FMS Scores Change With Performers’ Knowledge of the Grading Criteria—Are General Whole-Body Movement Screens Capturing “Dysfunction”? 

David M. Frost,1 Tyson A.C. Beach,2 Jack P. Callaghan,1 and Stuart M. McGill1

1Department of Kinesiology, University of Waterloo, Waterloo, Ontario, Canada; and 2Faculty of Kinesiology and Physical Education, University of Toronto, Toronto, Ontario, Canada

Abstract

Frost, DM, Beach, TAC, Callaghan, JP, and McGill, SM. FMS scores change with performers’ knowledge of the grading criteria—Are general whole-body movement screens capturing “dysfunction”? J Strength Cond Res 29(11): 3037–3044, 2015—Deficits in joint mobility and stability could certainly impact individuals’ Functional Movement Screen (FMS) scores; however, it is also plausible that the movement patterns observed are influenced by the performers’ knowledge of the grading criteria. Twenty-one firefighters volunteered to participate, and their FMS scores were graded before and immediately after receiving knowledge of the movement patterns required to achieve a perfect score on the FMS. Standardized verbal instructions were used to administer both screens, and the participants were not provided with any coaching or feedback. Time-synchronized sagittal and frontal plane videos were used to grade the FMS. The firefighters significantly (p < 0.001) improved their FMS scores from 14.1 (1.8) to 16.7 (1.9) when provided with knowledge pertaining to the specific grading criteria. Significant improvements (p < 0.05) were also noted in the deep squat (1.4 [0.7]–2.0 [0.6]), hurdle step (2.1 [0.4]–2.4 [0.5]), in-line lunge (2.1 [0.4]–2.7 [0.5]), and shoulder mobility (1.8 [0.8]–2.4 [0.7]) tests. Because a knowledge of a task’s grading criteria can alter a general whole-body movement screen score, FMS or otherwise, observed changes may not solely reflect “dysfunction.” The instant that individuals are provided with coaching and feedback regarding their performance on a particular task, the task may lose its utility to evaluate the transfer of training or predict musculoskeletal injury risk.

Key Words  firefighter, injury, prevention, prediction, risk

Introduction

Movement evaluations and preparticipation screens have been widely adopted by scientists and practitioners, given that links have been made between individuals’ movement behaviors and their risk of sustaining an injury (14,40). Although there is a wealth of information that can be acquired by observing the way people move, the interpretation regarding underlying mechanism(s) responsible for a specific behavior can be obscured and mitigated by many variables (11). One of these potential confounders is the issue of whether a knowledge of a task’s scoring criteria can change how individuals perform. If someone can influence their score based on their knowledge or understanding of the test, the outcomes of any strategy to prevent injury and improve performance, be it coaching or exercise related, could be compromised.

Consider the following example: when individuals exhibit lumbar spine flexion and frontal plane knee motion while performing a bodyweight squat, their movement behavior cannot be attributed to any 1 factor without using sophisticated equipment. Their hip range of motion, hamstring length, quadriceps and gluteal strength, or cardiovascular efficiency could all contribute to the motion pattern observed, yet it would be impossible to cite an exact reason as to why they adopted this movement pattern using visual observations alone. There is also evidence to suggest that their movement behavior may not be caused by any of these things; an individual’s perception of risk (35), prior experience (26), understanding of the task (15), focus of attention (36), motivation (7), and simply being aware of the fact that movement might matter (33) can also influence the way they squat. As a result, assuming that any pattern is the product of “movement dysfunction” and is in need of “corrective” exercise may be entirely unfounded.

When the aim of a movement screen is to predict one’s risk of sustaining a future musculoskeletal injury or guide recommendations for training, it may be more appropriate to acquire information pertaining to the individual’s habitual...
or preferred behaviors than those employed in the presence of instruction, coaching, or feedback. This is true, particularly, given that individuals may move in a manner that reflects their interpretation of what was instructed, demonstrated, or expected. Accordingly, it may be difficult to evaluate individuals’ engrained patterns if the performers are made aware of the screen’s objectives or the criteria being used for grading (i.e., Hawthorne effect). If capable, those persons being screened could adapt their movement patterns to perform in a manner that they perceive to be “good,” irrespective of their preferred habitual behavior. A similar outcome may arise when individuals are offered an opportunity to practice or rehearse the screening tasks; their performance on the test could vary over time. In fact, in

<table>
<thead>
<tr>
<th>Test</th>
<th>Subject</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deep Scoat</td>
<td>1</td>
</tr>
<tr>
<td>Hurdle Stop</td>
<td>2</td>
</tr>
<tr>
<td>In-line Lunge</td>
<td>3</td>
</tr>
<tr>
<td>Shoulder Mobility</td>
<td>4</td>
</tr>
<tr>
<td>Straight Leg Raise</td>
<td>5</td>
</tr>
<tr>
<td>Stability Push-up</td>
<td>6</td>
</tr>
<tr>
<td>Rotary Stability</td>
<td>7</td>
</tr>
</tbody>
</table>

**Figure 1.** The grades assigned to each individual before and immediately after receiving feedback regarding the movement patterns required to achieve a perfect score.
a previous study documenting movement screen changes before and after 12 weeks of exercise (10), “control” subjects exhibited as many changes (positive and negative) as did those involved with training. Clearly, in such cases, it is difficult to know if the changes observed have had any impact on injury risk.

