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

PALMER, Joe, CUFF, Andrew and LINDLEY, Melanie

Available from Sheffield Hallam University Research Archive (SHURA) at:
http://shura.shu.ac.uk/16758/

This document is the author deposited version. You are advised to consult the publisher's version if you wish to cite from it.

Published version


Copyright and re-use policy

See http://shura.shu.ac.uk/information.html
Intra-Rater reliability of the Functional Movement Screen (FMS) amongst NHS Physiotherapists

Joe L. Palmer1,2*, Andrew V. Cuff2-6 and Melanie Lindley2

Abstract

Objectives: The primary aim of this study was to evaluate the intra-rater reliability of the FMS amongst a group of novice National Health Service (NHS) Physiotherapists. The secondary objective was to establish if Intra-rater reliability differed between non-specialist rotational Physiotherapists and Physiotherapists working within the musculoskeletal (MSk) setting.

Design: Reliability Study.

Method: 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 intra-rater reliability of FMS composite scores was Excellent (mean ICC of 0.91 (95% CI 0.81-0.96)). The non-specialist rotational Physiotherapist group demonstrated Good-Excellent intra-rater reliability (ICC 0.9; 95% CI 0.79-0.95). Specialist MSk Physiotherapists demonstrated Excellent intra-rater reliability (ICC 0.92; 95% CI 0.84-0.96). Intra-rater reliability of the 7 component tests of the FMS ranged from Poor-Excellent (KW 95% CI 0.11-0.98).

Conclusion: Among novice NHS Physiotherapists, the FMS composite score demonstrated Excellent intra-rater reliability. MSk specialists were found to demonstrate a marginally superior level of intra-rater agreement compared to non-specialist rotational Physiotherapists; however this is likely to be negligible in a clinical context. Clinical specialism also appears to have little impact on the intra-rater reliability of FMS components with both groups of raters achieving a Poor-Excellent level of agreement.

Keywords: Intra-rater, reliability, functional movement screen, FMS

Introduction

Musculoskeletal (MSK) examination and treatment approaches are conventionally based on isolated components of movement, such as muscle power, muscle length and joint range of motion (ROM) [1]. These methods overlook the role of the central nervous system (CNS) in allowing complex movements to be executed whilst maintaining balance and equilibrium. Changes in muscle length and tone in relation to tasks are not isolated events; they are the product of highly co-ordinated patterns of muscle activation produced through interactions between the CNS and Musculoskeletal systems [2]. This interaction during functional activities therefore suggests that traditional methods of assessment and rehabilitation may fail to represent the true nature of movement and suggest the need for a more comprehensive assessment of movement during clinical examination, in addition to traditional bed based assessment methods.

The limitations of merely focusing on the assessment and treatment of anatomical structure have been demonstrated in the literature. Improvements in passive hip joint ROM did not correlate with changes during functional movements commonly performed during sporting activities [3]. The authors found there to be no significant changes in hip extension or rotation during...
dynamic movements despite significant improvements in passive hip ROM following a hip flexibility programme [3]. This suggests that simply assessing and treating specific ‘structures’ within the body has little influence on an individual’s ability to perform more complex, functional movements. As such, a greater emphasis has been placed on movement capacity as opposed to solely targeting specific anatomical structures [4]. This emphasis ensures that clinicians are assessing both the body structure and function, and activity domains of the ICF model (WHO 2014). The Functional Movement Screen (FMS) is a systematic method of observing and scoring individuals on the quality of their movement through specific patterns [2].

