INTERRATER AND INTRARATER RELIABILITY OF THE FUNCTIONAL MOVEMENT SCREEN

CRAIG A. SMITH,[1] **NICOLE J. CHIMERA,**[2] **NICHOLAS J. WRIGHT,**[1] **AND MEGHAN WARREN**[1]

[1]Department of Physical Therapy and Athletic Training, Northern Arizona University, Flagstaff, Arizona; and [2]Department of Athletic Training, Daemen College, Amherst, New York

ABSTRACT

Smith, CA, Chimera, NJ, Wright, NJ, and Warren, M. Interrater and intrarater reliability of the functional movement screen. *J Strength Cond Res* 27(4): 982–987, 2013—The purpose of this study was to investigate interrater and intrarater reliability of the Functional Movement Screen (FMS) with real-time administration with raters of different educational background and experience. The FMS was assessed with real-time administration in healthy injury-free men and women and included a certified FMS rater for comparison with other raters. A relatively new tool, the FMS, was developed to screen 7 individual movement patterns to classify subjects' injury risk. Previous reliability studies have been published with only one investigating intrarater reliability. These studies had limitations in study design and clinical applicability such as the use of only video to rate or the use of raters without comparison to a certified FMS rater. Raters ($n = 4$) with varying degrees of FMS experience and educational levels underwent a 2-hour FMS training session. Subjects ($n = 19$) were rated during 2 sessions, 1 week apart, using standard FMS protocol and equipment. Interrater reliability was good for session 1 (intraclass correlation coefficient [ICC] = 0.89) and for session 2 (ICC = 0.87). The individual FMS movements showed hurdle step as the least reliable (ICC = 0.30 for session 1 and 0.35 for session 2), whereas the most reliable was shoulder mobility (ICC = 0.98 for session 1 and 0.96 for session 2). Intrarater reliability was good for all raters (ICC = 0.81–0.91), with similar ICC regardless of education or previous experience with FMS. The results showed that the FMS could be consistently scored by people with varying degrees of experience with the FMS after a 2-hour training session. Intrarater reliability was not increased with FMS certification.

KEY WORDS clinical test, function, injury risk, movement pattern

INTRODUCTION

Musculoskeletal and soft tissue injuries are common occurrences during exercise and athletics (8,14). According to the National Collegiate Athletic Association Surveillance System, approximately 15.47 contact and noncontact injuries occurred per 1000 exposures, where exposure equaled 1 practice or game (8). Injury due to exercise has not been limited to the playing field. During the 8 weeks of U.S. Army basic training, between 23–28% of men and 42–67% of women suffered an injury that required medical care (14).

Previous literature (8,14) has classified injury risk factors into 2 categories: intrinsic and extrinsic (14). Intrinsic risk factors are inherent aspects within the person such as age, gender, fitness level, previous injury, and strength and flexibility imbalances (13). Extrinsic factors are features outside the body such as playing field, training program, and shoe wear. These 2 categories are further subdivided into potentially modifiable and nonmodifiable factors. Nonmodifiable factors such as age, gender, or previous injury are unchangeable, whereas modifiable are risk factors that are amenable to intervention. Therefore, detection of modifiable intrinsic and extrinsic risk factors is critical to decrease risk and rate of musculoskeletal injuries associated with exercise and athletics (8,14).

A relatively new tool, the Functional Movement Screen (FMS), was developed to qualitatively screen basic movement patterns used in athletics to assess intrinsic risk for injury based on the ability to perform certain movements with or without compensation (1,3–5,9,10). The FMS has been described as an injury predictor with a score below 14 associated with an increased risk of serious injury in professional football players (12). Since this initial study, there has been limited investigation into the use of the FMS to predict injury in active populations (2,16) and to determine the effectiveness of interventions (6,7,11), which may allow for the FMS to be used as screen for a modifiable risk factor for injury. The FMS has not been validated for prediction of injury among active populations, but the limited current literature has demonstrated an increase of injury with scores below 14 (2,12,16).

