Journal of Strength and Conditioning Research, 2005, 19(3), 583–586

MEASURING CORE STABILITY

WENDELL P. LIEMOHN,[1] TED A. BAUMGARTNER,[2] AND LAURA H. GAGNON[1]

[1]*Department of Exercise, Sport, and Leisure Studies, University of Tennessee, Knoxville, Tennessee 37996;* [2]*Department of Exercise Science, University of Georgia, Athens, Georgia 30602.*

ABSTRACT. Liemohn, W.P., T.A. Baumgartner, and L.H. Gagnon. Measuring core stability. *J. Strength Cond. Res. 19(3):583–586. 2005.*—In this study, a 4-item battery of core stability (CS) tests modeled on core stabilization activities used in training and rehabilitation research was developed, and a measurement schedule was established to maximize internal consistency and stability reliabilities. Specifically, we found that 4 test administrations on each of 4 days produced intraclass correlation coefficients that in most instances exceeded 0.90 and stability reliability coefficients on the third and fourth days of testing that exceeded 0.90 for 2 of the tests and 0.80 for the other 2. Thus, it is recommended that in future research, examiners administer the battery for at least 3 days and consider the data collected on day 3 as the best estimate of participant CS.

KEY WORDS. low back pain, neural spine, axial skeleton control, quantifying trunk coordination, maintaining spine balance

INTRODUCTION

The term core stability (CS) has attained a high degree of prominence in the past few years; quite possibly, it may have emanated from the exercises popularized by the San Francisco Spine Institute (SFSI) when the concept of the neutral spine was stressed in their 1989 manual titled *Dynamic Lumbar Stabilization Program* (32). During this era, stabilization training was used with both athletic and nonathletic populations (27, 29–31). Core stability remains a key component in (a) clinical rehabilitation (12, 23, 28), (b) the training of competitive athletes (15, 16, 21), and (c) the training programs of individuals who are endeavoring to improve their health and physical fitness (1, 5, 19).

Panjabi (26) presented a conceptualization of CS (he called it spinal stability) that is based on 3 subsystems: the (a) passive spinal column, (b) active spinal muscles, and (c) neural control unit. Drawing from Panjabi, we define CS as the functional integration of the passive spinal column, active spinal muscles, and the neural control unit in a manner that allows the individual to maintain the intervertebral neutral zones within physiologic limits while performing activities of daily living.

Following Panjabbi's (26) previously mentioned conceptualization scheme, our discussion will emphasize the active core muscles and the neural control unit, for the passive spinal column is the least amenable to training. Although the terms core stability and core strength are sometimes used interchangeably, we have chosen to subsume core strength within CS. Core stability requires coordination in addition to core strength and endurance.

In our discussion of core muscles, we will follow Bergmark's (4) classification scheme that groups core muscles into either the Global Stabilization System (GSS) or the Local Stabilization System (LSS). The larger and smaller muscles of the trunk are the chief contributors to the GSS and the LSS, respectively. The role of the LSS is related more to the coordination and control of motion segments than to the more forceful movements provided by the muscles of the GSS that have larger masses and longer moment arms of force (2). The LSS muscles also are closer to the spinal column and thus can provide varying degrees of segmental control. For example, the intertransversarii mediales, interspinales, and rotatores are extremely close to the center of rotation of the spinal segments. Their very small physiological cross-sectional area and their high density of muscle spindles (4.5–7.3 times richer than the number in the multifidus [24]) suggests that they may act primarily as position transducers of the spinal column (6, 9, 22). This would suggest that these LSS muscles would appear to be particularly important to the coordination required in CS.

The major purpose of this research was to develop a measurement schedule that would enable us to quantify CS and maximize internal consistency reliability and stability reliability. In our prior research, although internal consistency reliability was satisfactory, stability reliability was not (17, 18); this was attributed to the fact that our CS tests require balance and coordination (that can improve with repeated testing) in addition to strength.

METHODS

Experimental Approach to the Problem

When the SFSI's CS training activities (32) are analyzed, coordination and balance appear to be key elements, for even in the more difficult tasks, the thoracic extensors are used at but a fraction of their maximum voluntary contraction (7, 20). Although there are numerous ways to measure strength of the core musculature (10, 11, 14, 16), these tests emphasize strength and/or endurance, whereas performance on the SFSI's stability exercises also requires balance and coordination. Cosio-Lima et al. (8) used a standing balance test as an indicator of CS; we chose to try to replicate actual CS training postures in our balance tests.