The FMS consists of 3 screening tests and 7 individual movement patterns; 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) movement patterns [2]. These seven movements are then scored according to specific criteria on a categorical scale ranging from 0-3. A score of one is given if the individual is unable to perform/complete the movement pattern in accordance with its FMS test definition. A score of 2 indicates that the individual is able to perform the movement with a degree of compensation, poor improper form and/or loss of alignment when compared to FMS test definition. A maximal score of three is given to those who are able to successfully complete the movement pattern according to the pre-set criteria without compensation. Finally, a score of zero is given to those who report pain whilst performing the movement pattern; this is regardless of the quality of the movement performed. Completion of the seven component movements within the FMS produces a total (composite) score ranging from 0 (if all movements provoke pain) to 21 (if a participant is able to perform the movement with a maximal degree of compensation, poor improper form and/or loss of alignment when compared to FMS test definition). A score of three is given to those who are able to successfully complete the movement pattern according to the pre-set criteria without compensation. A maximal score of three is given to those who are able to successfully complete the movement pattern according to the pre-set criteria without compensation. Finally, a score of zero is given to those who report pain whilst performing the movement pattern; this is regardless of the quality of the movement performed. Completion of the seven component movements within the FMS produces a total (composite) score ranging from 0 (if all movements provoke pain) to 21 (if a participant is able to perform the movement with a maximal degree of compensation, poor improper form and/or loss of alignment when compared to FMS test definition). A score of three is given to those who are able to successfully complete the movement pattern according to the pre-set criteria without compensation. Finally, a score of zero is given to those who report pain whilst performing the movement pattern; this is regardless of the quality of the movement performed. Completion of the seven component movements within the FMS produces a total (composite) score ranging from 0 (if all movements provoke pain) to 21 (if a participant is able to perform the movement with a maximal degree of compensation, poor improper form and/or loss of alignment when compared to FMS test definition).

The intra-rater reliability of the FMS has been investigated with conflicting results; particularly when groups of raters with varying FMS experience have been compared [5-9]. This could however be related to the observed methodological heterogeneity within the current evidence base [10]. Although this particular psychometric property of the FMS has been evaluated in these studies, the reliability of Physiotherapists working within the public sector health care has yet to be investigated. As a result it is not possible to draw definitive conclusions as to the external validity of these findings within this particular clinical setting [4]. With a growing number of NHS clinicians working privately in amateur and semi-professional sport, it is possible that the FMS may begin to be used in NHS practice with people of varying activity levels. Furthermore, people participating in recreational activity may present to NHS clinics; as such, there is need to determine reliability within this professional setting.

It is not known whether clinical specialism impacts on the reliability of FMS scoring; a factor which could be significant within the NHS Physiotherapy outpatient departments whereby a vast range of clinical specialism amongst clinicians occurs. The primary aim of this study was to evaluate the intra-rater reliability of the FMS amongst NHS Physiotherapists with no prior experience of using the FMS. The secondary objective was to establish if intra-rater reliability differed between non-specialist rotational Physiotherapists and specialist MSK Physiotherapists.

Methods

Participants

A purposive convenience sampling method was used in order to achieve a recruitment target of 40 participants to be videoed completing the FMS; these were recruited from the South Yorkshire region using posters and University email. Prospective participants were approached, briefed on the purpose of the study and given an information sheet. Individuals were then asked to complete a Physical Activity Readiness Questionnaire (PAR-Q) to assess suitability for physical activity [11]. Potential participants completed the three standardised FMS clearing tests; consisting of the impingement, prone press up and posterior rocking clearing tests. Unlike the 7 FMS component tests, the three clearing tests are scored as positive or negative according to the presence or absence of pain; any participants who reported pain during the clearing tests were excluded [2]. For further information on how each of the clearing tests are performed the reader is directed to Cook et al (2010) [2]. Individuals who were identified as eligible then signed an informed consent form (Table 1).

Raters

Six NHS physiotherapists were recruited from NHS trusts within the Yorkshire and Derbyshire regions of the United Kingdom. This was achieved using a purposive judgemental sampling approach; a total of three specialist MSK physiotherapists and

Table 1. Participant Inclusion/Exclusion Criteria.

