9-1-2012

Endurance Tests Are the Most Reliable Core Stability Related Measurements

Andy Waldhelm
Louisiana State University

Li Li
Georgia Southern University, lili@georgiasouthern.edu

Available online at www.sciencedirect.com
SciVerse ScienceDirect
Journal of Sport and Health Science 1 (2012) 121–128
www.jshs.org.cn

Original article

Endurance tests are the most reliable core stability related measurements

Andy Waldhelm, Li Li*

Department of Kinesiology, Louisiana State University, 112 Long Fieldhouse, Baton Rouge, LA 70803, USA

Received 23 March 2012; revised 25 June 2012; accepted 13 July 2012

Abstract

Purpose: To determine the intra-tester reliability of clinical measurements that assess five components related to core stability: strength, endurance, flexibility, motor control, and function.
Methods: Participants were 15 college-aged males who had not suffered any orthopedic injury in the past year. Core strength measurements included eight isometric tests and a sit-up test. The four core endurance tests were the trunk flexor test, trunk extensor test, and bilateral side bridge tests. Flexibility tests included the sit-and-reach test and active range of the trunk and hip joint motions. Proprioception via passive reposition tests of the hips and a single limb balance test on an unsteady platform were used to evaluate core motor control. Functional measurements consisted of a squat test and a single leg hop test for time and distance. Measurements were performed during two data collection sessions with a week's rest between the sessions. Intra-class correlation coefficients were calculated to establish reliability.
Results: The overall intra-rater reliability for all core stability related measurements ranged from low (ICC = 0.35, left hip reposition) to very high (ICC = 0.98, sit-and-reach). As a group, the core endurance tests were observed to be the most reliable.
Conclusion: There are highly reliable tests in each of the five groups. Overall, core endurance tests are the most reliable measurements, followed by the flexibility, strength, neuromuscular control, and functional tests, respectively.

Keywords: Core stability; Endurance; Reliability; Strength

1. Introduction

Over the past decade, core stability has become a common concept in the field of sports medicine. The practice of measuring core stability has been used to identify athletes who may be at risk for injuries, to assess rehabilitation outcomes of an injured athlete, and in sports performance enhancement programs. Historically, the term "core stability" did not become popular until the 21st century, with the idea developing from the study of spinal stability by individuals, such as Manorah Panjabi.[1] Panjabi[1] was the first to introduce the three physiological subsystems responsible for stabilization: passive, active, and neural control. Although lack of core stability has been associated with low back pain[2] and athletic injuries,[3] defining and measuring core stability remains difficult.

Hodges[4] was believed to be the first to propose a thorough definition of core stability, when he presented a composite model of lumbopelvic stability. Hodges[4] defined lumbopelvic stability as the "dynamic process of controlling static position in the functional context, but allowing the trunk to move with control in other situations". Similarly, Bliss and Teeple[5] defined the dynamic stability of the spine as the ability to use muscular strength and endurance to control the spine beyond the neutral zone when performing functional and athletic activities.

Willson et al.[6] defined core stability as the ability of the lumbopelvic-hip complex to return to equilibrium following a perturbation without buckling of the vertebral column. Later

* Corresponding author.
E-mail address: lli3@lsu.edu (L. Li)
Peer review under responsibility of *Shanghai University of Sport*

Production and hosting by Elsevier

2095-2546
http://dx.doi.org/10.1016/j.jshs.2012.07.007

Kibler et al.[7] described core stability as being able to control the position and motion of the trunk over the pelvis and leg. This allows the core to produce, transfer, and control force and motion to the terminal segment during kinetic chain activities. Despite the lack of a universal definition, core stability remains a hot topic in the field of sports medicine. Google search of "Core stability" on March 21, 2012 yield more than 7 million results in 0.3 s.

Impairments in core stability have been associated with low back pain and lower extremity injuries in athletes. Nadler et al.[8] reported that female athletes who suffered from low back pain or sustained a lower extremity injury demonstrated a significant disparity in side-to-side maximum hip extension strength. Similarly, over an athletic season, Leetun et al.[3] observed individuals, among intercollegiate basketball and track athletes, with hip abduction and external rotation weakness were more likely to sustain a lower extremity injury. Although athletic injuries have been associated with impairments in core stability, assessing core stability remains difficult.

Although there is no consensus on the definition and measurement of core stability, several tests and measurements are available that claim to measure and assess components of core stability. Suggested core stability components include strength, endurance, flexibility, motor control, and function. Leetun et al.[3] assessed the core strength and endurance of 140 collegiate basketball and track athletes with the objective of identifying individuals at risk for injuries. They recorded maximum isometric hip abductor and external rotation strength and the muscular endurance capabilities of the anterior, posterior, and lateral trunk muscles. They observed that individuals with stronger core musculature were less likely to sustain a lower extremity injury. Gabbe et al.[9] measured the range of motion of the trunk and hip joints. Parkhurst and Burnett[10] assessed motor control of the core when they attempted to identify the relationship between lower back proprioception and injury. Along with two other tests, they used a trunk reposition test to measure low back proprioception.

