Review article

Criterion-Related Validity of Sit-And-Reach Tests for Estimating Hamstring and Lumbar Extensibility: A Meta-Analysis

Daniel Mayorga-Vega [1]✉, Rafael Merino-Marban [2] and Jesús Viciana [1]

[1] Department of Physical Education and Sport, University of Granada, Spain; [2] Department of Didactics of Musical, Plastic and Corporal Expression, University of Malaga, Spain

Abstract

The main purpose of the present meta-analysis was to examine the scientific literature on the criterion-related validity of sit-and-reach tests for estimating hamstring and lumbar extensibility. For this purpose relevant studies were searched from seven electronic databases dated up through December 2012. Primary outcomes of criterion-related validity were Pearson´s zero-order correlation coefficients (r) between sit-and-reach tests and hamstrings and/or lumbar extensibility criterion measures. Then, from the included studies, the Hunter-Schmidt´s psychometric meta-analysis approach was conducted to estimate population criterion-related validity of sit-and-reach tests. Firstly, the corrected correlation mean (r_p), unaffected by statistical artefacts (i.e., sampling error and measurement error), was calculated separately for each sit-and-reach test. Subsequently, the three potential moderator variables (sex of participants, age of participants, and level of hamstring extensibility) were examined by a partially hierarchical analysis. Of the 34 studies included in the present meta-analysis, 99 correlations values across eight sit-and-reach tests and 51 across seven sit-and-reach tests were retrieved for hamstring and lumbar extensibility, respectively. The overall results showed that all sit-and-reach tests had a moderate mean criterion-related validity for estimating hamstring extensibility ($r_p = 0.46$-0.67), but they had a low mean for estimating lumbar extensibility ($r_p = 0.16$-0.35). Generally, females, adults and participants with high levels of hamstring extensibility tended to have greater mean values of criterion-related validity for estimating hamstring extensibility. When the use of angular tests is limited such as in a school setting or in large scale studies, scientists and practitioners could use the sit-and-reach tests as a useful alternative for hamstring extensibility estimation, but not for estimating lumbar extensibility.

Key words: Concurrent validity, range of motion, flexibility, field test, lineal test, systematic review.

Introduction

Lack of hamstring muscles extensibility conditions a decrease of pelvic mobility (Kendall et al., 2005). This invariably leads to biomechanical changes in the pressure distribution of the spine and consequent spinal disorders (da Silva Días and Gómez-Conesa, 2008). Therefore, poor hamstring extensibility has been associated with thoracic hyperkyphosis (Fisk et al., 1984), spondylolysis (Standaert and Herring, 2000), disc herniation (Harvey and Tanner, 1991), changes in lumbopelvic rhythm (Esola et al., 1996; López-Miñarro and Alacid, 2009) and low back pain (Biering-Sorensen, 1984; Mierau et al., 1989). Additionally, individuals with shortened hamstring muscles

present gait limitations, increased risk of falls, and susceptibility to musculoskeletal injuries (Erkula et al., 2002; Jones et al., 1998).

Nowadays different kinds of tests are used to assess hamstring extensibility. Flexibility is typically characterized by the maximum range of motion in a joint or series of joints (McHugh et al., 1998). Thus, angular tests that specifically measure hip flexion with the knee extended (straight leg raise test), or the range of knee extension with the hip flexed to 90 degrees (knee extension or popliteal angle test), have been widely considered the criterion measures of hamstring extensibility (e.g., Ayala et al., 2011; Hartman and Looney, 2003; López-Miñarro and Rodríguez-García, 2010c). Nevertheless, due to the necessity of sophisticated instruments, qualified technicians, and time constraints, the use of these angular tests seem to be limited in several settings such as in a school context or large scale studies (Castro-Piñero et al., 2009b).

Unlike the angular tests, lineal tests have a simple procedure, are easy to administer, require minimal skills training for their application, and the equipment necessary to perform them is very affordable (Castro-Piñero et al., 2009b; López Miñarro et al., 2008c). Sit-and-reach (SR) tests in which a fingertips-to-tangent feet distance is measured are probably the most widely used lineal measures of flexibility (Holt et al., 1999; Castro-Piñero et al., 2009a). However, as the SR is a test which involves the movement of the whole body, it has been suggested that the position of the fingertips does not give valid information about hamstring extensibility (Hoeger et al., 1990). The main factors that seem to affect the validity of SR tests to estimate hamstring extensibility are the differences in length proportion between the upper and lower limbs (Hoeger et al., 1990), the position of the head (Smith and Miller, 1985) and the position of the ankles (Kawano et al., 2010; Liemohn et al., 1997). In addition, recent studies have also found that the levels of hamstring extensibility influence the criterion-related validity of SR tests (López-Miñarro et al., 2011; López-Miñarro and Rodríguez-García, 2010c).

The choice of a flexibility test must be based on its functionality and validity (López-Miñarro, 2010). Although the angular tests have the advantage of being the criterion measure to assess flexibility, due to several practical reasons they have the disadvantage of having a limited use in several settings (Castro-Piñero et al., 2009b). In these settings, as the SR tests have the advantage of allowing for an evaluation in a short amount of time with

minimal skills and instruments, potentially they could be a useful alternative to estimate flexibility. Nevertheless, as in the application of any fitness field test, the SR tests' results are a simple estimation and, therefore, the evaluators must be aware of validity coefficients in order to interpret the scores of these tests correctly. Unfortunately, the studies examining criterion-related validity of SR tests for estimating hamstring and lumbar extensibility have shown inconclusive results (Baltaci et al., 2003; Hui and Yuen, 2000; Hui et al., 1999; Jones et al., 1998).

Each primary study that is published about criterion-related validity of the SR tests only constitutes as a single piece of a constantly growing body of evidence (Cooper et al., 2009). For example, in some studies the correlation coefficient is statistically significant, while in others a statistically significant association is not found. In some cases the strength of the association is quite high, while low in others. To make sense of the often conflicting results found in the scientific literature, researchers have to conduct meta-analyses (Cooper et al., 2009; Hunter and Schmidt, 2004; Lipsey and Wilson, 2001). Hence, the meta-analyses remain a useful tool for the evaluation of evidence (Flather et al., 1997), forming a critical process for theory development in science (Hunter and Schmidt, 2004).

Unfortunately, to our knowledge there are not any meta-analyses addressing the criterion-related validity of SR tests. Beyond the simple but important function of describing and summarizing the scientific findings of a research area, the main contribution of a meta-analysis is to estimate as accurately as possible the population parameters (Hunter and Schmidt, 2004). Therefore, the results of a meta-analysis let us generalize the research findings, as well as test hypotheses that may have never been tested in primary studies. Finally, the meta-analyses permit us to examine today´s lack of knowledge in a specific area and to guide scientists in future research (Cooper et al., 2009).

Consequently, the main purpose of the present meta-analysis was to examine the scientific literature on criterion-related validity of SR tests for estimating hamstring and lumbar extensibility in apparently healthy individuals. More specifically, the objectives of this study were: (a) to describe and summarize the up-to-date scientific findings of criterion-related validity of SR tests for estimating hamstring and lumbar extensibility; (b) to estimate and compare the overall population mean of the criterion-related validity coefficients of each SR test for estimating hamstring and lumbar extensibility; and (c) to examine the influence of some study features (sex of the participants, age of participants, and level of hamstring extensibility) in criterion-related validity coefficients of SR tests.

Methods

Search strategy
The following seven electronic databases were searched from their inception through December 2012: SportDiscus, Scopus, Medline, Pubmed, Web of Science, ERIC, and Dissertations & Theses Database. The search terms used were based on two concepts. Concept one included terms for the SR test (sit and reach) and concept two included terms for validity (validity, related, relationship, correlation, comparison, hip, hamstring, flexibility, ROM, range of motion, range of movement, straight leg raise, knee extension, popliteal angle, lumbar, back, Macrae and Wright, Macrae & Wright, Schober, radiography, goniometer, and inclinometer). The terms of the same concept were combined together with the Boolean operator "OR" and then the two concepts were combined using the Boolean operator ''AND'' (Benito Peinado et al., 2007). The keywords that consisted of more than one word were enclosed in quotes. In addition, the reference lists of all included papers were manually searched.

Selection criteria
The selection criteria to identify studies that examined the criterion-related validity of SR tests for estimating hamstring and/or lumbar extensibility were: (a) studies with apparently healthy participants who did not present any injury, physical and/or mental disabilities; (b) studies with SR tests that yielded the values of the maximum reach of the fingertips; and (c) studies in which hamstring and/or lumbar extensibility criterion measurements used are widely accepted in the scientific literature (e.g., straight leg raise or knee extension tests for hamstring extensibility and Macrae & Wright or inclinometer methods for lumbar extensibility). In addition to papers, master/doctoral dissertations and conference proceedings were also accepted. No language or publication date restrictions were imposed.

Coding studies
For this meta-analysis, data were collected from studies that reported relationships between SR tests and hamstring and/or lumbar extensibility criterion measures with apparently healthy participants of any age. From each selected study the following data were coded: Study identity number, sample size (n), sex of participants (1 = males, 2 = females), age of participants (1 = children, < 18 years; 2 = adults, ≥ 18 years), SR test protocol (1 = Classic SR, 2 = Modified SR, 3 = Back-saver SR, 4 = Modified back-saver SR, 5 = V SR, 6 = Modified V SR, 7 = Unilateral SR, 8 = Chair SR), criterion-related validity value (Pearson´s r correlation coefficient), reliability of SR test (intraclass correlation coefficient), reliability of hamstring and/or lumbar extensibility criterion measures (intraclass correlation coefficient), and the average score of hamstring extensibility criterion measure. Because identification of study features is usually explicitly stated in each of the primary articles, the use of more than one rather was deemed unnecessary.