<table>
<thead>
<tr>
<th>Inclusion</th>
<th>Exclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;18 years of age</td>
<td>Participant refusal</td>
</tr>
<tr>
<td>Ability/willingness to adhere to trial procedures</td>
<td>Answered ‘Yes’ on any PAR-Q question</td>
</tr>
<tr>
<td>Cardiac History</td>
<td></td>
</tr>
<tr>
<td>Neurological Impairment</td>
<td></td>
</tr>
<tr>
<td>Pregnancy</td>
<td></td>
</tr>
<tr>
<td>Dizziness</td>
<td></td>
</tr>
<tr>
<td>Hypertension</td>
<td></td>
</tr>
<tr>
<td>Received treatment for OR reported having any musculoskeletal pathology within the previous 6 weeks</td>
<td></td>
</tr>
<tr>
<td>Reported pain on any of the 3 FMS Clearing Tests</td>
<td></td>
</tr>
</tbody>
</table>
three non-specialist rotational physiotherapists were recruited. Eligible specialist and non-specialist NHS Physiotherapists completed an informed consent form and were entered into the study (Tables 2 and 3).

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

<table>
<thead>
<tr>
<th>Inclusion</th>
<th>Exclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Ability to and willing to adhere to trial procedures</td>
<td>• Specialist Physiotherapist across all areas of clinical practice according to NHS Employers (2005) definition</td>
</tr>
<tr>
<td>• Non-specialised rotational Physiotherapists according to NHS Employers (2005) definition</td>
<td>• Previous experience of using the FMS in both clinical and non-clinical settings</td>
</tr>
<tr>
<td>• Currently employed in the NHS</td>
<td>• Non-qualified Physiotherapy staff according to NHS Employers (2005) definition</td>
</tr>
<tr>
<td></td>
<td>• Unable to attend CPD training session, rating session 1 OR rating session 2</td>
</tr>
<tr>
<td></td>
<td>• Significant, non-correctable visual impairment</td>
</tr>
</tbody>
</table>

Table 3. Specialist Raters Inclusion/Exclusion Criteria.

<table>
<thead>
<tr>
<th>Inclusion</th>
<th>Exclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Ability to and willing to adhere to trial procedures</td>
<td>• Specialist Physiotherapist in a non-Musculoskeletal clinical area according to NHS Employers (2005) definition</td>
</tr>
<tr>
<td>• Specialist Musculoskeletal Physiotherapists according to NHS Employers (2005) definition</td>
<td>• Previous experience of using the FMS in both clinical and non-clinical settings</td>
</tr>
<tr>
<td>• Currently employed in the NHS</td>
<td>• Non-qualified Physiotherapy staff according to NHS Employers (2005) definition</td>
</tr>
<tr>
<td></td>
<td>• Unable to attend CPD training session, rating session 1 OR rating session 2</td>
</tr>
<tr>
<td></td>
<td>• Significant, non-correctable visual impairment</td>
</tr>
</tbody>
</table>

Procedures
Participants were video recorded using two Sony HDR-XR260 (Sony Corporation, Minato, Tokyo, Japan) camcorders in both the sagittal and coronal planes whilst performing each of the seven FMS component tests under the guidance of a single instructor, using an official FMS Test Kit (Functional Movement Systems Incorporated, Chatham, Virginia, USA). Camera heights and distances from the participant were set to that of the previously validated video method [5].
livered by a certified FMS practitioner prior to starting rating session 1. The session consisted of an introduction to the FMS and test procedures of each of the seven movement components and three relevant modifications for the DS, TSPU and RS tests. Raters were also informed of the three FMS Clearing tests and how they had been applied as part of the exclusion criteria within the study. At this point, raters were given the opportunity to ask questions in an open discussion with the tutor to address any remaining queries. The session concluded with raters viewing two example videos which adopted the exact format used within the 40 rating videos to allow for familiarisation with test procedures.

Rating sessions
To minimise potential test-retest bias the ordering of participant videos was randomised for each rater in both rating sessions using an online software package \[12\]. This randomisation, coupled with raters being blinded to measurements made in session 1 aimed to reduce the chances of raters recalling scores for each participant. To further reduce the potential for rater recollection introducing bias into their scoring, a two-week washout period was introduced between rating sessions.