Assuming core stability contributes to different functions and activities, another option in assessing core stability indirectly is to observe an individual performing a relevant functional movement or activity. Kibler et al.[7] suggested evaluating the performance of a one leg squat or single leg balance activity for deviations. Deviations or difficulty performing the activity suggests possible core stability impairment.

We might be able to define, and/or understand, the concept of core stability if we have better understanding of the parameters that contribute to core stability, or related to core stability indirectly.

Despite the number of available core stability related measurements, the reliability of these tests can vary. Bohannon[11] observed very high intra-rater reliability for isometric trunk strength during a single session reliability study. Unlike Bohannon,[11] Moreland et al.[12] found very low inter-rater reliability when measuring trunk isometric forces. Testing core muscular endurance of athletes, Evans et al.[13] observed high to very high intra-tester reliability. Similarly, Gabbe et al.[9] found high to very high test-retest reliability of four parameters related to core flexibility measurements. Using a single limb dynamic balance assessment to evaluate core motor control, Cachupe et al.[14] reported very high reliability during a single day testing session. Loudon et al.[15] reported moderate to very high intra-rater reliability when performing five functional tests on individuals with knee pain.

Following a thorough review of the literature, 35 different tests that may relate to core stability were identified and classified in five different groups. All of these parameters could potentially help us understand core stability if we know they can be measured reliably. The objective of our study was to introduce, measure, and compare the reliability of these 35 tests, all which can be performed in a clinical setting. Most of these measures are used in clinics by the same clinician to evaluate training effects, rehabilitation progress, or other concerns over a period of time. We will evaluate the reliability of one rater over time as our first attempt. We hypothesized parameters in each of the five groups: strength, endurance, flexibility, motor control, and function, would be equally reliable.

2. Materials and methods

2.1. Participants and rater

Fifteen active, right lower extremity dominant, college-age males (age: 21.2 $\pm$ 1.3 year, weight: 74.1 $\pm$ 13.4 kg, height: 1.6 $\pm$ 0.1 m) recruited from a local university volunteered for the study. Lower extremity dominance was determined by asking the participant "if you were to kick a soccer ball as hard as you could, which leg would you use?" The leg chosen was classified as the dominant leg. All participants reported the absence of any orthopedic injury to their trunk and extremities within the past year. The participants provided informed consent, as approved by the local Institutional Review Board, prior to data collection. A physical therapist with 7 years of clinical experience, with an assistant, performed the tests.

2.2. Procedures

A test-retest design was used to assess the intra-rater reliability for all 35 core stability related measurements, with the examiner blinded from the results between sessions. All participants were required to attend two testing sessions separated by 7 days. For both sessions, all tests were performed in random order between and within the testing categories, except for the endurance tests. The endurance tests were performed in a within category random order last due to the fatiguing nature of the tests. Each participant's age, weight, and height were recorded prior to session one. A 5-min warm-up was performed by walking on a treadmill with self-selected speed before each testing session.

2.3. Strength tests

The strength tests were eight isometric tests and an isoinertial test. The isometric tests were performed on a Biodex System 3 Pro (Biodex Medical Systems, Inc., Shirley, NY,

USA). Isometric strength measurements followed modified protocols described by Essendrop et al.[16] and Nadler et al.[8] Maximal isometric strength for trunk flexion and extension, bilateral hip extension, abduction, and external rotation was recorded. A practice trial was performed for each test to allow the participant to become familiar with the protocol. The average of three maximum force measurements was recorded. The participants held each contraction for 5 s.

Trunk flexion and extension were performed while standing, with the pelvis stabilized, and without upper extremity support. The attachment was placed two inches below the participant's sternal notch for trunk flexion (Fig. 1) and between the scapulae for trunk extension (Fig. 2). Bilateral hip extension and abduction force was collected while standing, with the hips in neutral position, and without upper extremity support. The attachment was placed two inches above the posterior joint line of the knee for hip extension and two inches above the lateral joint line of the knee for abduction. The bilateral hip external rotation force was measured in sitting. The participant's hips and knees were flexed at 90° without upper extremity support. The attachment was placed two inches above the ankle joint.

The isoinertial strength test was a timed sit-up test. The protocol for the sit-up test was developed by the American Alliance of Health, Physical Education, Recreation, and Dance (AAHPERD[17]). The objective of the test was to perform as many full sit-ups as possible within 1 min. The sit-up test was initiated in the hook-lying position, with arms held across chest, knees flexed at 90°, and feet secured. To complete a full sit-up, the participant's scapulae touched the mat in the lying position and the elbows made contact with the knees in sitting.