Despite the published research using the FMS and increased use in clinical settings, the reliability has not been adequately substantiated for clinical use. Without a clear

Address correspondence to Meghan Warren, meghan.warren@nau.edu.

27(4)/982–987

Journal of Strength and Conditioning Research

determination of reliability, clinicians are limited in the ability to use this tool to measure change. Differences between scores within a person being tested could be because of measurement error in scoring vs. actual change. Furthermore, scores between clinicians could vary greatly because of the interpretation of the FMS scoring criteria. Therefore, understanding the reliability and consistency of this screening tool is critical if the FMS is to be used to detect an intrinsic modifiable risk factor in clinical settings.

Previous interrater reliability studies for the FMS using the standard scoring criteria have been published (15,18,19) with only 1 investigating intrarater reliability (19). Although scores were found to be good (15,18) to moderate (19), there were significant limitations to applicability in clinical settings. The use of video only to rate subjects lacks real-time administration (15), whereas the use of novice raters excludes comparison with a certified FMS rater (19). In response to these limitations, a study with real-time administration performed by raters of varied backgrounds and experiences with the FMS was necessary.

The purpose of this study was to investigate the interrater and intrarater reliability of the FMS with real-time administration with raters of different educational background and experience levels with healthy injury-free men and women. The first hypothesis for this study was that interrater and intrarater reliability for the FMS will be good during real-time administration. The second hypothesis was that FMS certification would result in increased intrarater reliability during real-time administration.

METHODS

Experimental Approach to the Problem

The purpose of this study was to investigate the interrater and intrarater reliability of the FMS between raters of different experience and education with real-time administration in healthy injury-free men and women. The first hypothesis of this study was that the FMS demonstrates good interrater and intrarater reliability. The second hypothesis was that FMS certification would result in increased intrarater reliability during real-time administration. Therefore, to determine interrater reliability of the FMS, the study was designed to have 4 raters simultaneously score the FMS with each subject using standard instructions as per FMS guidelines during real-time administration. To determine intrarater reliability, all subjects would return 1 week later for retesting by the same raters.

Subjects

A convenience sample recruited from a university physical therapy program consisted of 20 healthy, injury-free, and physically active men ($n = 10$) and women ($n = 10$) aged 26 (median [range]) [22–41] years and overall normal body mass index ([BMI]; mean $\pm$ $SD = 24.0 \pm 2.9\ \text{kg}\cdot\text{m}^{-2}$) during the fall and winter seasons. The subjects received a full explanation of the nature, purpose, and risks of the study and were given the opportunity to ask any questions. After the explanation, subjects signed an informed consent document approved by the Institutional Review Board at Northern Arizona University before participation in the study. Subjects then completed the Physical Activity Readiness Questionnaire (PAR-Q) (20) to ascertain if screening by a health care professional was required before participating in exercise. Potential subjects were excluded from the study if there were any "yes" answers on the PAR-Q. One female subject was excluded from participation on the day of the first testing session because of an unrelated musculoskeletal injury sustained the day before testing leaving 19 subjects in this study.

Functional Movement Screen

The movement screen was performed using standard FMS equipment (Perform Better, Cranston, RI, USA) in a series of 7 fundamental movements scored using a 4-point ordinal scale (3–0) to obtain a total score (21–0) and 3 clearing tests to determine the presence of pain (3–5). The 7 fundamental movements include deep squat, hurdle step, inline lunge, shoulder mobility, active straight leg raise, trunk stability push-up, and rotational stability. The 3 clearing tests of the FMS included the shoulder impingement test, spinal extension test, and spinal flexion test. The clearing tests were used to provoke pain that may not be uncovered in the 7 movements. The score for the FMS was determined by specific criteria for each of the 7 fundamental movements–movement with a score of 1 corresponding to inability to perform the movement, 2 corresponding to performing the movement with compensation, and 3 corresponding to the ability to correctly complete the movement without compensation. If there was pain during any portion of the movement or pain with a clearing test, a score of 0 was given. Asymmetry is noted in 5 movements performed bilaterally: hurdle step, inline lunge, shoulder mobility, active straight leg raise, and rotational stability. For example, the right inline lunge was performed completely without compensation, whereas the left side could not maintain balance during the movement. Asymmetry between each side was recorded by scoring the right inline lunge as a 3 and the left as 1. The lowest score of the movements performed bilaterally was used to calculate the total; all 7 movement scores were summed to determine the total FMS score with an overall range of 0–21.