Raters were each set up on two individual computers; one set up to show a continuous video of all non-modified movements and a second to give raters the opportunity to view modification videos. This was to allow easy transition between non-modified and modified videos. Raters were able to pause full length (non-modified) videos at any time to allow for breaks or to observe modification videos on computer two. However, in order to prevent raters from pausing videos and subsequently analysing still frames of individual movements, the pausing of videos was only permitted whilst on an introductory/transitional screen (Figure 1). This was to ensure observations replicated clinical practice where this facility would likely be unfeasible. Raters were also instructed to only view video modifications for participants DS, TSPU or RS if they felt that a participant did not score a 3 on their respective unmodified video. This is reminiscent of clinical practice whereby a rater would only ask a participant to perform a modified movement if they felt that they were unable to achieve a maximum score when performing a non-modified movement [2].

Statistical analysis
Weighted Kappa (KW) statistic and 95% Confidence Interval (95% CI) were calculated in order to evaluate intra-rater reliability of component scores \[13\].

Two-way mixed Intraclass Correlation Coefficients (ICC) and 95% CI were calculated in order to evaluate the intra-rater reliability of total scores \[13\]. ICC’s were assessed for normal distribution using the Kolmogorov-Smirnoff Test.

Statistics were calculated using the MedCalc (Version 10.4.0.0; MedCalc software. Mariakerke, Belgium) and SPSS 21.0 (SPSS Inc. Chicago, IL, USA) statistical packages. Reliability coefficients were interpreted according to criteria developed by Landis and Koch (1977) \[14\]; 0.0-0.2 Poor, 0.21-0.4 Fair, 0.41-0.6 Moderate, 0.61-0.8 Good and 0.81-1.0 Excellent.

Results
Forty participants (20 Males, 20 Females) met the eligibility criteria, consented and completed the study. Three participants (2 Males, 1 Female) withdrew prior to completing the study due to scheduling difficulties. The mean age of the participants was 28.9 (Range 18.9-60.3; SD+/-11.65) years. Participants demonstrated a broad range of physical activity levels according to University of California at Los Angeles (UCLA) Activity Scoring (Range 3-10; mean 7.75; SD+/-2.4).

The intra-rater reliability of FMS composite scores across the 6 raters resulted in a mean ICC of 0.91 (95% CI 0.81-0.96) and is therefore considered to be excellent (Figure 5) \[14\]. The non-specialist rotational physiotherapist group demonstrated good–excellent intra-rater reliability (ICC 0.9; 95% CI 0.79-0.95) (Figure 5) \[14\]. Specialist MSK physiotherapists demonstrated excellent intra-rater reliability (ICC 0.92; 95% CI 0.84-0.96) (Figure 5) \[14\]. Kolmogorov-Smirnoff testing showed the composite scoring of the FMS across raters to be not normally distributed (P<0.001) \[15\].

The six raters demonstrated a broad spectrum of intra-rater agreement levels across the 7 component tests of the FMS, ranging from poor-excellent (KW 95% CI 0.11-0.98). The most
reliable component movement across the 6 raters was the SM (KW 0.90, 95% CI 0.79-1.0) and the least reliable was the HS (KW 0.4, 95% CI 0.11-0.71) (Figure 6) [14].

![Figure 6](Image)

Intra-rater reliability of component scoring amongst non-specialist rotational Physiotherapists was found to be Poor-Excellent (KW 0.21-0.99). The most reliable FMS component test across non-specialist rotational raters was the SM (KW 0.91, 95% CI 0.78-1.0) and the least reliable being the RS (KW 0.37, 95% CI 0.26-0.62) [14].

Intra-rater reliability of component scoring amongst specialist MSK physiotherapists was found to be Poor – Excellent (KW -0.34-0.96). The most reliable FMS component test across specialist MSK physiotherapists was the SM (KW 0.90, 95% CI 0.80-1.0) and the least reliable being the HS (KW 0.21, 95% CI -0.34-0.56) [14].

