model. *British Journal of Sports Medicine* **46,** 193-197.

Worsley, P., Warner, M., Mottram, S., Gadola, S., Veeger, H.E.J., Hermens, H., Morrissey, D., Little, P., Cooper, C., Carr, A. and Stokes M. (2013) Motor control retraining exercises for shoulder impingement: effects on function, muscle activation and biomechanics in young adults. *Journal of Shoulder and Elbow Surgery* **22(4),** e11-e19.

Yeung, S.S., Suen, A.M. and Yeung, E.W. (2009) A prospective cohort study of hamstrings injuries in competitive sprinters: Preseason muscle imbalance as a possible risk factor. *British Journal of Sports Medicine* **(45),** 589-594.

Zazulak, B.T., Hewett, T.E, Reeves, N.P., Goldberg, B. and Cholewicki, J. (2007) Deficits in neuromuscular control of the trunk predict knee injury risk. *The American Journal of Sports Medicine* **35(7),** 1123-1130.

Key points

- The movement control tests of The Foundation Matrix had acceptable reliability between raters and within raters on different days
- Agreement between observations made on tests performed real-time and on video recordings was low, indicating poor validity of use of video recordings
- Some movement evaluation criteria related to specific tests that did not achieve excellent agreement could be modified to improve reliability

AUTHOR BIOGRAPHY

Carolina R. MISCHIATI
Employment
Senior lecturer in Sports Therapy at the University of Chichester.
Degree
BSc
Research interests
Movement and stability dysfunction and the restoration of optimal function, and injury prevention screening.
E-mail: c.mischiati@chi.ac.uk

Mark COMERFORD
Employment
Director of Movement Performance Solutions Ltd and a Visiting Research Fellow at the University of Southampton.
Degree
BSc
Research interests
The development of clinically relevant models of movement control muscle function with the aim of managing pain and enhancing performance.

Emma GOSFORD
Employment
A graduate Sports Therapist.
Degree
BSc
Research interests
Injury prevention and movement screening.

Jacqueline SWART
Employment
A Performance Matrix Accredited Tutor and Consultant and Kinetic Control Accredited Tutor.
Degree
BSc
Research interests
Injury prevention, treating and training athletes and has been using The Performance Matrix system for assessment and successfully integrating Pilates exercise in with athlete's rehabilitation.

Sean EWINGS
Employment
A Research Fellow in the Southampton Statistical Sciences Research Institute.
Degree
PhD
Research interests
The application of statistical methods to health research

Nadine BOTHA
Employment
A Senior Musculoskeletal Physiotherapist in the National Health Service and a National Institute for Health Research Intern.
Degree
MRes
Research interests
Movement control patterns; focusing on conservative management of musculoskeletal conditions including osteoarthritis with an emphasis on the role of exercise in both prevention and management.

Maria STOKES
Employment
Professor of Musculoskeletal Rehabilitation, Faculty of Health Sciences, at the University of Southampton.
Degree
PhD
Research interests
Active living for healthy musculoskeletal ageing and physical management of musculoskeletal disorders.

Sarah L. MOTTRAM
Employment
Director of Movement Performance Solutions Ltd and a Visiting Research Fellow at the University of Southampton.
Degree
MSc
Research interests
The influence of impaired movement control on activity limitations, recurrence of pain and performance.

✉ **Carolina Mischiati**
Department of Sport and Exercise Sciences, University of Chichester, West Sussex, PO19 6PE, UK