In addition, although various protocols for evaluating quality of single studies have been described, there is no widespread agreement on the validity of this type of evaluation approach. Thus, rejecting certain single studies and accepting others for inclusion in a meta-analysis on the basis of a quality score remains a controversial procedure (Flather et al., 1997). Hence, according to Flather et al. (1997), our approach has been to ensure that the design has not been flawed (e.g., conducted by scientifically

evidenced criterion measures), and that there has been a complete reporting of relevant outcomes. For a study to be included in this meta-analysis, sample size, SR test protocol, hamstring and/or lumbar criterion measures and Pearson's r were considered to be critical. In the event that the authors mixed subgroups of a study feature (e.g., males mixed with females), failed to identify a study feature (e.g., criterion measure or reliability scores) or were ambiguous (e.g., hamstring extensibility scores around 80º shown graphically) the data was omitted. When in the same study data for males and females were expressed both separately and together, only the separate data were coded. When in the same study data were expressed for both legs separately or for two different days from the same sample (i.e., such as in Mier, 2011), the average value of the coefficients was coded.

Finally, in the event that included studies used multiple validity coefficients for hamstring and/or lumbar extensibility, only the data relative to one criterion measure of each muscle group was coded. Regarding hamstring extensibility, all studies reported correlation values with the straight leg raise test, while only in a few articles the values with the knee extension test was also stated (Davis et al., 2008; García, 1995; Harman and Looney, 2003). Therefore, in order to avoid moderator effects issues by criterion measure test, only the correlation values of the straight leg raise test were coded. As regards lumbar extensibility, only Hartman and Looney (2003) performed more than one criterion measure test (Single inclinometer and Macrae & Wright methods). Due to the fact that the Macrae & Wright method has been used the most widely, the results with this test were coded.

Data analyses

In the present study, Pearson's zero-order correlation coefficient (r) was considered the unit of measure as an indication of criterion-related validity of SR tests, which represents the strength of associations between the estimates of SR tests and the criterion measures. Because several studies reported criterion-related validity results of different SR test protocols from the same sample, r values were extracted separately for each SR test to avoid dependency issues in the meta-analysis (Cooper et al., 2009). Similarly, criterion-related validity values were extracted separately for hamstring and lumbar extensibility. However, if a single study reported more than one r value within the same SR test protocol, but from different subsamples (e.g., males and females), we assumed each r value from different subsamples to be independent from each other and included them in a single meta-analysis (Lipsey and Wilson, 2001).

Publication bias: In addition to the followed search strategy and selection criteria to avoid availability bias, an examination of the selected studies was carried out to avoid a potential duplication of information retrieved. Since some selected studies had full or partial duplicated information, these particular r correlations values were not analyzed in the meta-analyses. Furthermore, before computing correlations, several exploratory analyses were also conducted to detect the presence of publication bias. Firstly, a file drawer analysis based on effect size was

performed to estimate the number of unlocated studies averaging null results ($r = 0$) that would have to exist to bring the mean effect size (r_p) down to the small mean r value (Rosenthal, 1979). Depending on the results of the file drawer analysis, we had to conclude if it was likely that there would be this particular number of "lost" studies to reduce the actual r to a small value. According to Cohen's guidelines (1992), the correlation coefficient was interpreted as small when $r < 0.30$.

Secondly, according to Light and Pillemer (1984), the scatter plots of correlations coefficients against sample size for each SR test protocol related to both hamstring and lumbar extensibility were analyzed. According to this graphic method, in the absence of publication bias, the resulting figure should take the form of an inverted funnel. However, based on the statistical significance of the studies, if there is publication bias the small-sample studies reporting small r values will be disproportionately absent because they are the studies that will fail to attain statistical significance. Finally, with the objective of quantifying the outcomes of the scatter plots, as suggested by Begg and Mazumdar (1994), a Spearman's rank order correlation between r values and sample size was calculated. In the presence of publication bias, this correlation should be statistically significant negative due to the absence of small-sample studies in the lower left hand corner.

Computation of correlations: The Hunter-Schmidt's psychometric meta-analysis approach was conducted to obtain the population estimates of the criterion-related validity of SR tests (Hunter and Schmidt, 2004). This approach estimates the population correlation by individually correcting the observed correlations due to various artefacts such as sampling error and measurement error. First, the "bare-bone" mean r (r_c), corrected for only sampling error, was calculated by weighting each r with the respective sample size when aggregating them into r_c. Then, we calculated the corrected mean r at the population level (r_p) that was unaffected by both sampling error and measurement error. The resulting mean correlation corrected for sampling error and measurement error is offered as the best estimate of the population parameter. In order to correct the measurement errors, the reliability coefficients (intraclass correlation coefficients) of each individual SR and criterion measure tests were used. Because the reliability coefficients were not available for all of the included studies, the unknown reliability values were previously estimated for each test. The median of the all reported reliability coefficients for each SR test protocol and criterion measure test was used. Finally, the 95% confidence intervals of r_p (95% CI) were calculated.

Moderator analysis: In the present meta-analysis, due to the low number of r values found, partially hierarchical analyses of moderator variables were carried out. According to Hunter and Schmidt (2004), to determine the presence of moderator effects which may affect overall criterion-related validity of SR tests (r_p), three different criteria were simultaneously examined: (a) the percentage of variance accounted for by statistical artefacts is less than 75% of the observed variance in r_p; (b) the Q homogeneity statistic is statistically significant ($p <$
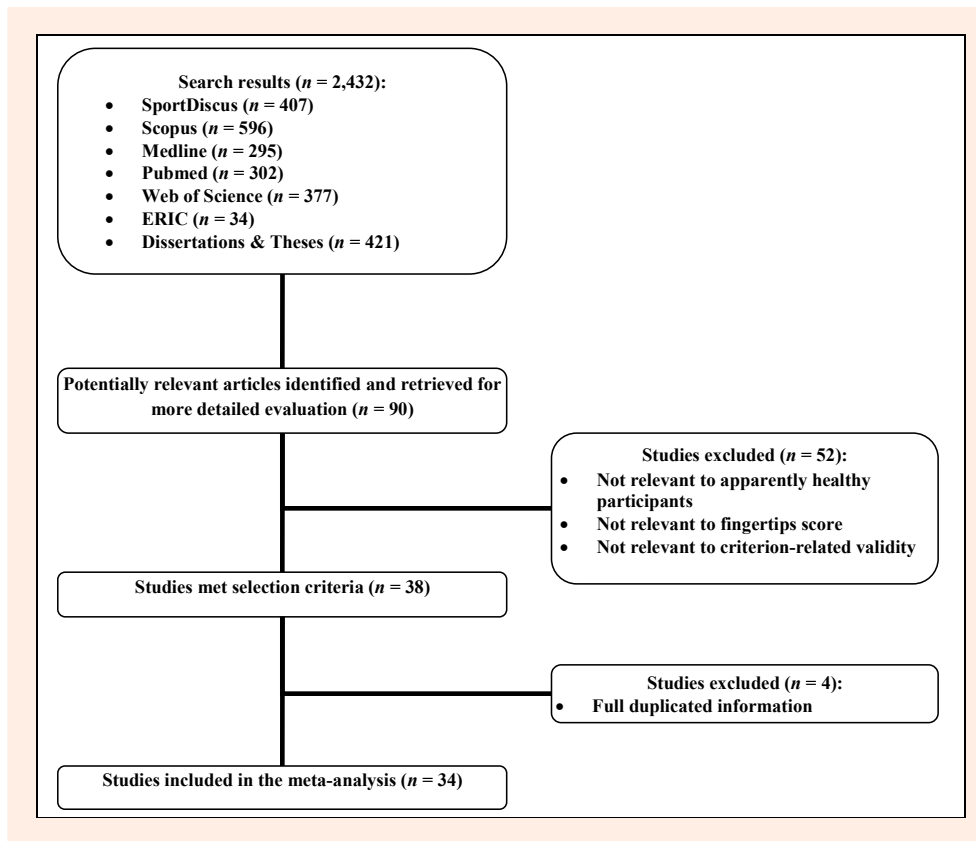
Figure 1. Flow chart of studies selection process.

0.05); and (c) the 95% credibly interval (95% CV) is relatively large or includes the value zero. If at least one of the three criteria were met, we concluded that the results could be affected by moderator effects. In case of the presence of moderator effects, criterion-related validity values of each SR test were analyzed separately by: (a) sex of participants (i.e., male and female); (b) age of participants (i.e., children and adults); and (c) level of hamstring extensibility (i.e., low average level, $< 80º$, and high average level, $\geq 80º$) (Kendall et al., 2005).

Results

Study description
Figure 1 shows a flow chart of the study selection process. Of the 2,432 literature search results, 90 potentially relevant publications were identified and retrieved for a more detailed evaluation. Finally, due to duplication issues, of the 38 studies that met the inclusion criteria, only 34 studies were included in the present meta-analysis. Apart from a few studies retrieved which were carried out with apparently non-healthy participants or lineal tests that did not yield the values of the maximum reach of the fingertips, other studies (or *r* values) were not included either in the present meta-analysis because they examined the relationship between the SR test and the pelvic tilt scores (e.g., Davis et al., 2008; Kawano et al., 2010; López-Miñarro, 2010; Rodríguez-García et al., 2008). The pelvic tilt is measured by the inclination angle of the sacrum with regard to the horizontal line at the point of maximal forward reach on the SR test. Therefore, although the pelvis position is influenced by the hamstring

extensibility, its measure must be considered as an estimation of hamstring extensibility (indirect measure), and not as a criterion measure to determinate it (direct measure) such as the straight leg raise or knee extension tests (Santonja Medina et al., 1995). However, nowadays some studies have suggested that the criterion measures of hamstring extensibility must be reexamined and readjusted (Cardoso et al., 2007; Hartman and Looney, 2003) (see strengths and limitations section).