Discussion

The aims of this study were to evaluate the intra-rater reliability of the FMS amongst two groups of NHS physiotherapists; representative of the skill mix encountered within a typical NHS physiotherapy outpatient department. The reliability of composite FMS scores were consistent with the findings of previous studies, with the 6 raters demonstrating an excellent level of agreement overall (ICC 0.91; 95% CI 0.81-0.96) [6-9]. However, as composite data sets were found to be not normally distributed the possibility of ICC inflation cannot be overlooked [16].

Gribble et al (2013) [4] and Smith et al (2013) [7] used a similar investigative approach to our study by assessing intra-rater reliability amongst raters of varying backgrounds, expertise and FMS experience. The authors reported mean ICC’s of 0.88 and 0.75 respectively, therefore indicating a good-excellent level of intra-rater agreement. This is therefore reflective of the results of our study (ICC 0.91, 95% CI 0.81-0.96). The studies [4,7] differed significantly however in their findings relating to intra-rater agreement amongst raters with varying clinical experience/expertise; all with no prior FMS experience. A physical therapy student and athletic trainer, both novice FMS raters demonstrated excellent intra-rater reliability in composite scoring, producing ICC’s of 0.88 and 0.91 respectively [7]. This echoes findings within our study where non-specialist and MSK specialist NHS physiotherapists, with no experience of using the FMS prior to participation, demonstrated an excellent level of intra-rater agreement (ICC’s of 0.90 and 0.92 respectively). In contrast, Gribble et al (2013) [4] found a significant difference in composite reliability between novice qualified and non-qualified athletic trainers; with qualified athletic trainers demonstrating superior reliability. Potential reasons for this could be due to the inclusion of non-qualified athletic trainers as raters, it has been shown that physiotherapy students demonstrate inferior reliability compared to qualified physiotherapists when using the FMS [17].

There is an apparent trend observed across the literature which suggests that clinical expertise may lead to improved intra-rater reliability of composite FMS scoring, as these individuals consistently demonstrate superior reliability. The results of this study do not suggest a clinically significant difference in intra-rater reliability between specialist and non-specialist clinicians.

Intra-rater agreement of the seven FMS component movements varied significantly with mean agreement levels of the 6 raters ranging from Fair to Excellent. Of the seven component movements, the HS was found to be the least reliable amongst the 6 raters (KW 0.4; 95% CI 0.11-0.71) and the SM test the most reliable (KW 0.9; 95% CI 0.79-1.00). These findings reflect similarly in the literature. Onate et al (2012) [8] and Teyhen et al (2012) [9] each found the HS to only reach Poor-Moderate levels of agreement (KW 0.16 and 0.59 respectively); below the acceptable level of reliability for clinical practice [18].

SM scores did not reach the level of agreement seen within this study, ranging from Good (KW 0.68) to Excellent (KW 0.84) for Teyhen et al (2012) [9] and Onate et al (2013) [8] respectively. This broad spectrum of intra-rater agreement between these two FMS components is likely due to their relative complexity and opportunities for compensation. The SM test criteria are succinct and easily interpreted as they are mainly based on a simple measure of the distance between the participant’s hands. In contrast, the HS requires a high level of proprioception, joint mobility/stability and balance; offering numerous opportunities for compensation. This could therefore result in the same observer missing previous/identifying new flaws in a participants HS across two separate rating sessions; resulting in poor levels of intra-rater agreement.

The secondary aim of this study was to establish if MSK specialism amongst NHS physiotherapists influenced the reliability of FMS scoring. Results are suggestive that this is unlikely to be so. In four of the seven FMS component movements (ILL, SLR, TSPU, RS), Specialist MSK physiotherapists did demonstrate superior reliability when compared to Rotational Physiotherapists. This was however not the case for the remaining three components tests (DS, HS, SM) where rotational physiotherapists showed the greatest level of agreement. This lack of distinction between rotational and specialist MSK
physiotherapists is also apparent in a lack of precision in the scoring for the majority of FMS component movements with wide 95% CI’s observed for both groups of raters [19].

