



Brazilian Journal of Physical Therapy

<https://www.journals.elsevier.com/brazilian-journal-of-physical-therapy>



ORIGINAL RESEARCH

Comprehensive movement system screening tool (MSST) for athletes: Development and measurement properties



Courtney M. Butowicz^{a,c,*}, Marisa Pontillo^{a,b}, David Ebaugh^a, Sheri P. Silfies^{a,c}

^a Physical Therapy and Rehabilitation Sciences Department, Drexel University, Philadelphia, PA, USA

^b Penn Sports Medicine Center, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA

^c Exercise Science Department, University of South Carolina, Columbia, SC, USA

Received 15 January 2019; received in revised form 2 October 2019; accepted 9 October 2019

Available online 1 November 2019

KEYWORDS

Athletic injuries;
Screening;
Movement system;
Validity

Abstract

Objectives: Non-traumatic extremity injuries are particularly common in sports, representing a significant economic, academic, and psychosocial burden on athletes. Proposed musculoskeletal risk factors for increased injury and decreased performance in athletes include movement pattern inefficiency, decreased regional stability, decreased mobility, and asymmetrical movement. The Movement System Screening Tool (MSST) is a comprehensive screening tool designed to assess these factors. Thus, the purpose of this study was to describe the development and determine the content and construct validity and inter-rater reliability of the MSST.

Methods: A modified Delphi panel of experts determined content validity. 80 athletes (40 with and 40 without a current non-traumatic shoulder injury) completed 21 clinical tests, with exploratory factor analysis and known group analysis performed to determine construct validity. Two independent raters were used to establish individual item and composite score inter-rater reliability.

Results: Exploratory factor analysis identified three of the four apriori constructs over 7 factors (14 tests), representing 63% of the variance. Known group analysis revealed a significantly lower composite score in athletes with vs. without a current non-traumatic shoulder injury (56.9 ± 5.8 vs. 62.7 ± 4.5 , respectively). A preliminary cut score of 62 was chosen with sensitivity of 85% and specificity of 44%. Composite score inter-rater reliability was excellent ICC (2,1) = 0.94, 95% CI (0.91, 0.96) and item reliability ranged from $\kappa = .57$ to 1.00.

Conclusions: The MSST possesses constructs representative of injury risk and measurement properties acceptable for use in clinical settings. Comprehensive screens with construct validity and known measurement error are needed not only to identify athletes at risk of injury, but

* Corresponding author at: 921 Assembly St, Room 308, Columbia, SC 29205, USA.

E-mail: c.butowicz@gmail.com (C.M. Butowicz).

also provide an instrument that can be used in studies that seek to validate training approaches proposed to change movement impairment and injury risk in athletes.

© 2019 Associação Brasileira de Pesquisa e Pós-Graduação em Fisioterapia. Published by Elsevier Editora Ltda. All rights reserved.

Introduction

Sports injuries are common and represent a significant economic, academic, and psychosocial burden on athletes.¹ Non-contact injuries account for nearly 40% of all injuries sustained during practices and 20% of injuries incurred during sanctioned games.² While nearly one-third of all non-contact injuries affect overhead athletes,³ most movement screens assessing injury risk predominately focus on lower extremity and/or only address upper extremity mobility.^{4–6} Risk factors suggested to decrease performance and increase injury risk include decreased regional stability (e.g., core stability), movement pattern inefficiency, decreased mobility, and asymmetrical movement.^{4,7–22} Movement screening tools are widely used to assess injury risk in athletic populations; however, few have established validity and reliability.^{23–28}

The Functional Movement Screen (FMS) and Nine Test Screen Battery (9TSB) are examples of screens that aim to identify functional movement limitations,²⁹ yet the definition of “fundamental” movement is disagreed upon and limited evidence exists regarding the development methods of these tools.³⁰ The FMS has demonstrated good inter-rater and test-retest reliability^{25,26}; however, predictive validity of the screen is inconclusive.^{31–35} Furthermore, construct factor analysis does not support use of the composite FMS score as low internal consistency and inconsistent factor loading was demonstrated among the 7 FMS tasks within different populations.^{36,37} The 9TSB has demonstrated good inter-rater reliability, with no evidence of predictive validity.²³ Additionally, these screening tools are heavily biased toward lower extremity stability and mobility tests, and lack more dynamic tests (e.g., plyometric movements). Therefore, there is a need for a valid and reliable comprehensive screening tool with adequate measurement properties and injury predictive validity based on proposed injury risk factor constructs.³⁰

To address limitations of current movement screening tools, a new tool was systematically developed. The Movement System Screening Tool (MSST) was designed to assess the constructs of *movement pattern efficiency*, *regional stability*, *mobility*, and *movement symmetry* across the upper and lower extremities and core. The purposes of this study were to: 1) describe the development of the MSST, and 2) determine the measurement properties (content validity, construct validity, inter-rater reliability) in an athletic population.

Methods

Participants

Eighty athletes participated in the cross-sectional study (40 with a current shoulder injury). Sixty-five percent of the athletes (26/40) with a shoulder injury were currently receiving physical therapy services. Inclusion criteria were athletes between 18 to 35 years of age with a minimum participation of 10 h per week in practice, games, and/or strength and conditioning workouts. An athlete was defined as any individual currently competing in any sport at a professional, semi-professional, collegiate varsity, junior varsity, or club level. Athletes were excluded if they presented with: concussion (current or within the last six months), current leg, trunk or neck injury, a diagnosed balance disorder, and/or a current head cold, sinus infection, or inner ear infection. Athletes with a shoulder injury experienced a non-traumatic shoulder injury within the previous six months that required medical treatment. Athletes with and without a shoulder injury were matched by age ($\pm$ five years), sex, sport group [a) overhead athletes; b) athletes who use their upper extremities in their sports, but are not overhead athletes, (e.g., lacrosse)]; and body mass index ($\pm$ 5 kg/m²). Athletes from the following sports participated: football (27), baseball/softball (10), CrossFit/Weightlifting (10), swimming/water polo (9), crew/sailing (5), wrestling/rugby (5), tennis (4), track and field (4), basketball (2), lacrosse (2), soccer (1), rock climbing (1). All athletes gave written informed consent to procedures approved by the Drexel University Institutional Review Board (Philadelphia, PA, USA, Approval #1307002208).