2.4. Endurance tests

Following protocols established by McGill et al.,[18] four core endurance tests were performed. The objective of the endurance tests was to hold a static position for as long as possible. The endurance tests were the trunk flexor test, trunk extensor test, and bilateral side bridge tests.

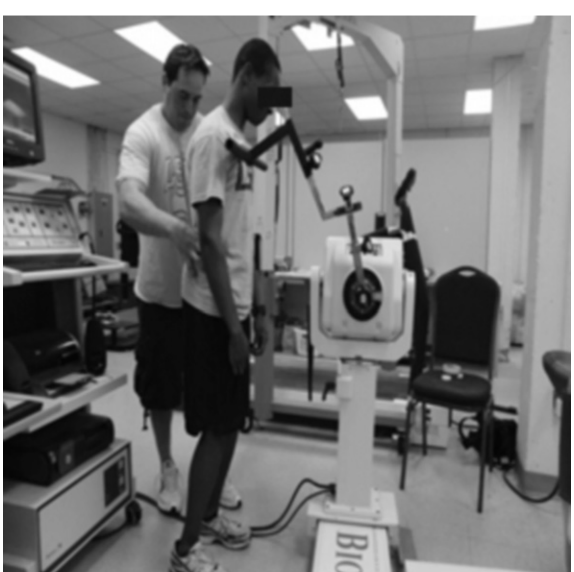

Fig. 1. Trunk flexion strength test.

Fig. 2. Trunk extension strength test.

The trunk flexor test began with the participant in the sit-up position with their trunk supported at 60° of trunk flexion. Knees and hips were flexed at 90°, arms crossed over chest, and feet secured. The support of the trunk was then removed, and the participant held the position for as long as possible. The test was terminated when the participant was no longer able to hold the position.

The trunk extensor test was performed with the participant lying prone on a treatment table. Their pelvis, hips, and knees were secured to the treatment table, while a chair at the same height as the surface of the table supported the trunk and upper extremities. The chair was removed, and the individual held a horizontal body position for as long as possible with arms crossed over chest. The test was discontinued when the participant fell below the horizontal position.

The side bridge tests were performed in the side-lying position on a treatment table. The participant's knees were extended with the top foot placed in front of the lower foot. The participant supported their weight only on their lower elbow and feet while lifting their hips off the mat. The test was stopped when the side-lying position was lost or when the hips returned to the mat.

2.5. Flexibility tests

Flexibility tests included active range of motion (ROM) measurements for the trunk and hip joints, as well as a sit-and-reach test. ROM measurements for trunk flexion, extension, rotation, and hip extension were based on Norkin and White.[19] Trunk flexion and extension ROM were assessed by measuring

the distance between C7 and S1 while standing in neutral position (neutral length). To locate C7 and S1, the examiner palpated the vertebrae and marked them with a pen. The participant forward flexed as far as possible with the pelvis stabilized. The length between C7 and S1 was remeasured, and the difference in lengths (neutral and flexed) was recorded as the trunk flexion ROM.

Similarly, for trunk extension, the participant extended as far as possible with the pelvis stabilized. The distance between C7 and S1 was remeasured and the length difference from neutral position was the trunk extension ROM. To measure trunk rotation, the participant sat on a chair with their feet on the floor and their trunk and head in neutral position. The participant rotated their trunk and head as far as possible in both directions. A 30-cm plastic goniometer was positioned so the fulcrum was above the center of the participant's head. The stationary arm was parallel to the imaginary line between the iliac crests, and the movement arm aligned with an imaginary line between the acromial processes of the shoulders.

Active hip extension was measured with the participant in the prone position, knees extended, and pelvis stabilized. The fulcrum of the 30-cm plastic goniometer was positioned over the greater trochanter, while the stabilizing arm was aligned with the lateral midline of the pelvis. Following maximal active hip extension, the movement arm was aligned with the lateral midline of the femur.

Active hip internal and external rotation ROMs were measured using the method described by Ellison et al.[20] The participant was positioned in the prone position with their hip positioned in neutral position and knee flexed at 90°. The non-testing leg was placed at 30° of hip abduction with the knee extended and pelvis stabilized. The 30-cm plastic goniometer was positioned with the stabilizing arm aligned vertically, while the movement arm was aligned along the shaft of the tibia.

The sit-and-reach test was performed using the protocol listed in the American College of Sports Medicine (ACSM) Guidelines.[21] The participants sat with their shoes on and the feet resting against a sit-and-reach box. The examiner extended and stabilized their knees. They positioned one hand on top of the other with palms down. They were requested to lean as far as possible along the measurement scale without flexing their knees. The furthest distance reached along the scale was recorded to the nearest 0.5 cm. The average of two trials was documented.