Table 1 presents a summary of studies of criterion-related validity of SR tests for estimating hamstring and lumbar extensibility. Regarding the criterion-related validity for estimating hamstring extensibility, a total of 99 *r* values across eight SR test protocols were retrieved, ranging from three values in the Chair SR and Modified V SR tests to 47 values in the Classic SR test. Total sample sizes for each SR test ranged from 182 in the Chair SR test to 3,481 in the Classic SR test. The individual criterion-related validity correlation coefficients of SR tests for estimating hamstring extensibility ranged from 0.19 to 0.93. Regarding criterion-related validity for estimating lumbar extensibility, a total of 51 *r* values across seven SR test protocols were retrieved, ranging from two values in the Unilateral SR test to 21 values in Classic SR test. Studies examining the criterion-related validity of the Chair SR test for estimating lumbar extensibility were not found. Total sample sizes for each SR test ranged from 158 in the Unilateral SR test to 1,762 in Classic SR test. The individual criterion-related validity correlation coefficients of SR tests for estimating lumbar extensibility ranged from 0.00 to 0.60.

Table 1. Summary of studies of criterion-related validity of sit-and-reach tests for estimating hamstring and lumbar extensibility.

Reference	Sample	n	Age (yrs)	Test	Hamstring extensibility			Lumbar extensibility		
					Criter	♂ (r)	♀ (r)	Criter	♂ (r)	♀ (r)
Ayala et al. (2011)	Professional futsal players	♂=55 ♀=48	26.0 (4.5) 23.0 (5.3)	CSR MSR BSSR	PSLR PSLR PSLR	.62* .76* .47	.93* .73* .91*			
Ayala et al. (2012)	Recreationally active university students	♂=156 ♀=87	21.3 (2.5) 20.7 (1.6)	CSR	PSLR		.79*			
Baker (1985)	High and Middle school students	♀=100	14.1 (.8)	CSR	PSLR		**.64***	MWM		**.28***
Baltaci et al. (2003)	University students	♀=102	22.0 (1.0)	CSR BSSR CHSR	PSLR PSLR PSLR		.63*-.53* .37*-.44* .22*-.16			
Book (1989)	Public and private schools students	♂=203 ♀=255	9-18	CSR	PSLR	.65*	.81*	MWM	.33*	.30*
Bozic et al. (2010)	Physically active sport and PE white students	♂=84	21.3 (2.6)	CSR	PSLR	.63*				
Castro-Piñero et al. (2009b)	Caucasian children and adolescents	♂=29 ♀=27 ♂=16 ♀=15	6-12 13-17	CSR MSR CSR MSR	PSLR PSLR PSLR PSLR	.38* .34* .38* .26				
Chung and Yuen (1999)	University students	♂=52	20.7 (1.3)	CSR MSR	PSLR PSLR	.77* .71*		MWM MWM	.23 .24	
Davis et al. (2008)	University students	♂=42 ♀=39	23.6 (4.1) 24.1 (4.3)	CSR CSR	PSLR PKE	.65* .57*				
García (1995)	High school students	♂=54 ♀=55	15-18	CSR CSR	PSLR AKE	.64* .63*	.67* .61*	MWM	.32*	.28*
Hartman and Looney (2003)	Schoolchildren	♂=85 ♀=88	6-12	CSR CSR BSSR BSSR	PSLR AKE PSLR AKE	.67*-.66* .40*-.40* .69*-.67* .50*-.47*	.47*-.49* .52*-.54* .42*-.48* .54*-.57*	MWM SIM MWM SIM	.05 .29* .00-.07 .26*-.28*	.07 .16 .06-.06 .10-.10
Hui et al. (1999)	University students	♂=62 ♀=96	21.1 (3.1) 20.6 (2.1)	CSR MSR BSSR MBSSR VSR MVSR	PSLR PSLR PSLR PSLR PSLR PSLR	.48*-.47* .45*-.45* .46*-.44* .44*-.45* .58*-.63* .57*-.62*	.46*-.53* .41*-.47* .39*-.50* .35*-.47* .44*-.52* .46*-.51*	MWM MWM MWM MWM MWM MWM	.27* .24 .24-.27* .17-.20 .42* .38*	.24* .22* .18-.15 .22*-.22* .24* .28*
Hui and Yuen (2000)	University students	♂=62 ♀=96	21.1 (3.1) 20.6 (2.1)	CSR BSSR VSR USR	PSLR PSLR PSLR PSLR	**.48*-.47*** **.46*-.44*** **.58*-.63*** .61*-.67*	**.46*-.53*** **.39*-.50*** **.44*-.52*** .50*-.54*	MWM MWM MWM MWM	**.27*** **.24-.27*** **.42*** .47*-.47*	**.24*** **.18-15** **.24*** .26*-.23*
Jackson and Baker (1986)	School PE students	♀=100	14.1 (.8)	CSR	PSLR		.64*	MWM		.28*
Jackson and Langford (1989)	?	♂=52 ♀=52	20-45	CSR	PSLR	.89*	.70*	MWM	.59*	.12
Jones et al. (1998)	Exercise classes at a retirement community	♂=32 ♀=48	74.5 (5.7) 74.0 (6.7)	CSR BSSR CHSR	PSLR PSLR PSLR	.74* .70* .76*	.71* .71* .81*			
Kanbur et al. (2005)	Non-prepubertal/ non-regularly exercised boys	♂=69	13-14	CSR	PSLR	.64*-.65*				
Langford (1987)	?	♂=52 ♀=52	20-45	CSR	PSLR	**.89***	**.70***	MWM	**.60***	**.12**
Lemmink et al. (2003)	Independently living people over 55	♂=49 ♀=71	67.7 (7.5) 65.6 (8.6)	CSR MSR	PSLR PSLR	.74* .54*	.57* .57*	AAOS AAOS	.13 .21	.31* .26*
Liemohn et al. (1994)	University students	♂=20 ♀=20	24.0 (4.6) 25.1 (6.3)	CSR BSSR	PSLR PSLR	.72* .76*	.70* .70*	SIM SIM	.29 .32	.40 .38

This table includes all studies that met selection criteria, however, full or partial information was not included in the meta-analysis (in bold) due to duplication issues; ♂, males; ♀, females; ?, information unavailable; Criter: Criterion, CSR, Classic sit-and-reach test; MSR, Modified sit-and-reach test; BSSR, Back-saver sit-and-reach test; MBSSR, Modified back-saver sit-and-reach test; VSR, V sit-and-reach test; MVSR, Modified v sit-and-reach test; USR, Unilateral sit-and-reach test; CHSR, Chair sit-and-reach test; PSLR, Passive straight leg raise test; ASLR, Active straight leg raise test; PKE, Passive knee extension test; AKE, Active knee extension test; SMM, Spinal Mouse method; SIM, Single Inclinometer method; MWM, Macrae & Wright method; AAOSM, American Academy of Orthopedic Surgeons method;, Pearson´s r for the left and right leg, respectively.
* Pearson´s r statistically significant at p < 0.05

Table 1. Continued.