MSST development

A systematic literature search generated a list of tests with reported athletic injury predictive capabilities and/or that are commonly used in musculoskeletal assessments of movement patterns, mobility, control of dynamic movements, or muscle capacity (e.g., strength, endurance). The literature search was performed within the CINAHL, MEDLINE, PUBMED, and Google Scholar databases, with supplemental searches conducted from review of the reference lists. See Appendix 1 for list of search terms and flow diagram of literature used for content validity. Tests were developed and/or modified for constructs in which a panel of expert

clinicians identified a gap in the literature. The identified gap was specific tests of core stability. As part of this process, the unilateral hip bridge endurance (UHBE) test was modified and validated as a clinical test of core control.³⁸ A seated clinical core control test (3CT) was developed as another potential assessment of core control.³⁹

Content validity

A modified Delphi approach⁴⁰ with an expert panel (N = 15) of experienced (mean experience: $15.1 \pm .9$ years) physical therapists with sports and orthopaedic specialty certifications, athletic trainers, certified strength and conditioning specialists, and biomechanists from different institutions determined the *content validity* and reduced the number of tests from the initial list. Content validity reflects whether or not the “domain of content” is sufficiently covered by the instrument or measure.⁴¹ An instrument possesses content validity if it comprises all components of the theoretical content and illustrates the relative importance of each component.⁴² For example, the proposed MSST would demonstrate content validity if the experts reached consensus that the theoretical constructs of the MSST are adequately represented, and the importance of each construct/component was reflected. Experts determined if the literature review results comprehensively addressed proposed injury risk factors and if each body region was represented in the first Delphi round. In the second round, experts rated their level of agreement on which of the proposed risk factors for injury/constructs each test assessed, the primary body region the test was assessing, and the importance of symmetry (Appendix 1, Supplementary Materials). A majority (13/15) agreed on the primary region each test assessed, the importance of asymmetry, and the proposed risk factor/construct.

In the third round, experts received the second round results, and the inter-rater reliability and validity of individual tests. Inter-rater reliability for individual tests came from a sub-study completed in our lab or from the published literature. Experts rank-ordered tests for each construct/proposed risk factor based on results of the previous round, test measurement properties, and their expert opinion. Third round results achieved consensus and revealed that experts selected 21 (15 tested bilaterally) of the 33 original tests for inclusion in the screen.

Construct validity

Construct validity is the ability of an instrument to measure an abstract concept or construct.⁴² Construct validity establishes theoretical relationships between the primary concept of interest (i.e., movement system impairment) and another concept (i.e., injury).⁴¹ Eighty athletes were recruited for the construct validity study. Age, height, mass, and hours of activity per week, were recorded (Table 1). For the purposes of describing the sample, verifying that healthy controls did not have an underlying shoulder injury, and those with an injury demonstrated impairments, all athletes underwent a clinical shoulder exam and completed the Penn Shoulder Score (PSS). Shoulder examination included contractile testing of shoulder musculature, range

of motion, and provocative testing. The PSS has test-retest reliability of $ICC_{(2,1)} = 0.94$ (95% CI: 0.89, 0.97), internal consistency (Cronbach’s alpha) of 0.93, $SEM = \pm 8.5$ scale points (based on 90% CI).⁴³

Athletes completed 21 tests identified through the Delphi process, in one session with a maximum of three trials on each test. The best performance was recorded. Raters recorded the location of any deviation (i.e., joint/segment), the degree of deviation (i.e., none, subtle, severe), and the presence of pain with movement. Test performance was scored as per documented test criteria. The original test criteria or measurements were converted to a 0–3 scale, similar to other published movement screens,^{29,44} with a maximum MSST composite score of 78. Tests performed bilaterally were rated/scored for each side, except glenohumeral internal rotation deficit (GIRD). GIRD determination and score was based on published scoring criteria, resulting in a single score for this bilateral assessment.⁴⁵

Inter-rater reliability sub-study

Two raters independently assessed each of the 80 athletes’ performance on the qualitative tests within the MSST (e.g., side bridge with active hip abduction, step down). Raters were not blind to athlete group status and included an exercise scientist and a physical therapist with a sports specialist certification. The second rater was trained via recordings and live performances of athletes performing MSST tests prior to data collection. During data collection, the primary rater read the test instructions to the athletes. For tests requiring a qualitative assessment, raters independently recorded performance observations. Quantitative tests within the MSST (e.g., single-leg hop) were not included in the sub-study given prior published inter-rater reliability.^{27,46–49}

Statistical analysis

Independent *t*-tests assessed athlete characteristics to confirm there were no group differences ($p < 0.05$) on the matched characteristics. All statistical analyses were conducted using SPSS (version 22.0; SPSS Inc, Chicago, IL).

Construct validity

Exploratory factor analysis

Prior to the exploratory factor analysis (EFA), test items were analyzed for multicollinearity. Randomly missing data (<1%) were imputed using single imputation while non-randomly missing data were imputed using multiple imputations. The average of five imputations derived data points for non-randomly missing data. Tests were examined for factor analysis appropriateness via the Kaplin-Meyer-Olkin (KMO) measure of sampling adequacy, Bartlett’s test of sphericity, and the anti-image correlation matrix. KMO and Bartlett’s test of sphericity determined if an adequate number of tests for each factor existed, and whether the original tests were sufficiently correlated, respectively.^{28,50} Tests with unacceptable sampling adequacy were removed, and an EFA with and without factor rotation (Varimax rotation⁵⁰)

Table 1 Athlete characteristics by group and sex (mean $\pm$ standard deviation).