Standardized FMS instructions and protocol were adapted from the most recent FMS text (3). To standardize the testing between subjects, the instructions were recorded on audio recorder (Olympus, Center Valley, PA, USA) with pictures taken from a digital camcorder (Sony, Tokyo, Japan) for each setup and starting position. The audio and pictures that were combined into videos using Photoshop Premiere (Adobe, San Jose, CA, USA) were shown to each subject before the movement was performed.

Raters

Four raters of varied experience and background completed the FMS screening with each subject. Raters included an entry-level physical therapy student who had completed over

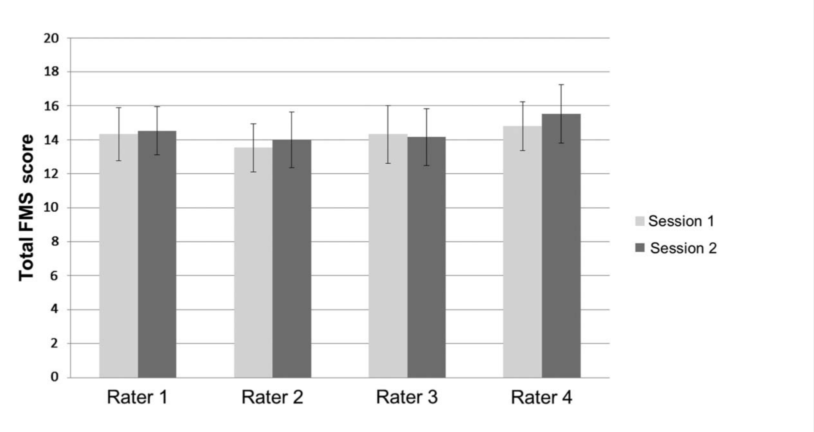

Figure 1. Mean and *SD*s for total Functional Movement Screen (FMS) scores for 4 raters for sessions 1 and 2.

100 FMS tests, but was not certified (rater 1); a certified FMS tester (rater 2); a faculty member in Athletic Training with a PhD in Biomechanics and Movement Science, but no previous experience with FMS (rater 3); and an entry-level physical therapy student with no previous experience with FMS (rater 4). A certification process has been developed by the creators of the FMS (3), which requires attendance of a 1-day seminar and successful completion of a certification exam.

A 2-hour training session was conducted using materials from the FMS developers (3). This session was led by the noncertified experienced FMS tester (rater 1) and covered the 7 movements, the 3 clearing tests, the verbal instructions, and the scoring criteria (3). All raters were provided scoring instructions as per the FMS guidelines for each of the movements and clearing tests that included the written and visual description of scoring criteria from 3 to 1.

Procedure

Each subject took part in 2 testing sessions (sessions 1 and 2) separated by 7 days. To limit any learned effects between sessions, subjects were instructed to not practice the FMS movements after session 1. To control any variations from session 1 to session 2, subjects returned for testing at the same time 1 week later and were also asked to wear the same self-selected shoes for both testing sessions. The subjects were instructed to follow the same routine from session 1 to session 2 without any changes in daily activities.

At session 1, height and weight were measured with a wall-mounted stadiometer and digital scale (Cardinal Scale, Webb City, MO, USA). Each subject began both sessions with a 5-minute warm-up on a cycle ergometer (Monark Exercise AB, Stockholm, Sweden). After warming up, all

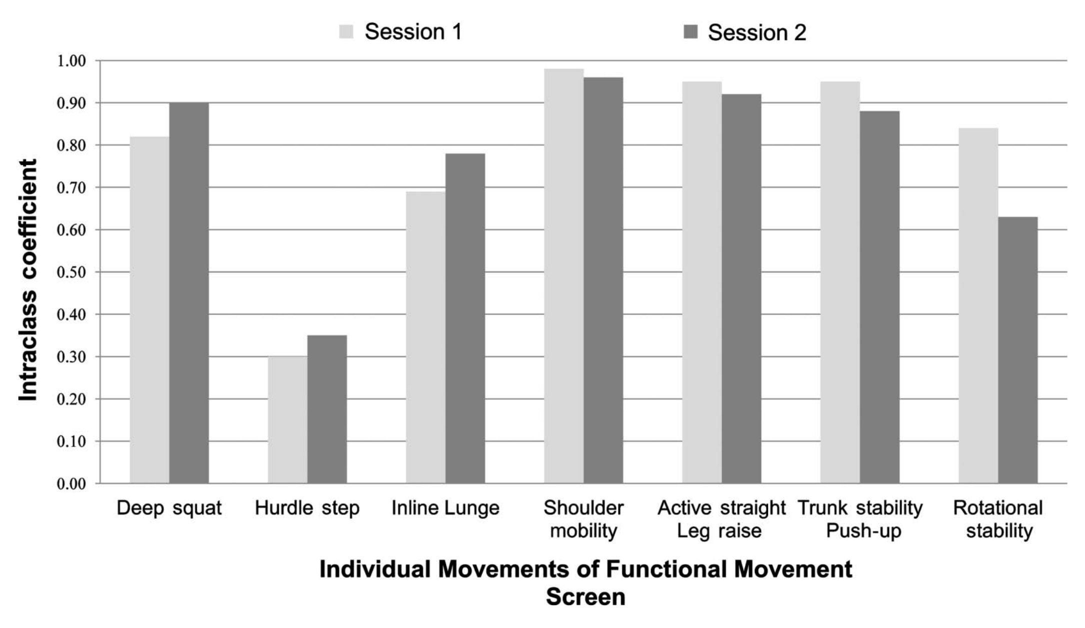

Figure 2. Interclass correlation coefficients for individual movements of the Functional Movement Screen.

TABLE 1. Intrarater reliability intraclass correlation coefficients for each rater for Functional Movement Screen.*

Rater number	Rater description	ICC	95% Confidence interval for ICC
1	Physical therapy student. Noncertified experienced FMS rater	0.90	0.76–0.96
2	Certified FMS rater	0.81	0.57–0.92
3	Athletic training faculty member. Noncertified inexperienced FMS rater	0.91	0.78–0.96
4	Physical therapy student. Noncertified inexperienced FMS rater	0.88	0.72–0.95

*ICC = intraclass correlation; FMS = Functional Movement Screen.

4 raters measured the tibial tuberosity height (from the floor to the tibial tuberosity) and hand length (from the distal point of the longest finger to the distal crease of the wrist) using the equipment from the FMS kit following standard FMS instructions (3). Tibial tuberosity and hand length measurements of rater 1 were used during all testing sessions for hurdle step, inline lunge, and shoulder mobility. The subject was instructed to listen to the videos, assume the proper starting position, and repeat the movement 3 times using the standard FMS board. An additional repetition was performed if needed by the raters to determine the score. After the last movement, the video for the next movement in the FMS was played. If the subject needed more instruction, the video was replayed; however, no additional information was verbally provided to the subjects. The same procedure was followed for session 2 to determine intrarater reliability; interrater reliability was computed from both session 1 and 2.

Statistical Analyses

Descriptive statistics were calculated as mean values with *SD* for normal interval data and medians with range or percent for non-normal and categorical data. Intraclass correlation coefficients (ICCs) from repeated-measures analysis of variance were calculated to determine intrarater and interrater reliability of overall FMS scores and individual components of the FMS. The clinical significance was defined as poor for an ICC below 0.50, moderate for 0.50–0.75, and good for 0.75 or higher (17). The interrater reliability was determined from comparison between the 4 individual raters. The intrarater reliability was determined from within rater comparison from session 1 and 2 separated by 7 days. All analyses were completed using SAS, version 9.2 (SAS Institute, Inc., Cary, NC, USA).

RESULTS

The overall FMS scores ranged from 11 to 17 with a mean ± *SD* of 14.3 ± 1.5 (using testing session 1 and rater 1 as an example). The mean values and *SD*s for total FMS scores of each rater for sessions 1 and 2 are shown in Figure 1.

Interrater reliability was good for session 1 (ICC = 0.89; 95% confidence interval [CI]: 0.80–0.95) and for session 2 (ICC = 0.87; 95% CI: 0.76–0.94). Additionally, there was 100% agreement with the 3 clearing tests among all the raters. Interrater reliability was high for measurements of hand length (ICC = 0.96 [95% CI: 0.92–0.98] for session 1 and 0.97 [95% CI: 0.94–0.99] for session 2) and tibia length (ICC = 0.99 [95% CI: 0.98–1.00] for sessions 1 and 2). Examination of the individual tests of the FMS showed hurdle step as the least reliable (ICC = 0.30 [95% CI: 0.08–0.57] for session 1 and 0.35 [95% CI: 0.12–0.61] for session 2), whereas the most reliable was shoulder mobility (ICC = 0.98 [95% CI: 0.96–0.99] for session 1 and 0.96 [95% CI: 0.92–0.98] for session 2) (Figure 2).

Intrarater reliability was good for each rater (Table 1) ranging from an ICC of 0.91 (95% CI: 0.78–0.96) for rater 3 to an ICC of 0.81 (95% CI: 0.57–0.92) for rater 2. Although good, the lowest ICC was the certified FMS rater (rater 2), whereas the highest ICC was found to be the faculty member with extensive experience in movement analysis (rater 3).

DISCUSSION

With the use of the FMS in different clinical settings by multiple health professionals (3), the determination of reliability with different raters during real-time administration is critical. The purpose of this study was to determine the reliability of the FMS during real-time administration. The first hypothesis of this study was that good interrater and intrarater reliability would be found. The second hypothesis was that FMS certification would result in increased intrarater reliability during real-time administration. To date, the reliability during real-time administration has been unknown; thus, this study adds an important clinically meaningful component to the current body of literature.

Previous literature (15,18,19) demonstrated good to moderate reliability, but limitations in study designs warranted further study (15,18,19). Minick et al. (15) used video recordings of subjects performing the FMS and tested the ability of 2 expert and 2 novice raters' scores based on anterior and lateral viewpoints. The use of video with set viewpoints does not replicate real-time FMS administration or scoring that requires the clinician to move and observe the subject from multiple viewpoints during a single trial (15). Furthermore, all 4 raters had significant instruction and practice with the FMS. The 2 novice raters had completed an entire standardized course in the FMS (15). The novices had used the FMS clinically for less than 1 year, but many clinicians have not had formal education into the administration of the FMS. After this initial research into reliability, 2 studies investigated the FMS without videos but did not include comparison to certified raters that limits the generalizability of the findings (18,19). These studies found good (15,18) to moderate (19) interrater reliability and good intrarater reliability (19), but the limitations have not allowed for clear determination of reliability with a real-time administration of the FMS.

The results of this study supported the hypothesis that the FMS would have good interrater and intrarater reliability. Interrater reliability for the total FMS score was good. The analysis of the separate components of the FMS showed that all movements had good to moderate reliability, expect for the hurdle step. Although contrary to previous data, which indicated lower reliability scoring the rotational stability movement (15), the poor reliability of the hurdle step when scored by multiple raters could be related to the scoring requirements. The subject must perform the following criteria to receive a score of 3: remain upright and balanced while keeping the ankle, knee, and hip in sagittal plane alignment. If the criteria are not achieved, then a lower score is given. The inherent difficulty with real-time administration and different viewpoints for separate criteria may have made scoring more difficult for the hurdle step movement, which may not have been encountered in previous study that used video-based analysis (15).

The intrarater reliability was good for all raters. However, contrary to the hypothesis that FMS certification would improve intrarater reliability, the certified FMS Rater (rater 2) had the lowest intrarater ICC. Rater 2 demonstrated good intrarater reliability (ICC = 0.81), but certification did not seem to improve scoring consistency from session 1 to session 2. Furthermore, the most consistent rater had no previous FMS experience but did have the most education and experience in movement analysis (rater 3; ICC = 0.91). In this small study, higher intrarater reliability appeared to be more related to education in movement analysis than FMS certification. The assumption underlying this conclusion was that subjects did not change from session 1 to session 2, which may not be true. Subjects were instructed not to practice the movements at home, and the 2 sessions were a week apart to minimize any learning effect. Despite these steps to limit any changes in movement during the FMS, subjects may have practiced or demonstrated a learned effect from session 1 to session 2. If rater 2 was more sensitive to these changes than the other raters, the ICC of rater 2 would have been lowered.

Because the goal of this study was to assess the ability to consistently score the FMS by raters of varied experience during real time to ensure clinical meaningfulness, sessions were not recorded by videotape. Thus, a limitation to our study was that we were unable to perform post hoc examination of each score for each movement. Because the criteria for scoring the movements of the FMS has to be met only once out of 3 trials, the scores between raters may have varied with more complex tasks. For example, the overhead squat movement required the dowel to remain overhead without moving forward past the toes and the femurs below horizontal while the knees remained aligned over the feet. The verification of all those criteria meant that the overhead squat had to be viewed from the front and side. One rater may have seen all the criteria met from a side view during the first repetition but not at the front view of the second repetition. Another rater during the first repetition determined the criteria were met from the front view on the first repetition and the side view on the second repetition. Therefore, the first rater would score a 2, whereas the second would score a 3. A video recording would have given insight into raters scoring and intrarater reliability if the movement actually changed from one repetition or one session to the next. Future research into the reliability of the FMS should incorporate video recording of the rating sessions combined with live administration.

Other limitations in the study design were the use of 4 individual raters to determine scores and the small sample size. The use of 4 individual raters does not represent all the professionals using the FMS; however, this study has the most diverse set of raters to date, and it is the only study to compare findings with a certified FMS rater. This allows for much greater applicability to the clinical setting than any previous study into reliability of the FMS. The relatively small sample was recruited from a single academic program at a University, but the subjects' total FMS score reflected the range reported in previous studies.

In conclusion, this study of physically active injury-free men and women showed that real-time administration of FMS could be consistently scored by people with varying degrees of experience with the FMS.

PRACTICAL APPLICATIONS

The results of this study suggest that various professionals who work with athletes and clients can reliably and consistently score he FMS. Because the FMS can be administered and scored by various members of a multidisciplinary team, the FMS could be used as athletes and clients move from rehabilitation to performance training. The hurdle step movement was one of the components of the FMS that was found to have poor reliability between raters. Therefore, more emphasis may need to be placed on the specific criteria of scoring the hurdle step as compared with the other movements during the training of raters to increase reliability.

ACKNOWLEDGMENTS

The authors declare they have no conflicts of interest. The results of the present study do not constitute endorsement by the authors or the National Strength and Conditioning Association. This study was conducted without extramural funding.

