The Reliability and Validity of a Chair Sit-and-Reach Test as a Measure of Hamstring Flexibility in Older Adults

C. Jessie Jones, Roberta E. Rikli, Julie Max, and Guillermo Noffal

The purpose of this study was to examine the test-retest reliability and the criterion validity of a newly developed chair sit-and-reach (CSR) test as a measure of hamstring flexibility in older adults. CSR performance was also compared to sit-and-reach (SR) and back-saver sit-and-reach (BSR) measures of hamstring flexibility. To estimate reliability, 76 men and women (M age = 70.5 years) performed the CSR on 2 different days, 2-5 days apart. In the validity phase of the study, scores of 80 men and women (M age = 74.2 years) were obtained on three field test measures of hamstring flexibility (CSR, SR, and BSR) and on a criterion test (goniometer measurement of a passive straight-leg raise). Results indicate that the CSR has good intraclass test-retest reliability (R = .92 for men; r = .96 for women), and has a moderate-to-good relationship with the criterion measure (r = .76 for men; r = .81 for women). The criterion validity of the CSR for the male and female participants is comparable to that of the SR (r = .74 and r = .71, respectively) and BSR (r = .70 and r = .71, respectively). Results indicate that the CSR test produces reasonably accurate and stable measures of hamstring flexibility. In addition, it appears that the CSR is a safe and socially acceptable alternative to traditional floor sit-and-reach tests as a measure of hamstring flexibility in older adults.

Key words: aging, field test, assessment, mobility

Lack of hamstring flexibility has been associated with low back pain, postural deviations, gait limitations, risk of falling, and susceptibility to musculoskeletal injuries (American College of Sports Medicine, 1995; Grabiner, Koh, Lundin, & Jahnigen, 1993; Kendall, McCreary & Poyance, 1993; Liemohn, Snodgrass, & Sharpe, 1988). In older adults, tight hamstrings especially can lead to reduced stride length and walking speed, which in turn can cause problems with dynamic balance (Brown, 1993). Due to the importance of hamstring flexibility, its measurement is included in most current fitness test programs including the AAHPERD Physical Best (American Alliance for Health, Physical Education, Recreation and Dance, 1988), the Prudential FITNESSGRAM (Cooper Institute for Aerobics Research, 1994), the Y's Way to Physical Fitness (Golding, Myers, & Sinning, 1989), the AAHPERD Functional Fitness Test for Adults Over 60 (Osness et al., 1996), and the President's Challenge Fitness Test (President's Council on Physical Fitness and Sports, 1990).

The most common method of assessing hamstring flexibility in the field setting has been the floor sit-and-reach (SR) test, originally reported by Wells and Dillon (1952). Recently, a modified one-leg version of the SR, the back-saver sit-and-reach (BSR), has been recommended as an alternative to the two-leg SR (Cooper Institute for Aerobics Research, 1994). The rationale for the BSR is based on the work of Cailliet (1988) who suggested that stretching one hamstring at a time, instead of both at once, results in less stress and risk of injury for the low back and spine. Studies have shown that both the SR and BSR tests are highly reliable, with R values consistently above .90, and they have at least moderate criterion validity relative to goniometer-measured hamstring flexibility (r values range from .51 to .89; Jackson & Baker, 1986; Jackson & Langford, 1989; Patterson, Wiksten, Ray, Flanders, & Sanphy, 1996).

Although the SR and BSR are generally considered acceptable field test measures of hamstring flexibility for...
To determine the test-retest reliability of the CSR, and (2) to evaluate the validity of the CSR by comparing CSR scores to a criterion (goniometer) measure of hamstring flexibility in older adults. CSR performance also was compared to other common field test measures of hamstring flexibility—the SR and BSR.

Methods

Participants

Seventy-six older adults (34 men and 42 women, M age = 70.5 years) were solicited from a university-based exercise program to participate in the reliability phase of the study. A different group of 80 volunteers (32 men and 48 women, M age = 74.2 years) were recruited from nutrition and exercise classes at a nearby retirement community. The criteria for inclusion in the study were that the participants be over the age of 60 years, have no musculoskeletal limitations which would prohibit their performance on the tests, and agree to sign a statement of informed consent.

Procedures

Prior to all testing, participants performed an 8-min warm-up and static stretch routine emphasizing the lower body. Participants in the reliability study, conducted approximately 4 weeks prior to the validity study, performed the CSR on 2 different days, 2–5 days apart. Testing protocols during reliability testing, including technician training procedures, were the same as those described below in the validity phase of the study. So that interrater reliability would be reflected in the reliability analysis, different technicians were used to conduct the tests on Day 1 and Day 2. Day 2 technicians were not aware of the scores obtained on Day 1.