Group	Sex	N	Age (yrs)	Height (cm)	Mass (kg)	Hours Per week	
						Activity	PSS
CTR	F	13	21.4 $\pm$ 4.2	168.5 $\pm$ 8.9	68.7 $\pm$ 11.6	18.5 $\pm$ 6.7	29.5 $\pm$ 1.5
	M	27	20.9 $\pm$ 3.1	179.8 $\pm$ 7.2	92.1 $\pm$ 16.5	19.4 $\pm$ 5.1	29.3 $\pm$ 1.3
INJ	F	13	22.2 $\pm$ 3.9	165.8 $\pm$ 6.5	64.4 $\pm$ 9.4	16.5 $\pm$ 4.9	20.6 $\pm$ 6.0
	M	27	20.8 $\pm$ 2.8	181.3 $\pm$ 7.0	94.3 $\pm$ 17.1	19.7 $\pm$ 7.5	23.9 $\pm$ 4.0

Abbreviations: CTR: controls; INJ: injured; PSS: Penn Shoulder Score.

was completed. Factor extraction was performed using the maximum likelihood method and factors with eigenvalues greater than 1.0 were extracted while test item communalities within each factor were suppressed at 0.4.⁵¹

Known group analysis

Since movement system impairment is associated with non-traumatic injuries,²⁷ a known group validity approach was used to determine if the MSST could discriminate between athletes with and without a current non-traumatic injury. An independent *t*-test was used to assess MSST composite scores between these groups. An *a priori* power analysis determined a sample size of 34 per group was needed to achieve significance at $p < 0.05$ and power of 0.80 for a two-tailed independent *t*-test with an effect size of 0.8.

In an attempt to protect athletes with current shoulder injury from potential further injury, the closed kinetic chain upper extremity stability test (CKCUEST) was not performed in that group. Analysis was first performed with a CKCUEST score of zero assigned to athletes with shoulder injury. However, it is impossible to know if all athletes with current shoulder injury would have received a zero score on this test. Therefore, to reduce potential bias of this approach, a second analysis was performed with the CKCUEST score removed from the composite score for all athletes.

A receiver operating characteristic (ROC) curve analysis was also performed to assess the ability MSST performance to discriminate between groups and to identify a preliminary cutoff score which maximized sensitivity and specificity for identifying movement system impairments associated with shoulder injury. The area under the curve (AUC) was interpreted as: no discrimination (0.5), acceptable (between 0.7 and 0.8), excellent (between 0.8 and 0.9), and outstanding (>0.9).⁵²

Inter-rater reliability sub study

Inter-rater reliability (Cohen's kappa [k]) was determined for qualitatively scored individual tests based upon scores determined through visual assessment of test performance. Inter-rater reliability of individual tests were interpreted according to Landis and Koch [<0 : no agreement, 0.1–0.20: slight agreement, 0.21–0.40: fair agreement, 0.41–0.60: moderate agreement, 0.61–0.80: substantial agreement, 0.81–0.99: almost perfect agreement].⁵³

Inter-rater reliability of the composite score

Individual test scores were converted to generate a composite score for the full screen. The composite score inter-rater reliability was assessed using an Intraclass cor-

relation coefficient ($ICC_{(2,1)}$) and interpreted according to Portney and Watkins⁴² [<0.40 : poor reliability, 0.41–0.74: moderate reliability, >0.75 : excellent reliability].

Results

Content validity & inter-rater reliability of individual tests

The modified Delphi approach identified 21 different tests (15 were bilateral) for potential inclusion. Inter-rater test reliability ranged from $\kappa = 0.03$ –1.00 (Table 2). The 21 proposed tests demonstrated content validity and represented assessments of each of the proposed risk factors in each body region.

Construct validity

The KMO (0.58) and Bartlett's test ($p = 0.001$) revealed that 15 of the remaining 21 MSST tests were appropriate for the EFA. Six tests were removed during examination of individual test KMO statistics within the anti-image correlation matrix secondary to unacceptable sampling adequacy: modified Thomas test, in-line lunge, trunk stability push-up, trunk extensor endurance, trunk flexor endurance, and 3CT.

EFA revealed 7 factors (Eigenvalues greater than 1.0; Table 3), representing 63% of the explained variance. Scapular dyskinesis did not meet the criteria for test item commonality within a factor and was therefore removed from the screen. The construct of movement pattern efficiency (MPE; 21.3% explained variance) loaded on two factors: lower extremity MPE with core stability, and dynamic MPE. The construct of regional stability represented 24.8% explained variance loading on three factors (dynamic, core, and core to extremity). Mobility (16.8% explained variance) loaded on two regional factors (extremity mobility; trunk mobility). For details see factor structure in Table 3. The factor structure was further examined using a Varimax rotation, which produced 8 distinct factors, explaining 67% of the variance. The 8 factors represent the same constructs as the original analysis; however, the rotation further separated the mobility construct into three regions (upper extremity, lower extremity and trunk).

The 14 remaining tests did not appear to equally address the construct across all regions. Therefore, the 3CT, representing isolated core stability (removed due to unacceptable sampling adequacy), and the scapular dyskinesis test (only

Table 2 Individual item inter-rater reliability (kappa) of qualitative assessments performed on all athletes.

Test	κ	Test	κ
Active hip ABD-L	.30	Active Straight Leg Raise-L	*
Active hip ABD-R	.62	Active Straight Leg Raise-R	*
Resisted active hip ABD-L	.73	Double leg hip bridge	*
Resisted active hip ABD-R	.40	Bridge-leg extension-L	.36
Side bridge-L	*	Bridge-leg extension-R	.61
Side bridge-R	*	Bridge-leg extension resisted-L	*
Side bridge hip ABD-L	.57	Bridge leg extension resisted-R	*
Side bridge hip ABD-R	.50	Prone hip extension-L	.75
Side bridge hip ABD resisted-L	.09	Prone hip extension-R	.87
Side bridge hip ABD resisted-R	.25	Prone hip extension-arm lift-L	.03
Modified Thomas test-L	*	Prone hip extension-arm lift-R	.50
Modified Thomas test-R	*	Step down-L	.82
Rotary stability-L	*	Step down-R	.73
Rotary stability-R	*	In-line lunge-L	*
Hurdle step-L	1.0	In-line lunge-R	*
Hurdle step-R	.87	Trunk stability push-up	*
Overhead squat	*		

Abbreviations: κ : kappa; L: left; R: right; ABD: abduction. * Indicates no covariance due to perfect agreement between raters, thus kappa could not be computed.

Table 3 Factor Loading for Non-Rotated Factor Matrix.