Reference	Sample	n	Age (yrs)	Test	Hamstring extensibility			Lumbar extensibility		
					Criter	♂ (r)	♀ (r)	Criter	♂ (r)	♀ (r)
López-Miñarro et al. (2008)	University students	♂=102	22.9 (3.2)	CSR	PSLR	.56*-.59*	.72*-.74*	SIM	.32*	.14
		♀=96	23.2 (4.5)	VSR	PSLR	.53*-.55*	.63*-.65*	SIM	.33*	.29*
López Miñarro et al. (2008a)	Canoeists	♂=44 ♀=22	13.3 (.6)	CSR	PSLR	.77*-.73*	.74*-.81*			
López Miñarro et al. (2008b)	Canoeists	?=66	13.3 (.6)	CSR	PSLR	**.70*-.68***				
López Miñarro et al. (2008c)	?	♂=120	22.8 (3.1)	CSR	PSLR	**.56*-.59***	**.72*-.74***			
		♀=100	23.1 (4.6)	BSSR	PSLR	.51*-.50*	.68*-.68*			
				USR	PSLR	.54*-.58*	.73*-.75*			
López-Miñarro et al. (2009)	University students	♂=76	23.5 (4.0)	CSR	PSLR	.56*-.59*	.75*-.73*			
		♀=67	23.9 (5.4)	BSSR	PSLR	.53*-.51*	.70*-.66*			
López-Miñarro et al. (2010a)	?	♂=73	23.0 (3.5)	CSR	PSLR	.44*-.48*	.75*-.73*			
		♀=71	23.1 (4.3)	MSR	PSLR	.28*-.32*	.63*-.64*			
López-Miñarro et al. (2010b)	University students	♂=130	22.9 (3.2)	CSR	PSLR	**.56*-.59***	**.72*-.74***			
		♀=110	23.2 (4.5)	MSR	PSLR	.41*-.45*	.62*-.63*			
				BSSR	PSLR	**.51*-.49***	**.68*-.68***			
				VSR	PSLR	**.53*-.55***	**.63*-.65***			
López-Miñarro and Rodríguez-García (2010c)	Recreationally active university students: Low and normal flexibility	♂=120 ♂=120	22.9 (3.6)	CSR CSR	PSLR PSLR	.31*-.41* .61*-.55*				
López-Miñarro et al. (2011)	Older women: Low, moderate and high flexibility	♀=36 ♀=35 ♀=35	65.3 (9.1)	CSR	PSLR		.43*-.41* .54*-.57* .73*-.70*			
López-Miñarro et al. (2012)	Canoeists Kayakers	♂=51 ♂=60	17.5 (6.3)	CSR	PSLR	.67*-.66* .59*-.59*				
Mier (2011)	Physically active adults	♂=30 ♀=30	25.0 (9.3) 23.7 (7.9)	CSR CSR	PSLR PSLR	.64*/.66*	.79*/.81*			
Minkler and Patterson (1994)	Regular PE activity classes practitioners (mainly Caucasians)	♂=48 ♀=51	24.3 (4.7) 21.5 (3.8)	MSR	PSLR	.75*	.66*	MWM	.40*	.25
Miyazaki et al. (2010)	Community-dwelling elderly	♂=42 ♀=119	72.6 (6.9)	CSR	PSLR	.60*		SMM	.18	
Orloff (1988)	Gymnasium practitioners	♂=47 ♀=28	19-54	CSR	ASLR	.52*				
Patterson et al. (1996)	Middle school students (various ethnic origins)	♂=40 ♀=44	13.0 (.9) 12.7 (.8)	BSSR	PSLR	.72*-.68*	.51*-.52*	MWM	.15-.10	.17-.25
Rodríguez-García et al. (2008)	Fit sports activities practitioners	♂=125 ♀=118	22.9 (3.2) 23.2 (4.5)	CSR	PSLR	**.56*-.59***	**.72*-.74***	SIM	**.32***	**.14**
Simoneau (1998)	Physically active university students	♀=34	20.3 (.9)	CSR	PSLR		.78*	MWM		.26
Yuen and Hui (1998)	University students	♂=19	?	CSR	PSLR	.52*-.57*		MWM	.18	
		♀=36		MSR	PSLR	.49*-.55*		MWM	.27	
				BSSR	PSLR	.46*-.44*		MWM	.16-.12	
				MBSSR	PSLR	.39*-.52*		MWM	.21-.17	
				VSR	PSLR	.54*-.60*		MWM	.19	
				MVSR	PSLR	.58*-.61*		MWM	.25	

Note. This table includes all studies that met selection criteria, however, full or partial information was not included in the meta-analysis (in bold) due to duplication issues; ♂, males; ♀, females; ?, information unavailable; Criter, Criterion, CSR, Classic sit-and-reach test; MSR, Modified sit-and-reach test; BSSR, Back-saver sit-and-reach test; MBSSR, Modified back-saver sit-and-reach test; VSR, V sit-and-reach test; MVSR, Modified v sit-and-reach test; USR, Unilateral sit-and-reach test; CHSR, Chair sit-and-reach test; PSLR, Passive straight leg raise test; ASLR, Active straight leg raise test; PKE, Passive knee extension test; AKE, Active knee extension test; SMM, Spinal Mouse method; SIM, Single Inclinometer method; MWM, Macrae & Wright method; AAOSM, American Academy of Orthopedic Surgeons method; Pearson´s *r* for the left and right leg, respectively.
* Pearson´s *r* statistically significant at $p < 0.05$

Publication bias

Due to some studies having fully or partially duplicated information, these *r* coefficients values were not analyzed in the present meta-analyses despite the fact that these studies met the selection criteria. For example, Baker (1985) and Langford´s (1987) doctoral dissertations were not included because the data were published later in a journal (although in Langford´s works there was a little difference in one *r* value, it was simply considered a typo because the other data were equal) (Jackson and Baker, 1986; Jackson and Langford, 1989). López Miñarro´s et al. (2008b) study information (males mixed with females) were not computed because the same data were also published with males and females separately (López Miñarro et al., 2008a). Additionally, full or partial information from a few studies of the same authors, sample character-

istics, and correlation results was not included either due to duplication issues (Hui and Yuen, 2000; López Miñarro et al., 2008c; López-Miñarro et al., 2010b; Rodríguez-García et al., 2008). Pearson´s *r* correlation values of selected studies that were excluded for meta-analysis are indicated (in bold) in Table 1.
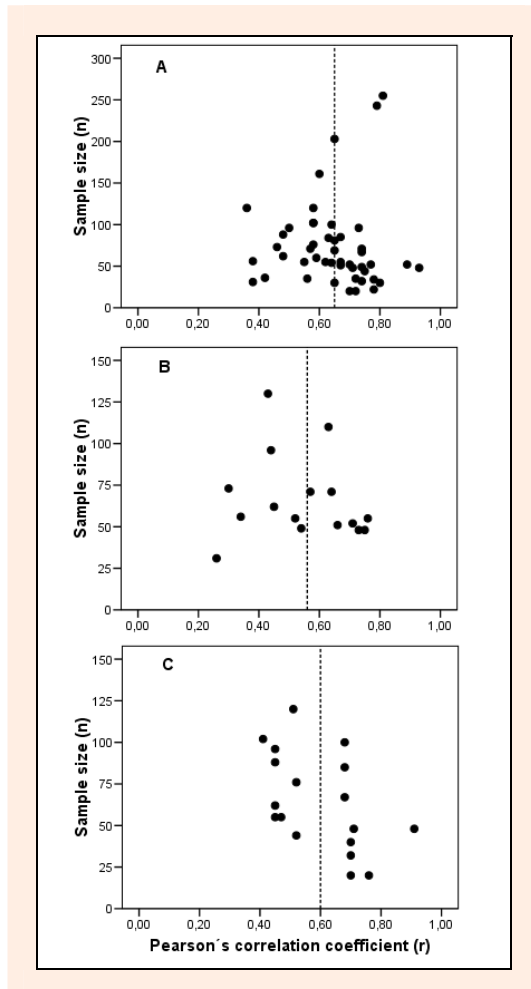
unlikely (67-133%). Hence, we concluded that it was unlikely that there would be this particular number of "lost" studies for each SR test protocol. On the other hand, regarding the lumbar extensibility, the file drawer analyses were not calculated because the actual *r* values were small.

Figure 2. Scatter plots of sample size and criterion-related validity coefficients (*r*) for estimating hamstring extensibility: (a) Classic sit-and-reach; (b) Modified sit-and-reach; and (c) Back-saver sit-and-reach. Dashed line represents median values of validity coefficients.
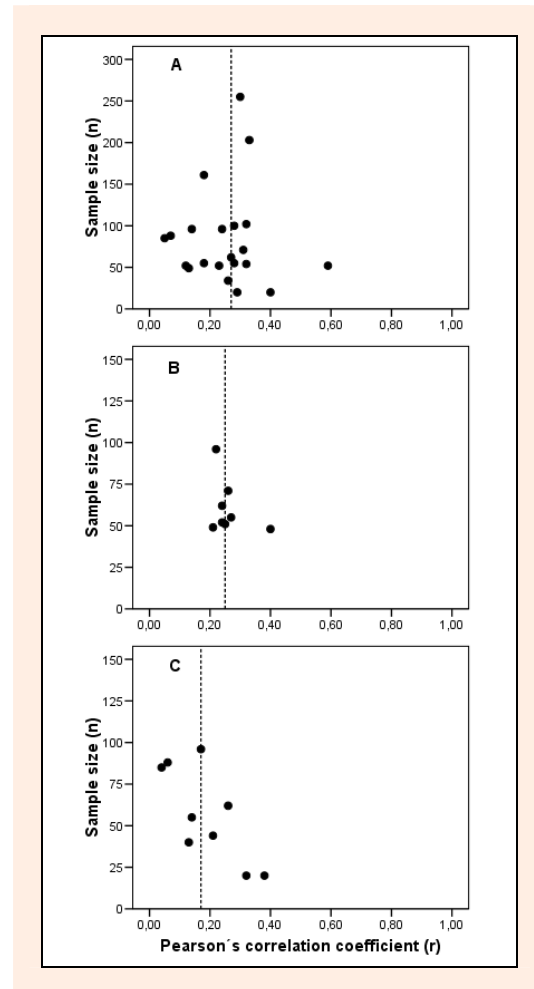

Figure 3. Scatter plots of sample size and criterion-related validity coefficients (r) for estimating lumbar extensibility: (a) Classic sit-and-reach; (b) Modified sit-and-reach; and (c) Back-saver sit-and-reach. Dashed line represents median values of validity coefficients.

Afterward, several exploratory analyses were conducted to detect the presence of publication bias. Regarding hamstring extensibility, the results of the file drawer analyses are based on effect size for estimating the number of unlocated studies averaging null results ($r = 0$) that would have to exist to bring the mean r_p down to 0.29. These results are shown in the following lines (in parenthesis the unlocated/located percentage): 63 for the Classic SR (134%), 15 for the Modified SR (94%), 19 for the Back-saver SR (106%), 2 for the Modified back-saver SR (67%), 6 for the V SR (120%), 4 for the Modified V SR (133%), 5 for the Unilateral SR (125%), and 3 for the Chair SR (100%). Although we are aware that there is not a large number of "lost" studies for the Modified back-saver SR, V SR, Modified V SR, Unilateral SR, and Chair SR, the percentage of unlocated/located studies was

Figures 2 and 3 show the scatter plots of sample size against criterion-related validity coefficients for estimating hamstring and lumbar extensibility, respectively. Due to the low number of *r* values for the most SR test protocols (2-5 *r* values), only the scatter plots for the Classic SR, Modified SR, and Back-saver SR tests were examined. According to this graphic method, the figures suggested that there was an absence of publication bias for the Classic SR and Modified SR tests. However, the two scatter plots of the Back-saver SR test suggested the presence of publication bias, because of the absence of *r* values in the lower left hand corner of the inverted funnel plot. In this line, for the Back-saver SR test, the results of Spearman´s rank order correlation between *r* values and sample size showed a statistically significant negative correlation for estimating hamstring extensibility (*r* = -

0.66, $p = 0.003$) and marginally significant for lumbar extensibility ($r = -0.61$, $p = 0.081$). Nevertheless, for the Classic SR and Modified SR tests the results did not show a statistically significant correlation for either estimating hamstring (Classic SR, $r = -0.29$, $p = 0.050$; Modified SR, $r = -0.33$, $p = 0.207$) or lumbar extensibility (Classic SR, $r = -0.02$, $p = 0.935$; Modified SR, $r = -0.22$, $p = 0.608$). Finally, although we aware that the results for the Classic SR test for estimating hamstring extensibility were marginally significant, the r value was considerably lower than for the Back-saver SR test.

