INTERRATER RELIABILITY OF THE FUNCTIONAL MOVEMENT SCREEN

KATE I. MINICK,1 KYLE B. KIESEL,1,2 LEE BURTON,3 AARON TAYLOR,3 PHIL PLISKY,1,2 AND ROBERT J. BUTLER1

1Department of Physical Therapy, University of Evansville, Evansville, Indiana; 2ProRehab PC, Evansville, Indiana; and 3Department of Physical Education, Averett University, Danville, Virginia

ABSTRACT

Minick, KI, Kiesel, KB, Burton, L, Taylor, A, Plisky, P, and Butler, RJ. Interrater reliability of the Functional Movement Screen. J Strength Cond Res 24(2): 479–486, 2010—The Functional Movement Screen (FMS) is a series of 7 tests that categorize fundamental movement. Each test is scored on an ordinal scale with 4 categories. The purpose of this study was to determine the interrater reliability of the FMS. Forty healthy subjects were videotaped while performing the FMS. The videos were independently scored by 4 raters, including 2 experts who instruct FMS training courses and 2 novices who completed a standardized training course on the FMS. Interrater reliability was analyzed using the weighted kappa statistic. The novice raters demonstrated excellent or substantial agreement on 14 of the 17 tests, whereas the expert raters did the same on 13 of the 17 tests. When the novice raters were paired with the expert raters, all 17 components demonstrated excellent or substantial agreement. These data indicate that the FMS can confidently be applied by trained individuals. This would suggest that the FMS can be confidently used to assess the movement patterns of athletes and to make decisions related to interventions for performance enhancement, and the FMS may assist in identifying athletes at risk for injury.

KEY WORDS flexibility, injury prevention, motor control

INTRODUCTION

The increased sports participation in recent decades has brought with it an increase in the risk for sustaining musculoskeletal injuries. It has long been thought that isolated muscle stretching would be an effective intervention to reduce muscle soreness or musculoskeletal injury; however, recent research has suggested that this is not the case (10,11,20). To reduce injury risk, sports medicine professionals have begun to focus on improving movement patterns as opposed to focusing on rehabilitation of a specific joint (5,13). Research has demonstrated that an isolated rehabilitation approach after injury is not sufficient to normalize performance that encompasses the entire body (17). Data have also suggested that an isolated injury will adversely affect regions away from the injury site (2–4,9,12,15,18,22,23,25,26). The term regional interdependence has recently been applied to conceptually explain why dysfunction in one body region may be contributing to weakness, tightness, or pain in another region (27). Thus, a valid and reliable measurement tool that assesses multiple domains of function simultaneously is in demand.

Current research has suggested that tests assessing multiple domains of function (e.g., balance, strength, and range of motion) simultaneously may improve the accuracy of identifying athletes at risk for injury (19). These tests can typically be carried out by strength and conditioning professionals in a variety of settings. Assessments by strength and conditioning professionals typically include testing to determine the athlete’s strength, speed, agility, and flexibility (1). Strength and conditioning assessments have met an important demand in the overall functional performance assessment of the athlete, but there is still a need to routinely assess the individual athlete’s fundamental movement characteristics (5,13). Recently, a tool to assess fundamental movement, The Functional Movement Screen (FMS), has been described by Cook et al. (6,7).

The FMS consists of a series of 7 fundamental movement tests designed to categorize functional movement patterns. The 7 movement tests use a variety of positions and movements closely related to normal growth and development. It is conceptualized that fundamental movements, such as those tested in the FMS, operate as the basis of more complex movement patterns used in common daily activities and sports. Preliminary data have established the relationship between an athlete’s functional movement characteristics, as measured by the FMS, and injury risk in professional football players (14). However, there is currently no data examining the interrater reliability of the FMS.

The overall goal of this study is to establish the interrater reliability of the FMS by comparing individuals with different levels of training on the FMS. Our study will aim to compare
Experimental Approach to the Problem
The FMS is a novel testing tool that is growing in popularity in the clinical setting; however, reliability between raters for the FMS has yet to be established. This information is important for clinical settings that may use more than 1 clinician to score the FMS. To determine the interrater reliability of the FMS, while controlling for previous experience, 4 raters each independently scored the subjects performing each component of the FMS. The raters consisted of 2 experts and 2 novices. An expert is defined as an individual who was instrumental in the development of the FMS with over 10 years of experience with the tool. Novice individuals were defined as having taken the standardized introductory training course and have used the FMS less than a year.

The scores of the 2 experts were compared, as were the scores of the 2 novices. Finally, scores comparing paired expert and novice raters were evaluated. The standardized scoring criteria originally described were used for this study (8). For each test, the level of agreement was calculated between each pair of raters. In this way, the need to understand the reliability of the FMS among clinicians will be addressed to determine if current training is adequate for scoring the FMS.

