Measuring Core Stability

Wendell P. Liemohn,1 Ted A. Baumgartner,2 and Laura H. Gagnon1

1Department of Exercise, Sport, and Leisure Studies, University of Tennessee, Knoxville, Tennessee 37996; 2Department 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 non-athletic 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 Panjabi'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 X 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
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>
<thead>
<tr>
<th>Test</th>
<th>Day</th>
<th>Trial 1</th>
<th>Trial 2</th>
<th>Trial 3</th>
<th>Trial 4</th>
<th>Trial 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kneeling arm raise</td>
<td>1</td>
<td>15.4</td>
<td>13.5</td>
<td>10.5</td>
<td>9.4</td>
<td>8.7</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>10.2</td>
<td>6.6</td>
<td>6.2</td>
<td>6.2</td>
<td>5.6</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>5.1</td>
<td>3.0</td>
<td>2.9</td>
<td>2.5</td>
<td>2.6</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1.6</td>
<td>2.1</td>
<td>2.1</td>
<td>1.8</td>
<td>2.1</td>
</tr>
<tr>
<td>Quad arm raise (parallel)</td>
<td>1</td>
<td>13.2</td>
<td>9.7</td>
<td>8.5</td>
<td>7.5</td>
<td>6.9</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>6.2</td>
<td>4.8</td>
<td>4.7</td>
<td>4.9</td>
<td>4.1</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>4.4</td>
<td>3.3</td>
<td>3.8</td>
<td>3.1</td>
<td>2.9</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>4.0</td>
<td>2.8</td>
<td>2.6</td>
<td>2.5</td>
<td>3.1</td>
</tr>
<tr>
<td>Quad arm raise (perpendicular)</td>
<td>1</td>
<td>10.4</td>
<td>7.1</td>
<td>5.8</td>
<td>4.0</td>
<td>4.0</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>4.4</td>
<td>3.3</td>
<td>2.7</td>
<td>2.1</td>
<td>3.1</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>2.8</td>
<td>2.8</td>
<td>1.6</td>
<td>2.6</td>
<td>1.6</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>2.1</td>
<td>2.0</td>
<td>1.8</td>
<td>2.3</td>
<td>2.1</td>
</tr>
<tr>
<td>Bridging</td>
<td>1</td>
<td>10.4</td>
<td>8.0</td>
<td>5.2</td>
<td>6.9</td>
<td>5.0</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>6.3</td>
<td>6.6</td>
<td>4.3</td>
<td>4.2</td>
<td>5.6</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>3.5</td>
<td>4.4</td>
<td>3.5</td>
<td>3.0</td>
<td>3.0</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>4.3</td>
<td>4.1</td>
<td>3.7</td>
<td>3.0</td>
<td>4.0</td>
</tr>
</tbody>
</table>
Kneeling arm raise
testing for each test.
Bridging
Quad arm raise (parallel)
Quad arm raise (perpendicular)

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) 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.

**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 Stabibility 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


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