Criterion-related validity

Table 2 reports the number of studies (K), the cumulative number of r values (n), the total sample size accumulated (N), the overall weighted mean of r corrected for sampling error only (r_c), the overall weighted mean of r corrected for both sampling error and measurement error (r_p), as well as the 95% CI for overall criterion-related validity correlation coefficients (r_p) separately for estimating hamstring and lumbar extensibility across each SR test protocol. In addition, to detect the presence of moderator effects which may affect overall criterion-related validity of SR tests, the 95% CV, the percentage of variance accounted for by statistical artefacts, and the Q homogeneity statistic were calculated.

Hamstring extensibility: The overall results showed that all SR test protocols had a moderate mean correlation coefficient of criterion-related validity for estimating hamstring extensibility (r_p range = 0.46-0.67) in which all 95% CI did not include the value zero. For five of the eight SR test protocols, the percentage of variance accounted for by statistical artefacts was less than 75%, the Q homogeneity statistic was statistically significant ($p < 0.05$), and the 95% CV was relatively large. Therefore, follow-up moderator analyses were conducted using predefined moderators as it was hypothesized in the present study.

Lumbar extensibility: The overall results showed

that all SR test protocols had a low mean correlation coefficient of criterion-related validity for estimating lumbar extensibility (r_p range = 0.16-0.35) in which the 95% CI of the Back-saver SR and the Modified back-saver SR tests included the value zero. Furthermore, studies addressing the criterion-related validity of the Chair SR test for estimating lumbar extensibility were not found. Finally, since none of the three criteria were met in the seven SR test protocols, moderator analyses were not conducted for lumbar extensibility.

Moderator analyses

Table 3 reports the results of moderator analyses to examine the effects of the sex of the participants (i.e., male and female), the age of participants (i.e., children and adults), and the level of hamstring extensibility (i.e., low average level, $< 80°$, and high average level, $\geq 80°$) on overall criterion-related validity correlation coefficients for estimating hamstring extensibility for each SR test protocol potentially affected by moderator effects (i.e., the Classic SR, Modified SR, Back-saver SR, Unilateral SR, and Chair SR). Collectively, slight differences in r_p values were detected in different categories of included moderators across the analyzed SR tests.

Gender of participants: The results showed that all SR test protocols had a moderate-to-high mean correlation coefficient of criterion-related validity for estimating hamstring extensibility for males (r_p range = 0.55-0.83) and moderate for females (r_p range = 0.41-0.70) in which all 95% CI did not include the value zero. There was a tendency of the mean correlation coefficient being slightly greater for females than for males on each SR test, except for the Chair SR test where the opposite results were found. However, we have to be aware that, except for the Chair SR test, all the 95% CI of mean correlation coefficients were overlapped. Moreover, we should also be cautious because the low numbers of r values over the analyses were supported. Additionally, according to moderator analysis criteria, at least one of

Table 2. Results of meta-analyses for overall criterion-related validity correlation coefficients across sit-and-reach test protocols.

Sit-and-reach test	K	n	N	r_c	r_p	95% CI[a]	95% CV[b]	% of variance[c]	Q statistic
Hamstring extensibility									
Classic sit-and-reach	28	47	3,481	.65	.67	.55, .80	.44, .91	22.55	208.39*
Modified sit-and-reach	9	16	1,058	.54	.56	.39, .73	.32, .80	33.14	48.28*
Back-saver sit-and-reach	10	18	1,158	.57	.59	.43, .75	.38, .80	36.65	49.12*
Modified back-saver sit-and-reach	2	3	213	.44	.46	.28, .65	.46, .46	100.00	.18
V sit-and-reach	3	5	411	.56	.60	.46, .74	.60, .60	100.00	3.23
Modified V sit-and-reach	2	3	213	.55	.59	.44, .74	.59, .59	100.00	1.14
Unilateral sit-and-reach	2	4	378	.61	.64	.52, .76	.51, .76	47.43	8.43*
Chair sit-and-reach	2	3	182	.45	.49	.29, .68	-.11, 1.00	9.50	31.57*
Lumbar extensibility									
Classic sit-and-reach	13	21	1,762	.25	.26	.05, .46	.19, .32	91.52	22.95
Modified sit-and-reach	5	8	484	.26	.26	.03, .50	.26, .26	100.00	1.39
Back-saver sit-and-reach	5	9	510	.15	.16	-.10, .41	.16, .16	100.00	4.72
Modified back-saver sit-and-reach	2	3	213	.20	.21	-.01, .44	.21, .21	100.00	.07
V sit-and-reach	3	5	411	.30	.31	.11, .51	.31, .31	100.00	2.42
Modified V sit-and-reach	2	3	213	.30	.32	.11, .53	.32, .32	100.00	.63
Unilateral sit-and-reach	1	2	158	.34	.35	.15, .54	.26, .43	83.73	2.39
Chair sit-and-reach	-	-	-	-	-	-	-	-	-

Note. K, number of studies; n, number of rs; N, total sample size; r_c, overall weighted mean of r corrected for sampling error only; r_p, overall weighted mean of r corrected for sampling error and measurement error; [a] 95% confidence interval; [b] 95% credibly interval [c] Percentage of variance accounted for by statistical artefacts including sampling error and measurement error of sit-and-reach tests. * $p < 0.05$

Table 3. Results of moderator analyses for criterion-related validity correlation coefficients for estimating hamstring extensibility across all sit-and-reach test protocols potentially affected by moderator effects†

Moderator	Effect	K	n	N	r_c	r_p	95% CI[a]	95% CV[b]	% of variance[c]	Q statistic
Gender of participants										
Classic sit-and-reach	Males	19	21	1,493	.62	.64	.50, .78	.46, .82	38.28	54.86*
	Females	18	20	1,361	.68	.70	.58, .82	.49, .91	25.06	79.80*
Modified sit-and-reach	Males	7	7	469	.53	.55	.38, .71	.26, .83	26.80	59.70*
	Females	6	6	447	.60	.62	.48, .76	.53, .72	68.72	23.28*
Back-saver sit-and-reach	Males	8	8	490	.57	.59	.42, .75	.48, .69	71.70	11.16
	Females	9	9	613	.58	.60	.45, .75	.33, .87	24.05	37.42*
Unilateral sit-and-reach	Males	2	2	182	.59	.61	.48, .74	.61, .61	100.00	.72
	Females	2	2	196	.63	.66	.55, .77	.47, .85	25.66	7.80*
Chair sit-and-reach	Males	1	1	32	.76	.83	.71, .94	.83, .83	100.00	.00
	Females	2	2	150	.39	.41	.22, .60	-.16, .99	9.79	20.42
Age of participants										
Classic sit-and-reach	Children	8	14	1,173	.66	.67	.55, .79	.47, .87	26.28	53.27*
	Adults	20	33	2,308	.64	.68	.55, .80	.43, .92	21.22	155.50*
Modified sit-and-reach	Children	1	2	87	.31	.32	.05, .59	.32, .32	100.00	1.34
	Adults	8	14	971	.56	.58	.42, .74	.37, .79	35.13	45.55*
Back-saver sit-and-reach	Children	2	4	257	.58	.59	.43, .75	.45, .74	55.88	7.16
	Adults	8	14	901	.56	.59	.43, .75	.36, .82	33.38	41.94*
Unilateral sit-and-reach	Children	-	-	-	-	-	-	-	-	-
	Adults	2	4	378	.61	.64	.52, .76	.51, .76	47.43	8.43*
Chair sit-and-reach	Children	-	-	-	-	-	-		-	-
	Adults	2	3	182	.45	.49	.29, .68	-.11, 1.00	9.50	31.57*
Level of hamstring extensibility										
Classic sit-and-reach	Low	15	16	1,129	.60	.63	.48, .77	.41, .84	30.79	51.97*
	High	19	25	1,984	.67	.70	.59, .81	.46, .94	18.39	135.91*
Modified sit-and-reach	Low	4	5	355	.51	.53	.36, .70	.24, .82	24.87	64.35*
	High	7	10	648	.55	.58	.41, .74	.36, .79	37.42	42.75*
Back-saver sit-and-reach	Low	5	6	433	.54	.57	.41, .72	.31, .82	27.37	21.92*
	High	7	11	670	.59	.61	.45, .77	.44, .79	45.34	24.26*
Unilateral sit-and-reach	Low	1	1	120	.56	.58	.47, .70	.58, .58	100.00	.00
	High	2	3	258	.63	.66	.55, .78	.51, .81	38.22	7.85*
Chair sit-and-reach	Low	2	2	134	.33	.35	.14, .56	-.13, .83	16.42	12.18*
	High	1	1	48	.81	.86	.79, .94	.86, .86	100.00	.00

Note. K, number of studies; n, number of rs; N, total sample size; r_c, overall weighted mean of r corrected for sampling error only; r_p, overall weighted mean of r corrected for sampling error and measurement error; [a] 95% confidence interval; [b] 95% credibly interval [c] Percentage of variance accounted for by statistical artefacts including sampling error and measurement error of sit-and-reach tests. † Because some studies mixed genders or hamstring extensibility levels were missed, the overall n for these categories is lower for some sit-and-reach tests. * $p < 0.05$

the three criteria was met in the SR test protocols (except for the Unilateral SR and Chair SR for males, because logically these had only two and one r values, respectively), indicating that the criterion-related validity of these SR tests separately for sex was still heterogeneous. Finally, because some studies grouped males and females together, in Table 3 overall n of the sex of participants is lower for some SR test protocols.

