



# Inter-Rater reliability of the Functional Movement Screen (FMS) amongst NHS Physiotherapists

Andrew Cuff<sup>1,2\*</sup>, Joe L. Palmer<sup>2,3</sup> and Melanie Lindley<sup>2</sup>

\*Correspondence: [andrewcuff@connecthealth.co.uk](mailto:andrewcuff@connecthealth.co.uk)



CrossMark

Click for updates

<sup>1</sup>Developing Consultant Physiotherapist & Clinical Lead, Wakefield and North Kirklees MSK Services, Connect Health, Trinity Medical Centre, Yorkshire, WF1 1PG, UK.

<sup>2</sup>Sheffield Hallam University, Faculty of Health and Wellbeing, Collegiate Crescent, Sheffield, South Yorkshire, S10 2BA, UK.

<sup>3</sup>Extended Scope Practitioner, Primary Care Sheffield, Darnall Primary Care Centre, Sheffield, South Yorkshire, S9 5DH, UK.

## Abstract

**Objectives:** The primary objective of this study was to determine the inter-rater reliability of the FMS when used by novice NHS (National Health Service) Physiotherapists. The secondary objective was to determine whether clinical specialism has any impact on the inter-rater reliability of the FMS.

**Design:** Reliability study.

**Methods:** Forty participants with no recent MSK injury were video recorded completing the 7 component FMS tests. Six NHS Physiotherapists with no previous experience using the FMS attended a 2 hour training programme delivered by a certified FMS practitioner. Raters then viewed and scored videos of the 40 participants completing the FMS.

**Results:** The inter-rater reliability of the FMS composite score was excellent (ICC of 0.82 (95% CI: 0.41-0.93)). Non-specialist rotational Physiotherapists demonstrated excellent inter-rater reliability (ICC = 0.89, 95%CI 0.78-0.94), whereas the specialist musculoskeletal Physiotherapists demonstrated good inter-rater reliability (ICC = 0.79, 95%CI 0.19-0.92) for FMS composite score. The seven individual movement tests of the FMS demonstrated poor to excellent inter-rater reliability. The Hurdle Step was the least reliable of the movement tests (kw = 0.15, 95% CI: -0.09-0.38), whereas Shoulder Mobility was the most reliable of the movement tests (kw = 0.85, 95%CI: 0.72-0.97). The seven individual movement tests of the FMS demonstrated moderate to excellent inter-rater reliability between non-specialist rotational Physiotherapists. In contrast, the seven individual movement tests of the FMS demonstrated poor to excellent inter-rater reliability between specialist musculoskeletal Physiotherapists.

**Conclusion:** The FMS represents a good attempt to objectify the subjective with the FMS composite score demonstrating excellent inter-rater reliability. Due to poor construct validity, it has been suggested only component scores should be utilised.

The results from this study suggest that five of the seven individual movement tests do not demonstrate acceptable reliability for clinical use.

With the composite score lacking construct validity and the majority of the component scores lacking both intra- and inter-rater reliability, the continued use of FMS within clinical practice is not supported.

**Keywords:** Inter-rater, Reliability, Functional Movement Screen, FMS

## Introduction

Traditional approaches to musculoskeletal examination and treatment have focused on isolated methods such as joint range of movement, muscle strength and muscle length [1]. This isolated approach however, fails to encompass the entire

kinetic or kinematic chain whilst also ignoring the role of the nervous system in movement; 'the brain knows nothing of individual muscle action, but knows only of movement' [2].

Approaches to rehabilitation are currently changing with a move away from attempting to target isolated muscle groups

or joints, in favour of moving towards a more comprehensive analysis and assessment of movement [3]. It has been proposed that rehabilitation approaches that attempt to maximise cortical neuroplastic change, may provide the greatest potential for rehabilitation success [4]. In order to achieve this, cognitive effort and motor skill training of meaningful movements or tasks is required; a potential limitation of the current bed-based methods of assessment and treatment [5].

This has been well demonstrated where static hip extension range measured using the Thomas' Test was not reflective of the peak hip extension seen during running, showing a low correlation between static and dynamic measures of hip extension [6]. It is therefore apparent that what happens during dynamic movement is not a reflection of what happens during clinical testing. Furthermore, improvements in hip flexibility did not transfer to increased mobility during dynamic movement [7]; changes in passive range of motion do not therefore automatically transfer to changes in movement.

An isolated rehabilitation approach following injury is not sufficient and an injury in a single area of the body, can adversely affect regions away from the injury site [8-10]; this is known as regional interdependence. Regional interdependence refers to how or why an injury or habitual movement patterns in one area of the body may be contributing to altered movement patterns in another region of the body [11].

Previous pain or injury can affect future movement due to the proprioceptive representation of the involved body part in the primary sensory cortex changing [12]. This in turn may have future implications for motor output and motor control as these representations are the maps utilised to plan and execute movement; if the representation of a body part is inaccurate, then motor control may be compromised, potentially creating a feedback loop of mutual degradation [12,13].

Utilising standardised, validated outcome measures is an explicit requirement of the Quality Assurance Standards [14] and the HCPC Standards of Proficiency for Physiotherapists [15]. It is therefore apparent that a valid and reliable outcome measure that assesses multiple aspects of function simultaneously is required [8]. The Functional Movement Screen (FMS) is gaining notoriety among clinicians as an outcome assessment tool to quantify movement patterns [3,16] with the developers of the FMS describing the tool as a 'ranking and grading system that documents movement patterns that are deemed key to normal function [17]. The FMS consists of 3 screening tests and 7 component tests; the Deep Squat (DS), Hurdle Step (HS), Inline Lunge (ILL), Shoulder Mobility (SM), Active Straight Leg Raise (ASLR), Trunk Stability Push-UP (TSPU) and Rotary Stability (RS). These are then scored using a categorical scale ranging from 0-3, resulting in a potential overall score ranging from 0-21. The scoring method is described in detail in our previous study [20] and for a detailed description of how each component/ screening movement is performed, the reader is directed to Cook et al (2010) [17].