Because the surface area of our force platform is not large enough to accommodate postures such as the quadruped used in the SFSI's stabilization exercises, we chose to use the Stability Platform (Lafayette Instrument Co., Lafayette, IN). The Stability Platform is a very sensitive instrument that was designed for the measurement of standing balance with the feet typically placed parallel to the tilt axis (25, 33); however, the large size of its platform (~66 × 106 cm) accommodates postures used in the SFSI's CS training program, including Bridging 1.12 and Quadruped Arm Raises 2.9 (32). We used 2 versions of the latter test (Figure 1) in our battery, 1 with the body parallel and 1 with the body perpendicular to the tilt axis. We also added a test in which from a kneeling posture on

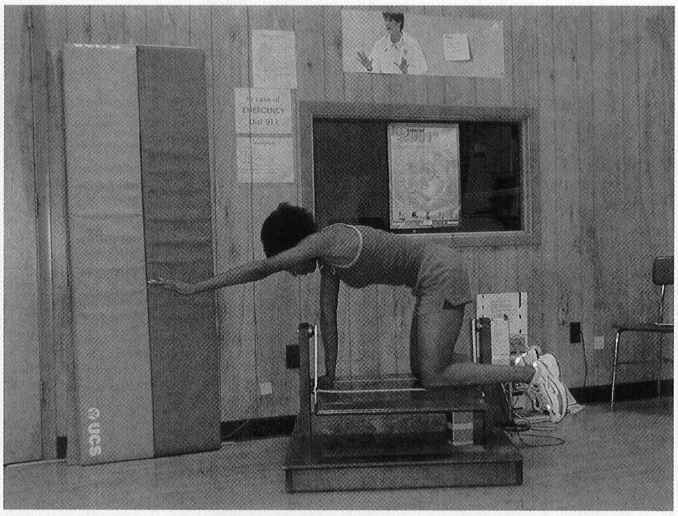

FIGURE 1. Quadruped arm raise (body parallel to tilt axis). In this test, the subject alternately raises each arm in concert with the metronome (i.e., each arm is raised 20 times to shoulder level in each 30-second test).

the Stability Platform, subjects alternately raised their arms in time with a metronome; Hodges and Richardson (13) found that there was delay in activation of the transversus abdominis as back patients performed this activity. For the latter test, the arm raising was done with the metronome set at 60 b·min^{-1}; for the 2 quadruped arm raise tests, the metronome was set at 40 b·min^{-1}. All our balance tasks were of 30-second duration, and the tilt limits of the balance board were set at 5° to either side; with this arrangement, the clock counter counts the number of seconds within the 30-second test that the subject did not maintain balance within the 10° arc. Additional data collected but not reported here included (a) the number of times that the subject was outside the 10° arc in each 30-second trial and (b) ratings of perceived exertion for each test.

Subjects

Sixteen university students (9 men, 7 women) free of any orthopedic disability that would have precluded their participation volunteered to participate. Men and women were combined because our objective was to determine the reliability of our test protocol, and our prior research (17, 18) did not suggest that gender was an important issue. The procedures were reviewed and approved by our university's Institutional Review Board; each subject signed a written informed consent before participating in this study.

Procedures

Based on our prior investigations (17, 18), we knew that collecting multiple trial scores on each day of multiple days was necessary to identify a measurement schedule that would yield reliable test scores because subjects tended to improve performance with practice. For this investigation, we selected 4 tests from our 6-balance-test battery and administered each of these tests on 4 days with 5 trials administered on each day. The 4 days of test administration were typically spread out over an 8–12-day period. Each day a test was administered, the test was explained, and a 20-second trial was given, and then the participants were administered each test once; trials 2–5 were similarly administered without the 20-second practice period. The score of a participant was the number of seconds out of balance, so the smaller a score, the better the score. We hypothesized that the scores from the first day of testing would be much worse than the scores from subsequent days of testing, so the first-day trials were considered practice trials for the participants to learn to perform the test. Also, we hypothesized that the first trial for days 2–4 of testing should also be considered a practice trial.

Statistical Analyses

Intraclass reliability coefficients for the trial scores of a participant using a 1-way analysis of variance design (3) were calculated for each testing day of each test. Stability reliability for the score of a participant on a single day using the data from days 2 and 3, days 3 and 4, and days 2–4 was estimated using an intraclass reliability coefficient formula from Baumgartner et al. (3).