Age of participants: The results showed that all SR test protocols had a low-to-moderate mean correlation coefficient of criterion-related validity for estimating hamstring extensibility for children (r_p range = 0.32-0.67) and moderate for adults (r_p range = 0.49-0.68) in which all 95% CI did not include the value zero. Of all the examined SR test protocols, only studies of the Classic SR, Modified SR, and Back-saver SR tests were found for both children and adults. The results of the present meta-analysis showed that there was a trend in the mean correlation coefficient reported to be greater for adults than for children in the Classic SR and Modified SR, but not in the Back-saver SR test where the r average values were equal. However, in any case, all 95% CI of mean correlation coefficients were overlapped. Furthermore, we should also be cautious because the low numbers of r

values over the analyses were supported. Finally, according to moderator analysis criteria, at least one of the three criteria was met in most SR test protocols (except for the Modified SR for children, because logically these had only two r values), indicating that the criterion-related validity of these SR tests separately for age were still heterogeneous.

Level of hamstring extensibility: The results showed that all SR test protocols had a low-to-moderate mean correlation coefficient of criterion-related validity for participants with low level of hamstring extensibility (< 80º in the average score of the straight leg raise test) (r_p range = 0.35-0.63) and moderate-to-high for participants with a high level of hamstring extensibility (≥ 80º in the average score of the straight leg raise test) (r_p range = 0.58-0.86) in which all 95% CI did not include the value zero. For all examined SR test protocols, there was a trend of the mean correlation coefficient to being greater for participants with high levels of hamstring extensibility than for those with low levels.

However, we have to be aware that, except for the Chair SR test, all the 95% CI of mean correlation coefficients were overlapped, as well as the low numbers of r values over the analyses were supported. Additionally,

according to moderator analysis criteria, at least one of the three criteria was met in all SR test protocols (except for the Unilateral SR for low levels and Chair SR for high levels because logically these had only one *r* value), indicating that the criterion-related validity of these SR tests separately for level of hamstring extensibility were still heterogeneous. Finally, because several studies failed to identify the level of hamstring extensibility or were ambiguous (i.e., hamstring extensibility scores around 80º shown graphically), in Table 3 overall *n* of level of hamstring extensibility is lower for some SR tests.

Discussion

From its conception, the Classic SR test has been subjected to numerous modifications, often with the aim of improving its validity. However, according to the results of the present meta-analysis, and although we are aware that all the 95% CI of mean correlation coefficients were overlapped, the Classic SR test showed a greater average criterion-related validity coefficient. Hence, if our purpose is to assess hamstring extensibility, it seems that the use of a modification of the classic protocol is not justified.

Specifically, it has been suggested for several years that the Classic SR test did not consider limb length differences (Hoeger et al., 1990). To solve this methodological "problem", Hoeger et al. (1990) proposed the Modified SR, which incorporates a finger-to-box distance to account for proportional differences between legs and arms. In this line, these authors found that adolescents with longer legs relative to arms had poorer performance on the Classic SR test, and the Modified SR negated the concern about disproportionate limb length bias by establishing a relative zero point for each person. Unfortunately, this study failed to address the very important issue of criterion-related validity.

The present meta-analysis showed a greater overall mean criterion-related validity for the Classic SR than for the Modified SR. In addition, for other modifications of SR tests that incorporated fingers-to-box distance (i.e., the Modified back-saver SR and Modified V SR), the average criterion-related validity coefficients were higher for the end scores version than for the modified one. In this line, in most primary studies in which the criterion-related validity of end and differences scores of SR tests was studied among the same sample, coefficients values were slightly greater for traditional protocols (e.g., Ayala et al., 2011; Castro-Piñero et al., 2009b; Lemmink et al., 2003; López-Miñarro et al., 2010a; López-Miñarro et al., 2010b).

Regarding the criterion-related validity for estimating lumbar extensibility, in addition to the low correlation coefficient found, we have to be aware that the Pearson´s zero-order correlation coefficient was considered; therefore, because of the common explanation for hamstring and lumbar extensibility, the "real" criterion-related validity values for estimating lumbar extensibility could be even lower.

Finally, in line with the results of the present meta-analysis, previous primary studies carried out with young adults (López-Miñarro and Rodríguez-García, 2010c) and elderly women (López-Miñarro et al., 2011) found that the level of hamstring extensibility influenced the criterion-related validity of the Classic SR and Toe touch tests. However, due to the fact that in the present meta-analysis the *n* was classified based on the average scores of the straight leg raise test, we were aware that several participants with low hamstring extensibility could be classified as high flexibility and vice versa. This fact could reduce drastically the difference reported in the results of the present meta-analysis.

Strengths and limitations

The meta-analysis is a useful tool to assess the scientific evidence, but an understanding of its strengths and limitations is needed for most appropriate use of this method (Flather et al., 1997). Overall, the main strength of a meta-analysis is that it lets us obtain more reliable population estimates of findings than those of the constituent studies. Therefore, the results of a meta-analysis let us generalize the research findings, as well as test hypotheses that may have never been tested in primary studies. Likewise, the meta-analysis represents the best up-to-date approach to describe and summarize the scientific findings of a research area (Hunter and Schmidt, 2004). Lastly, meta-analysis methods can advance in an entire discipline by addressing more general questions in the area (Cooper et al., 2009).

Regarding the strengths of the present meta-analysis, we followed several measures to avoid (or at least to reduce) publication bias. A lot of research studies fail to be published at all, while others are published only in abstract form, conference proceeding, or dissertation but not as scientific articles. Furthermore, research studies with favorable results are far more likely to be published than those with inconclusive results. Likewise, identification of relevant studies may also be difficult because of their publication in less accessible journals. Thus, performing a meta-analysis when a proportion of the relevant data is missing can provide misleading results, and publication bias may spuriously support a hypothesis by continuously selecting favorable results and rejecting unfavorable ones (Flather et al., 1997).

Therefore, to avoid availability bias, we conducted a wide literature search. The potential inclusion of all relevant single studies in the present meta-analysis (i.e., published and unpublished or English and non-English language) by extent and careful searching might clearly help reduce the impact of publication bias in the present meta-analysis. Hence, the inclusion of unpublished and non-English language studies in the literature search is an important strength of the present meta-analysis. Multiple publication bias also exists when the same researchers responsible for multiple publications report the same validity coefficients, derived from the same participants under the same experimental conditions. Thus, in the present meta-analysis all studies by the same authors were thoroughly cross-referenced with each other. Since some selected studies had fully or partially duplicated information, these particular correlations values were not ana-

lyzed in the meta-analyses. Lastly, several exploratory analyses were also conducted to detect the presence of publication bias.

Finally, the Hunter-Schmidt´s psychometric meta-analysis approach (2004) was conducted in the present study to obtain the population estimates of criterion-related validity of SR tests. Because sample sizes are never infinite and measures are never perfectly reliable, sampling error and measurement error are always present in all real data. The psychometric meta-analysis approach corrects the observed correlations due both to sampling error and measurement error. Thus, this method is probably one of the best approaches to estimate the population correlation coefficients.

On the other hand, there were some limitations that should be considered when examining the results of the present study. The main limitations of the present meta-analysis were related to the small number of criterion-related validity coefficients found. Firstly, estimating the population parameters based on small samples is simply less accurate than in a large-sized meta-analysis. Secondly, a partially hierarchical breakdown had to be used. The main problem in this kind of analysis is that it might produce quite misleading results due to confounding and interaction effects. We are aware that a fully hierarchical moderator analysis approach may be a more appropriate method to resolve this problem. However, more correlations coefficients would be needed for each level of moderators. For these reasons, the results of the present study should be considered with caution, especially for those SR test protocols from which only a few studies were retrieved. Firmer conclusions should await the accumulation of a larger number of studies (Hunter and Schmidt, 2004).

Another limitation of the present meta-analysis is related to the criterion measures used in the included studies. Joint(s) range of motion measured through radiography seems to be the best criterion measurement to assess flexibility (Gajdosik and Bohannon, 1987), but due to several practical reasons such us high cost, necessity of sophisticated instruments, qualified technicians, or time constraints, the use of this method is limited (Castro-Piñero et al., 2009b). On the other hand, goniometers are relatively easy to obtain, valid and highly accurate instruments to measure joint range of motion; therefore, joint(s) range of motion measured through goniometers has been widely considered a valid and suitable criterion measure of hamstring extensibility (e.g., Ayala et al., 2011; Hartman and Looney, 2003; López-Miñarro and Rodríguez-García, 2010c). In this line, all the previous studies found considered the angular tests measured by goniometers as the criterion measures. However, nowadays some studies have suggested that the criterion measures of hamstring extensibility must be reexamined and readjusted (Cardoso et al., 2007; Hartman and Looney, 2003). Similarly, although none of the previous studies has used radiography as the criterion measure of lumbar extensibility, they administered tests with a demonstrated high reliability and validity (Macrae and Wright, 1969; Williams et al., 1993).