The Functional Movement Screen (FMS) is a 7-task test that was developed as a low-cost means to “red-flag” potential problems within individuals’ movement system that may predispose them to future injury (3). As an instrument to discriminate who will and who will not become injured, it has shown some promise; a number of studies have shown a relationship between FMS scores and injury (2,20,27,30), as has been reported for hamstring flexibility (21,23), cardiovascular efficiency (17,23), and body mass index (22,25). Although there is also contradictory evidence to suggest that FMS scores are not linked with injury (1,16,38,39), it is important to recognize that even if relationships exist, the FMS has not been evaluated as an assessment or diagnostic tool nor was it intended for such purposes (4). At present, there is no evidence to suggest that a particular FMS score accurately or reliably reflects the presence of movement “impairments” or “dysfunction.” Deficits in joint mobility and stability could certainly impact individuals’ FMS scores, and it is also plausible that the scores are influenced by the performers’ awareness and appreciation for the criteria being used to grade their performance. With this in mind, the objective of this investigation was to compare individual FMS scores before and immediately after providing them with feedback regarding the movement patterns required to achieve a perfect score on the test. It was hypothesized that scores would change as individuals adapted their movement in an attempt to meet the grading criteria.

**METHODS**

**Experimental Approach to the Problem**

A repeated measures study design was used to examine whether a knowledge of the grading criteria influences FMS task performance. The FMS was administered using published verbal instructions (5) to a group of professional firefighters. No feedback was given, nor were the objectives of the screen described during this first screen.
3 minutes of completing the initial test, the participants were asked to perform the FMS for the second time but were provided with a verbal description of the criteria used to grade each of the 7 screening tasks immediately before performing each task. The instructions provided were standardized across the participants, and no specific feedback was offered regarding any individual’s original FMS score. A research assistant blinded to the testing procedures graded the prescreen and postscreen using video, and comparisons were made between the task and total FMS scores.

Subjects
Twenty-one firefighters (19 men and 2 women) from Waterloo Fire Rescue (Waterloo, ON, Canada) and the Kitchener Fire Department (Kitchener, ON, Canada) were recruited to participate in this investigation. All men and women were free of musculoskeletal injury and pain at the time of testing and on full active duty. Their mean (SD) age, height, and body mass were 33.4 years (7.0), 1.81 m (0.05), and 89.7 kg (14.6), respectively. The University’s Office of Research Ethics approved the investigation, and all the participants gave their informed consent before the data collection began.

Functional Movement Screen
The FMS is a 7-task test comprising fundamental movement patterns that require a balance of joint mobility and neuromuscular control (3). It was designed as a simple tool that could be used to identify compensatory motions, imbalances, or asymmetries before the onset of exercise. The 7
screening tasks are as follows: (a) deep squat (SQT)—a dowel is placed over the head with the arms outstretched and the individual squats as low as possible; (b) hurdle step (HRD)—a dowel is placed across the shoulders and the individuals step over a hurdle placed directly in front of them; (c) in-line lunge (LNG)—with the feet aligned and a dowel contacting the head, back, and sacrum the individual performs a split squat; (d) shoulder mobility (SHR)—the individuals attempts to touch their fists together behind their back (internal and external shoulder rotation); (e) active straight leg raise—while lying supine with their head on the ground the individuals actively raise 1 leg as high as possible; (f) trunk stability push-up (PSH)—the individuals perform a push-up with their hands shoulder width apart; (g) rotary stability (ROT)—the individuals assume a quadruped position and attempt to touch their knee and elbow, first on the same side of the body and then on the opposite side. Additional tests are also included with the SHR, PSH, and ROT to expose other potential sources of pain that may be overlooked during the primary task performance. Further details of each task have been published previously (3–6,13,28,32).

**Testing Procedures**

Upon arrival, the participants were told that they would be asked to perform a series of tasks to observe the way they move. The FMS was administered by an FMS certified instructor using the verbal instructions outlined by Cook et al. (5). No further feedback was given, and the participants were blinded to the test objectives, scoring criteria, and their screen results. Individuals were made aware that they were not being “tested,” but rather they were simply being “observed,” to ensure that their performance was as natural as possible. Four repetitions (2 forwards and 2 backwards) of each task were performed, and as recommended (3,4), the “best” repetition was graded after the completion of the study using synchronized video collected from the sagittal and frontal planes.