RESULTS

Presented in Table 1 are the trial means for each test and each day of testing. The trial means for the first day of a

TABLE 1. Trial means for each day of testing for each test.

		Mean				
Test	Day	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5
Kneeling arm raise	1	15.4	13.5	10.5	9.4	8.7
	2	10.2	6.6	6.2	6.2	5.6
	3	5.1	3.0	2.9	2.5	2.6
	4	1.6	2.1	2.1	1.8	2.1
Quad arm raise (parallel)	1	13.2	9.7	8.5	7.5	6.9
	2	6.2	4.8	4.7	4.9	4.1
	3	4.4	3.3	3.8	3.1	2.9
	4	4.0	2.8	2.6	2.5	3.1
Quad arm raise (perpendicular)	1	10.4	7.1	5.8	4.0	4.0
	2	4.4	3.3	2.7	2.1	3.1
	3	2.9	2.8	1.6	2.6	1.6
	4	2.1	2.0	1.8	2.3	2.1
Bridging	1	10.4	8.0	5.2	6.9	5.0
	2	6.3	6.6	4.3	4.2	5.6
	3	3.5	4.4	3.5	3.0	3.0
	4	4.3	4.1	3.7	3.0	4.0

TABLE 2. Reliability coeffcients and means for each day of testing for each test.

Test	Day	Reliability	Mean
Kneeling arm raise	2	0.95	6.1
	3	0.94	2.8
	4	0.95	2.0
Quad arm raise (parallel)	2	0.91	4.6
	3	0.91	3.3
	4	0.89	2.7
Quad arm raise (perpendicular)	2	0.71	2.8
	3	0.78	2.1
	4	0.94	2.0
Bridging	2	0.87	5.2
	3	0.90	3.5
	4	0.91	3.7

test are markedly higher (worse) than the trial means for the second day of a test as hypothesized. Usually the mean for the first trial of a day for any of the 4 tests was the highest mean for the day. Based on the means reported in Table 1, the decision was made to consider day 1 of testing as a practice day and the first trial of a test for days 2–4 as a practice trial. The score of a participant for a day would then be the mean or sum of the trial 2–5 scores.

Intraclass reliability coefficients for the sum or mean of the trial scores of a participant using a 1-way analysis of variance design (3) were calculated for each testing day of each test. Intraclass reliability coefficients for the sum or mean of the trial 2–5 scores within a day and the mean score for a day are reported in Table 2. The reliability coefficients are high, usually at least 0.90. The 2 reliability coefficients in the 0.70s for test 3 are low, but reliability of the test scores for test 3 increased daily. Means for testing days decreased considerably from day 2 to day 3 but decreased less from day 3 to day 4.

Another decision in establishing a measurement schedule is the number of days to administer the test in order to obtain a reliable score. Reliability of the scores within a day (internal consistency reliability) must be high in order to obtain high reliability between days (stability reliability). Stability reliability for the score of a participant on a single day using the data from days 2 and 3, days 3 and 4, and days 2–4 was estimated using an intraclass reliability coefficient formula from Baumgartner et al. (3). Reliability coefficients for the score of a participant on a single day using the scores from days 2–4 were low, ranging from 0.56 to 0.76. Means for days and stability reliability coefficients using data from 2 days are reported in Table 3. Reliability coefficients for the score of a participant on a single day are low using the scores from testing days 2 and 3 but are higher using the scores from testing days 3 and 4. Reliability coefficients in the 0.80s are considered good and in the 0.90s are considered high.

DISCUSSION

For each test, 5 trials of the test were administered on each of 4 days. Participants received an explanation and demonstration of a test before being tested each day. We hypothesized that the test scores from the first day of testing would have to be considered practice to learn to perform the tests. This hypothesis was supported by the data since for each test the mean trial scores were markedly higher (worse) on the first day than on other days of testing. Also, we hypothesized that the test scores from the first trial of a day would have to be considered practice/warm-up to perform the test. This hypothesis was supported by the data since typically the mean for trial 1 was the highest (worst) mean. In fact, for the first 3 days of testing, the mean for trial 1 was always the highest mean. Thus, internal consistency reliability was calculated for days 2–4 of testing using the scores for trials 2–5. Internal consistency reliability was acceptable on testing days 2–4 for all 4 tests. Internal consistency reliability tended to be best on the fourth day of testing. Trial means for a day were fairly consistent for trials 2–5.

TABLE 3. Stability reliability of a day sum and means of day sums.