Participants in the validity study were assessed on the CSR, SR, and BSR in a counterbalanced (rotating) order determined prior to the test day and indicated on the score card. The CSR, SR, and BSR tests were conducted by a team of six graduate students and six older adult volunteer technicians, all of whom had participated in a group training session led by the study coordinators. During the training, technicians practiced on each other until they demonstrated proper procedures to the study coordinators. The goniometer assessment of hamstring flexibility, administered after completion of the other three tests, was administered by three experienced clinicians (two physical therapy aides and one athletic trainer). A pilot study, utilizing a subsample of 19 participants, indicated that the interrater reliability of the 3 clinicians was .92. The clinicians administering the goniometer tests were unaware of scores received on the CSR, SR, and BSR tests.

Only the preferred leg score (the leg yielding the best score) was used for the CSR, BSR and goniometer tests. Once the preferred leg was determined, that score was held constant throughout all three of the single-leg measures. Following a demonstration of each test, 2 practice trials and two test trials were given for each of the measures. Participants were reminded to exhale as they bent forward, avoid bouncing or rapid, forceful movement, and never stretch to the point of pain. On all sit-and-reach measures, if the knee(s) started to bend, the participants were asked to slowly sit back until the knee(s) were straight before scoring. The best of the two test trials1 (scored to the nearest 1/2 in.) was used for subsequent analysis on the SR, BSR, and CSR. The average of the two test trials (scored to the nearest degree) was used for the goniometer test. All measures were administered on the same day, and all tests were conducted with the participants’ shoes on. With older adults, the time required for removing shoes can be extensive and is often prohibitive when testing groups within the field setting.

Measures

Chair Sit-and-Reach. Following a demonstration, participants sat on a folding chair (17-in. high seat) and moved forward until they were sitting near the front edge. (The chair was placed against a wall and checked to see that it would remain stable throughout the testing). Participants were asked to extend their preferred leg in front of their hip, with the heel on the floor and foot dorsiflexed (at approximately a 90° angle), and bend the other leg so that the sole of the foot was flat on the floor about 6–12 in. to the side of the body’s midline. With the extended leg as straight as possible and hands on top of each other with palms down (tips of the middle fingers
even), participants were to "slowly bend forward at the hip joint, keeping the spine as straight as possible and the head in normal alignment with the spine (not tucked)." Participants were instructed to reach down the extended leg in an attempt to touch the toes. The participant held a brief static position (for 2 s), while the administrator recorded the "reached score" using an 18-
in. ruler positioned parallel to the lower leg (shin; see Figure 1). The middle of the toe at the end of the shoe represented a "zero" score. Reaches short of the toes were recorded as minus scores, and reaches beyond the toes were recorded as plus scores. Test-retest reliability estimates for the CSR are reported in the results section of this paper.

Sit-and-Reach. The SR test was administered using the procedures outlined in the. Osness et al. manual (1995). A yardstick was placed on the floor, with a 12-in. strip of masking tape positioned at the 20-in. mark (6 in. on each side of the yardstick). Following a demonstration, participants sat on the floor with their shoes on, legs fully extended, a yardstick between their legs ("0" mark of yardstick toward the hips), and their heels 12 in. apart. Throughout testing, the administrator checked to ensure that the heels remained at the 20-in. mark. With the extended leg as straight as possible, hands on top of each other (tips of the middle fingers even), and palms down, the participant slowly reached forward sliding the hands along the yardstick as far as possible. In this study, the 20-in. mark represented a zero score, with reaches short of the mark recorded as minus scores and reaches beyond the mark as plus scores. Scores were recorded to the nearest 1/2 inch. Previous studies indicate that reliability estimates for the SR are consistently high (.96 < R < .99; Bravo et al., 1994; Jackson & Baker, 1986; Jackson & Langford, 1989; Shaulis, Golding, & Tandy, 1994).

Back Saver Sit-and-Reach. The procedures for the BSR were similar to those described in the Prudential FITNESSGRAM (Cooper Institute for Aerobics Research, 1994), with the major exception being that the foot was not positioned against a sit-and-reach box. Instead, foot placement was similar to that used for the SR (i.e., heel positioned even with the 20-in. mark on a yardstick and 6 in. to the side). Participants were asked to extend their preferred leg only and bend the other leg so that the sole of the foot was flat on the floor, 6–12 in. to the side of the yardstick. The 20-in. mark represented a zero score, with reaches short of the mark recorded as minus scores and reaches beyond the mark as plus scores. BSR measures have been found to be highly reliable for both male and female participants (R = .99) (Patterson et al., 1996).