Upon completion of the initial screen, the participants were asked to complete the FMS for the second time within 3 minutes. The same verbal instructions were given in addition to a verbal description of the specific movement criteria that are used to assign scores for each of the 7 FMS tasks (5). The movement criteria were described to the participants precisely as outlined by Cook et al. (5). At no time were the participants given feedback regarding their initial task performance, nor were they “coached” while completing the first or second screen.

**Statistical Analyses**

Video was used to objectively assign FMS task scores using the methods outlined in (3,4). Briefly, a score of 0–3 was assigned to each task whereby a 3, 2, 1, and 0 represented “performed without compensation” (according to relevant criteria), “performed with compensation,” “could not perform” (according to relevant criteria), and “pain,” respectively (3,4). Tasks requiring performances of the left and right sides of the body were scored independently but were given a task grade equal to that of the lowest score. The cumulative sum of all 7 tasks represented the total FMS score (21 was the highest score possible). All the screens were assigned alphanumeric codes by an independent researcher and graded 3 months after the investigation by an independent rater so that the order of testing could not be differentiated. In addition to assigning a grade to each task, a record of all observations was documented such that any changes could be attributed to specific criteria.

The number of participants exhibiting each of the “compensatory” patterns listed as a scoring criterion was documented, and Wilcoxon signed-rank tests (given nonparametric data) were used to investigate the between-session differences in each screening task and the total FMS score. An alpha level of 0.05 was identified a priori as necessary to achieve statistical significance.

**Results**

Participants’ mean (SD) FMS score increased (p < 0.001) from 14.1 (1.8) to 16.7 (1.9) when they were provided with a knowledge of the scoring criteria. Significant improvements (p < 0.05) were also noted in the SQT (1.4 [0.7]–2.0 [0.6]), HRD (2.1 [0.4]–2.4 [0.5]), LNG (2.1 [0.4]–2.7 [0.5]), and SHR (1.8 [0.8]–2.4 [0.7]) tests (Table 1). The scores assigned to each participant on both tests are given in Figure 1.
The number of participants exhibiting changes in each of the relevant scoring criterion during the second test is illustrated in Figure 2. With the exception of the SHR screen in which there is only 1 criterion, participants’ improvements could not be attributed to a single observation. Further, not all changes were positive; a number of individuals exhibited “compensatory” motion during the second screen when originally they had not (e.g., femur below horizontal during the SQT).

**Discussion**

As was originally hypothesized, participants’ FMS score improved in minutes simply by providing them with a knowledge regarding the specific criteria required to achieve a perfect score. In fact, with the exception of 1 participant, every firefighter achieved a higher FMS score when they performed the screen the second time. Significant changes were also observed in the SQT, HRD, LNG, and SHR tests, though these changes could not be attributed to any 1 criterion. The participants seemed to have adapted their movement in a variety of ways, which may reflect their understanding or interpretation of the instructions that were provided, or their familiarity with the tasks.

Individuals’ FMS scores can certainly be influenced by their mobility, flexibility, stability, strength, endurance, cardiorespiratory efficiency, and any other measure that might be used to characterize their level of “fitness,” but it is important to recognize the fact that movement screens of this nature cannot distinguish between these abilities when the raters’ interpretations are based solely on visual observation. Previous work has suggested that individuals having a total FMS score <14 may be at an increased risk of sustaining an injury because their movement is constrained by “impairments” such as limited joint mobility and stability (19). In this study, 9 participants received a grade <14 on the initial screen, which based on the above mentioned findings, would be reason to introduce “corrective” exercises to improve the supposed movement “dysfunction” (18,19). But perhaps, instead, the firefighters’ movement behaviors were influenced by several factors including their interpretation of the task instructions, previous experience, and motivation to perform; alternatively, they may not have been aware of the criteria being used to grade their performance. After receiving information regarding the movement patterns required to achieve a perfect score, only 1 of the 9 “at-risk” individuals obtained a score <14.

When seeking to identify at-risk individuals, there is arguably more value in assessing performers’ habitual or preferred movement behaviors as opposed to what they can do when provided with task-specific feedback. This implies that participants’ preliminary FMS score may in fact have been a better indicator of risk, irrespective of the reasons why they moved in a particular manner. However, if this true, and the FMS can be used to establish risk, the reaction to a low FMS score should not be to improve one’s performance on the screen. The results from this study showed that this could be accomplished immediately by providing a knowledge and further instruction. Citing the specific criteria that are used to assign FMS scores was able to influence participants’ performance without addressing any specific movement pattern. Reducing one’s injury risk will likely require an intervention be designed to change their habitual movement behaviors across a range of tasks relevant to their life’s demands, and to date, there is no evidence to suggest that this can be accomplished by improving their FMS. Kiesel et al. (18) did report an average increase of 3 points after a 7-week exercise intervention designed to improve the participants’ FMS score. More specifically, the stated objective was to correct the identified movement deficits and normalize “dysfunction” by employing specific “corrective exercises.” However, it is difficult to interpret these findings given that no control group was included in the study and the results were near identical to those reported here, wherein participants’ scores increased irrespective of any potential deficits in their movement system.