Factor	Test	Loading	% Explained Variance
LE Movement Pattern Efficiency	Bridge-Hip Abduction	0.53	15.0%
	Hurdle Step	0.64	
	Double Leg Lowering	0.55	
	Step Down	0.67	
Mobility	Shoulder Mobility	0.77	12.1%
	Straight Leg Raise	0.64	
	Shoulder Internal Rotation	0.52	
Dynamic Stability	Single Leg Hop	0.77	10.3%
	CKCUEST	0.53	
Core Stability	Unilateral Hip Bridge	0.65	8.2%
Core to Extremity Stability	Overhead Squat	0.44	6.3%
	Rotary Stability	0.47	
Dynamic Movement Pattern Efficiency	Y-Balance Test	0.52	6.2%
Core Mobility	Trunk Flexion/Extension	0.50	4.8%
Total			62.9%

Abbreviations: LE: lower extremity; TP: trunk/pelvis; CKCUEST: closed kinetic chain upper extremity stability test.

assessment of scapular movement pattern) were kept in the final 16-test version of the screen (Table 4). Notably, both of these tests had good inter-rater reliability. Inter-rater reliability of the qualitative tests in the final version demonstrated fair to perfect agreement ($\kappa = 0.57$ – 1.00). Using data from our cohort study, we established within session reliability for the unilateral hip bridge endurance ($ICC_{(2,1)} = 0.94$ – 0.95) and 3CT ($ICC_{(2,1)} = 0.85$).

Known group analysis

There were no significant differences in demographics between athletes with and without non-traumatic shoulder injury. Athletes with a shoulder injury scored significantly lower on the Penn Shoulder Score ($\bar{x} = 78.2 \pm 11.7$;

$t = 10.0$, $p < 0.0001$, Cohen's $d = 2.2$) than athletes without an injury ($\bar{x} = 97.5 \pm 3.9$). Athletes with a shoulder injury (MSST composite score = 59.3 ± 5.9) scored significantly lower than healthy athletes (MSST composite score = 65.0 ± 4.8) with the CKCUEST included in the analysis ($t = 4.77$, $p < 0.0001$, Cohen's $d = 1.06$). Mean difference between groups (5.7 points) resulted in a large effect size. The findings did not change with removal of the CKCUEST from the composite score of both groups (shoulder injury group $\bar{x} = 56.9 \pm 5.8$; control group $\bar{x} = 62.2 \pm 4.8$, $t = 2.43$, $p = 0.02$, Cohen's $d = 0.54$). ROC curve analysis revealed the MSST's ability to identify movement system impairment associated with shoulder injury was acceptable ($AUC = 0.76$, $p < 0.0001$, 95% CI: 0.65, 0.86). A preliminary cut score of 62 was chosen with sensitivity of 85% and specificity of 44%.

Table 4 Movement System Screening Tool (MSST) Descriptions and Scoring Criteria for the Tests.

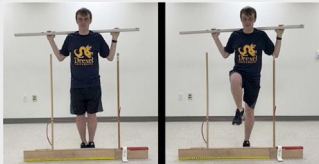
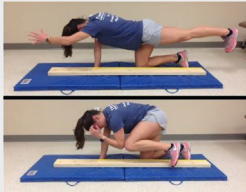
Test	Description	Scoring
Lower Extremity Tests Side bridge with active hip abduction 	Participant lies on his/her side, propped up on the forearm with the shoulder over the elbow, and the bottom knee bent to 90°. Top and bottom thighs should be in line with one another. Top leg should be straight, toes lifted towards the shin and pointed forward. Participant lifts pelvis off the floor/table until head, spine, and bottom leg are in a straight line, then raises the top leg.	0: Pain present during test 1: Obvious deviations were noted 2: Subtle deviations were noted 3: No compensations or deviations were present Deviations: - Pelvis or spine did not remain in a neutral position - Lack of symmetry in performance between sides - Any trunk movement Sides scored separately. Total Possible Points = 6
Overhead squat¹⁷ 	Standing with feet shoulder width apart, toes pointing forward and holding the PVC overhead, he/she squats as deeply as possible 3 times. The rater evaluates from the front, side, and back. If participant was unable to complete the full motion in an error-free fashion, a wooden 2 x 6 board was placed under the heels and the squat was re-evaluated.	0: Pain present during test 1: Unable to squat with thighs below parallel without compensations using 2 x 6 board 2: Able to squat to below parallel with 2 x 6 board 3: Able to squat to below parallel without 2 x 6 board and no compensations. Total Possible Points = 3
Hurdle step¹⁷ 	Participant stands with feet together and toes touching the board. Grasping the dowel with both hands, he/she places the dowel behind his/her neck and across the shoulders. From this position, the participant maintains an upright posture, raises the right leg and step over the hurdle, making sure to keep the moving foot toes towards the shin. Once hurdle is cleared, participant touches the floor with his/her heel and returns to the starting position. Alignment of the ankle, knee and hip should be maintained.	0: Pain present during test 1: Contact between the moving foot and hurdle or loss of balance was observed 2: Alignment was lost between hips, knees and ankles or the dowel did not remain horizontal 3: Test was performed without any compensation *Leg stepping over the hurdle is the leg being scored Sides scored separately. Total Possible Points = 6
Rotary stability¹⁷ 	Participant assumes a quadruped position with the hips and knees at 90° and a 2 x 6 board between the hands and knees. With the ankles dorsiflexed, toes, knees, and thumbs touching the board, participant raises the arm and extends the ipsilateral hip and knee simultaneously. After achieving this position, participant brings the elevated elbow and knee towards the midline of the body to make contact above the board and then return to the starting position.	0: Pain present during test 1: Unable to perform the diagonal pattern 2: Unable to perform the ipsilateral pattern but able to perform the diagonal pattern 3: Able to perform the ipsilateral pattern Sides scored separately. Total Possible Points = 6

Table 4 (Continued)