Goniometer-Measured Flexibility. The goniometer assessment of hamstring flexibility was administered after completion of the other three tests by experienced examiners who were unaware of the scores participants received on the earlier tests. Following procedures outlined by the American Academy of Orthopedic Surgeons (1966), a goniometer was used to measure hamstring flexibility during a passive straight-leg raise. This test was selected because of its prevalent acceptance as a criterion measure for hamstring flexibility and its high reliability (.95 < R < .99) (Jackson & Baker, 1986; Jackson & Langford, 1989; Patterson et al., 1996). As indicated earlier, the interrater reliability for the examiners in this study was .92, based on a subsample of 19 participants. The test protocol involved aligning the axis of the goniometer with the axis of the hip joint. The stationary arm was placed in line with the trunk, with the movable arm positioned in line with the femur. With the knee held straight, the participant's preferred leg was passively moved into hip flexion until tightness was felt. A technician assisted with moving the leg through flexion and keeping the participant in the correct position, while the clinician recorded the scores to the nearest degree.

Data Analysis

Test-retest reliability was estimated by calculating the intraclass coefficient (R) using one-way analysis of variance (ANOVA) procedures appropriate for a single trial (Baumgartner & Jackson, 1995). Pearson correlation analysis was used to determine the relationships between the CSR, SR, BSR, and the criterion goniometer measurement. Ninety-five percent confidence intervals were computed for all correlation coefficients using Fisher's "Z transformation" procedures (Glass & Hopkins, 1984; Morrow & Jackson, 1993).

Results

Descriptive statistics of participants in the reliability phase of the study are presented in Table 1. Test-retest
means and standard deviations, intraclass R values, and reliability confidence intervals are presented in Table 2. The high intraclass correlations (R = .92 for men and R = .96 for women), together with a nonsignificant change in scores from Day 1 testing to Day 2 (p > .05), indicate that CSR measurements are highly stable. Although data initially were recorded in, (for ease in interpreting results to older adults), scores were transformed to cm for data analysis and reporting.

Descriptive characteristics and mean flexibility scores of the validity study participants are presented in Table 3. Independent t test analyses of the data indicate that the women were more flexible than the men on all four hamstring measures (p < .0001). Also, ANOVA analysis indicates significant differences in flexibility scores produced by the CSR, SR, and BSR procedures, F(2, 158) = 98.2, p < .0001. Post hoc comparisons suggest that flexibility scores are better on the CSR than either the SR or BSR (p < .001). Scores on the SR and BSR tests are not significantly different and, in fact, are quite similar within participant groups (see Table 3 for comparison of all flexibility measures). The superior CSR scores, however, should be interpreted with the realization that the measurement protocol differs somewhat from the SR and BSR. On the CSR, scores represent the distance reached relative to the tip of the toes. On the SR and BSR, scores represent the distance reached relative to a line on the floor even with the heel(s) of the foot.

Table 4 contains the correlation r values and 95% confidence intervals indicating the relationship between the field sit-and-reach tests and the criterion measure. As indicated in the table, the correlations between the CSR and the criterion (goniometer-measured flexibility) for both male and female participants (r = .76 and .81, respectively) was comparable to and, in fact, slightly greater than the correlations of the SR (r = .74 and .71, respectively) and BSR with the criterion measures (r = .70).

Table 3. Means and standard deviations of descriptive characteristics and flexibility scores for validity study participants

<table>
<thead>
<tr>
<th></th>
<th>Men (n = 32)</th>
<th>Women (n = 48)</th>
<th>t ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>Age (years)</td>
<td>74.53</td>
<td>(6.59)</td>
<td>74.02</td>
</tr>
<tr>
<td>Height (cm)</td>
<td>175.18</td>
<td>(11.67)</td>
<td>159.38</td>
</tr>
<tr>
<td>Weight (kg)</td>
<td>79.36</td>
<td>(10.71)</td>
<td>62.07</td>
</tr>
<tr>
<td>Chair sit-and-reach (cm)</td>
<td>-9.69</td>
<td>(13.43)</td>
<td>3.18</td>
</tr>
<tr>
<td>Sit-and-reach (cm)</td>
<td>-20.84</td>
<td>(12.81)</td>
<td>-5.73</td>
</tr>
<tr>
<td>Back saver (cm)</td>
<td>-20.90</td>
<td>(14.44)</td>
<td>-5.37</td>
</tr>
<tr>
<td>Goniometer (°)</td>
<td>74.72</td>
<td>(14.24)</td>
<td>91.29</td>
</tr>
</tbody>
</table>

Note. M = mean; SD = standard deviation. Chair sit-and-reach and back-saver scores represent only the preferred leg (defined as the leg which results in the better score).