Test	Day	Mean	Reliability: Days 2 and 3	Reliability: Days 3 and 4
Kneeling arm raise	2	24.4	0.52	
	3	11.2		0.94
	4	8.0		
Quad arm raise (parallel)	2	18.4	0.68	
	3	13.2		0.93
	4	10.8		
Quad arm raise (perpendicular)	2	11.2	0.51	
	3	8.4		0.84
	4	8.0		
Bridging	2	20.8	0.68	
	3	14.0		0.83
	4	14.8		

Stability reliability coefficients for the score of a participant collected on a single day were calculated using the scores for testing days 2 and 3 and for testing days 3 and 4. We had hoped to find sufficiently high stability reliability coefficients to be able to suggest that in the future only testing on days 1 and 2 would be necessary. However, stability reliability was low using the test scores for days 2 and 3 and sufficiently high using the test scores for days 3 and 4. Means for the days decreased more from day 2 to day 3 than from day 3 to day 4. Thus, test scores are fairly stable from day 3 to day 4 of testing. The intraclass correlation coefficient used to estimate stability reliability is affected by changes in the mean score from day to day. The high stability reliability coefficients obtained suggest that the score and rank of a participant in a group is fairly stable from day 3 of testing to day 4 of testing.

Based on the findings in this study, administering 5 trials on each of 3 days of the Stability Platform tests used in this study is sufficient to obtain a test score with good internal and stability reliability. The test score of a participant would be the sum or mean of trials 2–5 on day 3 of testing.

PRACTICAL APPLICATIONS

Now that we are comfortable with both the internal and the stability reliability of our measurement schedule for our 4 CS tests on the Stability Platform, we are testing subjects on the previously mentioned tests as well as on their performance on a series of core strength activities that require varying degrees of coordination (e.g., floor vs.

Swiss-ball strength training activities/tests). We also plan to use one of our CS tests as the primary dependent variable when we examine the effectiveness of a Pilates training program against a more conventional one in the rehabilitation of patients with low back pain.