Test	Description	Scoring
Trunk flexion/extension mobility ¹⁷	For the trunk flexion test, participant assumes a quadruped position and are instructed to rock his/her hips toward his/her heels, lower the chest to his/her knees, and reach his/her hands in front of the body as far as possible. For the trunk extension test, participant lies prone on an exercise mat with the hands directly under the shoulders and the feet together. From this position, participant presses his/her chest off the mat as much as possible by straightening the elbows and without any lower body movement. Raters observe for limited or excessive motion in the shoulder (flexion test only), thoracic and lumbar spines, and hips.	0: Pain present during either test 1: Limited or excessive motion observed on both tests 2: Limited or excessive motion observed on only one test 3: No limited or excessive motion was observed on either test Total Possible Points = 3
Step down ⁶⁰	Participant stands on a 20-cm stool, crosses his/her arms across the chest, and squats down as far as possible (attempting to touch the reaching heel to the ground) five times consecutively. The movement is performed in a slow and controlled manner, while maintaining balance. If unable to touch the reaching heel to the ground, he/she is instructed to stop at approximately 60° of knee flexion. Raters provide verbal feedback, if needed, to those unable to reach the heel to the ground. This ensures that any deviations are not a function of step height. Adapted from Crossley et al 2011. ⁶⁰	0: Pain present during test 1: Obvious deviations were noted 2: Subtle deviations were noted 3: No compensations or deviations were present Deviations: - Pelvis and trunk did not remain neutral - Knee collapsed toward midline of the body - Stance heel lifted from the step or loss of balance - Inability to achieve at least 60° knee flexion Sides scored separately. Total Possible Points = 6
Y-Balance Test Anterior ⁴⁹	Participant stands with stance foot in the middle of the platform so that toes are behind the start line, hands on hips, and reach forward pushing the indicator box as far as possible in one smooth movement. After box is pushed, the distance is recorded, and box is returned to the start position for the next trial. Trials are unsuccessful if at any point the hands came off the hips, the stance heel is lifted, or a loss of balance is observed. The average of 3 trials is normalized to height. This test was adapted from Plisky et al 2009 ⁴⁹	0: Pain present during test 1: >2 SD from normative mean by sex 2: <2 and >1 SD from normative mean by sex 3: <1 SD from normative mean by sex Women's scores were based on means of 0.43 ± 0.06 (left) and 0.43 ± 0.07 (right) Men's scores were based on means of 0.45 ± 0.06 (left) and 0.45 ± 0.06 (right) Sides scored separately. Total Possible Points = 6
Active straight leg raise ¹⁷	Participant lies supine on a mat table with a 2 × 6 board directly under the knees, toes pulled towards the shin, and the arms by the side. From this position, participant raises the one leg as high as possible, while keeping the raised leg straight and the opposite knee against the board.	0: Pain present during test 1: Vertical line of the malleolus of the moving leg is below the knee joint line of the non-moving leg 2: Vertical line of malleolus of moving leg is between the mid-thigh and knee joint line of the non-moving leg 3: Vertical line of the malleolus of the moving leg is above the mid-thigh of the non-moving leg Sides scored separately. Total Possible Points = 6

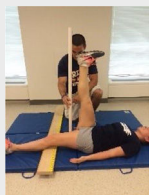
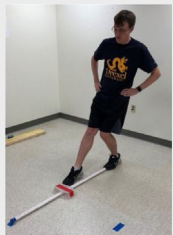
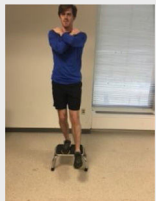


Table 4 (Continued)

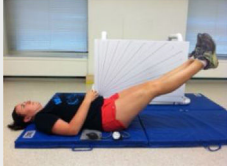
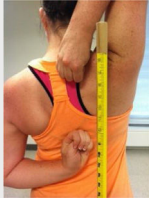
Test	Description	Scoring
Double-leg lowering test⁶¹ 	Participant lies on his/her back with legs straight and hips flexed to 90°. From this position, slowly lower the legs towards the floor without changing the pressure in the blood pressure cuff under their lumbar spine. Once the pressure changes 10 mm Hg, the angle of the legs relative to the horizontal is recorded. This test and scoring procedure was adapted from Lanning et al 2006⁶¹	0: Pain present during test 1: > 2 SD from normative mean by sex 2: <2, >1 SD from normative mean by sex 3: <1 SD from normative mean by sex Women's scores based on means of 52.0° ± 5.0° Men's scores based on means of 46.0° ± 3.0°. ⁶³ Total Possible Points = 3
Scapular dyskinesis⁶² 	Test uses 5 repetitions of bilateral, active shoulder flexion and abduction. Each motion is demonstrated, and participants can perform a few practice trials. Participant performs the movements with a 2-pound (lb) (< 150 lbs) or 4-lb (> 150 lbs) weight in each hand. Participant begins standing in a neutral position, arms at the side of the body, elbows straight, and thumbs pointing up. Raters stand 2 to 3 m behind the participant to observe scapular motion. This test was adapted from McClure et al 2009. ⁶²	0: Pain present during test 1: Obvious winging or dysrhythmia observed on 3/5 trials in either flexion or abduction 2: Subtle winging or dysrhythmia observed on 3/5 trials in either flexion or abduction 3: No winging or dysrhythmia was observed Sides scored separately Total Possible Points = 6
Shoulder mobility¹⁷ 	In one motion while standing, participant moves the right fist overhead and down the back as far as possible while simultaneously taking the left fist up the back as far as possible. Do not "creep" the hands closer after initial placement. Measure the distance between the two closest bony prominences on the hands (measured to nearest 0.5 in.. Measure hand length prior to testing (most distal wrist crease to the tip of the third digit). Raters record the distance of the closest reach. If the participant can touch the hands together, then a distance of zero is recorded for that side.	0: Pain present during test 1: >1.5 hand lengths apart 2: > 1 hand length apart 3: < 1 hand length apart Sides scored separately. Total Possible Points = 6
Closed kinetic chain upper extremity stability⁵⁹ 	Two lines of tape are placed 36 inches apart on the floor. Participant starts the test in a standard push-up position, with one hand on each tapeline. Participant is allowed a practice trial to ensure proper form, defined as: feet shoulder width apart; shoulders, hips, knees and ankles aligned in the coronal plane; and each hand must touch the opposite line to count as a repetition. To perform, bring one hand over to the opposite tape line, return to the starting tapeline, and then repeated the task with the opposite hand. Test and scoring were adapted from Goldbeck and Davies. ⁵⁹	Score: Average number of touches in 15 seconds 0: Pain present during test. 1: > 2 SD from normative mean by sex 2: < 2, > 1 SD from normative mean by sex 3: < 1 SD from normative mean by sex. Normative means of 21.8 ± 3.9⁶⁴ Total Possible Points = 3

Table 4 (Continued)