2.6. Motor control tests

The group of motor control tests included a passive reposition test for each hip and a single limb balance assessment with and without a blindfolded. The passive hip reposition tests were performed using methods modified from those described by Zazulak et al.[22] Our tests differed from Zazulak et al.[22] who performed repositioning tests of the lumbar spine rather than the hips. The objective of the reposition tests was for a participant to stop their passively moving leg at a target degree of hip ROM. The hip repositioning tests were performed on the Biodex System 3 Pro using the Passive Mode. The lower extremity was moved between 10° of hip flexion and extension at a rate of 2°/s. The participant was positioned in standing, with a blindfold on, where they were allowed to use their upper extremities for support. The hip attachment was positioned two inches above the knee to allow the testing limb to be off the ground. The participant's thigh was first passively moved from neutral (starting) position to a randomized target position and held for 5 s. The thigh was then returned to the neutral position. The participant's thigh was again passively moved, and the participant manually stopped his limb at the perceived target position using the emergency stop button. The degrees away from the target position were recorded, and the average of two trials was documented.

The single limb athletic test (Fig. 3) performed on the Biodex Balance System SD (Biodex Medical Systems, Inc.,) was used to assess single limb stability. The single limb athletic test is a dynamic stability test performed on an unstable platform without upper extremity support. Levels of difficulty range from 1 (hardest) to 12 (easiest), and level 10 was used in our assessment. Level 10 was used after a pilot study revealed it was a safe level to perform when the participant was blindfolded and the participants were required to use the hip strategy to maintain balance. Four different conditions were performed: right (dominate) limb with eyes open; left (non-dominate) limb with eyes open; right limb blindfolded; and left limb blindfolded. Each test was performed for three 10-s trials.

Fig. 3. Single limb balance test.

2.7. Functional tests

The last group of measurements had three functional tests: squat test, timed single leg hop test, and the single leg hop test for distance. The protocol for the bilateral squat test was performed using the protocol described by Loudon et al.[15] The goal of the test was to perform the maximum number of squats during the 30-s test. The participant started from a sitting position with their hips and knees flexed at 90° in a chair without armrests. To perform one repetition, the participant rose to full knee extension and returned to the chair. They kept their arms crossed over their chest during the test, and the number of repetitions performed was recorded.

The timed and distance single limb hop tests were performed according to the methods outlined by Reid et al.[23] The goal of timed hop test was to hop on one leg as quickly as possible over a distance of 30 feet (9.14 m). The participant performed one trial of the timed hop test on each limb. The single leg hop for distance test was performed by hopping and landing on the same leg. The distance hopped was measured from toe to toe, and the participants were required to hold their landing for at least 2 s for a successful trial. Three hops from each leg were performed, and the longest hop was recorded.

2.8. Statistical analyses

At the completion of testing, all results were analyzed using SPSS for Windows (version 17.0; SPSS Inc., Chicago, IL, USA). Descriptive statistics (mean ± SD) were used to report the daily testing results. The range of testing results was first evaluated using coefficient of variance (CV) and differences between the two days tests. Further, intra-class correlation coefficient (ICC (2,1)) was used to estimate reliability, and 95% confidence intervals (CI) were provided. The ICC (2,1) was performed using the following equation[24]:

$$\mathrm{ICC}(2,1) = (\mathrm{BMS} - \mathrm{EMS})/[(\mathrm{BMS} + (k-1)\mathrm{EMS} + k(\mathrm{JMS} - \mathrm{EMS}))/n]$$

where BMS is between mean square; EMS is residual mean square; JMS is between judges mean square; k is number of scores; and n is numbers of persons observed. ICC (2,1) was used since it includes the variability of measurements for any session on any participant.[24] Munro and Page's[25] ICC classification system was used for determining acceptable reliability. This system classified ICC values as little, if any (0.00−0.25), low (0.26−0.49), moderate (0.50−0.69), high (0.70−0.89), and very high (0.90−1.00). CI with $\alpha = 0.05$ was developed using the following equation[26]:

$$\mathrm{CI} = (F_\mathrm{L} - 1)/(F_\mathrm{L} + (k-1))$$

where $F_\mathrm{L} = F_\mathrm{obs}/F_\mathrm{tabled}$ for the lower limit; $F_\mathrm{L} = F_\mathrm{obs} \times F_\mathrm{tabled}$ for the upper limit; F_obs is row effects (session); F_tabled = the $(1-0.5\alpha) \times 100$th percentile of the distribution with $n-1$ representing the numerator degrees of freedom; and $(n-1)(k-1)$ representing the denominator degrees of freedom.

3. Results

Table 1 presents descriptive results of mean ± SD, CV, and relative difference of all dependent variables between the two testing sessions. The overall CV ranged from 6% to 87% in session one and 5% to 80% in session two. The CV for the strength tests ranged from 16% to 42% in session one and 14% to 46% in session two. The CV for the endurance tests in session one ranged from 35% to 52%, while they ranged from 29% to 46% in session two. The CV ranged from 8% to 66% in session one and 7% to 62% in session two for the flexibility tests. The CV for the motor control tests ranged from 24% to 87% in session one and 28% to 80% in session two. For the functional tests, the CV in session one ranged from 6% to 15% and from 5% to 11% in session two.

The relative difference between sessions for all core stability related measurements ranged from 0 to 41.4%. The lowest relative difference for the strength tests was observed for left hip external rotation (0.4%), while the highest was trunk extension (19.4%). For the endurance tests, the left-side bridge had the smallest relative difference (0.3%), while right-side bridge had the largest difference (8.9%). The relative differences for the flexibility tests ranged from left hip internal rotation (1.4%) to trunk extension (11.0%). For the motor control tests, no relative difference was witnessed for the left hip reposition test between sessions. The highest relative difference of the group was for the right hip reposition test (41.4%). The functional tests had the lowest range of relative differences of the five groups. They ranged from the squat test (0.4%) to the left hop for distance test (4.3%).

The overall intra-rater reliability for all core stability related measurements ranged from low (−0.35) to very high (0.98). Nineteen (54%) of the 35 measurements were considered to have high (0.70−0.89) or very high (0.90−1.00) reliability, 12 (34%) of the tests were considered to have moderate (0.50−0.69) reliability, while four (11%) of the tests were considered to have low (0.26−0.49) reliability.

Table 2 presents the intra-rater reliability of the individual parameters. All strength tests, except the right hip abduction test (0.45), had moderate to very high reliability, with the sit-up test having the highest (0.92). The endurance tests obtained moderate to very high reliability (0.66−0.96), with the left-side bridge test having the highest (0.96). The flexibility tests were observed to have moderate to very high reliability (0.62−0.98), with the traditional sit-and-reach test having the highest reliability (0.98). The motor control measurements were identified to have moderate to high reliability (0.52−0.90), with the exception of the left hip reposition test, which was not reliable (−0.35). The functional tests had the greatest amount of discrepancy (0.42−0.92) among the five groups. Within the group, right (0.45) and left (0.42) hop tests for time had low reliability, the squat test had moderate reliability (0.55), with the right (0.91) and left (0.92) hop tests for distance having very high reliability.

Table 1
Descriptive statistics for core stability related measurements ($n = 15$).

	Session one		Session two		
	Mean ± SD	CV (%)	Mean ± SD	CV (%)	Relative difference (%)
Strength tests					
Trunk flexion (N)	57.75 ± 19.50	34	63.49 ± 29.10	46	9.9
Trunk extension (N)	72.31 ± 30.60	42	86.31 ± 39.60	46	19.4
Right hip extension (N)	59.42 ± 18.00	30	69.19 ± 27.10	39	16.4
Left hip extension (N)	68.29 ± 17.10	25	69.29 ± 19.80	29	1.5
Right abduction (N)	69.89 ± 20.50	29	73.62 ± 15.50	21	5.3
Left hip abduction (N)	63.33 ± 18.00	28	72.71 ± 17.00	23	14.8
Right hip ER (N)	56.89 ± 11.00	19	55.00 ± 12.10	22	3.3
Left hip ER (N)	54.07 ± 16.10	30	54.31 ± 17.20	32	0.4
Sit-up test	45.60 ± 7.33	16	49.13 ± 7.01	14	7.7
Endurance tests					
Trunk flexion (s)	57.87 ± 29.10	50	58.60 ± 17.00	29	1.3
Trunk extension (s)	83.27 ± 29.40	35	81.27 ± 25.00	31	2.4
Right side bridge (s)	82.20 ± 32.20	39	74.87 ± 25.20	34	8.9
Left side bridge (s)	77.20 ± 39.90	52	76.93 ± 35.70	46	0.3
Flexibility tests					
Sit-and-reach (cm)	4.20 ± 2.77	66	4.62 ± 2.87	62	10.0
Trunk flexion (cm)	12.42 ± 1.51	12	13.39 ± 1.58	12	7.8
Trunk extension (cm)	5.98 ± 1.03	17	6.64 ± 0.99	15	11.0
Right trunk rotation (°)	91.93 ± 8.90	10	89.33 ± 6.15	7	2.8
Left trunk rotation (°)	89.20 ± 7.26	8	88.07 ± 5.35	6	1.3
Right hip extension (°)	27.53 ± 7.73	28	29.67 ± 6.11	21	7.8
Left hip extension (°)	27.40 ± 7.37	27	28.73 ± 8.51	30	4.9
Right hip IR (°)	45.47 ± 10.00	22	46.80 ± 8.61	18	2.9
Left hip IR (°)	47.60 ± 10.10	21	48.26 ± 6.86	14	1.4
Right hip ER (°)	47.80 ± 6.28	13	50.00 ± 9.09	18	4.6
Left hip ER (°)	48.33 ± 6.57	14	52.00 ± 9.96	19	7.6
Motor control tests					
Eyes open right	1.63 ± 0.77	47	1.29 ± 0.52	40	20.9
Eyes open left	1.833 ± 0.77	42	1.55 ± 0.58	37	15.4
Blindfolded right	5.20 ± 1.27	24	4.37 ± 1.57	36	16.0
Blindfolded left	4.94 ± 1.57	32	4.77 ± 1.35	28	3.4
Right hip reposition (°)	2.27 ± 1.97	87	1.33 ± 1.06	80	41.4
Left hip reposition (°)	2.26 ± 1.16	51	2.26 ± 1.66	73	0
Functional tests					
Squat test	30.07 ± 4.65	15	30.20 ± 4.44	15	0.4
Right hop distance (cm)	148.93 ± 10.10	7	151.60 ± 7.45	5	1.8
Left hop distance (cm)	145.62 ± 8.88	6	151.91 ± 8.76	6	4.3
Right hop timed (s)	3.08 ± 0.43	14	2.98 ± 0.26	9	3.2
Left hop timed (s)	3.14 ± 0.42	13	3.02 ± 0.33	11	3.8

Abbreviations: CV = coefficient of variance (SD/mean); ER = external rotation; IR = internal rotation.
Note: Relative difference is calculated as $|\text{mean } 2 - \text{mean } 1|/\text{mean } 1 \times 100\%$.

4. Discussion

The purpose of our study was to introduce, measure, and compare the reliability of 35 different tests identified as being related to core stability. These tests examined five different components that contribute to core stability. Contrary to our hypothesis, core endurance tests were the most reliable measurements among the five groups, with flexibility tests the second most reliable, followed by strength, motor control, and functional assessments, respectively.

Some descriptive results observed in this study compared well with previous parameters reported in the literature, but others did not. Comparing to Moreland et al.,[12] our observations of trunk strength and endurance were similar with theirs. Among the variables that are different, differences could stem from the differences of testing population, methods and equipment. Some of the differences can be explained by other research. For example, females have been observed to have longer trunk extension endurance times compared with men.[18]

Three other possible core stability related measurements resulted in different outcomes from previous studies: hip internal and external active ROM and the squat test. Testing active hip internal and external ROM as part of a lower extremity screen, Gabbe et al.[9] recorded smaller degrees of flexibility compared with our study (internal rotation 27°/46°, external rotation 22°/78°). The most noticeable difference between the studies was the participant's testing position. Gabbe et al.[9] performed their ROM tests in the sitting position, while we tested in prone. The sitting position requires the participant to move against gravity, while in the prone

Table 2
Intra-rater reliability for core stability related measurements.

	ICC (2,1)	95% CI
Strength tests		
Trunk flexion[c]	0.62	−0.13−0.87
Trunk extension[b]	0.81	0.43−0.94
Right hip extension[b]	0.73	0.19−0.91
Left hip extension[c]	0.68	0.05−0.89
Right hip abduction[d]	0.45	−0.64−0.82
Left hip abduction[c]	0.61	−0.16−0.95
Right hip ER[b]	0.71	0.15−0.90
Left hip ER[b]	0.85	0.55−0.95
Sit-up test[a]	0.92	0.77−0.97
Endurance tests		
Trunk flexion[c]	0.66	0.01−0.89
Trunk extension[b]	0.79	0.38−0.93
Right side bridge[b]	0.74	0.30−0.92
Left side bridge[a]	0.96	0.87−0.99
Flexibility tests		
Sit-and-reach[a]	0.98	0.95−0.99
Trunk flexion[b]	0.71	0.13−0.90
Trunk extension[b]	0.79	0.37−0.93
Right trunk rotation[c]	0.67	0.01−0.89
Left trunk rotation[c]	0.69	0.07−0.90
Right hip extension[c]	0.64	−0.07−0.88
Left hip extension[b]	0.84	0.52−0.95
Right hip IR[b]	0.74	0.22−0.91
Left hip IR[c]	0.65	−0.05−0.88
Right hip ER[c]	0.62	−0.12−0.87
Left hip ER[c]	0.68	0.03−0.89
Motor control tests		
Eyes open right[b]	0.87	0.60−0.96
Eyes open left[b]	0.76	0.27−0.92
Blindfolded right[a]	0.90	0.72−0.97
Blindfolded left[b]	0.80	0.41−0.93
Right hip reposition[c]	0.52	−0.42−0.84
Left hip reposition[d]	−0.35	−3.02−0.55
Functional tests		
Squat test[c]	0.55	−0.35−0.85
Right hop distance[a]	0.91	0.74−0.97
Left hop distance[a]	0.92	0.76−0.97
Right hop timed[d]	0.45	−0.65−0.81
Left hop timed[d]	0.42	−0.72−0.81

Abbreviations: CI = confidence interval; ER = external rotation; ICC = intra-class correlation coefficient; IR = internal rotation.
[a] Very high reliability (0.90−1.00).
[b] High reliability (0.70−0.89).
[c] Moderate reliability (0.50−0.69).
[d] Low reliability (0.26−0.49).

position−gravity−assists the movement. Furthermore, in the sitting position, a mechanical block of the joint could limit the hip flexibility.