Another area of concern is that moderator analyses

showed that there was still a large amount of unexplained variance after controlling for artefacts and predefined moderators. Studies included in a meta-analysis are expected to vary in a number of ways. Thus, beyond the sampling error and other statistical artefacts, differences between studies (e.g., sample, study design, or tests procedure) undoubtedly affect these results. For example, the straight raise leg test can be measured by different kinds of movements (i.e., active or passive), instruments (e.g., radiography, goniometer or inclinometer), number of researchers, number of repetitions, time of rest between repetitions, and criteria of maximum extensibility. Additionally, in the present meta-analysis different criterion measures were used to estimate the lumbar extensibility. This statistical heterogeneity can be quantified, but there is usually uncertainty about how important the differences really are. Thus, quantifying and accounting for differences between component studies in a meta-analysis remains a substantial methodological problem and a continuing source of debate (Flather et al., 1997).

Finally, coding some study features was problematic due to different reasons. The moderator analysis had missing data in sex categories because some authors mixed males with females in their studies. Hamstring extensibility also had missing data because several authors failed to identify it or it was ambiguous. In addition, because in the present meta-analysis the hamstring extensibility was classified based on the average scores, we are aware that several participants with low hamstring extensibility could be classified as high flexibility and vice versa. Lastly, although participant characteristics such as physical activity levels or sports practice were potentially moderating features, coding for them was not possible because most studies did not identify them.

Conclusion

Overall the SR tests have a moderate mean correlation coefficient of criterion-related validity for estimating hamstring extensibility, but they have a low mean criterion-related validity for estimating lumbar extensibility. The Classic SR test shows the greater average criterion-related validity for estimating hamstring extensibility. The results of the present meta-analysis suggest that the end scores of the classic versions of the SR tests (e.g., the Classic SR) are a better indicator of hamstring extensibility than the modifications that incorporate the fingers-to-box distance (e.g., the Modified SR). Regarding the three potential moderators examined (sex of participants, age of participants, and level of hamstring extensibility), generally females, adults, and participants with high levels of hamstring extensibility tended to have greater mean values of criterion-related validity for estimating hamstring extensibility. However, due to the low number of r values found, the fact that almost all the 95% CI of mean correlation coefficients were overlapped, and that criterion-related validity of SR tests within each category was still heterogeneous, we should be cautious with the results of the present meta-analysis.

Therefore, when angular tests such as the straight leg raise or knee extension tests cannot be used, the SR

tests seem to be a useful alternative to estimate hamstring extensibility; however, to assess lumbar extensibility other widely used tests such as the Macrae & Wright or Single/Double inclinometer methods should be used. Nevertheless, as in the application of any field fitness test, evaluators must be aware that the results of SR tests are simply an estimation and, therefore, not a direct measure of the hamstring extensibility. On the other hand, when there are a higher number of studies accumulated, a large-sized meta-analysis with a fully hierarchical analysis approach should be carried out. Future research should further study the criterion-related validity of SR tests, especially in modifications of SR tests such as the SR with plantar flexion, among populations such as children or athletes, and go deeply into other related aspects such as the level of hamstring extensibility.

Acknowledgments

References

Ayala, F., Sainz de Baranda, P., De Ste Croix, M. and Santonja, F. (2011) Criterion-related validity of four clinical tests used to measure hamstring flexibility in professional futsal players. *Physical Therapy in Sport* **12,** 175-181.

Ayala, F., Sainz de Baranda, P., De Ste Croix, M. and Santonja, F. (2012) Reproducibility and criterion-related validity of the sit and reach test and toe touch test for estimating hamstring flexibility in recreationally active young adults. *Physical Therapy in Sport* **13,** 219-226.

Baker, A.A. (1985) The relative contribution of flexibility of the back and hamstring muscles in the performance of the sit and reach component of the AAHPERD health related fitness test in girls thirteen to fifteen years of age. Doctoral thesis, University of North Texas, United States.

Baltaci, G., Tunay, V., Besler, A. and Gerçeker, S. (2003) Comparison of three different sit and reach tests for measurement of hamstring flexibility in female university students. *British Journal of Sports Medicine* **37,** 59-61.

Begg, C.B. and Mazumdar, M. (1994) Operating characteristics of a rank order correlation for publication bias. *Biometrics* **50,** 1088-1101.

Benito Peinado, P.J., Díaz Molina, V., Calderón Montero, F.J., Peinado Lozano, A.B., Martín Caro, C., Álvarez Sánchez, M. and Pérez Tejero, J. (2007) Literature review in exercise physiology: Practical recommendations. *Revista Internacional de Ciencias del Deporte* **6,** 1-11.

Biering-Sorensen, F. (1984) Physical measurements as risk indicator for low-back trouble over a one year period. *Spine* **9,** 106-119.

Book, C. B. (1989) *Validation of the sit-and-reach test*. Doctoral thesis, University of Minnesota, United States.

Bozic, P. R., Pazin, N., Berjan, B. B., Planic, N. M. and Cuk, I. D. (2010) Evaluation of the field tests of flexibility of the lower extremity: Reliability, and the concurrent and factorial validity. *Journal of Strength and Conditioning Research* **24,** 2523-2531.

Cardoso, J.R., Azevedo, N.C.T., Cassano, C.S., Kawano, M.M. and Âmbar G. (2007) Intra and interobserver reliability of angular kinematic analysis of the hip joint during the sit-and-reach test to measure hamstring length in university students. *Revista Brasileña de Fisioterapia* **11,** 133-138.

Castro-Piñero, J., Artero, E.G., España-Romero, V., Ortega, F.B., Sjöström, M. and Ruiz, J.R. (2009a) Criterion-related validity of field-based fitness tests in youth: A systematic review. *British Journal of Sports Medicine* **44,** 934-943.

Castro-Piñero, J., Chillón, P., Ortega, F.B., Montesinos, J.L., Sjöström, M. and Ruiz, J.R. (2009b) Criterion-related validity of sit-and-reach and modified sit-and-reach test for estimating hamstring flexibility in children and adolescents aged 6-17 years. *International Journal of Sports Medicine* **30,** 658-662.

Chung, P. K. and Yuen, C. K. (1999) Criterion-related validity of sit-and-reach test in university men in Hong Kong. *Perceptual and Motor Skills* **88,** 304-316.

Cohen, J.A. (1992) Power primer. *Psychological Bulletin* **112,** 155-159.

Cooper, H., Hedges, L.V. and Valentine, J.C. (2009) *The handbook of research synthesis and meta-analysis*. 2nd edition. Sage, New York.

Da Silva Díaz, R. and Gómez-Conesa, A. (2008) Shortened hamstring syndrome. *Fisioterapia* **30,** 186-193.

Davis, D.S., Quinn, R.O., Whiteman, C.T., Williams, J.D. and Young, C.R. (2008) Concurrent validity of four clinical tests used to measure hamstring flexibility. *Journal of Strength and Conditioning Research* **22,** 583-588.

Erkula, G., Demirkan, F., Kilic, B.A. and Kiter, E. (2002) Hamstring shortening in healthy adults. *Journal of Back and Musculoskeletal Rehabilitation* **16,** 77-81.

Esola, M.A., McClure, P.W., Fitzgerald, G.K. and Siegler, S. (1996) Analysis of lumbar spine and hip motion during forward bending in subjects with and without a history of low back pain. *Spine* **21,** 71-78.

Fisk, J.W., Baigent, M.L. and Hill, P.D. (1984) Scheuermann´s disease. Clinical and radiological survey of 17 and 18 year olds. *American Journal of Physical Medicine* **63,** 18-30.

Flather, M.D., Farkouh, M.E., Pogue, J.M. and Yusuf, S. (1997) Strengths and limitations of meta-analysis: Larger studies may be more reliable. *Controlled Clinical Trials* **18,** 568-579.

Gajdosik, R.L. and Bohannon, R.W. (1987) Clinical measurement of range of motion: Review of goniometry emphasizing reliability and validity. *Physical Therapy* **67,** 1867-1872.

García, S.C. (1995) *Validity of the sit-and-reach test for male and female adolescents*. Doctoral thesis, University of Oregon, United States.

Hartman, J.G. and Looney, M. (2003) Norm-referenced and criterion-referenced reliability and validity of the back-saver sit-and-reach. *Measurement in Physical Education and Exercise Science* **7,** 71-87.

Harvey, J. and Tanner, S. (1991) Low back pain in young athletes: A practical approach. *Sports Medicine* **12,** 394-406.

Hoeger, W.W., Hopkins, D.R., Button, S. and Palmer, T.A. (1990) Comparing the sit and reach with the modified sit and reach in measuring flexibility in adolescents. *Pediatric Exercise Science* **2,** 156-162.

Holt, L.E., Pelma, T.W. and Burke, D.G. (1999) Modifications to the standard sit-and-reach flexibility protocol. *Journal of Athletic Training* **34,** 43-47.

Hui, S.S. and Yuen, P.Y. (2000) Validity of the modified back-saver sit-and-reach test: A comparison with other protocols. *Medicine & Science in Sports & Exercise* **32,** 1655-1659.

Hui, S.C., Yuen, P.Y., Morrow, J.R. and Jackson, A.W. (1999) Comparison of the criterion-related validity of sit-and-reach tests with and without limb length adjustment in Asian adults. *Research Quarterly for Exercise and Sport* **70,** 401-406.