Table 4. Correlations and 95% confidence intervals of chair sit-and-reach, sit-and-reach, and back-saver sit-and-reach scores with goniometer-measured flexibility

<table>
<thead>
<tr>
<th></th>
<th>Men (n = 32)</th>
<th>Women (n = 48)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>r</td>
<td>CI</td>
</tr>
<tr>
<td>CSR</td>
<td>.76</td>
<td>(.57-.88)</td>
</tr>
<tr>
<td>SR</td>
<td>.74</td>
<td>(.54-.86)</td>
</tr>
<tr>
<td>BSR</td>
<td>.70</td>
<td>(.48-.84)</td>
</tr>
</tbody>
</table>

Note. r = correlation; CI = 95% confidence interval; CSR = chair sit-and-reach; SR = sit-and-reach; BSR = back-saver sit-and-reach.
.70 and .71, respectively). However, none of the differences between the correlation values reached statistical significance at the .05 level.

Discussion

Although the SR and the BSR are the most commonly used field measures of hamstring flexibility in current fitness test batteries, both tests have inherent limitations for older adults who may have difficulty getting down and up from the floor or have difficulty sitting on a level surface with legs extended. Therefore, a chair sit-and-reach was proposed as an alternative procedure for assessing hamstring flexibility—a procedure which would presumably enable more older adults to participate and minimize the possibility of injury during testing. The purpose of this study was to examine the relationship of the CSR to other measures of hamstring flexibility in older adults, particularly with respect to the test-retest reliability and criterion validity of the CSR.

Results indicate that the CSR test has good stability reliability for both men ($R = .92$) and women ($R = .96$) and that its criterion validity correlations ($r = .76$ for men and .81 for women) are slightly greater (although not statistically different) than for SR or BSR procedures (see Table 4). The validity coefficients found in this study for CSR, SR, and BSR ($R > .81$) are similar to those found in other studies with other age groups. Jackson and Baker (1986) and Jackson and Langford (1989) reported validity coefficients for the SR test ranging from .64 to .88 in studies involving teenage and middle-age participants, respectively. Also, Patterson et al. (1996), in a study involving 11–15-year-olds, reported fairly comparable BSR coefficients for male participants (left leg = .68; right leg = .72), but somewhat lower values for female participants (left leg = .51; right leg = .52). The high CSR reliability values for the older adults in this study were also similar to the SR and BSR values reported in other studies, with $R$ coefficients in all cases consistently above .90.

Although the findings of this study and others indicate that the CSR, SR, and BSR all have comparable reliability and validity coefficients for participants who can perform the tests, 8 (approximately 10%) of the original volunteer participants in the validity study had to be excluded because they either could not or would not get down on the floor for sit-and-reach testing. No participants, on the other hand, were eliminated due to their inability to perform the chair sit-and-reach test. Also, in spite of our emphasis on proper spotting of the participants in this study, one female participant did fall backward during the SR testing, hitting her head on the gymnasium floor. As indicated earlier, weakened trunk muscles or tight hamstrings, or both, make it difficult for some older adults to maintain a straight-leg sitting position on the floor. No injuries occurred during testing on the CSR. However, careful spotting is recommended when assessing frail participants or individuals with balance problems.

In conclusion, the measurement of hamstring flexibility is an important component of health-related fitness. Results of this study indicate that the CSR is highly reliable and has moderate validity as a measure of hamstring flexibility. Further, the CSR appears to be a safe and socially acceptable assessment procedure for older adults and can measure each leg separately to detect any bilateral differences in hamstring flexibility. Early detection of shortened hamstrings would be valuable for the practitioner in providing feedback for exercise prescription to help correct or reduce imbalances that may lead to mobility problems and potential injuries. Furthermore, in an era of assessment and accountability for health care (Russek, Wooden, Ekedahl & Bush, 1997), the CSR could potentially provide practitioners an excellent outcome measure for assessing the benefits of therapeutic intervention with this population, although additional studies are needed to test the ability of the CSR to detect change over time. Also, more studies are recommended to investigate the reliability and validity of the CSR with physically frail and disabled populations. The relationship of the CSR in this study to other hamstring flexibility measures must be delimited to the population studied, apparently healthy older adults with no major orthopedic limitations.

References


Note

1. Test theory suggests that with multiple trials the average score is normally a more reliable indication of performance than the best score. However, because SR, BSR, and CSR performance has been found to be quite reliable using the “best score” protocol (above .90), we believe this method is justified and, in fact, recommended in group-testing situations where efficiency in testing time is critical.

Authors’ Notes

This research was supported by a grant from PacifiCare Health Systems. The authors thank the staff and all volunteer participants from the Leisure World and Morningside Retirement Communities and from the Lifespan Wellness Clinic for their assistance with this study. Please address all correspondence regarding this article to C. Jessie Jones, Division of Kinesiology and Health Promotion, California State University–Fullerton, Fullerton, CA 92834.

E-mail: jjones@fullerton.edu