Loudon et al.[15] performed the squat test on 11 healthy adults as part of a functional performance assessment. Using the same protocol, the participants in their study performed fewer squats (20/30). This could be attributed to different populations tested in the studies. They used volunteers who were mostly female with a mean age of 30 years, while our participants were male with an average age of 21 who could have been in better physical condition. Different testing protocols and testing populations could explain the differences between our observations and the literature.

Despite the differences in the testing scores, many of the core stability related measurements used in our study had similar reliability compared with earlier studies. Two of the tests included the sit-and-reach test and the single leg stance. Gabbe et al.[9] found corresponding sit-and-reach intra-rater reliability, ICC 0.97−0.98, when compared with our results. This can be contributed to the simplicity of the testing equipment and protocol. Cachupe et al.[14] also recorded similar reliability for the single leg balance test: ICC 0.81 compared with an ICC that ranged from 0.76 to 0.90 for the four tests we performed. Both of the tests used comparable protocols and participants.

While some of the core stability related measurements had a similar reliability, other tests were observed to have lower reliability when compared with earlier reports. Compared with our observations, Essendrop et al.[16] found higher intra-rater reliability for trunk flexion strength: ICC 0.62−0.97, and trunk extension strength: ICC 0.81−0.93. Differences in reliability could be attributed to the testing position. Both studies tested in the standing position with the pelvis stabilized, but Essendrop and associates[16] also stabilized the shoulders of their participants. Although this position could isolate the trunk muscles, it limits the need for muscle coordination, which is essential in functional and athletic activities.

Measuring core endurance, Evans et al.[13] observed a more reliable trunk flexion test: ICC 0.66−0.95, compared with our study. This could be explained by the 2 weeks between testing session in their study compared with the 1 week in ours. With the 2 weeks between sessions, the learning effect could be decreased. Loudon et al.[15] had a different reliability outcome when compared with our study. The reliability of the squat test they performed was greater: ICC 0.55−0.79, compared with the results we observed. Having only 2−3 days between sessions and the testing order not changing could contribute to the higher reliability. One of the factors could explain the differences in intra-rater reliability used to describe the differences in the descriptive statistics (i.e., testing protocol).

There were differences observed between the relative differences and the ICC of several measurements. For example, the squat test had a small relative difference, 0.4%, but only moderate reliability: ICC 0.55. The opposite was observed for trunk extension strength, where a high relative difference was recorded (19.4%), but the measurement had high reliability, an ICC 0.81. Disparity in the range of the scores may contribute to the inconsistencies between the relative difference and the ICC. With a small range, the relative difference may also be small, but the tests may not be reliable and vice versa.

Our observations provided valuable information on the reliability of several core stability related measurements. Please note the confidence interval of the ICC estimation. For a parameter with ICC 0.85, it still can have a wide 95% CI from 0.55 to 0.95. Please keep this in mind when interpreting these results. Caution also must be taken when attempting to generalize the results beyond the population of healthy, college-aged males without recent orthopedic injury. Although inter-rater reliability was not performed, we were able to

identify four tests that had poor reliability. In the future, we can then eliminate these measures when we analyze inter-rater reliability. Furthermore, many of the measurements used in our study could be performed using a different protocol or instrumentation.

One thing puzzles us is the results of left and right hip repositioning tests. The result of the left hip was moderately reliable (0.52) but that of the right hip was not reliable at all (−0.35). One possible explanation is leg dominant since all of our participants were right limb dominant. Dominant limb could be stronger and associated with more acute proprioceptive sensibility.

Overall, the results in this study are beneficial to the practice of assessing core stability. Core stability is a complicated concept that relates to different components, including strength, endurance, flexibility, motor control, and function. Therefore, partial evaluation will result in an incomplete assessment of core stability. Our results showed the reliability of core stability related measurements could vary. It is especially true when a thorough evaluation of core stability is performed. We have identified the intra-rater reliability of 35 core stability related measures. Our results were slightly lower, but we selected testing positions that required the participant to be in a functional posture. For comparing observations from different research or clinic, inter-rater reliability should be assessed based on our results. Future studies will also explore how the reliable core stability related measures correlate with athletic performance or injury.