Consistent with previous interpretations (13), the results of this study raise concerns about the test–retest reliability of the FMS. There is some evidence to suggest that FMS scores are reliable between days (32,37), but it is not clear if or how much augmented feedback was provided in the investigations cited. It is plausible that “cueing” stabilizes FMS scores, possibly leading to acceptable levels of test–retest reliability. However, Frost et al. (13) elected against such practices given that the study objective was to investigate the transfer of training to general movement tasks. In their study, the participants indicated that they understood the standardized instructions provided and performed FMS tasks in the absence of additional feedback or instruction. It was found that control subjects (i.e., those not participating in the exercise programs) exhibited bidirectional changes when screened for a second time. Interestingly, had Frost et al. (13) used only the FMS to gage the effectiveness of the 2 training interventions examined (fitness training and movement-oriented fitness training), they may have concluded that neither training program was effective because each group’s mean FMS score did not change. However, further analyses of the data revealed a number of movement-related training adaptations that were undetected by the FMS (10); the authors found that objective kinematic data (i.e., spine and frontal plane knee motion) were able to differentiate between the 2 training and control groups. A case can be made against using the FMS as a transfer test, but it is not yet clear that training to increase FMS scores is any more warranted. Results of this study suggest that increased FMS scores could be achieved artificially without necessarily altering injury risk, and there is little evidence that increasing FMS scores would improve athletic or occupational performance (31,34).

There is obviously merit in using qualitative interpretations of movement patterns to assist with the prediction of
injury risk or to personalize training recommendations; however, when deciding on an observation strategy, it may be prudent to first identify the possible mechanisms for the injuries of interest so that “key features” of each motion pattern being tested can be used as grading criteria (24). Second, a task’s demands (e.g., load) have been shown to impact individuals’ movement strategies (11). Therefore, using visual observations of a low-demand screen to predict how individuals will perform a more challenging activity that reflects the demands of their sport, occupation, or life could be misleading. This notion is illustrated in Figure 3. As part of a larger project (9), the firefighters in this investigation were asked to perform several tasks to simulate the demands of their occupation while instrumented for biomechanical analyses. Despite obtaining FMS scores of 16 and 20 on his prefeedback and postfeedback screens, respectively, the individual depicted in Figure 3 exhibited substantial frontal plane knee motion and a large frontal plane knee moment while performing a simulated hose advance. Both of these joint motions have been cited as risk factors for sustaining an anterior cruciate ligament injury (14). The inclusion of higher demand tasks such as this may not provide clarity as to the underlying mechanism responsible for the movement behavior employed, but could help to establish an improved sensitivity and specificity for estimating risk, and therefore more appropriate recommendations for training.

Movement evaluations and preparticipation screens are being used and investigated by scientists and practitioners from across the world (3,8,29). However, it is important to appreciate the assumptions that are made when individuals are evaluated solely by visual observation (the FMS is only one such application). In this study, the improvements observed were influenced by changing participants’ focus of attention to specific features of their movement patterns not previously considered, but it would be irresponsible to suggest that this is the only reason that the observed changes occurred. It is also possible that the firefighters were more motivated the second time they were screened or that they better understood the original instructions. The results may have also been influenced in part by the inherent variability in individuals’ movement behavior (12). That said, regardless of the exact reason as to why participants changed, the fact is that they improved their FMS score within minutes. The moment that an individual is provided with coaching, knowledge, or feedback regarding their performance on a particular task, the task may lose its utility to evaluate the transfer of training or predict one’s risk of injury if the performers have some knowledge of the tasks’ grading criteria. Whether or not the FMS becomes a viable method to predict future injury, the results of this study suggest that future efforts should not be directed to improve individuals’ performance on the test itself given that this objective could be accomplished artificially without actually impacting injury risk or athletic and occupational performance.

ACKNOWLEDGMENTS

The authors would like to extend their gratitude to each member of Waterloo Fire Rescue and the Kitchener Fire Department for their involvement with this work. The funding for this project was provided by the Centre of Research Expertise for the Prevention of Musculoskeletal Disorders and the Natural Sciences and Engineering Research Council of Canada. Dr. Jack Callaghan is supported by the Canada Research Chair in Spine Biomechanics and Injury Prevention. The results of this study do constitute endorsement by the National Strength and Conditioning Association.

REFERENCES

Interpreting a Movement Screen Score