The implications of these findings suggest that the intra-rater reliability of the FMS amongst NHS Physiotherapists does differ significantly between composite and component scoring. Composite scoring demonstrated excellent reliability across all raters; however, due to the significant variation in the levels of intra-rater agreement for FMS components, the potential clinical value of composite scores is questionable. A recent study highlights this point through factor analysis of the component movements of the FMS [20]. Due to the significant heterogeneity of the FMS components the validity of interpreting composite scores as a unitary construct is highly questionable and the use of the component scores only is suggested. Five of the component tests did not demonstrate a clinically acceptable level of reliability and in turn, further questions the clinical utility of the FMS.

The cost of FMS accreditation is approximately £225 [21]. The training programme delivered to raters in this study was representative of that typically received within a clinical setting, where a senior physiotherapist may undergo formal certification before training other team members. The raters within this study demonstrated excellent intra-rater reliability. Throughout the literature FMS certification has not demonstrated superior reliability [4,8,9] while one study found the only FMS certified rater to have the poorest level of intra-rater reliability [7]. Therefore, it appears that FMS certification has little influence on intra-rater reliability, and in turn the clinical utility of the FMS, questioning the need for formal certification.

A strength of this study was the recruitment of a sample of participants and Physiotherapists reflective of that encountered within a typical NHS Outpatient Physiotherapy Department. The sample of 40 participants and the multiple comparisons between raters enhance the external validity of findings, allowing them to be confidently generalised within this clinical setting [4,10,22].

A potential limitation of this study is the lack of instructor variance. As participants were guided through the FMS by one individual, it does not take into consideration the potential influence of different FMS administrators on participant response.

This study aimed to utilise the demographic of raters and participants likely to be found within a typical NHS physiotherapy outpatient department. It can therefore be concluded that the intra-rater reliability of FMS composite and component does not differ significantly between non-specialist rotational and specialist MSK physiotherapists working within the NHS. It does not however provide any insight as to whether the FMS offers any benefit to patient care and questions the reproducibility and clinical utility of the FMS.

Inter-rater reliability accounts for any potential error between clinicians as well as all potential errors encountered on assessing the intra-rater reliability of a measure [22]; inter-rater reliability has not been determined in this population. There is a need to investigate the inter-rater reliability of the FMS within this population to further inform clinical utility and should be addressed in future studies.

Conclusion
Among NHS physiotherapists, the FMS composite score demonstrated excellent intra-rater reliability. There was no clinically significant difference between specialist and non-specialist physiotherapists. There is therefore a need to investigate the inter-rater reliability within this clinical population. Only the SM and TSPU component tests demonstrated an acceptable level of intra-rater reliability for clinical use. Recent literature has brought into question the utility of composite scoring of the FMS whilst this study questions the intra-rater reliability of the component scores. The FMS should therefore be utilised with caution within this clinical setting.

Competing interests
The authors declare that they have no competing interests.

Authors’ contributions

<table>
<thead>
<tr>
<th>Authors’ contributions</th>
<th>JLP</th>
<th>AVC</th>
<th>ML</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research concept and design</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Collection and/or assembly of data</td>
<td>✓</td>
<td>✓</td>
<td>--</td>
</tr>
<tr>
<td>Data analysis and interpretation</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Writing the article</td>
<td>✓</td>
<td>✓</td>
<td>--</td>
</tr>
<tr>
<td>Critical revision of the article</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Final approval of article</td>
<td>✓</td>
<td>✓</td>
<td>--</td>
</tr>
<tr>
<td>Statistical analysis</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

Acknowledgement
The authors wish to thank all who participated in the data collection process within this study.

Publication history
Editor: Gordon John AlderinK, Grand Valley State University, USA.
Received: 19-Jan-2017 Final Revised: 23-Feb-2017
Accepted: 10-Mar-2017 Published: 24-Mar-2017

References
5. Care Quality Commission. The essential standards. 2012. | Website


Citation: