
ORIGINAL RESEARCH

REAL-TIME TEST-RETEST AND INTERRATER RELIABILITY OF SELECT PHYSICAL PERFORMANCE MEASURES IN PHYSICALLY ACTIVE COLLEGE-AGED STUDENTS

Daniel T. Tarara, EdD, LAT, ATC¹
Eric J. Hegedus, PT, DPT, MHSc²
Jeffrey B. Taylor, PT, DPT, OCS, SCS, CSCS²

ABSTRACT

Background: A new 16-item physical performance measure screening battery (16-PPM) was developed in order to expand on established movement based qualitatively scored functional screening batteries to encompass a broader spectrum of quantitatively scored functional constructs such as strength, endurance, and power.

Purpose/Hypothesis: The purpose of this study was quantify the real-time tests-retest and expert versus novice interrater reliability of the 16-PPM screen on a group of physically active college-aged individuals. The authors' hypothesized that the test-retest and interrater reliability of quantitatively-scored performance measures would be highly correlated ($ICC \geq 0.75$) and that qualitatively-scored movement screening tests would be moderately correlated ($K_w = 0.41-0.60$).

Study Design: Cohort reliability study

Methods: Nineteen (8 males, 11 females) healthy physically active college-aged students completed the 16-PPM on two days, one week apart.

Results: The majority of the quantitatively scored components of the 16-PPMs demonstrated good expert-novice interrater reliability ($ICC > 0.75$), while qualitatively scored tests had moderate ($K_w = 0.41-0.60$) to substantial ($K_w = 0.61-0.80$) agreement. Test-retest reliability was consistent between raters, with most quantitatively scored PPMs exhibiting superior reliability to the qualitatively scored PPMs.

Conclusions: The 16-PPM test items showed good test-retest and interrater reliability. However, results indicate that expert raters may be more reliable than novice raters for qualitatively scored tests. The validity of this 16-PPM needs to be determined in future studies.

Clinical Relevance: Physical performance screening batteries may be used to help identify individuals at risk for future athletic injury; however, current PPMs that rely on qualitatively scored movement screens have exhibited inconsistent and questionable injury prediction validity. The addition of reliable quantitatively scored PPMs may complement qualitatively scored PPMs to improve the battery's predictive ability.

Key Words: Functional screen, physical performance measures, reliability

Level of Evidence: Level III

CORRESPONDING AUTHOR

Daniel T. Tarara, EdD, LAT, ATC
High Point University
School of Health Sciences
833 Montlieu Avenue
High Point, NC 27272
(336) 841-4604
Fax: (336) 888-6394
E-mail: dtarara@highpoint.edu

¹ Department of Exercise Science High Point University, School of Health Sciences, High Point, NC, USA

² Department of Physical Therapy High Point University, School of Health Sciences, High Point, NC, USA

INTRODUCTION

Musculoskeletal injuries are commonly associated with athletic participation.¹ In addition to pain, athletic injuries may also result in emotional stress, activity limitation, financial costs, or an increased risk for future injury.²⁻⁴ Some injuries are traumatic and potentially unavoidable, yet others result from a non-contact or repetitive mechanism. Non-contact injuries may be influenced by faulty biomechanics, asymmetry, or poor neuromuscular strength and control. If these factors are detected through injury risk screening batteries prior to participation, sports medicine professionals may be able to intervene by designing and implementing targeted prevention strategies to reduce injury risk.⁵

Established injury-risk screening exams, such as the Functional Movement Screen™ (FMS™)⁶ and Frohm-9⁷, are typically a collection of physical performance measures (PPMs) that aim to measure various constructs of performance by qualitatively scoring an individual's symmetry and quality of movement.⁸ Quantifying movement patterns may identify dysfunctional patterns that limit function and subsequently predispose an athlete to injury. While these tests of movement quality are useful tools in highlighting the importance of assessing function, questionable validity and sensitivity of these test batteries suggest that they may not be optimal for assessing injury risk.⁹⁻¹¹ Although movement patterns are an important aspect of athleticism, the measurement of function is complex and may necessitate a more comprehensive approach to functional testing than relying on a single construct.¹² Function encompasses a broad range of interdependent constructs, including functional movement patterns, muscle flexibility, balance, proprioception, speed, agility, aerobic and anaerobic conditioning, and muscular strength, power, and endurance,⁸ which may not be measurable with qualitative analyses of movement patterns. In this context, current injury risk screening batteries may not adequately quantify the wide spectrum of required function within athletic populations. The limitations of current screening batteries would suggest that the best combination of PPMs has yet to be identified for assessing function in athletes.

As part of a larger validity study, the authors are evaluating a 16 item PPM battery (16-PPM) that

takes advantage of the work represented by the FMS™ and Frohm-9, yet was expanded to include a number of different tests that quantitatively measure the constructs of strength, endurance, power, and motor control. A number of the included PPMs employ identical methodology to what has been established in the literature, while others have been modified from their original description or are newly-designed based on the authors' clinical practice. Though established screening batteries have reported adequate reliability,^{9,13-19} it is unknown whether an alternate group of quantitatively-scored tests can be performed with similar consistency. Thus, it is crucial to establish the reliability of the PPMs that compose this new screening battery.

Additionally, the 16-PPM battery was designed with the flexibility to be administered as a mass-screening tool for use with multiple athletic teams in order to facilitate efficiency of data collection. The testing conditions associated with screening large groups of athletes often necessitate the use of testing stations and the assistance of multiple raters. Raters, in an ideal context, should be experienced clinicians with a familiarity of the PPM battery; however, it is often more practical and cost-effective to incorporate novice raters in test administration.^{9,15,16} Thus, the test-retest and interrater reliability of the proposed 16-PPM screen needs to be established for both expert and novice raters before addressing issues related to validity and clinical utility. Therefore, the purpose of this study was to quantify the real-time test-retest and expert versus novice interrater reliability of a 16-PPM screen on a group of physically active college-aged individuals. The authors hypothesized that the test-retest and interrater reliability of quantitatively-scored performance measures would be highly correlated ($ICC \geq 0.75$) and that qualitatively-scored movement screening tests would be moderately correlated ($K_w = 0.41-0.60$).

METHODS

Subjects

Subjects were openly recruited from a university's club sports program. Inclusionary criteria required subjects to be currently active in a club sport season or having participated in a club sport within the past calendar year and to have maintained a physically

active lifestyle of cardiorespiratory endurance exercise (e.g. running, cycling, rowing) for three or more days per week or two or more days of a progressive resistive exercise program. Subjects were excluded if they had sustained an orthopedic injury within the past six months or had a pre-existing cardiovascular, pulmonary, or metabolic condition which contraindicated neuromuscular strength and endurance exercise. Twenty-four subjects were initially eligible for the study, 19 of whom completed the test-retest trial: Eight males aged 20.3 ± 1.2 years, height 180.3 ± 6.3 cm, weight 77.0 ± 6.3 kg, and 11 females aged 19.8 ± 1.0 years, height 166.9 ± 5.5 cm, weight 60.4 ± 4.9 kg. The five remaining subjects completed the initial intake forms, but elected not to complete the testing because of time constraints. All subjects voluntarily signed an informed consent form prior to participation.

Study design

The subjects completed the testing sequence on two occasions separated by a minimum of four days and a maximum of seven days. Subjects were instructed to maintain their normal physical activity routine between sessions. Each testing session consisted of the subjects completing the 16-PPM battery, which took approximately 50 minutes to complete. Testing was administered in a university biomechanics laboratory, with identical physical conditions during both testing sessions. The study was approved by the High Point University Human Participants Institutional Review Board.

Raters

One experienced and two novice raters were used for this investigation. The experienced rater had 22 years of clinical practice and was familiar with a broad range of PPMs used in evaluating function specific to athletic populations. The two novice raters were hand selected from an undergraduate exercise science program based on having no prior experience with any type of physical performance testing. All raters voluntarily signed an informed consent prior to participation.

Testing Procedure

The novice raters received an electronic copy of the 16-PPM scoring sheet (see Appendix) 24 hours prior

to starting data collection. The scoring sheet consisted of photos of each test and a written description of how to perform and score each test. The novice raters arrived to the laboratory one hour prior to testing of the first subject. A member of the investigation team (D.T.) provided an overview of each test and scoring criteria respectively. The novice raters were given the opportunity to ask clarifying questions as needed during the orientation session, as they could not ask questions once trials began. Raters were explicitly instructed that once subject trials began they could not ask clarifying questions about the 16-PPM or the scoring criteria. In addition, raters were instructed to independently score each PPM and to not communicate with each other during the testing sequence.

The subjects reported to the lab and completed three blinded copies of the 16-PPM intake form. The testing order of the 16 tests were randomized as follows: Full Squat, Downward Dog, Broad Jump, Closed Kinetic Chain Upper Extremity Stability Test (CKCUEST), Lower Extremity Y-Balance, In-Line Lunge for Distance, Lumbar Endurance, Single Leg Squat, Shoulder Mobility Test, Active Straight Leg Test, Side Plank Hip Abduction, Beighton Hypermobility, Triple Hop for Distance, Nordic Hamstring, Lateral Lunge for Distance, and Side Plank Hip Adduction. The subjects received uniform verbal descriptions and physical demonstrations of each test from the experienced rater. A series of verbal cues were given to prompt subjects for each test: "set" to assume the starting position; "go" to initiate the test; and "ok/stop" when the test was completed and scored. The raters received the following verbal cues for each test; "raters ready" before cueing the subject and "good" to affirm scoring. Raters were blinded to each other's score sheet and independently scored each subject's 16-PPM test battery.

16-PPM BATTERY

Quantitatively-Scored Tests

Broad Jump

The broad jump was administered as reported by Robertson et al.²⁰ Subjects were instructed to stand with their toes behind a taped line on the floor, squat and jump as far as possible. Arm swing was allowed during the jump and distance was measured from

the start line to the heel of the back foot. Three trials were performed and the best trial was used for scoring. The broad jump was scored to the nearest inch as measured by a tape measure secured to the ground.

CKCUEST

The CKCUEST was administered as previously reported by Goldbeck and Davies,²¹ with modification to the starting position. Two pieces of tape, 36 inches apart and parallel to each other were placed on the floor. Subjects were instructed to assume a push-up position with hands roughly shoulder width apart between the pieces of tape to accommodate for differences in shoulder width and arm length. Subjects were instructed to use an alternating pattern of reaching across their body and touching the opposite piece of tape as many times as possible in 15 seconds. The closed kinetic chain upper extremity stability test was scored as the total number of successful touches of both sides combined.

Lower Extremity Y-Balance

The Lower Extremity Y-Balance Test was administered as reported by Plisky et al.²² Subjects wore shoes and were instructed to stand on one leg in the center of the device and reach with the opposite toe to push an indicator in any one of three directions (anterior, posteromedial, posterolateral) in any order of their choosing. Hands were free during testing and subjects were told that they could not touch the non-stance leg down until all three directions were completed and they could not kick the indicator. The Lower Extremity Y-Balance Test was scored as the sum of the best effort from three successful trials in each direction.

In-Line Lunge for Distance

The In-Line Lunge for Distance was administered as previously reported by Crill et al.²³ Subjects were instructed to place the toes of both feet behind a line on the floor with hands on hips. Subjects lunge stepped forward as far as possible with one leg while the toe of the other leg had to stay behind the start line. Subjects needed to stick the landing for the trial to count. Distance was measured from the start line to the heel of the lead leg in cm. The in-line lunge was scored using the best of three trials for each leg.

Lateral Lunge for Distance

The Lateral Lunge for Distance was administered as previously reported by Crill et al.²³ Subjects were instructed to stand with their feet parallel to the start line with the medial border of the trail foot against the line and hands on hips. Subjects lunge stepped in a lateral direction as far as possible with one leg while the foot of the other leg was required to stay behind the start line and remain flat to the floor. Subjects needed to stick the landing for the trial to count. Distance was measured from the start line to the most medial aspect of the lead foot in cm. The in-line lunge was scored as a quantitative test by taking the best of three trials for each leg.

Lumbar Endurance

The Lumbar Endurance Test was a modified isometric back endurance test as reported by Wilkerson.²⁴ Subjects were instructed to begin on all fours with knees under hips and hands under shoulders. With a partner holding the subject's feet, the subject lifted their hands off of the floor and placed them, interlocked, behind the head with the result being a 90/90 position at the hips and knees. Subjects were asked to hold this position as long as possible and were given only one verbal cue to encourage holding the 90/90 position. The lumbar endurance test was scored as time to either failure or to an inability to hold the 90/90 position after one warning.

Side Plank Hip Abduction

The Side Plank Hip Abduction Test is a novel PPM designed specifically for this testing battery. Subjects were instructed to assume a side plank position with elbow under the shoulder and feet together. No body part was permitted to touch the ground between the elbow and the feet. The opposite arm was placed on the hip. The subject was then instructed to lift the top leg (hip abduction) 20 cm (the height of a plyometric box) and return to the starting position. Leg lifts less than 20cm were discounted. The side plank hip abduction test was scored as the number of leg lifts achieved in 30 seconds.

Side Plank Hip Adduction

The Side Plank Hip Adduction Test was also designed specifically for this testing battery. Subjects were instructed to assume a side plank position with

elbow under the shoulder and the top leg on a 20 cm tall plyometric-box. No body part was permitted to touch the ground between the elbow and the feet. The top hand was placed on the ipsilateral hip. The subject was then instructed to lift the bottom leg (hip adduction) approximately eight inches and return to the starting position. The side plank hip adduction test was scored as the number of leg lifts achieved in 30 seconds.

Triple Hop for Distance

The Triple Hop for Distance Test was administered as previously established in the literature.²⁵ Subjects started with toes behind a starting line on the floor and were instructed to hop three consecutive times on the same leg as far as possible with their hands free. All subjects were allowed one practice trial with the subsequent trial scored for analysis. Subjects needed to stick the landing for the trial to count. The triple hop was measured as the total distance hopped along a tape measure from the starting line to the subject's heel position after the third hop.

Nordic Hamstring

The Nordic Hamstring Test was designed specifically for this testing battery as a modification of the Nordic Hamstring exercise used for rehabilitation of lower extremity injuries.²⁶ The subject was instructed to kneel on both knees with hips extended and arms held at the ready at chest height. With a partner holding the subject's ankles, the subject was instructed to lean forward at the knees as far as possible, maintaining the hips at neutral (0 degrees extension) and slowly lower to the ground. The rater used a 180 degree extendable arm goniometer (Baseline Gollehon, Lafayette Instrument Company, Lafayette IN, USA) to measure forward trunk lean based on the tibial-femoral knee angle. The Nordic hamstring test was scored quantitatively as the knee angle at last moment of controlled descent.

Qualitatively-Scored Tests

Full Squat

The Full Squat was administered as previously established by Cook et al.⁶ Subjects were instructed to grasp a plastic rod in both hands and raise their arms overhead with elbows fully extended. With feet shoulder width apart and toes pointing straight

ahead, the subject performed a squat as deeply as possible. If the subject could not squat with thighs parallel to the floor, knees aligned over feet, upper torso parallel with tibia, or plastic rod over feet, a plastic piece of 2 x 4 was placed under the heels of the subject and they were asked to repeat the squat. Three trials were performed and the best of the three trials was used for scoring. The scoring system was modified to a 0-5 scale (see Appendix) based on the subjects' form and movement quality.

Downward Dog

The Downward Dog Test was a novel test designed specifically for this testing battery. Subjects were instructed to begin on all fours with knees under hips and hands one hand length in front of shoulders. Subjects were instructed to straighten their arms and legs and flatten their scapula to their back to form an inverted "V". Subjects were given time to get comfortable in this position but were not allowed to adjust the position of the hands and feet. The Downward Dog was scored on a 0-5 scale (see Appendix) based on the subjects' form and movement quality.

Single Leg Squat

The Single Leg Squat was administered with slight modification from that previously reported by Frohm et al.⁷ Subjects were instructed to stand on one leg 4-6 inches from an adjustable height bench. The bench was at a height equal to the subject's popliteal fossa. The subject's hands were placed on the hips as the subject squats to the bench, touches without resting and stands back up. Five repetitions were performed with each leg and the worst of the last two repetitions was scored. The single leg squat was scored on a 0-5 scale (see Appendix) based on the number of observed movement errors observed.

Shoulder Mobility Test

The Shoulder Mobility Test was administered as previously established by Cook et al.⁶ The distance from the distal most palmar wrist crease to the tip of the middle finger was measured on one hand. Subjects were instructed to make a fist and place one arm behind their head as far distally as possible while concurrently placing the opposite hand behind the low back and reach as far superiorly as possible. Sub-

jects could not “crawl” their fists up or down their back. The shoulder mobility test was measured as the distance between the two fists, then qualitatively assessed on a 0-5 scale (see Appendix) based on the presence of scapular winging.

Active Straight Leg Raise Test

The Active Straight Leg Raise Test was administered as previously established by Cook et al.⁶ Subjects were instructed to lie on their back with legs straight and toes pointing toward the ceiling. The examiner placed a 2x4 under the subject's knees. With arms flat on the ground, subjects lifted one leg keeping it straight. The active straight leg raise test was scored qualitatively assessed on a 0-5 scale (see appendix) based on how far the straight leg was lifted in reference to landmarks on the opposite leg.

Beighton Hypermobility

The Beighton Hypermobility Scale was administered as previously reported.²⁷ Subjects were asked to perform five screening movement to the best of their ability: 1) touch the palms of the hands to the floor with straight legs, 2) extend the elbows as much as possible, 3) extend the knees as much as possible, 4) touch each thumb to the palmar surface of the forearm, and 5) extend the little finger at the metacarpal-phalangeal joint. The original scoring metric for the test was a 0 – 9 point scale, one point awarded for each positive test with a higher score indicating a greater degree of hypermobility. Specifically for this testing battery, Beighton Hypermobility scores were converted to a 0-5 scale (see Appendix) with a lower score indication a greater degree of hypermobility. This scale modification was performed in order to maintain uniform scoring metrics for all qualitatively scored tests.

Statistical Analysis

The PPMs analyzed in this study were either scored quantitatively (continuous scale) or qualitatively (ordinal scale). Descriptive statistics for the qualitatively scored test data were calculated as mean \pm standard deviation and qualitative test data was described using frequency distributions of the score given by the experienced rater on session one for each gender. These values were chosen because they theoretically provide the most accurate score,

reflect clinical practice, and contain the base scoring metrics observed and recorded for each respective test.

The reporting of reliability and agreement was based on the guidelines of Kottner et al.²⁸ Test-retest reliability measures of each rater (expert, novice A, novice B) were calculated based on session one and two scores. Interrater reliability, comparing each of the three raters, was calculated using session one scores, as these were deemed the most clinically relevant. Test-retest and interrater reliability for each continuous variable were evaluated with interclass correlational coefficients (ICC), 95% confidence intervals (CI), and standard error of measurement (SEM) based on a two-way analysis of variance (ANOVA) ICC model (3, 1). ICC values range between 0.0 to 1.0, where values closer to 0.0 reflect poor reliability and values closer to 1.0 suggest high reliability.²⁹ ICC values were interpreted as follows: 0.4-0.74 = poor to moderate, and ≥ 0.75 = good.²⁹ SEM provides an indication of the precision of measurement, independent of the population, and reflects the scoring constancy within the unit of measure.³⁰ All of these descriptive and reliability analyses was performed using IBM® SPSS® V20.0 (SPSS, INC., Chicago IL).

Test-retest and interrater reliability for each ordinal variable by rater were evaluated with weighted kappa (K_w), 95% CI, and standard error (SE). The K_w was used to best capture the level of disagreement across ordinal scales that are not binary.³¹ K_w values of agreement were interpreted as follows: < 0 = poor, 0.0 to 0.2 = slight, 0.21 to 0.40 = fair, 0.41-0.60 = moderate, 0.61 to 0.80 = substantial, 0.81 to 1.0 = almost perfect.³² Weighted kappa statistics was performed using MedCalc for Windows, version 12.7.7 (MedCalc Software, Mariakerke, Belgium).

RESULTS

Descriptive statistics of all PPMs are reported in Tables 1 (quantitative data) and 2 (qualitative data) as measured by the expert rater on test session one.

INTERRATER RELIABILITY

Interrater reliability (expert vs novice 1, expert vs novice 2, and novice 1 vs novice 2) measures for continuous level data are reported in Table 3. Reliability statistics were consistent for all combinations

Table 1. Descriptive statistics of quantitatively scored tests measured on session one. All results are reported as mean \pm standard deviation.

Test	Side	Expert Rater	Novice 1 Rater	Novice 2 Rater
Broad Jump (cm)	-	76.5 \pm 14.4	77.2 \pm 15.2	76.3 \pm 14.4
CKCUEST (count)	-	25.9 \pm 4.4	25.3 \pm 4.0	25.4 \pm 4.0
LEYB (total)	R	294.0 \pm 37.2	292.8 \pm 32.9	292.5 \pm 34.6
	L	290.5 \pm 40.9	291.2 \pm 37.2	287.2 \pm 37.2
In-Line Lunge (cm)	R	111.2 \pm 16.9	114.8 \pm 15.4	113.0 \pm 16.5
	L	111.8 \pm 17.8	112.0 \pm 18.1	111.5 \pm 17.4
Lumbar Endurance (sec)	-	87.2 \pm 60.0	89.0 \pm 59.6	90.2 \pm 59.5
Side Plank Hip Abduction (count)	R	40.8 \pm 13.9	38.3 \pm 12.6	41.2 \pm 13.5
	L	36.5 \pm 10.9	36.2 \pm 10.2	37.2 \pm 11.0
Triple Hop (cm)	R	447.4 \pm 80.2	486.6 \pm 86.6	481.1 \pm 94.4
	L	453.5 \pm 96.0	483.8 \pm 88.4	476.2 \pm 87.7
Nordic Hamstring (degrees)	-	39.4 \pm 2.6	37.7 \pm 2.1	43.2 \pm 2.7
Lateral Lunge (cm)	R	91.1 \pm 5.2	91.1 \pm 5.4	90.9 \pm 4.1
	L	90.4 \pm 5.2	91.1 \pm 5.4	90.8 \pm 4.4
Side Plank Hip Adduction (count)	R	37.3 \pm 17.3	35.8 \pm 18.4	35.4 \pm 16.9
	L	34.5 \pm 18.4	35.0 \pm 18.2	31.9 \pm 16.6

CKCUEST= Closed kinetic chain upper extremity stability test; LEYB= Lower extremity Y-Balance test

of raters, with good reliability (ICC \geq 0.75) exhibited for most tests. Poor to moderate interrater reliability (ICC < 0.75) was measured in the Nordic Hamstring (ICC 0.03-0.74), left-sided Triple Hop for Distance (ICC 0.67-0.99), and right-sided Lateral Lunge for Distance (ICC 0.66-0.82). Interrater reliability for ordinal level data are reported in Table 4. Interrater reliability data were moderate to substantial for the Downward Dog (K_w 0.73 -0.93), Beighton Hypermobility (K_w 0.64 -0.69), Active Straight Leg Test (K_w 0.53 -0.71), and Full Squat (K_w 0.45 -0.60). Interrater reliability agreement ranged from fair to substantial for Single Leg Squat (K_w 0.33 -0.70) and Shoulder Mobility Test (K_w 0.24 -0.46).

TEST-RETEST RELIABILITY

Table 5 displays test-retest reliability statistics (ICC and SEM) as measured by the expert and each novice rater.

Overall, ICC's between each rater were consistent, with a majority of tests showing good reliability. However, certain tests exhibited poor to modest test-retest reliability measures (ICC < 0.75), including the Nordic Hamstring (ICC 0.05-0.29), CKCUEST (ICC 0.73-0.78), left (ICC 0.70-0.81) and right Lower Extremity Y-Balance (ICC 0.74-0.82), left-sided Triple Hop for Distance (ICC 0.63-0.69), left (ICC 0.70-0.78) and right Side Plank Hip Abduction (ICC 0.36-0.88), and left (ICC 0.64-0.83) and right Lateral Lunge for Distance (ICC 0.68-0.81). Test-retest reliability for qualitative ordinal level data are reported in Table 6. K_w values ranged from 0.32 to 0.81 for the expert rater, -0.09 to 0.73 for Novice rater A, and 0.25 to 0.78 for Novice rater B.

DISCUSSION

The purpose of this study was to analyze the reliability of a 16 item PPM screening battery that collectively evaluated athletes' movement strategies and performance. This group of tests was chosen based on tests available from published research on injury and recovery from injury, as well as a desire to have a mixture of qualitative and quantitative tests, and inclusion of tests representing components of sporting activities in the upper extremities, lower extremities, and trunk. As examples, triple hop symmetry has been found to be different in patients with anterior cruciate ligament reconstruction compared to a healthy control group³³ and asymmetrical performance on the Lower Extremity Y-balance test predicts lower extremity injury.³⁴ The unloaded, full squat is an example of a qualitative test based on examiner judgment of movement quality, symmetry, and substitution patterns while the broad jump is based quantitatively on the distance traversed in a 2-legged leap. Finally, the authors' felt that the Single Leg Squat would capture lower extremity motor

Table 2. Descriptive statistics of qualitatively scored tests as measured by the experienced rater on session one. All data reported as frequency percentages of assigned scores and mean \pm standard deviation.

TEST		0	1	2	3	4	5	$\bar{X}\pm SD$
Full Squat		0%	10.5%	31.6%	5.3%	0%	52.6%	3.5 \pm 1.6
Downward Dog		0%	26.3%	26.3%	5.3%	0%	47.4%	3.1 \pm 1.8
Single Leg Squat	R	0%	0%	5.3%	10.5%	42.1%	42.1%	4.1 \pm 0.8
	L	0%	0%	5.3%	21.1%	15.8%	57.9%	4.2 \pm 0.9
Shoulder Mobility Test	R	0%	5.3%	15.8%	47.4%	26.3%	2.3%	3.1 \pm 0.9
	L	0%	5.3%	10.5%	52.6%	21.1%	10.5%	3.2 \pm 0.9
Active Straight Leg Raise	R	0%	0%	10.5%	15.8%	15.8%	57.9%	4.2 \pm 1.0
	L	0%	0%	10.5%	15.8%	26.3%	47.4%	4.1 \pm 1.0
Beighton Hypermobility		0%	0%	10.5%	26.3%	26.3%	36.8%	3.8 \pm 1.0

Table 3. Intra-class correlation coefficients for interrater reliability test session one.

Test	Expert vs. Novice A			Expert vs Novice B			Novice A vs. Novice B		
	ICC	95%CI	SEM	ICC	95%CI	SEM	ICC	95%CI	SEM
Broad Jump (cm)	0.99	(0.98, 0.99)	21.0	0.99	(0.97, 0.99)	21.0	0.99	(0.98, 0.99)	20.5
CKCUEST (count)	0.97	(0.93, 0.99)	6.0	0.96	(0.91, 0.98)	6.0	0.96	(0.91, 0.98)	5.7
LEYB - left (total)	0.86	(0.68, 0.94)	55.3	0.88	(0.73, 0.95)	55.4	0.98	(-0.95, 0.99)	52.7
LEYB- right (total)	0.87	(0.69, 0.94)	49.7	0.86	(0.69, 0.94)	50.9	0.94	(0.94, 0.99)	47.8
In-line Lunge - left (cm)	0.97	(0.93, 0.99)	25.4	0.98	(0.95, 0.99)	25.0	0.95	(0.89, 0.98)	25.2
In-line Lunge - right (cm)	0.95	(0.88, 0.98)	23.1	0.94	(0.85, 0.97)	23.7	0.94	(0.87, 0.98)	22.7
Lumbar Endurance (sec)	0.98	(0.96, 0.99)	84.6	0.98	(0.96, 0.99)	84.8	0.99	(0.99, 0.99)	81.8
Side Plank Hip Abduction - left (count)	0.94	(0.84, 0.97)	15.0	0.94	(0.85, 0.97)	15.4	0.94	(0.86, 0.98)	15.1
Side Plank Hip Abduction – right (count)	0.96	(0.89, 0.98)	18.9	0.94	(0.85, 0.97)	20.4	0.95	(0.87, 0.98)	18.1
Triple Hop - left (cm)	0.68	(0.35, 0.86)	132.4	0.67	(0.32, 0.85)	131.1	0.99	(0.98, 0.99)	124.7
Triple Hop- right (cm)	0.77	(0.50, 0.90)	122.0	0.76	(0.49, 0.90)	127.5	0.96	(0.90, 0.98)	128.2
Nordic Hamstring Test (degrees)	0.74	(0.44, 0.89)	15.1	0.03	(-0.41, 0.46)	17.5	0.14	(-0.32, 0.55)	16.4
Lateral Lunge - left trail (cm)	0.91	(0.80, 0.96)	21.9	0.92	(0.80, 0.96)	25.3	0.87	(0.70, 0.94)	24.3
Lateral Lunge – right trail (cm)	0.80	(0.55, 0.91)	20.0	0.82	(0.59, 0.92)	23.6	0.66	(0.30, 0.85)	22.0
Side Plank Hip Adduction – left (count)	0.97	(0.93, 0.99)	25.9	0.90	(0.72, 0.96)	24.9	0.91	(0.80, 0.96)	24.9
Side Plank Hip Adduction – right (count)	0.92	(0.80, 0.97)	25.4	0.90	(0.75, 0.96)	24.3	0.90	(0.76, 0.96)	25.4

ICC = Inter-class Correlation Coefficients with 95% confidence interval; SEM = standard error of measure; CKCUEST=Closed kinetic chain upper extremity stability test; LEYB= Lower extremity Y-balance test.

Table 4. Weighted Kappa for interrater reliability on test session one.

Test	Expert vs. Novice A			Expert vs. Novice B			Novice A vs. Novice B		
	K_w	95%CI	SE	K_w	95%CI	SE	K_w	95%CI	SE
Full Squat	0.60	(0.34, 0.86)	0.13	0.55	(0.32, 0.78)	0.11	0.45	(0.21, 0.69)	0.12
Downward Dog	0.93	(0.82, 1.00)	0.05	0.71	(0.50, 0.92)	0.10	0.73	(0.52, 0.94)	0.11
Single Leg Squat - left	0.62	(-0.05, 0.53)	0.15	0.47	(0.18, 0.76)	0.14	0.52	(0.26, 0.78)	0.13
Single Leg Squat- right	0.33	(-0.12, 0.80)	0.23	0.71	(0.44, 0.98)	0.13	0.39	(0.06, 0.73)	0.16
Shoulder Mobility Test - left	0.44	(0.15, 0.74)	0.15	0.26	(0.01, 0.53)	0.13	0.46	(0.17, 0.76)	0.15
Shoulder Mobility Test - right	0.28	(0.04, 0.52)	0.12	0.35	(0.13, 0.56)	0.11	0.24	(0.01, 0.48)	0.12
Active Straight Leg Raise - left	0.71	(0.52, 0.90)	0.09	0.53	(0.32, 0.74)	0.11	0.63	(0.32, 0.94)	0.15
Active Straight Leg Raise - right	0.67	(0.49, 0.85)	0.09	0.57	(0.38, 0.76)	0.09	0.54	(0.22, 0.86)	0.16
Beighton Hypermobility	0.69	(0.55, 0.83)	0.07	0.64	(0.44, 0.85)	0.10	0.72	(0.62, 0.82)	0.05

K_w = weighted kappa with 95% confidence interval; SE = standard error

control, the Side Plank Hip Abduction test would capture trunk and hip stability and endurance, and the CKCUEST would capture upper extremity stability and quickness.

Collectively, the majority of individual tests exhibited good to excellent reliability, with better inter- and test-retest reliability for the tests scored quantitatively (performance components) than qualitatively (movement strategy components). In addition, comparable reliability was found for both expert and novice raters. These findings support the hypothesis that the real-time test-retest and expert versus novice interrater reliability of quantitatively-scored performance measures would be highly correlated ($ICC \leq 0.75$) and that qualitatively-scored

movement screening tests would be moderately correlated ($K_w = 0.41-0.60$).

Interrater reliability for Broad Jump, CKCUEST, Lower Extremity Y-balance test, In-line Lunge for Distance, Lumbar Endurance, and the Triple Hop for Distance was excellent, though this was expected, considering that these tests were highly objective, requiring the rater to only count or read a tape measure. The test-retest reliability of each test was also considered good to excellent, yet not as high as previously reported for the CKCUEST,³⁵ Lumbar Endurance,³⁶ and Triple Hop for Distance.³⁷ The reason that the current reported reliability measures were slightly lower than previous reports may be because the individual tests in the current study were part of

Table 5. Intra-class correlation coefficients for test-retest reliability for each rater.

Test	Expert			Novice A			Novice B		
	ICC	95%CI	SEM	ICC	95%CI	SEM	ICC	95%CI	SEM
Broad Jump (cm)	0.95	(0.89, 0.98)	21.43	0.89	(0.74, 0.96)	21.10	0.97	(0.92, 0.98)	21.09
CKCUEST (count)	0.73	(0.41, 0.89)	8.00	0.78	(0.47, 0.92)	6.92	0.77	(0.50, 0.91)	7.12
LEYB - left (total)	0.70	(0.36, 0.87)	58.5	0.79	(0.53, 0.91)	55.4	0.81	(-0.57, 0.92)	56.7
LEYB- right (total)	0.74	(0.44, 0.89)	56.6	0.82	(0.59, 0.93)	52.9	0.81	(0.57, 0.92)	54.9
In-line Lunge - left (cm)	0.78	(0.51, 0.91)	24.22	0.81	(0.56, 0.92)	24.69	0.78	(0.50, 0.91)	24.53
In-line Lunge - right (cm)	0.89	(0.74, 0.96)	24.34	0.83	(0.61, 0.93)	22.26	0.87	(0.68, 0.95)	24.08
Lumbar Endurance (sec)	0.77	(0.47, 0.91)	79.71	0.76	(0.36, 0.92)	83.82	0.76	(0.47, 0.90)	84.10
Side Plank Hip Abduction - left (count)	0.70	(0.35, 0.88)	17.19	0.75	(0.39, 0.91)	17.61	0.78	(0.51, 0.91)	18.39
Side Plank Hip Abduction – right (count)	0.88	(0.68, 0.95)	18.28	0.36	(-0.21, 0.75)	22.16	0.80	(0.55, 0.92)	18.65
Triple Hop - left (cm)	0.63	(0.25, 0.84)	139.94	0.63	(0.25, 0.84)	138.06	0.69	(0.34, 0.87)	134.81
Triple Hop- right (cm)	0.75	(0.45, 0.90)	124.04	0.78	(0.50, 0.91)	126.81	0.73	(0.41, 0.89)	130.47
Nordic Hamstring Test (degrees)	0.05	(-0.41, 0.49)	14.90	0.06	(-0.40, 0.50)	13.70	0.29	(-0.19, 0.67)	16.85
Lateral Lunge - left trail (cm)	0.83	(0.61, 0.93)	22.25	0.78	(0.50, 0.91)	21.78	0.64	(0.27, 0.85)	24.52
Lateral Lunge – right trail (cm)	0.81	(0.57, 0.92)	21.36	0.78	(0.52, 0.91)	20.43	0.68	(0.32, 0.86)	22.63
Side Plank Hip Adduction – left (count)	0.78	(0.43, 0.92)	24.84	0.85	(0.64, 0.94)	25.16	0.80	(0.55, 0.92)	23.96
Side Plank Hip Adduction – right (count)	0.78	(0.46, 0.92)	24.58	0.76	(0.46, 0.90)	25.60	0.83	(0.60, 0.93)	24.74

ICC = Inter-class Correlation Coefficients with 95% confidence interval; SEM = standard error of measure; CKCUEST=Closed kinetic chain upper extremity stability test; LEYB= Lower extremity Y-balance test.

Table 6. Weighted Kappa for test-retest reliability for each rater.

Test	Expert			Novice A			Novice B		
	K_w	95%CI	SE	K_w	95%CI	SE	K_w	95%CI	SE
Full Squat	0.79	(0.57, 1.0)	0.11	0.68	(0.40, 0.95)	0.14	0.63	(0.34, 0.92)	0.14
Downward Dog	0.47	(0.14, 0.80)	0.17	0.30	(-0.04, 0.66)	0.18	0.60	(0.34, 0.85)	0.12
Single Leg Squat - left	0.54	(0.25, 0.82)	0.14	0.46	(0.01, 0.92)	0.23	0.41	(-0.01, 0.83)	0.21
Single Leg Squat- right	0.43	(0.19, 0.68)	0.12	0.71	(0.50, 0.93)	0.11	0.25	(-0.14, 0.65)	0.20
Shoulder Mobility Test - left	0.32	(0.07, 0.58)	0.13	0.32	(-0.03, 0.64)	0.87	0.71	(0.41, 1.0)	0.15
Shoulder Mobility Test - right	0.49	(0.19, 0.80)	0.15	-0.09	(-0.26, 0.25)	0.13	0.73	(0.46, 0.99)	0.13
Active Straight Leg Raise - left	0.81	(0.61, 1.00)	0.10	0.59	(0.29, 0.88)	0.14	0.54	(0.23, 0.68)	0.16
Active Straight Leg Raise - right	0.73	(0.54, 0.91)	0.09	0.69	(0.46, 0.92)	0.11	0.78	(0.48, 1.0)	0.15
Beighton Hypermobility	0.69	(0.46, 0.92)	0.18	0.73	(0.58, 0.89)	0.08	0.72	(0.53, 0.90)	0.09

K_w = weighted kappa with 95% confidence interval; SE = standard error

a larger screening battery, allowing only one trial per test, whereas previous researchers have performed reliability analyses across three or more averaged trials. Also, performing the tests on two separate occasions may have contributed lower reliability due to score variability.

Results indicate that, with the exception of the Nordic Hamstring Test, novice raters demonstrated excellent interrater reliability with the expert rater when assessing the performance based, quantitative tests, yet were less consistent assessing the movement

screening tests. This finding is inconsistent with previous research, which has reported excellent interrater reliability between expert and novice raters of the FMS™^{9,38,39} although, as stated earlier, this may be a function of the modified scoring system used in this study. These findings indicate that novice raters can be a valuable component of the sports medicine team in administering PPM tests to an athletic population. When the opportunity to train novice raters is compressed due to time constraints, it may be best to use them in tasks that require simple quantitative tasks such as timing, counting repetitions, and measuring

distances. Further training may be necessary for novice raters to understand qualitative movement patterns and to reliably evaluate them.

The PPMs that focused on movement quality were all established tests taken from the FMS™, Frohm-9, or ACL prevention literature, except for the Downward Dog, which was designed specifically to assess the global flexibility of the posterior half of the body from the latissimus dorsi proximally to the triceps surae group distally. The established tests (Full Squat, Single Leg Squat, Shoulder Mobility Test, Active Straight Leg Raise, and Beighton Hypermobility) scored in this study using a 0-5 scale, exhibited considerably lower interrater and test-retest reliability than the more quantitative tests. This scoring differential was not surprising, considering the subjective nature of the assessment; however, interrater reliability for the novel Downward Dog test was substantially better than the other qualitative movement screening tests, although test-retest reliability was only moderate.

The previously established tests of performance for the lower extremity used in this screening battery had all been designed to assess strength and/or power during sagittal plane activities, effectively targeting the hip and knee extensor musculature; however, multi-directional sports with a high-risk of injury require frequent deceleration, change of direction, and significant lower extremity demands outside of the sagittal plane. Therefore, three lower extremity PPMs were designed and added to the test battery in order to assess frontal plane core and hip stability, strength (Plank Hip Abduction and Adduction tests), and functional range of motion (Lateral Lunge for Distance) and hamstrings function (Nordic Hamstring). Of these novel tests, the Side Plank Hip Adduction, and Lateral Lunge for Distance exhibited excellent interrater and test-retest reliability. Both the Side Plank Hip Abduction and Lateral Lunge for Distance showed excellent interrater reliability, yet test-retest reliability was subtly lower, especially when measured by the novice raters. The Nordic Hamstring test exhibited poor interrater and test-retest reliability. Test-retest reliability for the Nordic Hamstring was poor for all raters. This may be because identifying the last point at which the athlete was able to control their downward descent

was largely subjective, and may have been influenced by a learning effect between the two sessions. These results suggest that at this point, the Nordic Hamstring should not be used as an outcome measure, until further modification of the methodology has been refined to improve both the interrater and test-retest reliability.

The reliability results for the movement quality tests in the 16-PPM are comparable, yet lower than previously reported in the literature. This may be due to the assessment scale, seeing that it was modified from the conventional 0-3 scale used for the FMS™ and Frohm-9 to a 0-5 scale. Because the 0-3 scale often ends up in a bell-shaped distribution with significantly more athletes scoring a 2 than a 0, 1 or 3,⁴⁰ the 0-5 scale was designed to provoke, and successfully elicited, a wider range of scores (Table 2), hoping to ultimately more accurately differentiate injury risk. However, the wider range of scores also generated lower inter- and test-retest reliability scores than the traditional FMS™ and Frohm-9 tests.^{7,39-41}

PPMs may be a valuable component of injury risk or return to play screening procedures. Conventional screening batteries (e.g. FMS™ and Frohm-9) and individual tests (e.g. drop vertical jump, tuck jump test) focus on primary and compensatory neuromuscular movement strategies to assess injury risk, but do not encompass the broader spectrum of performance. While these tests effectively evaluate mobility and stability, the construct of power is not assessed during movement screening tests. Power is an integral component of most sports with a high risk of injury. Side-to-side asymmetries in the ability to produce power may put an athlete at higher risk of injury. Consequently, these batteries may not provide the most comprehensive representation of an athlete's risk of injury or ability to compete. In this respect, the authors believe that the tests which comprise the FMS™ and Frohm-9 may have their respective place in the hierarchical levels for the global assessment of function; however, there are alternative PPM screening batteries that may specifically add benefit to the assessment of function and performance in athletic populations. Tests that evaluate performance can be challenging to administer, requiring the athlete to consistently put forth maximal effort to ensure optimal test reliability. While more research needs to be

conducted to establish the construct validity, responsiveness, and criterion validity of the proposed 16-PPM these results suggest that the majority of the included PPMs examined exhibited acceptable reliability and may be a beneficial part of future injury risk and return to play screening procedures.

The primary limitation of this study was the methodology used in examining intrarater reliability. While each rater independently scored the performance of each test, the instructions were given only by the expert examiner. True intrarater reliability would involve each rater independently setting up, instructing, and scoring the results of each test. However, the reliability of performance measures is influenced by a rater's ability to consistently score a test and an athlete's ability to consistently perform a test. Thus, requiring the athlete to perform the same screen separately for each of the three raters on two occasions may have induced more error into the intrarater reliability results.

CONCLUSIONS

In summary, this study assessed the reliability of 16-PPM that, in combination, provide a comprehensive screen of an athlete's movement quality and performance, examining such constructs as strength, stability, mobility, and power. The proposed battery of tests has dynamically evolved into a preliminary screen of the 16 tests described in this manuscript, which is distinctly different from the FMS™ and Frohm-9, due to the presence of tests that more specifically measure power, function, and performance. The combination of movement and performance based tests is important in understanding the full picture of an athlete's movement strategies and neuromuscular deficits. More work is needed to investigate the validity of the entire battery and each individual test to predict injury risk. While the authors believe that this screening battery is comprehensive, further work may ultimately reduce the number of tests needed for injury risk prediction or return to play decisions.

REFERENCES

1. Hootman JM, Dick R, Agel J. Epidemiology of collegiate injuries for 15 sports: summary and recommendations for injury prevention initiatives. *J Athl Train*. 2007;42(2):311-319.

2. Mather RC, 3rd, Koenig L, Kocher MS, et al. Societal and Economic Impact of Anterior Cruciate Ligament Tears. *J Bone Joint Surg Am*. 2013;95(19):1751-1759.
3. Tracey J. The emotional response to the injury and rehabilitation process. *J Appl Sport Psychol*. 2003;15:279-293.
4. Oiestad BE, Engebretsen L, Storheim K, Risberg MA. Knee osteoarthritis after anterior cruciate ligament injury: a systematic review. *Am J Sports Med*. 2009;37(7):1434-1443.
5. Finch C. A new framework for research leading to sports injury prevention. *J Sci Med Sport*. 2006;9(1-2):3-9; discussion 10.
6. Cook G, Burton L, Kiesel K, Rose G, Bryant MF. *Movement: Functional Movement Systems: Screening, Assessment, Corrective Strategies*. Aptos, CA: On Target Publications, 2010.
7. Frohm A, Heijne A, Kowalski J, Svensson P, Myklebust G. A nine-test screening battery for athletes: a reliability study. *Scand J Med Sci Sports*. 2012;22(3):306-315.
8. Reiman MP, Manske RC. *Functional Testing in Human Performance*. Champaign, IL: Human Kinetics; 2009.
9. Teyhen DS, Shaffer SW, Lorensen CL, et al. The Functional Movement Screen: a reliability study. *J Orthop Sports Phys Ther*. 2012;42(6):530-540.
10. Kazman BJ, Galecki J, Lisman P, Deuster PA, F.G. OC. Factor Structure of the Functional Movement Screen in Marine Officer Candidates. *J Strength Cond Res*. 2013.
11. Lisman P, O'Connor FG, Deuster PA, Knapik JJ. Functional movement screen and aerobic fitness predict injuries in military training. *Med Sci Sports Exerc*. 2013;45(4):636-643.
12. Reiman MP, Manske RC. The assessment of function: How is it measured? A clinical perspective. *J Man Manip Ther*. 2011;19(2):91-99.
13. Elias JE. The Inter-rater Reliability of the Functional Movement Screen within an athletic population using Untrained Raters. *J Strength Cond Res*. 2013;1533-4287 (Electronic).
14. Gribble PA, Brigle J, Pietrosimone BG, Pfile KR, Webster KA. Intrarater reliability of the functional movement screen. *J Strength Cond Res*. 2013;27(4):978-981.
15. Minick KI, Kiesel KB, Burton L, Taylor A, Plisky P, Butler RJ. Interrater reliability of the functional movement screen. *J Strength Cond Res*. 2010;24(2):479-486.
16. Onate J, Cortes N, Welch C, Van Lunen BL. Expert versus novice interrater reliability and criterion validity of the landing error scoring system. *J Sport Rehabil*. 2010;19(1):41-56.

17. Onate JA, Dewey T, Kollock RO, et al. Real-time intersession and interrater reliability of the functional movement screen. *J Strength Cond Res.* 2012;26(2):408-415.
18. Shultz R, Anderson SC, Matheson GO, Marcello B, Besier T. Test-retest and interrater reliability of the functional movement screen. *J Athl Train.* 2013;48(3):331-336.
19. Smith CA, Chimera NJ, Wright NJ, Warren M. Interrater and intrarater reliability of the functional movement screen. *J Strength Cond Res.* 2013;27(4):982-987.
20. Robertson DGE, Fleming D. Kinetics of standing broad and vertical jumping. *Can J Sport Sci.* 1987;12(1):19-23.
21. Goldbeck TG, Davies GJ. Test-Retest Reliability of the Closed Kinetic Chain Upper Extremity Stability Test: A Clinical Field Test. *J Sport Rehabil.* 2000;9(1):35-45.
22. Plisky PJ, Gorman PP, Butler RJ, Kiesel KB, Underwood FB, Elkins B. The reliability of an instrumented device for measuring components of the star excursion balance test. *N Am J Sports Phys Ther.* 2009;4(2):92-99.
23. Crill MT, Kolba CP, Chleboun GS. Using Lunge Measurements for Baseline Fitness Testing. *J Sport Rehabil.* 2004;13(1):44-53.
24. Wilkerson G. Prediction of Core & Lower Extremity Sprains & Strains in College Football Players 2009 – 2011 3-Season Analysis. Paper Present at: 2012 Big Sky Sports Medicine Conference; February 2012; Big Sky, Montana.
25. Noyes FR, Barber SD, Mooar LA. A rationale for assessing sports activity levels and limitations in knee disorders. *Clin Orthop Relat Res.* 1989(246):238-249.
26. Mjolsnes R, Arnason A, Osthaugen T, Raastad T, Bahr R. A 10-week randomized trial comparing eccentric vs. concentric hamstring strength training in well-trained soccer players. *Scand J Med Sci Sports.* 2004;14(5):311-317.
27. Myer GD, Ford KR, Paterno MV, Nick TG, Hewett TE. The effects of generalized joint laxity on risk of anterior cruciate ligament injury in young female athletes. *Am J Sports Med.* 2008;36(6):1073-1080.
28. Kottner J, Audige L, Brorson S, et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *J Clin Epidemiol.* 2011;64(1):96-106.
29. Portney LG, Watkins MP. *Foundations of Clinical Research: Applications to Practice.* 3rd ed. Upper Saddle River, NJ: Pearson Printice Hall; 2009.
30. Vincent WJ, Weir JP. *Statistics in Kinesiology.* 4th ed. Champaign, IL: Human Kinetics; 2012.
31. Tooth LR, Ottenbacher KJ. The kappa statistic in rehabilitation research: an examination. *Arch Phys Med Rehabil.* 2004;85(8):1371-1376.
32. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33(1):159-174.
33. Myer GD, Schmitt LC, Brent JL, et al. Utilization of modified NFL combine testing to identify functional deficits in athletes following ACL reconstruction. *J Orthop Sports Phys Ther.* 2011;41(6):377-387.
34. Plisky PJ, Rauh MJ, Kaminski TW, Underwood FB. Star Excursion Balance Test as a predictor of lower extremity injury in high school basketball players. *J Orthop Sports Phys Ther.* 2006;36(12):911-919.
35. Tucci HT, Martins J, Sposito Gde C, Camarini PM, de Oliveira AS. Closed Kinetic Chain Upper Extremity Stability test (CKCUES test): a reliability study in persons with and without shoulder impingement syndrome. *BMC Musculoskelet Disord.* 2014;15(1):1.
36. Udermann BE, Mayer JM, Graves JE, Murray SR. Quantitative assessment of lumbar paraspinal muscle endurance. *J Athl Train.* 2003;38(3):259-262.
37. Ross MD, Langford B, Whelan PJ. Test-retest reliability of 4 single-leg horizontal hop tests. *J Strength Cond Res.* 2002;16(4):617-622.
38. Gulgin H, Hoogenboom B. The functional movement screening(FMS)[™]: An inter-rater reliability study between raters of varied experience. *Int J Sports Phys Ther.* 2014;9(1):14-20.
39. Minick KI, Kiesel KB, Burton L, Taylor A, Plisky P, Butler RJ. Interrater reliability of the functional movement screen. *J Strength Cond Res.* 2010;24(2):479-486.
40. Schneiders AG, Davidsson A, Horman E, Sullivan SJ. Functional movement screen normative values in a young, active population. *Int J Sports Phys Ther.* 2011;6(2):75-82.
41. Teyhen DS, Shaffer SW, Lorensen CL, et al. The Functional Movement Screen: a reliability study. *J Orthop Sports Phys Ther.* 2012;42(6):530-540.

Quantitatively- Scored Tests

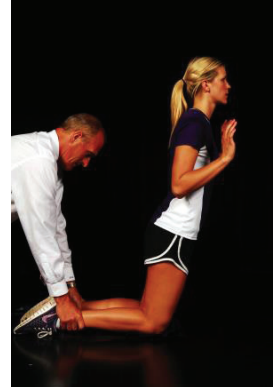
Broad Jump



Triple Hop for Distance



Nordic Hamstring



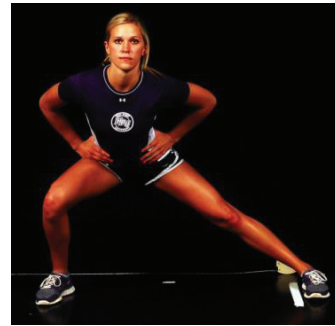
Lower Extremity Y-Balance



In-Line Lunge for Distance



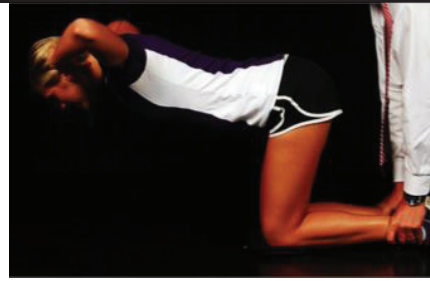
Lateral Lunge for Distance



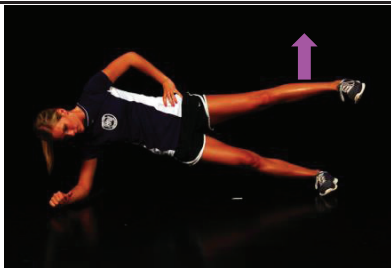
CKCUEST



Lumbar Endurance



Side Plank Hip Abduction

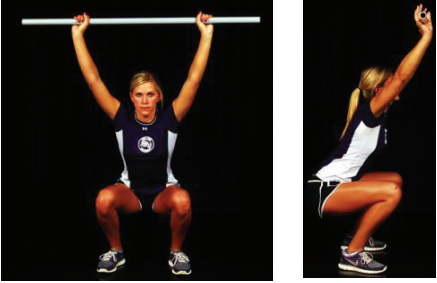
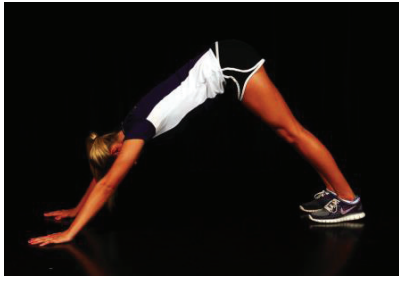







Side Plank Hip Adduction



Appendix 1. 16-PPM Battery (continued)

Qualitatively- Scored Tests

Full Squat		Downward Dog	
 <p><i>Scoring:</i> 0: Pain with test 1: Unable to squat without substitutions to parallel with a 2x4 2: Squat to parallel with 2x4, no substitutions 3: Full squat with 2x4, no substitutions 4: Squat to parallel, no substitutions 5: Full squat (below 90°), no substitutions</p>		 <p><i>Scoring:</i> 0: Pain with test 1: Unable to get heel to floor with one leg lifted straight back 2: One heel to floor with other leg lifted straight back 3: Both heels to floor with adjusted hand position and rounded back 4: Both heels to floor with adjusted hand position, and flat back 5: Patient able to touch heels to floor and assume inverted V position</p>	
Single Leg Squat	Active Straight Leg Raise	Shoulder Mobility	
 <p><i>Scoring:</i> 0: Pain with test 1: 5+ errors observed 2: 4 errors observed 3: 3 errors observed 4: 2 errors observed 5: 0-1 error observed</p>	 <p><i>Scoring:</i> 0: Pain with test 1: Dowel lines up below mid-shin 2: Dowel lines up between patella and mid-shin 3: Dowel lines up between mid-thigh and patella 4: Dowel lines up between greater trochanter and mid-thigh 5: Dowel lines up with greater trochanter or above</p>	 <p><i>Scoring:</i> 0: Pain with test 1: Fists greater than 2-hand lengths apart with or without scapular winging 2: Fists between 1- and 2-hand lengths apart with or without scapular winging 3: Fists within 1-hand length with scapular winging 4: Fists within 1-hand length with without scapular winging 5: Fists touch without scapular winging</p>	
Beighton Hypermobility			
			
		<p><i>Scoring:</i> 0: Pain with test 1: 8-9 points 2: 6-7 points 3: 4-5 points 4: 2-3 points 5: 0-1 points</p>	