References

1. Burton, L, Kiesel, K, and Cook, G. Mobility screening for the core: Interventions. *Athl Ther Today* 9: 52–57, 2004.
2. Chorba, RS, Chorba, DJ, Bouillon, LE, Overmyer, CA, and Landis, JA. Use of a functional movement screening tool to determine injury risk in female collegiate athletes. *N Am J Sports Phys Ther* 5: 47–54, 2010.
3. Cook, G. *Movement: Functional Movement Systems: Screening, Assessment, and Corrective Strategies.* Santa Cruz, CA: On Target Publications, 2010.
4. Cook, G, Burton, L, and Hoogenboom, B. Pre-participation screening: The use of fundamental movements as an assessment of function–Part 1. *N Am J Sports Phys Ther* 1: 62–72, 2006.
5. Cook, G, Burton, L, and Hoogenboom, B. Pre-participation screening: The use of fundamental movements as an assessment of function–Part 2. *N Am J Sports Phys Ther* 1: 132–139, 2006.
6. Cowen, VS. Functional fitness improvements after a worksite-based yoga initiative. *J Bodyw Mov Ther* 14: 50–54, 2010.
7. Goss, DL, Christopher, GE, Faulk, RT, and Moore, J. Functional training program bridges rehabilitation and return to duty. *J Spec Oper Med* 9: 29–48, 2009.
8. Hootman, JM, Dick, R, and Agel, J. Epidemiology of collegiate injuries for 15 sports: Summary and recommendations for injury prevention initiatives. *J Athl Train* 42: 311–319, 2007.
9. Kiesel, K, Burton, L, and Cook, G. Mobility screening for the core, part 3: Implications for athletic low back pain. *Athl Ther Today* 10: 36–39, 2005.
10. Kiesel, K, Burton, L, Cook, G, and Mattacola, CG. Mobility screening for the core. *Athl Ther Today* 9: 38–41, 2004.
11. Kiesel, K, Plisky, P, and Butler, R. Functional movement test scores improve following a standardized off-season intervention program in professional football players. *Scand J Med Sci Sports* 21: 287–292, 2011.
12. Kiesel, K, Plisky, PJ, and Voight, ML. Can serious injury in professional football be predicted by a preseason functional movement screen? *N Am J Sports Phys Ther* 2: 147–158, 2007.
13. Knapik, JJ, Bauman, CL, Jones, BH, Harris, JM, and Vaughan, L. Preseason strength and flexibility imbalances associated with athletic injuries in female collegiate athletes. *Am J Sports Med* 19: 76–81, 1991.
14. Jones, BH and Knapik, JJ. Physical training and exercise-related injuries. Surveillance, research and injury prevention in military populations. *Sports Med* 27: 111–125, 1999.
15. 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: 479–486, 2010.
16. O'Connor, FG, Deuster, PA, Davis, J, Pappas, CG, and Knapik, JJ. Functional movement screening: Predicting injuries in officer candidates. *Med Sci Sports Exerc* 43: 2224–2230, 2011.
17. Portney, LG and Watkins, MP. *Foundations of Clinical Research* (3rd ed.). Upper Saddle River, NJ: Pearson Prentice Hall, 2009.
18. Schneiders, AG, Davidsson, A, Horman, E, and Sullivan, SJ. Functional movement screen normative values in a young, active population. *Int J Sports Phys Ther* 6: 75–82, 2011.
19. Teyhen, DS, Donofry, DF, Shaffer, SW, Walker, MJ, Lorenson, CL, Dugan, JL, Halfpap, JP, and Childs, JD. Functional movement screen: A reliability study in service members. *US Army Med Dep J* 33: 71, 2010.
20. Thomas, S, Reading, J, and Shephard, RJ. Revision of the Physical Activity Readiness Questionnaire (PAR-Q). *Can J Sport Sci* 17: 338–345, 1992.