5. Conclusion

The objective of our study was to introduce and evaluate the reliability of 35 core stability related measurements, which examined five different components of core stability. There were highly reliable tests in each of the five groups. Overall, core endurance tests were the most reliable measurements, followed by the flexibility, strength, motor control, and functional tests, respectively. Therefore, when assessing core stability, it is critical to understand that the reliability of the related measurements may vary.

References

1. Panjabi MM. The stabilizing system of the spine. Part 1. Function, dysfunction, adaptation, and enhancement. *J Spinal Disord* 1992;**5**:383−9.
2. McGill SM. Low back pain: from formal description to issues for performance and rehabilitation. *Exerc Sport Sci Rev* 2001;**29**:26−31.
3. Leetun DT, Ireland ML, Willson JD, Ballantyne BT, Davis IM. Core stability measures as risk factors for lower extremity injury in athletes. *Med Sci Sports Exerc* 2004;**36**:926−34.
4. Hodges PW. Lumbopelvic stability: a functional model of biomechanics and motor control. In: Richardson HPC, editor. *Therapeutic exercise for lumbopelvic stabilization: a motor control approach for the treatment and prevention of low back pain*. 2nd ed. Edinburgh, UK: Churchill Livingstone; 2004. p. 13−28.
5. Bliss LS, Teeple P. Core stability: the centerpiece of any training program. *Curr Sports Med Rep* 2005;**4**:179−83.
6. Willson JD, Dougherty CP, Ireland ML, Davis IM. Core stability and its relationship to lower extremity function and injury. *J Am Acad Orthop Surg* 2005;**13**:316−25.
7. Kibler WB, Press J, Sciascia A. The role of core stability in athletic function. *Sports Med* 2006;**36**:189−98.
8. Nadler SF, Malanga GA, DePrince M, Stitik TP, Feinberg JH. The relationship between lower extremity injury, low back pain, and hip muscle strength in male and female follegiate athletes. *Clin J Sport Med* 2000;**10**:89−97.
9. Gabbe BJ, Bennell KL, Wajswelner H, Finch CF. Reliability of common lower extremity musculoskeletal screening tests. *Phys Ther Sport* 2004;**5**:90−7.
10. Parkhurst TM, Burnett CN. Injury and proprioception in the lower back. *J Orthop Sports Phys Ther* 1994;**19**:282−95.
11. Bohannon RW. Test-retest reliability of hand-held dynamometry during a single session strength assessment. *Phys Ther* 1986;**66**:206−9.
12. Moreland J, Finch E, Stratford P, Balsor B, Gill C. Interrater reliability of six tests of trunk muscle function and endurance. *J Orthop Sports Phys Ther* 1997;**26**:200−8.
13. Evans K, Refshauge KM, Adams R. Trunk muscle endurance tests: reliabiltiy and gender differences in athletes. *J Sci Med Sport* 2007;**10**:447−55.
14. Cachupe WJ, Shifflett B, Kahanov L, Wughalter EH. Reliability of biodex balance system measure. *Meas Phys Educ Exerc Sci* 2001;**5**:97−108.
15. Loudon JK, Wiesner D, Goist-Foley HL, Asjes C, Loudon KL. Intrarater reliability of functional performance tests for subjects with patellofemoral pain syndrome. *J Athl Train* 2002;**37**:256−61.
16. Essendrop M, Schibye B, Hansen K. Reliability of isometric muscle strength tests for the trunk, hands, and shoulders. *Int J Ind Ergon* 2001;**28**:379−87.
17. American Alliance of Health, Physical Education, Recreation, and Dance. *Health-related physical fitness test manual*. Reston, VA: American Alliance for Health, Physical Education, Recreation and Dance; 1980.
18. McGill SM, Childs A, Liebenson C. Endurance times for low back stabilization exercises: clinical targets for testing and training from a normal datatbase. *Arch Phys Med Rehabil* 1999;**80**:941−4.
19. Norkin CC, White DJ, editors. *Measurement of joint motion: a guide to goniometry*. 2nd ed. Philadelphia, PA: F.A. Davis Company; 1995.
20. Ellison JB, Rose SJ, Sahrmann SA. Patterns of hip rotation range of motion: a comparison between healthy subjects and patients with low back pain. *Phys Ther* 1990;**70**:537−41.
21. American College of Sports Medicine. *Guidelines for exercise testing and prescription*. 6th ed. Baltimore, MD: Lippincott, Williams, & Walkins; 2000.
22. 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**:368−73.
23. Reid A, Birmingham TB, Stratford PW, Alcock GK, Giffin JR. Hop testing provides a reliable and valid outcome measure during rehabilitation after anterior cruciate ligament reconstruction. *Phys Ther* 2007;**87**:337−49.
24. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979;**86**:420−8.
25. Munro BH, Page EB. *Statistical methods for health care research*. 2nd ed. Philadelphia, PA: J.B. Lippincott Co.; 1993.
26. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychol Methods* 1996;**1**:30−46.