REFERENCES

1. ALLEN, S., G.A. DUDLEY, M. IOSIA, D. STANFORTH, AND B. STEUERWALD. Core strength training. *Gatoraide Sports Science Institute,* 2002. 13(1): 14.
2. AROKOSKI, J.P., M. KAAUKAANPAA, T. VALTA, I. JUVONEN, J. PARTANEN, AND S. TAIMELA. Back and hip extensor muscle function during therapeutic exercise. *Arch. Phys. Med. Rehabil.* 80:842–850. 1999.
3. BAUMGARTNER, T.A., A.S. JACKSON, M.T. MAHAR, AND D.A. ROWE. *Measurement for Evaluation in Physical Education and Exercise Science* (7th ed.). Madison, WI: McGraw-Hill, 2003.
4. BERGMARK, A. Stability of the lumbar spine: A study in mechanical engineering. *Acta Orthop. Scand.* 230(Suppl.):20–24. 1989.
5. BLACK, S.A. Training the core. *Fitness Manage.* (October):60–63. 2001.
6. BOGDUK, N. *Clinical Anatomy of the Lumbar Spine and Sacrum* (3rd ed.). London: Churchill Livingstone, 1998.
7. CHOLEWICKI, J., M. PANJABI, AND A. KHACHATRYAN. Stabilizing function of trunk flexor-extensor muscles around a neutral spine posture. *Spine* 22:2207–2212. 1997.
8. COSIO-LIMA, L.M., K.L. REYNOLDS, C. WINTER, V. PAOLONE, AND M.T. JONES. Effects of physioball and conventional floor exercises on early phase adaptations in back and abdominal core stability and balance in women. *J. Strength Cond. Res.* 17: 721–725. 2003.
9. CRISCO, J.J., AND M.M. PANJABI. The intersegmental and multisegmental muscles of the spine: A biomechanical model comparing lateral stabilising potential. *Spine* 16:793–799. 1991.
10. FRITZ, J.M., AND G.E. HICKS. Exercise protocols for low back pain. In: *Exercise Prescription and the Back.* W. Liemohn, ed. New York: McGraw-Hill Medical Publishing Division, 2001. pp. 167–181.
11. GRAVES, J.E., M.L. POLLOCK, AND C.X. BRYANT. Assessment of muscular strength and endurance. In: *ACSM's Resource Manual for Guidelines for Exercise Testing and Prescription.* J.L. Roitman, ed. Baltimore: Lippincott Williams & Wilkins, 2001. pp. 448–455.
12. HIDES, J.A., G.A. JULL, AND C.A. RICHARDSON. Long-term effects of specific stabilizing exercises for first-episode low back pain. *Spine* 26:E243–E248. 2001.
13. HODGES, P.W., AND C.A. RICHARDSON Inefficient muscular stabilisation of the lumbar spine associated with low back pain: A motor control evaluation of transversus abdominis. *Spine* 21: 2640–2650. 1996.
14. ITO, T., O. SHIRADO, H. SUZUKI, M. TAKAHASHI, K. KANEDA, AND T.E. STRAX. Lumbar trunk muscle endurance testing: An inexpensive alternative to a machine for evaluation. *Arch. Phys. Med. Rehabil.* 77:75–79. 1996.
15. JEFFREYS, I. Developing a progressive core stability program. *Strength Cond. J.* 24(5):65–66. 2002.
16. LEETUN, D.T., M.L. IRELAND, J.D. WILLSON, B.T. BALLANTYNE, AND I.M. DAVIS. Core stability measures as risk factors for lower extremity injury in athletes. *Med. Sci. Sports Exerc.* 36:926–934. 2004.
17. LIEMOHN, W., T.A. BAUMGARTNER, AND K.A. KELLY. Core stability: The genesis of a test battery. *Med. Sci. Sports Exerc.* 35(Suppl. 5):S275.– 2003.
18. LIEMOHN, W., R.C. JOHNSON, J.C. SANDERS, AND J. WOODS. Measuring cores strength and core stability: A pilot study. *Med. Sci. Sports Exerc.* 34(Suppl. 5):S153. 2002.
19. LIEMOHN, W., AND G. PARISER. Core strength: Implications for fitness and low back pain. *ACSM's Health & Fitness J.* 6(5):10–16. 2002.
20. MCGILL, S.M. *Low Back Disorders: Evidence-Based Prevention and Rehabilitation.* Champaign, IL: Human Kinetics, 2002.
21. MCGILL, S.M. Low back stability: From formal description to issues for performance and rehabilitation. *Exerc. Sport Sci. Rev.* 29:26–31. 2001.
22. MCGILL, S.M. Mechanics and pathomechanics of muscles acting on the lumbar spine. In: *Kinesiology: The Mechanics and Pathomechanics of Human Movement.* C.A. Oatis, ed. Philadelphia: Lippincott Williams & Wilkins, 2004. pp. 563–575.
23. MCGILL, S.M., AND J. CHOLEWICKI. Biomechanical basis for stability: An explanation to enhance clinical utility. *J. Orthop. Sports Phys. Ther.* 31:96–100. 2001.
24. NITZ, A.J., AND D. PECK. Comparison of muscle spindle concentrations in large and small human epaxial muscles acting in parallel combinations. *Am. Surg.* 52:273–277. 1986.
25. NORDAHL, S.H., T. AASEN, B.M. DYRKORN, S. EIDSVIK, AND O. MOLVAER. Static stabilometry and repeated testing in a normal population. *Aviat. Space Environ. Med.* 71:889–893. 2000.
26. PANJABI, M.M. The stabilizing system of the spine, part I. Function, dysfunction, adaptation, and enhancement. *J. Spinal Disord.* 5:383–389. 1992.
27. SAAL, J.A. Dynamic muscular stabilization in the nonoperative treatment of lumbar pain syndromes. *Orthop. Rev.* 19:691–700. 1990.
28. SAAL, J.A. Natural history and nonoperative treatment of lumbar disc herniation. *Spine* 21(Suppl. 24):S2–S9. 1996.
29. SAAL, J.A. Rehabilitation of football players with lumbar spine injury (part 1 of 2). *Physician Sports Med.* 16(9):61–68. 1988.
30. SAAL, J.A. Rehabilitation of football players with lumbar spine injury (part 2 of 2). *Physician Sports Med.* 16(10):117–125. 1988.
31. SAAL, J.A., AND J.S. SAAL. Nonoperative treatment of herniated lumbar intervertebral disc with radiculopathy—An outcome study. *Spine* 14:431–437. 1989.
32. SAN FRANCISCO SPINE INSTITUTE. *Dynamic Lumbar Stabilization Program.* San Francisco: San Francisco Spine Institute, 1989.
33. SHEA, C.H., AND G. WULF. Enhancing motor learning through external-focus instructions and feedback. *Hum. Mov. Sci.* 18: 553–571. 1999.

Address correspondence to Dr. Wendell Liemohn, wliemohn@utk.edu.