Subjects
Forty healthy college students (23 women, 17 men, average age 20.8, 13 varsity athletes) were recruited by word of mouth. By self-report, all subjects were free from injury and able to participate in desired physical activities. The purpose of the study was described to all potential subjects and each signed an informed consent approved by the University’s Institutional Review Board. Participants qualified for the study if they were at least 18 years of age with no self-reported spinal or extremity pain and were not currently under the care of a medical professional for any musculoskeletal complaint. One subject’s data were lost due to a video malfunction. The videos of the remaining 39 subjects were then viewed and scored by all the raters individually.

Procedures
The FMS is designed so that the 7 movement patterns are considered together as a comprehensive cross section of functional movement. These 7 tests, used to assess overall functional movement ability, include the deep squat (Figure 1), the hurdle step (Figure 2), the in-line lunge (Figure 3), the shoulder mobility test (Figure 4), the active straight leg raise (ASLR) (Figure 5), the trunk stability push-up (Figure 6), and the rotary stability test (Figure 7). There are 3 clearing tests, each associated with one of the individual FMS tests, which
**Figure 2.** Description of the scoring criteria used for the hurdle step component of the FMS.

**Figure 3.** Description of the scoring criteria used for the in-line lunge component of the FMS.
check for pain accompanying shoulder internal rotation/flexion and end-range spinal flexion and extension pain (Figure 8). All subjects performed each of the 7 tests while being videotaped from both anterior and lateral views. Two Sony Handycam (Sony Electronics Inc., Tokyo, Japan) video camcorders were used to record the participants performing the FMS movements. The Dartfish Connect software (Dartfish, Fribourg, Switzerland) was used to record and organize the videos.

The FMS is scored on a 0–3 ordinal scale. A score of 3 represents the subject's ability to perform the functional movement pattern as described, a score of 2 indicates that some type of compensation is present when completing the pattern, and a score of 1 is given when the subject is unable to perform the movement pattern. A zero is recorded if there is pain associated with any portion of the test including the clearing tests.

**Statistical Analyses**

The weighted Kappa statistic was calculated for each test between the 2 pairs of raters. The Kappa statistic is a measure of “true” agreement, beyond that which is expected by chance. It is a ratio of the proportion of time that raters agree, corrected for chance agreement, to the maximum proportion of times that the raters could agree (24). It takes the following form:

**Figure 4.** Description of the scoring criteria used for the shoulder mobility component of the FMS.

**Figure 5.** Description of the scoring criteria used for the active straight leg raise component of the FMS.
**Figure 6.** Description of the scoring criteria used for the trunk stability push-up component of the FMS.

**Figure 7.** Description of the scoring criteria used for the rotary stability component of the FMS.
Weighted Kappa ($\kappa_w$) reflects the degree of disagreement between raters by attaching greater emphasis to large differences between ratings than to small differences (24). It takes the following form:

$$\kappa_w = \frac{\sum w_f o - w_f c}{n - \sum w_f c},$$

with $\sum w_f o$ being the sum of the weighted observed frequencies and $w_f c$ being the sum of the weighted frequencies predicted by chance. Quadratic weights were used for the frequencies, with the quadratic weight of no difference equaling 1, disagreement by 1 category equaling 0.89, and disagreement by 2 categories equaling 0.56. The frequency of each score combination in the contingency table was multiplied by its respective quadratic weight to indicate the more significant effect of larger disagreement (24).

**RESULTS**

The tests were analyzed by addressing the kappa values of both individual right and left sides and the final score for each.
of the 7 tests. The pair of novice raters demonstrated excellent agreement on 6 of the 17 test components, including the deep squat and shoulder mobility tests, and portions of the trunk stability push-up and ASLR tests (21,24). Substantial agreement was evident on 8 of the 17 test components. The right and left components of the lunge and the final component of the rotary stability test each demonstrated moderate agreement (Table 1) (21,24).

The pair of expert raters varied more in scoring, with excellent agreement on 4 of the 17 test components, including the shoulder mobility test and the final component of the ASLR. Substantial agreement was seen in 9 of the 17 test components. Two components of the lunge and 2 components of the rotary stability tests demonstrated moderate agreement (Table 2) (21,24).

When comparing the average scores of the paired novice and expert raters, 14 of the 17 tests demonstrated excellent agreement. Substantial agreement was evident in 1 component of the rotary stability test ($k_w = 0.74$) and 2 components of the in-line lunge ($k_w = 0.74$, $x_w = 0.79$) (Table 3) (21,24).

**Discussion**

These data indicate that the FMS has high interrater reliability and can confidently be applied by trained individuals when the standard procedure is used. The majority of the averaged tests were ranked in the category of excellent agreement according to Portney and Watkins (21). However, there were noticeable differences between the 2 pairs of raters and between different tests. The variance in these results may be due to the experience of the raters, testing protocol, or unclearly defined scoring criteria.

Neither pair of raters fell below a moderate level of agreement. The novice raters had excellent agreement on 2 more test components when compared with the expert raters (6 vs. 4), although the percent agreement was similar (89.6% and 86.7% for novice and expert raters, respectively). When the novice raters were compared with expert raters, 14 of the 17 tests had excellent agreement. These results suggest that reliable scores should be obtained by individuals who have been trained in the standardized FMS program.

Additional scoring variability may be attributable to the inherent movement that was assessed. Some movements may benefit from a 3-dimensional approach as opposed to the 2-dimensional approach used in this protocol. The tests demonstrating the lowest $k_w$, the lunge and rotary stability tests, are best scored by evaluating all 3 planes of motion. In this study, one camera captured a straight frontal plane view and the second camera recorded the sagittal plane. Due to the 2-dimensional limitations of the video setup, it was difficult to accurately assess the extent of a participant’s extraneous movement in the transverse plane during these trials, and thus, a third camera may assist in completing the analysis of the movement. However, this protocol would be difficult to complete in the field setting and thus may not be practical.

A final component of the testing that may have contributed to the variance in scores may be associated with the scoring criteria. Some of the tests have less clearly defined descriptors of midrange performance. This is most appreciable in the lunge, hurdle step, and rotary stability tests. For these tests, the dichotomous extremes of performance are easily distinguishable; however, the division of the intermediate scores is less apparent. The results of this study are being used to review the current scoring criteria and determine if alterations are needed. The ability to clearly identify differential performance levels will assist coaches and trainers in improving interpretation of scores on the subsets to target specific areas of poor functional movement. Despite these limitations, the FMS remains a reliable measurement tool for functional movement analysis.

Future study designs should evaluate the reliability of real-time scoring of the FMS as compared with reviewing a video of the tests. Establishing the reliability of a real-time scoring protocol would allow for more rapid feedback on the test performance and reduce time spent on analysis. In the field setting, most professionals will not use video in administering and scoring the FMS. To provide efficient and immediate feedback to a large number of athletes, real-time analysis is most often used. This setting is an area for future research regarding the reliability of the FMS.

**Practical Applications**

The FMS is growing in popularity but has yet to be assessed for interrater reliability. The test is commonly used to assess

---

**Table 3.** Kappa values as determined by comparing the averaged scores of novice vs. expert raters (L = left, R = right; n = 39).*

<table>
<thead>
<tr>
<th>Test</th>
<th>Agreement (%)</th>
<th>Kappa</th>
<th>Level of agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Squat</td>
<td>100</td>
<td>1.0</td>
<td>Excellent</td>
</tr>
<tr>
<td>Hurdle R</td>
<td>93.6</td>
<td>0.87</td>
<td>Excellent</td>
</tr>
<tr>
<td>Hurdle L</td>
<td>94.9</td>
<td>0.89</td>
<td>Excellent</td>
</tr>
<tr>
<td>Hurdle final</td>
<td>92.3</td>
<td>0.83</td>
<td>Excellent</td>
</tr>
<tr>
<td>Lunge R</td>
<td>90.8</td>
<td>0.83</td>
<td>Excellent</td>
</tr>
<tr>
<td>Lunge L</td>
<td>87.2</td>
<td>0.74</td>
<td>Substantial</td>
</tr>
<tr>
<td>Lunge final</td>
<td>89.8</td>
<td>0.79</td>
<td>Substantial</td>
</tr>
<tr>
<td>Shoulder R</td>
<td>100</td>
<td>1.0</td>
<td>Excellent</td>
</tr>
<tr>
<td>Shoulder L</td>
<td>100</td>
<td>1.0</td>
<td>Excellent</td>
</tr>
<tr>
<td>Shoulder final</td>
<td>100</td>
<td>1.0</td>
<td>Excellent</td>
</tr>
<tr>
<td>ASLR R</td>
<td>97.0</td>
<td>0.94</td>
<td>Excellent</td>
</tr>
<tr>
<td>ASLR L</td>
<td>98.5</td>
<td>0.97</td>
<td>Excellent</td>
</tr>
<tr>
<td>ASLR final</td>
<td>97.0</td>
<td>0.94</td>
<td>Excellent</td>
</tr>
<tr>
<td>Push-up</td>
<td>97.5</td>
<td>0.96</td>
<td>Excellent</td>
</tr>
<tr>
<td>Rotary R</td>
<td>91</td>
<td>0.84</td>
<td>Excellent</td>
</tr>
<tr>
<td>Rotary L</td>
<td>83.4</td>
<td>0.74</td>
<td>Substantial</td>
</tr>
<tr>
<td>Rotary final</td>
<td>91</td>
<td>0.84</td>
<td>Excellent</td>
</tr>
</tbody>
</table>

*ASLR = active straight leg raise.
the movement patterns of athletes and to make decisions related to interventions for performance enhancement. The results of this study suggest that individuals who have undergone the standardized training protocol will score the FMS in a similar manner.

ACKNOWLEDGMENTS

The researchers would like to thank the University of Evansville Honor’s Program grant and the University of Evansville’s College of Education and Health Science for funding the research project. The 2 expert raters were originally involved in the conception of the FMS. The results of the present study do not constitute endorsement by the authors or the National Strength and Conditioning Association.

REFERENCES