Prior to considering the validity of an assessment tool, it

is important to first determine whether it is reliable [18]. The FMS consists of three clearing tests to determine the presence of pain, graded as positive or negative and seven individual movement tests that are each graded on a scale of 0-3, with a composite score calculated when all seven movement tests are completed with a total available score of 21 [3]. A score of 3 corresponds with the individual being able to correctly complete the movement without compensation; a score of 2 corresponds with the individual being able to perform the movement with compensation; a score of 1 corresponds with the individual being unable to perform the movement; a score of 0 corresponds with the individual reporting pain at any time when performing the movement [19].

A recent study conducted by Palmer, Cuff and Lindley (2017) questioned the intra-rater reliability of the FMS amongst NHS clinicians; both specialist musculoskeletal (MSK) and rotational Physiotherapists [20]. Furthermore, the inter-rater reliability is unclear with reports ranging from fair-excellent [16,21] and has not been determined amongst UK public sector clinicians [20].

Currently, the inter-rater reliability of the FMS has not been determined amongst clinicians working specifically within the UK public health sector, limiting the generalisability of previous studies where raters have either been Athletic Trainers, Private Physiotherapists or Students. Additionally, only one study has investigated the reliability of the FMS amongst novice raters [22], however, the 20-hour intensive rater training and use of only Physiotherapy students limits application to clinical practice. The samples of populations used in these studies to date are grossly homogenous utilising healthy physiotherapy students [19,21], military recruits [22], or high level athletes [16,23], further limiting application to clinical practice and potentially affecting the reliability coefficient produced [18].

When assessing the reliability of a scale, it is important that intra-rater reliability is determined prior to assessing inter-rater reliability [24]. Inter-rater reliability represents all potential errors encountered within intra-rater reliability as well as the potential for error between raters. Therefore, a scale with excellent inter-rater reliability would be suggestive of high intra-rater reliability however; a scale with excellent intra-rater reliability does not necessarily suggest high inter-rater reliability [24].

The intra-rater reliability of the FMS when used by novice National Health Service (NHS) Physiotherapist raters has been shown to be excellent (ICC 0.91; 0.89-0.93 95%CI) as part of a concurrent study [20].

Therefore, the primary objective of this study was to determine the inter-rater reliability of the FMS when used by novice NHS Physiotherapist raters. The secondary objective was to determine whether clinical specialism has any impact on the inter-rater reliability of the FMS.

## Methods

### Participants

A purposive convenience sample of 40 participants was

obtained through word of mouth, recruitment posters and University email that were recruited over a three-month period from Sheffield, South Yorkshire. Participants were considered eligible for inclusion if they were over 18 years of age, had the ability to, and were willing to adhere to trial procedures. Participants were excluded if they refused participation, reported pain on any of the three FMS clearing tests, answered 'Yes' to any questions on the PAR-Q, had received treatment for, or reported having any musculoskeletal pathology in the last six weeks, were pregnant, had a cardiac medical history, hypertension or neurological impairment (**Table 1**).

**Table 1. Participant Inclusion/ Exclusion Criteria.**

Inclusion	Exclusion
<ul style="list-style-type: none"> <li>• &gt;18 years of age</li> <li>• Ability/ willingness to adhere to trial procedures</li> </ul>	<ul style="list-style-type: none"> <li>• Participant refusal</li> <li>• Answered 'Yes' on any PAR-Q question</li> <li>• Cardiac History</li> <li>• Neurological Impairment</li> <li>• Pregnancy</li> <li>• Dizziness</li> <li>• Hypertension</li> <li>• Received treatment for OR reported having any musculoskeletal pathology within the previous 6 weeks</li> <li>• Reported pain on any of the 3 FMS Clearing Tests</li> </ul>

Potential participants had the study explained to them and were informed of the entrance criteria. Each individual completed a PAR-Q questionnaire to assess safety and suitability for inclusion; the PAR-Q is designed to identify adults for whom physical activity might be inappropriate without seeking prior medical advice [25]. Potential participants were then asked to complete the three FMS clearing tests consisting of the impingement, prone press-up and posterior-rocking clearing tests [17].

Those individuals who met the entrance criteria were then provided with an information sheet and given the opportunity to volunteer, before signing the informed consent form and being entered into the study. Participants provided informed consent for inclusion in both the intra-rater [20] and inter-rater reliability studies.

**Raters**

A purposive judgmental sample of six NHS Physiotherapists without prior FMS experience was obtained through word of mouth and recruitment posters from three NHS trusts in South Yorkshire, West Yorkshire and Derbyshire, UK. The raters consisted of three specialist musculoskeletal and three non-specialist rotational Physiotherapists. Eligible Physiotherapists who met the relevant inclusion/ exclusion criteria according to their level of specialism (**Tables 2 and 3**) completed an

**Table 2. Non-Specialist Raters Inclusion/ Exclusion Criteria.**

Inclusion	Exclusion
<ul style="list-style-type: none"> <li>• Ability to and willing to adhere to trial procedures</li> <li>• Non-specialised rotational Physiotherapists according to NHS Employers (2005) definition</li> <li>• Currently employed in the NHS</li> </ul>	<ul style="list-style-type: none"> <li>• Specialist Physiotherapist across all areas of clinical practice according to NHS Employers (2005) definition</li> <li>• Previous experience of using the FMS in both clinical and non-clinical settings</li> <li>• Non-qualified Physiotherapy staff according to NHS Employers (2005) definition</li> <li>• Unable to attend CPD training session, rating session 1 OR rating session 2</li> <li>• Significant, non-correctable visual impairment</li> </ul>

**Table 3. Specialist Raters Inclusion/ Exclusion Criteria.**

Inclusion	Exclusion
<ul style="list-style-type: none"> <li>• Ability to and willing to adhere to trial procedures</li> <li>• Specialist Musculoskeletal Physiotherapists according to NHS Employers (2005) definition</li> <li>• Currently employed in the NHS</li> </ul>	<ul style="list-style-type: none"> <li>• Specialist Physiotherapist in a non-Musculoskeletal clinical area according to NHS Employers (2005) definition</li> <li>• Previous experience of using the FMS in both clinical and non-clinical settings</li> <li>• Non-qualified Physiotherapy staff according to NHS Employers (2005) definition</li> <li>• Unable to attend CPD training session, rating session 1 OR rating session 2</li> <li>• Significant, non-correctable visual impairment</li> </ul>

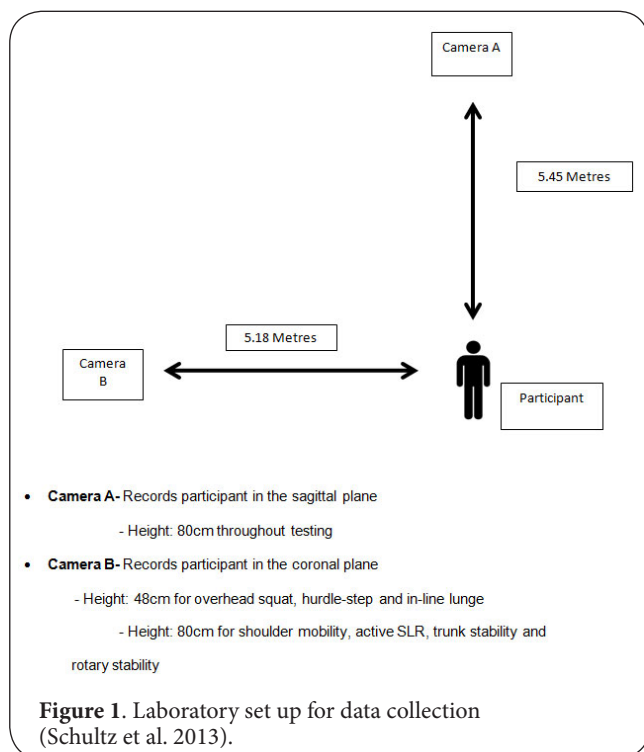
informed consent form and were entered into the study.

**Training**

The six raters attended a two-hour training session in order to introduce and train them in using the FMS. This session was delivered by a certified FMS practitioner and consisted of an introduction to the FMS, the seven individual movement tests and the three relevant modifications for the RS, DS and PU tests. The raters were shown the three FMS clearing tests and informed that these tests were part of the exclusion criteria for eligible participants. The raters were given a detailed explanation of the 0-3 categorical scoring system used for each individual movement test and were shown two example videos to gain familiarity with the FMS; Both videos were in the format to be used in the rating session to ensure familiarisation with study procedures and the scoring system.

## Procedures

Participants were recorded from both sagittal and coronal views using two Sony HDR-XR260 (Sony Corporation, Minato, Tokyo, Japan) camcorders whilst completing the seven individual test movements. Camera placement is illustrated in **Figure 1** and followed the video method previously validated by Shultz et al (2013) [16].



The seven individual test movements were then explained and demonstrated to the participants by a clinician experienced in the use of the FMS using an official FMS Test Kit (Functional Movement Systems Incorporated, Chatham, Virginia, USA).

Each participant was also filmed completing the relevant modifications for the DS, PU and RS tests. Participants repeated each movement test three times in accordance with the FMS protocol as described by Cook et al (2010) [17]; participants were not coached during their movements.

Raw video footage was then edited using Windows Live Movie Maker (Microsoft Corporation, Redmond, Washington, USA) to ensure that the sagittal views were followed by the coronal views of the same movement test. Any pertinent information required by the raters was added to the footage i.e., participant number, movement test being performed, pain reporting and hand span measurement during the SM test. Example screen shots of the videos shown to raters are displayed as Figures in Palmer Cuff and Lindley (2017) [20].

## Rating sessions

Each rater was set up on two individual computers. Computer

one showed a continuous video of the all (non-modified) FMS movements for each participant; Computer two contained the modifications of the DS, PU and RS tests. This was done to allow the raters to easily transition between the continuous (non-modified) FMS video and viewing the modification videos when applicable. The rater was instructed to only view the DS, PU or RS modification video if they felt that the participant did not score a '3' in that movement test. To prevent raters from analysing still images of FMS component tests, they were only permitted to pause the continuous FMS videos during introductory/ transitional screens [20]. This was to allow for breaks as required and to allow transition to observe modification videos on computer two.

Raters completed two rating sessions, two weeks apart. This 2 week wash-out period was used to minimise the potential for recollection of previous scoring which could have biased results. To further minimise bias, the order in which raters viewed the participant's videos between the two rating sessions was randomised using online software [26] and raters were blinded to their previous scores as well as those of the other raters.

## Statistical analysis

Weighted kappa ( $k_w$ ) statistic with quadratic weighting and 95% Confidence Interval (95% CI) was used [27] to determine inter-rater reliability of components scores.

The inter-rater reliability of composite FMS scores was assessed using an Intraclass Correlation Coefficient (ICC) and 95% CI and were assessed for normal distribution using the Kolmogorov-Smirnoff test [27].

Data was analysed using MedCalc for Windows (Version 10.4.0.0; MedCalc Software, Mariakerke, Belgium) and SPSS 21.0 (SPSS, Inc., Chicago, IL, USA). Inter-rater reliability was interpreted based on Landis and Koch (1977) criteria: Excellent reliability, 0.81-1.0; Good reliability, 0.61-0.80; Moderate reliability, 0.41-0.60; Fair reliability, 0.21-0.40; Poor reliability,  $\leq 0.20$  [28].

## Results

Forty participants (20 Males, 20 Females) met inclusion/exclusion criteria and completed the study. Three participants (2 Males, 1 Female) withdrew from the study before rating had commenced due to scheduling difficulties. The mean age of the participants was 28.9 (Range 18.9-60.3; SD +/-11.65) years. Participants included demonstrated a wide range of routine physical activity on University of California at Los Angeles (UCLA) Activity Scoring [29] (range 3-10; mean 7.75 +/-2.4) (Table 4).

The total FMS scores of all participants were within a range of 5-19. The Kolmogorov-Smirnoff test showed that the FMS composite scores were not normally distributed across the six raters ( $p < 0.001$ ).

The inter-rater reliability of the FMS composite scores re-



sulted in an ICC of 0.82 (95% CI: 0.41-0.93) and was considered excellent (**Table 5**).

The seven individual movement tests of the FMS demonstrated poor to excellent inter-rater reliability (**Table 6**). The HS was the least reliable of the movement tests ( $k_w=0.15$ , 95% CI: -0.09-0.38), whereas the SM was the most reliable of the movement tests ( $k_w=0.85$ , 95%CI: 0.72-0.97).

The non-specialist rotational Physiotherapists demonstrated excellent inter-rater reliability (ICC=0.89, 95%CI 0.78-0.94), whereas the specialist musculoskeletal Physiotherapists

demonstrated good inter-rater reliability (ICC=0.79, 95% CI 0.19-0.92) for FMS composite score (**Table 7**).

The seven individual movement tests of the FMS demonstrated moderate to excellent inter-rater reliability between non-specialist rotational Physiotherapists (**Table 8**). The HS was the least reliable of the movement tests ( $k_w=0.25$ , 95% CI: 0.06-0.44), whereas the PU was the most reliable of the movement tests ( $k_w=0.84$ , 95%CI: 0.69-0.98).

In contrast, the seven individual movement tests of the FMS demonstrated poor to excellent inter-rater reliability between

**Table 4. Demographics of Participants.**

Demographics	Mean +/- SD
<b>Participants Included</b>	
Sex (Male:Female)	20:20
Age	28.94 +/- 11.65
UCLA	7.75 +/- 2.4
<b>Participants Withdrawn</b>	
Sex (Male:Female)	2:1
Age	35.1 +/- 21.1
UCLA	8 +/- 2.64

**Table 5. Inter-Rater Reliability comparisons for FMS Composite Score.**

Rater Comparison	ICC	Landis & Koch (1977)	95% CI
1 vs 2	0.80	Excellent	-0.72 - 0.94
1 vs 3	0.91	Excellent	0.83 - 0.96
1 vs 4	0.89	Excellent	0.76 - 0.94
1 vs 5	0.88	Excellent	0.78 - 0.94
1 vs 6	0.76	Good	-0.04 - 0.92
2 vs 3	0.81	Excellent	0.48 - 0.91
2 vs 4	0.83	Excellent	0.54 - 0.93
2 vs 5	0.71	Good	-0.13 - 0.90
2 vs 6	0.93	Excellent	0.87 - 0.96
3 vs 4	0.87	Excellent	0.75 - 0.93
3 vs 5	0.83	Excellent	0.67 - 0.91
3 vs 6	0.8	Excellent	0.45 - 0.91
4 vs 5	0.78	Good	0.56 - 0.89
4 vs 6	0.84	Excellent	0.51 - 0.93
5 vs 6	0.72	Good	-0.17 - 0.91
<b>Mean</b>	<b>0.824</b>	<b>Excellent</b>	<b>0.41 - 0.93</b>

**Table 6. Inter-Rater Reliability comparisons for FMS Component Scores.**

Component Test	$k_w$	Landis & Koch (1977)	95% CI
Deep Squat	0.26	Fair	0.07-0.45
Hurdle Step	0.15	Poor	-0.09 - 0.38
Inline Lunge	0.39	Fair	0.16-0.61
Shoulder Mobility	0.85	Excellent	0.72-0.97
SLR	0.57	Moderate	0.39-0.76
Press Up	0.79	Good	0.65-0.93
Rotary Stability	0.57	Moderate	0.31-0.84

**Table 7. Inter-Rater Reliability of FMS Composite Scores for Non-Specialist Rotational and Specialist Musculoskeletal Physiotherapists.**

<b>Non-Specialist Rotational Physiotherapists</b>			
Rater Comparison	ICC	Landis and Koch (1977)	95% CI
1 vs 3	0.91	Excellent	0.83 - 0.96
1 vs 4	0.89	Excellent	0.76 - 0.94
3 vs 4	0.87	Excellent	0.75 - 0.93
<b>Mean</b>	<b>0.89</b>	<b>Excellent</b>	<b>0.78 - 0.94</b>
<b>Specialist Musculoskeletal Physiotherapists</b>			
Rater Comparison	ICC	Landis and Koch (1977)	95% CI
2 vs 5	0.71	Good	-0.13 - 0.90
2 vs 6	0.93	Excellent	0.87 - 0.96
5 vs 6	0.72	Good	-0.17 - 0.91
<b>Mean</b>	<b>0.79</b>	<b>Good</b>	<b>0.19 - 0.92</b>

**Table 8. Non-Specialist Rotational Physiotherapists - Inter-Rater Reliability of FMS Component Scores.**

Component Test	$k_w$	Landis & Koch (1977)	95% CI
Deep Squat	0.49	Moderate	0.28-0.69
Hurdle Step	0.25	Fair	0.06-0.44
Inline Lunge	0.45	Moderate	0.2-0.64
Shoulder Mobility	0.79	Good	0.63-0.95
SLR	0.65	Good	0.46-0.84
Press Up	0.84	Excellent	0.69-0.98
Rotary Stability	0.46	Moderate	0.19-0.75

specialist musculoskeletal Physiotherapists (**Table 9**). The DS was the least reliable of the movement tests ( $k_w=0.15$ , 95% CI: -0.01-0.32), whereas the SM was the most reliable of the movement tests ( $k_w=0.88$ , 95%CI: 0.77-0.99).

## Discussion

The primary aim of this study was to determine the inter-rater reliability of the FMS when used by novice NHS Physiotherapists. The findings of this study suggest that the FMS has an acceptable level of reliability when used by NHS Physiotherapists [30]. The inter-rater reliability point estimate of the FMS composite score was excellent with the 95% CI suggestive of

**Table 9. Specialist Musculoskeletal Physiotherapists - Inter-Rater Reliability of FMS Component Scores.**

Component Test	$k_w$	Landis & Koch (1977)	95% CI
Deep Squat	0.15	Poor	-0.01 to 0.32
Hurdle Step	0.19	Poor	-0.1 to 0.49
Inline Lunge	0.42	Moderate	0.21 - 0.66
Shoulder Mobility	0.88	Excellent	0.77-0.99
SLR	0.45	Moderate	0.26-0.63
Press Up	0.73	Good	0.58-0.88
Rotary Stability	0.67	Good	0.42-0.91

moderate-excellent reliability [28].

The findings of this study correspond with the current evidence base with regard to the inter-rater reliability of the FMS composite score. Previous studies have demonstrated inter-rater reliability for the composite score to be fair-excellent [21,23,31]; and Smith et al. (2013) demonstrated excellent inter-rater reliability [19].

Fair inter-rater reliability has been demonstrated by Shultz et al (2013) ( $K\alpha=0.38$ , 95%CI 0.35-0.41). The participants used in this study represent a small, homogenous sample of elite athletes with the total scores ranging from 14-20 [16]. The Krippendorff (K $\alpha$ ) statistic calculates the reliability based upon the differences in the observed range of scores included in the analysis. The use of the ICC however, which is ubiquitous elsewhere in the literature investigating inter-rater reliability (Table 10) focuses on the ratio of total variance and variance between groups; thus the use of the K $\alpha$  may explain this anomalous result in the literature [33].

The clinical utility of the FMS composite score has been questioned [32]. The authors conducted a factor analysis of the FMS amongst a large sample of 934 high-level athletic participants and demonstrated that the FMS does not represent a unidimensional construct and that each individual movement test may be measuring a separate multidimensional construct. When the heterogeneity of the individual movements is considered, these findings are not surprising. The authors concluded therefore that each individual movement test may offer greater clinical utility than the composite score, recommending the use of the composite score with caution as it is not clear what the FMS composite is aiming to measure.

Results from Kazman et al. (2014) are interesting to consider within the context of this study and the findings of previous studies of FMS inter-rater reliability [32]. The inter-rater reliability of the FMS component scores ranged from poor to excellent (Table 6), with five of the seven individual movement tests demonstrating less than an acceptable level of reliability; acceptable reliability defined as >0.70 [30]. With the utility of the FMS composite score being questioned and some authors suggesting use of the composite score with caution [32], the findings of this study, which are in accordance with

**Table 10. Inter-Rater Reliability Studies.**

Item	Reference	Statistic	Variance (95% CI)
<b>Composite Score</b>	Schultz et al. 2013	$K\alpha = 0.38$	0.35-0.41
	Schneiders et al. 2011	ICC = 0.97	
	Smith et al. 2013	ICC = 0.87	0.76-0.94
	Onate et al. 2012	ICC = 0.98	
	Teyren et al. 2012	ICC = 0.76	0.63-0.85
	Elias 2013	ICC = 0.90	
<b>Deep Squat</b>	Schultz et al. 2013	$K\alpha = 0.41$	
	Schneiders et al. 2011	ICC = 1.0	
	Smith et al. 2013	ICC = 0.90	
	Onate et al. 2012	ICC = 1.0	
	Teyren et al. 2012	ICC = 0.68	0.51-0.85
	Minick et al. 2010	$k_w = 1.0$	
<b>Hurdle Step</b>	Schultz et al. 2013	$K\alpha = 0.95$	
	Schneiders et al. 2011	ICC = 0.80	
	Smith et al. 2013	ICC = 0.35	0.12-0.61
	Onate et al. 2012	ICC = 0.31	
	Teyren et al. 2012	ICC = 0.67	0.45-0.88
	Minick et al. 2010	$k_w = 0.83$	
<b>Inline Lunge</b>	Schultz et al. 2013	$K\alpha = 0.10$	
	Schneiders et al. 2011	ICC = 0.86	
	Smith et al. 2013	ICC = 0.78	
	Onate et al. 2012	ICC = 0.88	
	Teyren et al. 2012	ICC = 0.45	0.25-0.65
	Minick et al. 2010	$k_w = 0.79$	
<b>Shoulder Mobility</b>	Schultz et al. 2013	$K\alpha = 0.64$	
	Schneiders et al. 2011	ICC = 0.94	
	Smith et al. 2013	ICC = 0.96	0.92-0.98
	Onate et al. 2012	ICC = 0.90	
	Teyren et al. 2012	ICC = 0.73	0.57-0.89
	Minick et al. 2010	$k_w = 1.0$	
<b>Straight Leg Raise</b>	Schultz et al. 2013	$K\alpha = 0.63$	
	Schneiders et al. 2011	ICC = 0.94	
	Smith et al. 2013	ICC = 0.91	
	Onate et al. 2012	ICC = 0.88	
	Teyren et al. 2012	ICC = 0.69	0.51-0.85
	Minick et al. 2010	$k_w = 0.94$	
<b>Press Up</b>	Schultz et al. 2013	$K\alpha = 0.31$	
	Schneiders et al. 2011	ICC = 1.0	
	Smith et al. 2013	ICC = 0.88	
	Onate et al. 2012	ICC = 0.75	
	Teyren et al. 2012	ICC = 0.82	0.73-0.90
	Minick et al. 2010	$k_w = 0.96$	
<b>Rotary Stability</b>	Schultz et al. 2013	$K\alpha = 0.25$	
	Schneiders et al. 2011	ICC = 1.0	
	Smith et al. 2013	ICC = 0.62	
	Onate et al. 2012	ICC = 1.0	
	Teyren et al. 2012	ICC = 0.77	0.57-0.96
	Minick et al. 2010	$k_w = 0.84$	

the variability (**Table 10**) of the inter-rater reliability of the component scores seen in the current literature, propose some important considerations.

Excellent inter-rater reliability has been shown for all seven component tests [31]; whilst Onate et al. (2012) demonstrated excellent inter-rater reliability for five of the seven component tests [21]. This is in contrast to the findings of this present study; despite both studies being moderate quality, they are limited by the inclusion of only one pair of raters and unclear blinding of raters, and the utilisation of a un-weighted kappa statistic [31]. This represents a strong limitation of the study as the un-weighted kappa statistic is appropriate for determining the reliability of nominal data, and as such will only demonstrate whether raters agree or disagree, and do not take the level of disagreement into consideration [18].

The current literature (**Tables 6 and 10**) does not demonstrate any of the component movement tests to consistently demonstrate an acceptable level of reliability [34]. With the apparent variability and inconsistency in the reliability coefficients the clinical utility of the individual movement tests is advised with caution.

The secondary aim of this study was to determine whether clinical specialism impacts on the inter-rater reliability of the FMS when used by NHS Physiotherapists. In contrast to our findings relating on the intra-rater reliability of the FMS [20], non-specialist rotational Physiotherapists showed excellent inter-rater reliability for composite scores (ICC=0.89, 95%CI 0.78-0.94) in comparison to musculoskeletal specialists who demonstrated good inter-rater reliability (ICC=0.79, 95%CI 0.19-0.92). The clinical implication of this is that the FMS can be used by both specialist and non-specialist Physiotherapists with acceptable reliability.

There was no consistent difference between the non-specialist rotational Physiotherapists and the specialist musculoskeletal Physiotherapists with regard to the component scores (**Tables 8 and 9**). This is again in keeping with our findings of the intra-rater reliability of FMS component scores [20].

The developers of the FMS propose that the FMS is a 'ranking and grading system that documents movement patterns that are key to normal function', labelling the individual movement tests as 'functional' [17]. However, what is functional for one individual may not be functional for another individual; the FMS is limited as it attempts to define function for the individual rather than letting the individual define their function to the clinician [35]. Furthermore, the FMS assumes that there is a correct way to move and any deviation from this is labelled as abnormal, a dysfunction that needs 'correction' [17] this is in contrast to good quality evidence suggesting that variability is a common feature of human movement and is intrinsic to all biological systems [36] whilst omitting the important role that individual psychology and cognition has upon human movement [37]; Subsequently, it would appear that the FMS is engrained within the biomedical model.

It has been suggested that it is currently not possible to

identify the complete optimal solution for a given movement and that clinicians need to demonstrate more individualised clinical assessment procedures [38]. Furthermore, variability of movement is inherent and unavoidable in individuals due to the constraints that shape behaviour [39]. Variations of movement within individuals and deviations in movement between individuals may replicate attempts to maximise the variability to help the individual adapt to the demands of any given situation depending on the environment. The recognition of this variability and the understanding that movement tests are not always achieved in the same way both within and between individuals suggests the need to re-evaluate the application of the biomedical model in assessing movement [39].

The question is therefore, is the FMS testing the function of the individual as the developers propose, or is the FMS merely testing the seven component movement tests? All movement is a skill [40] and extensive motor practice can reorganise movements so that individuals acquire that skill and in turn become better at that movement [41]. By removing context, the environment and therefore defining function for the individual, a low score on one of the individual FMS movement tests may actually represent an unpractised skill as opposed to a faulty movement pattern or dysfunction. Consequently, does change in FMS score within individuals over time or following 'corrective' exercise prescription [17] demonstrate correction of the limitations that the FMS is proposed to identify or learned effects as the individual becomes more practiced at the individual movement test?

The need for a valid and reliable outcome measure that assesses multiple aspects of function simultaneously is apparent [8]. Currently, the literature does not support the use of the FMS as this outcome measure. The clinical application therefore is the need for clinicians to demonstrate a more individualised clinical assessment when analysing an individual's movement to understand the context, psychology and biomechanics of an individual's function in order to determine how the individual interacts and performs their functional movement [39].

This study has contributed to the reliability literature of the FMS by addressing limitations in previous research investigating inter-rater reliability. Previous studies participants have been grossly homogenous, whereas participants in this study were more representative of the population seen in routine clinical practice [42] and demonstrated broad variability in both age and routine activity levels (**Table 4**). Although the sample of participants used should be regarded as a strength of this study, the exclusion of symptomatic individuals limits the representation of clinical practice; Future studies should consider the inclusion of individuals with injuries.

This study recruited 40 participants; a sample of 40 participants is required to achieve an ICC >0.6 if the measure is truly reliable [43] and only one previous study has included more participants when investigating the inter-rater reliability of

the FMS [22] however, the raters did not rate all participants (n=64), each rater only measured between 14-18 participants. Therefore, with all raters in this present study rating each participant it can be concluded that this is the largest reliability study investigating the inter-rater reliability of the FMS.

Another limitation within this investigation is that a standardised video method was utilised to record participants completing the FMS, and though this video method has previously been validated [16] the use of this method may not replicate routine clinical practice and therefore represents a potential limitation of the study.

Ordinal data is regarded as non-parametric data, non-parametric data can behave like parametric data if the data set is normally distributed [44,45]. The data set in this study was not normally distributed (KS  $p < 0.001$ ) and therefore the use of an ICC may potentially represent an over inflation of the inter-rater reliability of FMS composite scores. It is not apparent from the current literature whether any previous investigators have assessed for normal distribution of the data before utilising an ICC so this suggests a potential limitation of the inter-rater reliability evidence as a whole.

It is apparent that for the FMS to be continued to be used clinically, further research is required. Future studies should aim to replicate the results of Kazman et al. (2014) with regard to determining the dimensionality of the FMS within a more heterogeneous sample population [32]. With the use of the composite score appearing to lack clinical utility, there is a need to examine the construct validity of the FMS. Additionally, the reliability literature is limited by small sample sizes and as such, wide confidence intervals and in turn poor precision with regard to reliability coefficients. Future research may consider investigating the reliability of each individual movement test and should consider the utilisation of guidelines to improve the quality of reporting [43,46]. Investigations of the reliability of each individual movement test would be time efficient and allow the use of large sample sizes not currently evident in the literature [43,46].

## Conclusion

The FMS represents a good attempt to objectify the subjective with the FMS composite score demonstrating excellent inter-rater reliability. However, it does not appear that the FMS composite score offers much in the way of clinical utility and therefore should be interpreted and utilised with caution.

The inter-rater reliability of the FMS component scores is varied and inconsistent in the literature with the results from this study suggesting that five of the seven individual movement tests do not demonstrate acceptable reliability for clinical use; supporting the findings from our previous paper [20].

With the composite score lacking construct validity and the majority of the component scores lacking both intra- and inter-rater reliability, the continued use of FMS within clinical practice is not supported.

## Competing interests

The authors declare that they have no competing interests.

## Authors' contributions

Authors' contributions	AVC	JLP	ML
Research concept and design	✓	✓	--
Collection and/or assembly of data	✓	✓	--
Data analysis and interpretation	✓	✓	--
Writing the article	✓	✓	--
Critical revision of the article	✓	✓	✓
Final approval of article	✓	✓	✓
Statistical analysis	✓	✓	✓

## Publication history

Editor: Catherine Ortega, University of Texas Health Science Center, USA.  
Received: 27-Aug-2018 Final Revised: 15-Oct-2018  
Accepted: 22-Oct-2018 Published: 14-Nov-2018

## References

1. Butler RJ, Plisky PJ, Southers C, Scoma C and Kiesel KB. **Biomechanical analysis of the different classifications of the Functional Movement Screen deep squat test.** *Sports Biomech.* 2010; 9:270-9. | [Article](#) | [PubMed](#)
2. Beevor C. **The croonian lectures on muscular movements and their representation in the central nervous system.** *British medical journal.* 1903; 2:1480-1484. | [Article](#)
3. Gribble PA, Brigle J, Pietrosimone BG, Pfile KR and Webster KA. **Intrater reliability of the functional movement screen.** *J Strength Cond Res.* 2013; 27:978-81. | [Article](#) | [PubMed](#)
4. Boudreau SA, Farina D and Falla D. **The role of motor learning and neuroplasticity in designing rehabilitation approaches for musculoskeletal pain disorders.** *Man Ther.* 2010; 15:410-4. | [Article](#) | [PubMed](#)
5. Takeuchi N and Izumi S. **Rehabilitation with poststroke motor recovery: a review with a focus on neural plasticity.** *Stroke Res Treat.* 2013; 2013:128641. | [Article](#) | [PubMed Abstract](#) | [PubMed FullText](#)
6. Schache AG, Blanch PD and Murphy AT. **Relation of anterior pelvic tilt during running to clinical and kinematic measures of hip extension.** *Br J Sports Med.* 2000; 34:279-83. | [Article](#) | [PubMed Abstract](#) | [PubMed FullText](#)
7. Moreside JM and McGill SM. **Improvements in hip flexibility do not transfer to mobility in functional movement patterns.** *J Strength Cond Res.* 2013; 27:2635-43. | [Article](#) | [PubMed](#) | [PubMed FullText](#)
8. Minick KI, Kiesel KB, Burton L, Taylor A, Plisky P and Butler RJ. **Interrater reliability of the functional movement screen.** *J Strength Cond Res.* 2010; 24:479-86. | [Article](#) | [PubMed](#)
9. Cholewicki J, Greene HS, Polzhofer GK, Galloway MT, Shah RA and Radebold A. **Neuromuscular function in athletes following recovery from a recent acute low back injury.** *J Orthop Sports Phys Ther.* 2002; 32:568-75. | [Article](#) | [PubMed](#)
10. Silvers HJ, Giza ER and Mandelbaum BR. **Anterior cruciate ligament tear prevention in the female athlete.** *Curr Sports Med Rep.* 2005; 4:341-3. | [PubMed](#)
11. Wainner RS, Whitman JM, Cleland JA and Flynn TW. **Regional interdependence: a musculoskeletal examination model whose time has come.** *J Orthop Sports Phys Ther.* 2007; 37:658-60. | [Article](#) | [PubMed](#)
12. Moseley L. **Reconceptualising pain according to modern pain science.** *Physical therapy reviews.* 2007; 12:169-178.
13. Blakeslee S, and Blakeslee M. **The body has a mind of its own: how body maps in your brain help you do (almost) everything better.** New York: Random House Incorporated. 2007.
14. Chartered Society of Physiotherapy. **Quality Assurance Standards.** 2012. | [Website](#)



15. Health Care Professions Council. **Standards of proficiency for Physiotherapists**. 2018. | [Website](#)
16. Shultz R, Anderson SC, Matheson GO, Marcello B and Besier T. **Test-retest and interrater reliability of the functional movement screen**. *J Athl Train*. 2013; **48**:331-6. | [Article](#) | [PubMed Abstract](#) | [PubMed FullText](#)
17. Cook G, Burton L, Kiesel K, Rose G and Bryant M. **Movement: Functional Movement Systems**. USA: On Target Publications. 2010.
18. Streiner D and Norman G. **Human measurement scale**. 3rd ed. Oxford: Oxford University Press. 2003.
19. Smith CA, Chimera NJ, Wright NJ and Warren M. **Interrater and intrarater reliability of the functional movement screen**. *J Strength Cond Res*. 2013; **27**:982-7. | [Article](#) | [PubMed](#)
20. Palmer J, Cuff A and Lindley M. **Intra-Rater reliability of the Functional Movement Screen (FMS) amongst NHS Physiotherapists**. *Physical Therapy and Rehabilitation*. 2017; **4**:1-7. | [Pdf](#)
21. Onate JA, Dewey T, Kollock RO, Thomas KS, Van Lunen BL, DeMaio M and Ringleb SI. **Real-time intersession and interrater reliability of the functional movement screen**. *J Strength Cond Res*. 2012; **26**:408-15. | [Article](#) | [PubMed](#)
22. Teyhen DS, Shaffer SW, Lorenson CL, Halfpap JP, Donofry DF, Walker MJ, Dugan JL and Childs JD. **The Functional Movement Screen: a reliability study**. *J Orthop Sports Phys Ther*. 2012; **42**:530-40. | [Article](#) | [PubMed](#)
23. Leeder JE, Horsley IG and Herrington LC. **The Inter-rater Reliability of the Functional Movement Screen Within an Athletic Population Using Untrained Raters**. *J Strength Cond Res*. 2016; **30**:2591-9. | [Article](#) | [PubMed](#)
24. Greenhalgh T. **How to read a paper: The basis of evidence-based medicine**. 5th ed. Oxford: BMJ Books; 2014.
25. Shephard RJ. **PAR-Q, Canadian Home Fitness Test and exercise screening alternatives**. *Sports Med*. 1988; **5**:185-95. | [Article](#) | [PubMed](#)
26. Urbaniak G, Plous S. **Research Randomiser**. 20084eredrftg. | [Website](#)
27. Sim J and Wright C. **Research in Health Care: Concepts, Designs and Methods**. United Kingdom: Stanley Thornes; 2000.
28. Landis JR and Koch GG. **The measurement of observer agreement for categorical data**. *Biometrics*. 1977; **33**:159-74. | [PubMed](#)
29. Naal FD, Impellizzeri FM and Leunig M. **Which is the best activity rating scale for patients undergoing total joint arthroplasty?** *Clin Orthop Relat Res*. 2009; **467**:958-65. | [Article](#) | [PubMed Abstract](#) | [PubMed FullText](#)
30. Nunally J. **Psychometric Theory**. New York: McGraw-Hill. 1978.
31. Schneiders AG, Davidsson A, Horman E and Sullivan SJ. **Functional movement screen normative values in a young, active population**. *Int J Sports Phys Ther*. 2011; **6**:75-82. | [PubMed Abstract](#) | [PubMed FullText](#)
32. Kazman JB, Galecki JM, Lisman P, Deuster PA and O'Connor FG. **Factor structure of the functional movement screen in marine officer candidates**. *J Strength Cond Res*. 2014; **28**:672-8. | [Article](#) | [PubMed](#)
33. Stanish W and Taylor N. **Estimation of the intraclass correlation coefficient for the analysis of covariance model**. *The American Statistician*. 1983; **37**:221-224.
34. Frohm A, Heijne A, Kowalski J, Svensson P and Myklebust G. **A nine-test screening battery for athletes: a reliability study**. *Scand J Med Sci Sports*. 2012; **22**:306-15. | [Article](#) | [PubMed](#)
35. Hargrove T. **A guide to better movement: the science and practice of moving with more skill and less pain**. Seattle: Better Movement. 2014.
36. Stergiou N and Decker LM. **Human movement variability, nonlinear dynamics, and pathology: is there a connection?** *Hum Mov Sci*. 2011; **30**:869-88. | [Article](#) | [PubMed Abstract](#) | [PubMed FullText](#)
37. Araujo D and Davids K. **Embodied cognition and emergent decision-making in dynamic movement systems**. *The journal of thematic dialogue*. 2004; **2**:45-57.
38. Glazier PS and Davids K. **Constraints on the complete optimization of human motion**. *Sports Med*. 2009; **39**:15-28. | [Article](#) | [PubMed](#)
39. Davids K, Glazier P, Araujo D and Bartlett R. **Movement systems as dynamical systems: the functional role of variability and its implications for sports medicine**. *Sports Med*. 2003; **33**:245-60. | [Article](#) | [PubMed](#)
40. Bartlett R, Wheat J and Robins M. **Is movement variability important for sports biomechanists?** *Sports Biomech*. 2007; **6**:224-43. | [Article](#) | [PubMed](#)
41. Furuya S, Nakamura A and Nagata N. **Transfer of piano practice in fast performance of skilled finger movements**. *BMC Neurosci*. 2013; **14**:133. | [Article](#) | [PubMed Abstract](#) | [PubMed FullText](#)
42. May S, Littlewood C and Bishop A. **Reliability of procedures used in the physical examination of non-specific low back pain: a systematic review**. *Aust J Physiother*. 2006; **52**:91-102. | [Article](#) | [PubMed](#)
43. Brunton LK and Bartlett DJ. **Validity and reliability of two abbreviated versions of the Gross Motor Function Measure**. *Phys Ther*. 2011; **91**:577-88. | [Article](#) | [PubMed](#)
44. Knapp TR. **Treating ordinal scales as interval scales: an attempt to resolve the controversy**. *Nurs Res*. 1990; **39**:121-3. | [PubMed](#)
45. Jamieson S. **Likert scales: how to (ab)use them**. *Medical education*. 2004; **38**:1212-1218.
46. Kottner J, Audige L, Brorson S, Donner A, Gajewski BJ, Hrobjartsson A, Roberts C, Shoukri M and Streiner DL. **Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed**. *J Clin Epidemiol*. 2011; **64**:96-106. | [Article](#) | [PubMed](#)

**Citation:**

Cuff A, Palmer JL and Lindley M. **Inter-Rater reliability of the Functional Movement Screen (FMS) amongst NHS Physiotherapists**. *Phys Ther Rehabil*. 2018; **5**:17.  
<http://dx.doi.org/10.7243/2055-2386-5-17>