Test	Description	Scoring
Glenohumeral internal rotation deficit ⁴⁵ 	The measurement includes passive range of motion of shoulder internal and external rotation. Participant is supine. The start position (0°) for both internal and external rotation measures is as follows: shoulder in 90° of abduction, elbow flexed to 90°, forearm in neutral with the participant's hand/fingers pointing up towards the ceiling. From this position, the tester moves the glenohumeral joint into external or internal rotation.	0: Pain present during test 1: If GIRD is present (non-dominant - dominant IR difference > 14° AND non-dominant - dominant total arc difference > 10°) 2: > 10° difference in total arc motion between sides, but IR difference < 14° 3: < 10° difference in total arc motion between sides
Unilateral hip bridge endurance ³⁸ 	Participant lies supine with the arms across the chest, knees in 90° of flexion, and feet flat on the table (Figure 3c). Participant performs a single-leg hip bridge and holds this position as long as possible. The test is terminated when he/she is no longer able to maintain a neutral pelvic position as noted by 10° change in transverse or sagittal plane alignment. Three trials are performed on each side. Sides are scored individually.	0: Pain present during test. 1: > 1 SD from normative mean by sex 2: <1, > 1/2 SD from normative mean by sex 3: <1/2 SD from normative mean by sex Scores were based off the following means and standard deviations: females = 33.45 ± 29.70 s; males = 26.01 ± 18.44 s Total Possible Points = 3
Clinical Core Control Test (3CT) 	Participant sits atop a 75 cm Swiss ball with arms crossed and feet on ground, so knees, hips, and ankles are at 90°. Lift the feet from the ground, keeping the heels in contact with the ball, and attempt to maintain the testing position without allowing the feet to touch the ground or ball to touch the wall 6 inches behind. Test is terminated when feet touch ground or ball touches the wall.	0: Pain present during test. 1: > 1 SD from normative mean by sex 2: < 1, > 1/2 SD from normative mean by sex 3: <1/2 SD from normative mean by sex Scores based off means and standard deviations: females = 50.12 ± 45.01 s; males = 57.54 ± 59.92 s Total Possible Points = 3

Inter-rater MSST composite score reliability

Inter-rater reliability for the composite MSST score was $ICC_{(2,1)} = 0.94$, 95% CI: 0.91, 0.96), representing excellent reliability between raters during real time, simultaneous observation of athlete performance.

Discussion

The purpose of this study was to develop a comprehensive movement screen tool and determine its measurement properties. Content validity was verified using a modified Delphi approach. Using factor loading as a proxy for constructs, results of the EFA construct validity study supported three of the four proposed constructs. Regional stability and mobility were represented by body regions. Movement pattern efficiency loaded on two factors, representing 4 lower extremity movement tests. Combined, these 7 factors explain 62.9% of the variance in the MSST suggesting that these factors represent and validate hypothesized constructs.⁵⁴ Factor structure did not significantly change after using a Varimax rotation. Yet, it provided another level of detail that may be useful in future work if the MSST is divided into region dependent subsets, such as an upper-extremity screen for use in upper-extremity dominant sports. EFA did not identify movement asymmetry given no tests directly represented side-to-side performance asymmetry. However, unlike the FMS, the MSST scores each side separately for bilateral tests, allowing the evaluation of bilateral movements independent of the composite score.^{16,54,55}

Construct validity was further supported by the known group analysis with a significantly lower MSST composite score in athletes with a shoulder injury. The calculated effect sizes for the difference scores represent a large to medium effect. While the composite score indicates impaired movement system performance, it does not provide region specific impairment, or construct-specific information. The composite score serves as an alert that further evaluation of the individual test performance may be warranted.

The MSST composite score correctly identified athletes with a shoulder injury 75% of the time. This suggests that lower scores on the MSST are associated with movement system impairment associated with shoulder injury. A preliminary cut score of 62 (chosen with a bias toward sensitivity) indicates movement system impairment, suggesting that of the athletes with an injury, 85% (sensitivity) scored below a 62, while athletes that were not injured, 44% (specificity) scored higher than a 62. It is acknowledged that the higher sensitivity chosen assumes the acceptance of a potentially greater number of false positives. However, given that screening tools are not meant to be diagnostic, a higher sensitivity is acceptable.⁵⁶ We suggest the MSST composite score be used in conjunction with other pertinent athlete information affecting injury risk, such as injury history. While the data presented are specific to shoulder injuries, the utility of the MSST to identify movement system impairments associated with lower extremity injury is unknown. Future work is needed to determine prospective injury risk in athletes. Of note, the results of the ROC should be interpreted

in the context of the study design and should not be used as a predictive score for prospective injury risk.

As expected, the PSS of the injured group was significantly lower than the healthy controls. During performance of the movement screening tool, only 10% of athletes with a shoulder injury reported pain during any upper extremity test. No athlete with a shoulder injury reported an increase in symptoms that would have resulted in testing cessation. Collectively, this suggests that shoulder pain was not likely a major factor accounting for the lower MSST composite score in this group. It is possible that athletes with a current injury may have developed compensatory movement strategies following onset of injury in order to avoid pain, and as such, the screen may have captured these strategies.

Inter-rater reliability of the qualitatively scored tests in the MSST was investigated in this study as the scoring criteria used differs from previously published work. Inter-rater reliability on qualitative tests indicated fair to perfect agreement for those tests included in the MSST. The active straight leg raise and rotary stability tests demonstrated 100% agreement between raters and therefore did not have data variance necessary for a kappa computation. The excellent composite score reliability suggests the overall scoring mechanism is reliable between trained raters. Although it is possible that this may be inflated due to simultaneous performance assessment of the raters, these tests have published and/or established inter-rater reliability that would support the data presented here.^{27,46-49}

This study is not without limitations. Typically, EFA requires large sample sizes; however, the minimum sample size to variable ratio depends on the study design.⁵⁷ Lawley and Maxwell suggest 51 more cases than the number of variables is acceptable for EFA, in which case our data meet the criteria.⁵⁸ It is acknowledged that the observed measurement properties cannot be extrapolated to other musculoskeletal injuries and future work is needed as this is a limitation of the current study. The current screen does not include an open chain upper extremity strength assessment, which may be important for upper extremity dominant sports that require open chain tasks. Additionally, exclusion of the CKQUEST, for protection of athletes with a current shoulder injury, may have introduced a bias to our findings that was not controlled for in the two approaches to the known group analysis. Finally, we acknowledge that the current structure of the MSST is not yet field expedient, thereby posing a challenge for large scale movement screening efforts. Test-retest reliability, and minimal detectable change of the MSST is also currently unknown and should be addressed in future work.

Conclusions

The MSST is a comprehensive movement system screen that preliminarily demonstrates acceptable content and construct validity, as well as inter-rater reliability when experienced raters administer the screening tool. Comprehensive screening tools with construct validity and known measurement error are needed to identify athletes at risk of injury and provide an instrument that can be used in studies that seek to validate training approaches proposed to change movement impairment and injury risk in athletes.

Conflicts of interest

The authors declare no conflicts of interest.

Acknowledgement

Funded in part by the Legacy Fund of the American Academy of Sports Physical Therapy.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:<https://doi.org/10.1016/j.bjpt.2019.10.002>.

References

- Hewett TE, Myer GD, Ford KR. Reducing knee and anterior cruciate ligament injuries among female athletes: A systematic review of neuromuscular training interventions. *J Knee Surg.* 2005;18(1):82–88.
- 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.
- Laudner K, Sipes R. The incidence of shoulder injury among collegiate overhead athletes. *J Intercollegiate Sport.* 2009;2:260–268.
- Chorba RS, Chorba DJ, Bouillon LE, Overmyer CA, Landis JA. Use of a functional movement screening tool to determine injury risk in female collegiate athletes. *N Am J Sports Phys Ther.* 2010;5(2):47–54.
- 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.
- Shanley E, Rauh MJ, Michener LA, Ellenbecker TS, Garrison JC, Thigpen CA. Shoulder range of motion measures as risk factors for shoulder and elbow injuries in high school softball and baseball players. *Am J Sports Med.* 2011;39(9):1997–2006.
- Mandelbaum BR, Silvers HJ, Watanabe DS, et al. Effectiveness of a neuromuscular and proprioceptive training program in preventing anterior cruciate ligament injuries in female athletes: 2-year follow-up. *Am J Sports Med.* 2005;33(7):1003–1010.
- Hewett TE, Torg JS, Boden BP. Video analysis of trunk and knee motion during non-contact anterior cruciate ligament injury in female athletes: Lateral trunk and knee abduction motion are combined components of the injury mechanism. *Br J Sports Med.* 2009;43(6):417–422.
- Zazulak BT, Hewett TE, Reeves NP, Goldberg B, Cholewicki J. Deficits in neuromuscular control of the trunk predict knee injury risk: A prospective biomechanical-epidemiologic study. *Am J Sports Med.* 2007;35(7):1123–1130.
- Zazulak BT, Hewett TE, Reeves NP, Goldberg B, Cholewicki J. The effects of core proprioception on knee injury: A prospective biomechanical-epidemiological study. *Am J Sports Med.* 2007;35(3):368–373.
- Hewett TE, Myer GD, Ford KR, et al. Biomechanical measures of neuromuscular control and valgus loading of the knee predict anterior cruciate ligament injury risk in female athletes: A prospective study. *Am J Sports Med.* 2005;33(4):492–501.
- Hagglund M, Walden M, Ekstrand J. Previous injury as a risk factor for injury in elite football: A prospective study over two consecutive seasons. *Br J Sports Med.* 2006;40(9):767–772.
- Messina DF, Farney WC, DeLee JC. The incidence of injury in Texas high school basketball. A prospective study among male and female athletes. *Am J Sports Med.* 1999;27(3):294–299.
- Boden BP, Griffin LY, Garrett WE Jr. Etiology and prevention of noncontact ACL injury. *Phys Sportsmed.* 2000;28(4):53–60.
- Griffin LY, Agel J, Albohm MJ, et al. Noncontact anterior cruciate ligament injuries: Risk factors and prevention strategies. *J Am Acad Orthop Surg.* 2000;8(3):141–150.
- Croisier JL, Ganteaume S, Binet J, Genty M, Ferret JM. Strength imbalances and prevention of hamstring injury in professional soccer players: A prospective study. *Am J Sports Med.* 2008;36(8):1469–1475.
- Cook G, Burton L, Hoogenboom B. Pre-participation screening: The use of fundamental movements as an assessment of function — part 1. *North Am J Sports Phys Ther.* 2006;1(2):62–72.
- Cholewicki J, Silfies SP, Shah RA, et al. Delayed trunk muscle reflex responses increase the risk of low back injuries. *Spine.* 2005;30(23):2614–2620.
- Garrison JC, Arnold A, Macko MJ, Conway JE. Baseball players diagnosed with ulnar collateral ligament tears demonstrate decreased balance compared to healthy controls. *J Orthop Sports Phys Ther.* 2013;43(10):752–758.
- Ireland ML, Willson JD, Ballantyne BT, Davis IM. Hip strength in females with and without patellofemoral pain. *J Orthop Sports Phys Ther.* 2003;33(11):671–676.
- Silfies SP, Ebaugh D, Pontillo M, Butowicz CM. Critical review of the impact of core stability on upper extremity athletic injury and performance. *Braz J Phys Ther.* 2015;19(5):360–368.
- Tate A, Turner GN, Knab SE, Jorgensen C, Strittmatter A, Michener LA. Risk factors associated with shoulder pain and disability across the lifespan of competitive swimmers. *J Athl Train.* 2012;47(2):149–158.
- 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.
- Garrison M, Westrick R, Johnson MR, Benenson J. Association between the functional movement screen and injury development in college athletes. *Int J Sports Phys Ther.* 2015;10(1):21–28.
- Onate JA, Dewey T, Kollock RO, et al. Real-time intersession and interrater reliability of the functional movement screen. *J Strength Cond Res/Natl Strength Cond Assoc.* 2012;26(2):408–415.
- 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.
- Tarara DT, Hegedus EJ, Taylor JB. Real-time test-retest and interrater reliability of select physical performance measures in physically active college-aged students. *Int J Sports Phys Ther.* 2014;9(7):874–887.
- Hegedus EJ, McDonough S, Bleakley C, et al. Physical performance tests predict injury in National Collegiate Athletic Association athletes: A three-season prospective cohort study. *Br J Sports Med.* 2016, bjsports-2015-094885.
- Cook G, Burton L, Hoogenboom BJ, Voight M. Functional movement screening: The use of fundamental movements as an assessment of function-part 2. *Int J Sports Phys Ther.* 2014;9(4):549–563.
- Bennett H, Davison K, Arnold J, Slattery F, Martin M, Norton K. Multi-component musculoskeletal movement assessment tools: A systematic review and critical appraisal of their development and applicability to professional practice. *J Strength Cond Res.* 2017;31(10):2903–2919.
- Burton SL. *Performance and Injury Predictability During Firefighter Candidate Training (PhD Dissertation)*. Blacksburg, VA: Virginia Tech; 2006.
- Butler RJ, Contreras M, Burton LC, Plisky PJ, Goode A, Kiesel K. Modifiable risk factors predict injuries in firefighters during training academies. *Work.* 2013;46(1):11–17.

33. Kiesel K, Plisky PJ, Voight ML. Can serious injury in professional football be predicted by a preseason functional movement screen? *N Am J Sports Phys Ther.* 2007;2(3):147–158.
34. Knapik JJ, Cosio-Lima LM, Reynolds KL, Shumway RS. Efficacy of functional movement screening for predicting injuries in coast guard cadets. *J Strength Cond Res/Natl Strength Cond Assoc.* 2015;29(5):1157–1162.
35. Warren M, Smith CA, Chimera NJ. Association of the functional movement screen with injuries in division I athletes. *J Sport Rehabil.* 2015;24(2):163–170.
36. Li Y, Wang X, Chen X, Dai B. Exploratory factor analysis of the functional movement screen in elite athletes. *J Sports Sci.* 2015;33(11):1166–1172.
37. Kazman JB, Galecki JM, Lisman P, Deuster PA, O'Connor FG. Factor structure of the functional movement screen in marine officer candidates. *J Strength Cond Res/Natl Strength Cond Assoc.* 2014;28(3):672–678.
38. Butowicz CM, Ebaugh DD, Noehren B, Silfies SP. Validation of two clinical measures of core stability. *Int J Sports Phys Ther.* 2016;11(1):15–23.
39. Cholewicki J, Polzhofer GK, Radebold A. Postural control of trunk during unstable sitting. *J Biomech.* 2000;33(12):1733–1737.
40. Eubank BH, Mohtadi NG, Lafave MR, et al. Using the modified Delphi method to establish clinical consensus for the diagnosis and treatment of patients with rotator cuff pathology. *BMC Med Res Methodol.* 2016;16(1):56.
41. Sim J, Arnell P. Measurement validity in physical therapy research. *Phys Ther.* 1993;73(2):102–110.
42. Portney LG, Watkins MP. *Foundations of Clinical Research: Applications to Practice.* 3rd ed. Saddle River, New Jersey, USA: Pearson Prentice Hall; 2009.
43. Leggin BG, Michener LA, Shaffer MA, Brenneman SK, Iannotti JP, Williams GR Jr. The Penn shoulder score: reliability and validity. *J Orthop Sports Phys Ther.* 2006;36(3):138–151.
44. Cook G, Burton L, Hoogenboom BJ, Voight M. Functional movement screening: The use of fundamental movements as an assessment of function — part 1. *Int J Sports Phys Ther.* 2014;9(3):396–409.
45. Shanley E, Thigpen CA, Clark JC, et al. Changes in passive range of motion and development of glenohumeral internal rotation deficit (GIRD) in the professional pitching shoulder between spring training in two consecutive years. *J Shoulder Elbow Surg/Am Shoulder Elbow Surg.* 2012;21(11):1605–1612.
46. Johnsen MB, Eitzen I, Moksnes H, Risberg MA. Inter- and intrarater reliability of four single-legged hop tests and isokinetic muscle torque measurements in children. *Knee Surg Sports Traumatol Arthrosc: Off J ESSKA.* 2015;23(7):1907–1916.
47. Christiansen DH, Moller AD, Vestergaard JM, Mose S, Maribo T. The scapular dyskinesis test: Reliability, agreement, and predictive value in patients with subacromial impingement syndrome. *J Hand Ther.* 2017;30(2):208–213.
48. Kevern MA, Beecher M, Rao S. Reliability of measurement of glenohumeral internal rotation, external rotation, and total arc of motion in 3 test positions. *J Athl Train.* 2014;49(5):640–646.
49. 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. *North Am J Sports Phys Ther.* 2009;4(2):92–99.
50. Field A. *Discovering Statistics Using SPSS.* third edition ed. London: Sage Publications, Ltd; 2009.
51. Leech NL, Barrett KC, Morgan GA. *SPSS for Intermediate Statistics.* New York: Lawrence Erlbaum Associates; 2008.
52. Mandrekar JN. Receiver operating characteristic curve in diagnostic test assessment. *J Thorac Oncol.* 2010;5(9):1315–1316.
53. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33(1):159–174.
54. Devan MR, Pescatello LS, Faghri P, Anderson J. A prospective study of overuse knee injuries among female athletes with muscle imbalances and structural abnormalities. *J Athl Train.* 2004;39(3):263–267.
55. Smith CA, Chimera NJ, Warren M. Association of y balance test reach asymmetry and injury in division I athletes. *Med Sci Sports Exerc.* 2015;47(1):136–141.
56. Wright AA, Stern B, Hegedus EJ, Tarara DT, Taylor JB, Dischiavi SL. *Potential Limitations of the Functional Movement Screen: A Clinical Commentary.* BMJ Publishing Group Ltd and British Association of Sport and Exercise Medicine; 2016.
57. MacCallum R, Widaman K, Zhang S, Hong S. Sample size in factor analysis. *Psychol Methods.* 1999;4(1):84–89.
58. Lawley DN, Maxwell AE. *Factor Analysis as a Statistical Method.* American Elsevier Pub. Co.; 1971.
59. Goldbeck T, Davies G. Test-retest reliability of the closed kinetic chain upper extremity stability test: a clinical field test. *J Sport Rehabil.* 2000;9:35–45.
60. Crossley KM, Zhang WJ, Schache AG, et al. Performance on the single-leg squat task indicates hip abductor muscle function. *Am J Sports Med.* 2011;39(4):866–873.
61. Lanning CL, Uhl TL, Ingram CL, et al. Baseline values of trunk endurance and hip strength in collegiate athletes. *J Athl Train.* 2006;41(4):427–434.
62. McClure P, Tate AR, Kareha S. A clinical method for identifying scapular dyskinesis, part 1: reliability. *J Athl Train.* 2009;44(2):160–164.