Hunter, J.E. and Schmidt, F.L. (2004) *Methods of meta-analysis: Correcting error and bias in research findings*. 2nd edition. Sage, Newbury Park.

Jackson, A.W. and Baker, A.A. (1986) The relationship of the sit and reach test to criterion measures of hamstring and back flexibility in young females. *Research Quarterly for Exercise and Sport* **57,** 183-186.

Jackson, A. and Langford, N.J. (1989) The criterion-related validity of the sit and reach test: Replication and extension of previous findings. *Research Quarterly for Exercise and Sport* **60,** 384-387.

Jones, C. J., Rikli, R. E., Max, J. and Noffal, G. (1998) The reliability and validity of a chair sit-and-reach test as a measure of hamstring flexibility in older adult. *Research Quarterly for Exercise and Sport* **69,** 338-343.

Kanbur, N. O., Düzgün, I., Derman, O. and Baltaci, G. (2005) Do sexual maturation stages affect flexibility in adolescent boys aged 14 years? *The Journal of Sports Medicine and Physical Fitness* **45,** 53-57.

Kawano, M.M., Ambar, G., Oliveira, B.I.R., Boer, M.C., Cardoso, A.P. R.G. and Cardoso, J.R. (2010) Influence of the gastrocnemios muscle on the sit-and-reach test assessed by angular kinematic analysis. *Revista Brasileña de Fisioterapia* **14,** 10-15.

Kendall, F.P., McCreary, E.K., Provance, P.G., Rodgers, M.M. and Romani, W.A. (2005) *Muscles: Testing and function with pos-*

ture and pain. 5th edition. Lippincott, Williams, & Wilkins, Baltimore.

Langford, N.J. (1987) The relationship of the sit and reach test to criterion measures of hamstring and back flexibility in adult males and females. Doctoral thesis, North Texas State University, United States.

Lemmink, K.A.P.M., Kemper, H.C.G., de Greef, M.H.G., Rispens, P. and Stevens, M. (2003) The validity of the sit-and-reach test and the modified sit-and-reach test in middle-aged to older men and women. *Research Quarterly for Exercise and Sport* **74,** 331-336.

Liemohn, W., Martin, S.B. and Pariser, G.L. (1997) The effect of ankle posture on sit-and-reach test performance. *Journal of Strength and Conditioning Research* **11,** 239-241.

Liemohn, W., Sharpe, G. L. and Wasserman, J. F. (1994) Criterion related validity of the sit-and-reach test. *Journal of Strength and Conditioning Research* **8,** 91-94.

Light, R.J. and Pillemer, D.B. (1984) *Summing up: The science of reviewing research*. Harvard University Press, Cambridge.

Lipsey, M.W. and Wilson, D.B. (2001) *Practical meta-analysis*. Sage, Newbury Park.

López Miñarro, P.A., Ferragut Fiol, C., Alacid Cárceles, F., Yuste Lucas, J.L. and García Ibarra, A. (2008a) Validity of sit-and-reach and toe-touch tests in the evaluation of hamstring muscle length in young paddlers. *Apunts* **43,** 24-29.

López Miñarro, P.A., Rodríguez García, P.L., Yuste, J.L., Alacid, F., Ferragut, C. and García Ibarra, A. (2008b) Validity of the lumbo-sacral position in bending as measure of hamstring muscle extensibility on young athletes. *Archivos de Medicina del Deporte* **25,** 103-110.

López Miñarro, P.A., Sainz de Baranda Andújar, P., Yuste Lucas, J.L. and Rodríguez García, P.L. (2008c) Validity of the unilateral sit-and-reach test as measure of hamstring muscle extensibility. Comparison with other protocols. *Cultura, Ciencia y Deporte* **3,** 87-92.

López-Miñarro, P.A. (2010) Criterion-related validity of the lumbo-horizontal angle in flexion as a measure of hamstring muscle extensibility in young adults. *Cultura, Ciencia y Deporte* **5,** 25-31.

López-Miñarro, P.A. and Alacid, F. (2009) Influence of hamstring muscle extensibility on spinal curvatures in young athletes. *Science & Sports* **25,** 188-193.

López-Miñarro, P.A., Alacid, F., Muyor, J.M. and López, F.J. (2010a) Criterion-related validity of the modified sit-and-reach test as hamstring muscle extensibility measure in young adults. *Kronos* **9,** 39-46.

López-Miñarro, P.A., García Ibarra, A. and Rodríguez García, P.L. (2010b) Comparison between sit-and-reach tests for measuring hamstring muscle extensibility. *Apunts* **99,** 56-64.

López-Miñarro, P.A., Muyor, J.M. and Alacid, F. (2011) Validity of sit-and-reach tests as measures of hamstring extensibility in older women. *Revista Internacional de Medicina y Ciencias de la Actividad Física y el Deporte* **11,** 564-572.

López-Miñarro, P.A. and Rodríguez-García, P.L. (2010c) Hamstring muscle extensibility influences the criterion-related validity of sit-and-reach toe-touch tests. *Journal of Strength & Conditioning Research* **24,** 1013-1018.

López-Miñarro, P. A., Sainz de Baranda Andújar, P. and Rodríguez-García, P. L. (2009) A comparison of the sit-and-reach test and the back-saver sit-and-reach test in university students. *Journal of Sports Science and Medicine* **8,** 116-122.

López-Miñarro, P. A., Sáinz de Baranda, P., Rodríguez-García, P. L. and Yuste, J. L. (2008) Comparison between sit-and-reach test and V sit-and-reach test in young adults. *Gazzetta Medica Italiana Archivio per le Scienze Mediche* **167,** 135-142.

López-Miñarro, P. A., Vaquero-Cristóbal, R., Muyor, J. M., Alacid, F. and Isorna, M. (2012) Criterion-related validity of the sit-and-reach test as measure of hamstring extensibility in paddlers. *Cultura, Ciencia y Deporte* **7,** 95-101.

Macrae, I.F. and Wright, V. (1969) Measurement of back movement. *Annals of Rheumatic Disease* **28,** 584-589.

McHugh, M.P., Kremenic, I.J., Fox, M. B. and Gleim, G.W. (1998) The role of mechanical and neural restraints to joint range of motion during passive stretch. *Medicine and Science in Sports and Exercise* **30,** 928-932.

Mier, C.M. (2011) Accuracy and feasibility of video analysis for assess-ing hamstring flexibility and validity of the sit-and-reach test. *Research Quarterly for Exercise and Sport* **82,** 617-623.

Mierau, D., Cassidy, J.D. and Yong-Hing, K. (1989) Low-back pain and straight in children and adolescents. *Spine* **14,** 526-528.

Minkler, S. and Patterson, P. (1994) The validity of the modified sit-and-reach test in college-age students. *Research Quarterly for Exercise and Sport* **65,** 189-192.

Miyazaki, J., Murata, S., Horie, J. and Suzuki, S. (2010) Relationship between the sit-and-reach distance and spinal mobility and straight leg raising range. *Rigakuryoho Kagaku* **25,** 683-686.

Orloff, H. A. (1988) *Standardization, reliability and validation of flexibility field testing protocol*s. Doctoral thesis, University of Kansas, United States.

Patterson, P., Wiksten, D. L., Ray, L., Flanders, C. and Sanphy, D. (1996) The validity and reliability of the back saver sit-and-reach test in middle school girls and boys. *Research Quarterly for Exercise and Sport* **67,** 448-451.

Rodríguez-García, P.L., López-Miñarro, P.A., Yuste, J.L. and Sáinz de Baranda, P. (2008) Comparison of hamstring criterion-related validity, sagittal spinal curvatures, pelvic tilt and score between sit-and-reach and toe-touch tests in athletes. *Medicina dello Sport* **61,** 11-20.

Rosenthal, R. (1979) The "file drawer problem" and tolerance for null results. *Psychological Bulletin* **33,** 1005-1008.

Santonja Medina, F., Ferrer López, V. and González-Moro, I.M. (1995) Clinical examination of hasmtring tightness. *Selección* **4,** 81-91.

Simoneau, G. G. (1998) The impact of various anthropometric and flexibility measurements on the sit-and-reach test. *Journal of Strength and Conditioning Research* **12,** 232-237.

Smith, J.F. and Miller, C.V. (1985) The effect of head position on sit-and-reach performance. *Research Quarterly for Exercise and Sport* **56,** 84-85.

Standaert, C. J. and Herring, S. A. (2000) Spondylolysis: A critical review. *British Journal of Sports Medicine* **34,** 415-422.

Williams, R., Binkley, J., Bloch, R., Goldsmith, C.H. and Minuk, T. (1993) Reliability of the Modified-modified Schöber and Double inclinometer methods for measuring lumbar flexion and extension. *Physical Therapy* **73,** 26-37.

Yuen, P. Y., and Hui, S. C. (1998) Are difference scores better predictor of flexibility than end scores in sit-and-reach tests? *Medicine & Science in Sports & Exercise* **30,** S125.

Key points

- Overall sit-and-reach tests have a moderate mean criterion-related validity for estimating hamstring extensibility, but they have a low mean validity for estimating lumbar extensibility.
- Among all the sit-and-reach test protocols, the Classic sit-and-reach test seems to be the best option to estimate hamstring extensibility.
- End scores (e.g., the Classic sit-and-reach test) are a better indicator of hamstring extensibility than the modifications that incorporate fingers-to-box distance (e.g., the Modified sit-and-reach test).
- When angular tests such as straight leg raise or knee extension tests cannot be used, sit-and-reach tests seem to be a useful field test alternative to estimate hamstring extensibility, but not to estimate lumbar extensibility.

AUTHORS BIOGRAPHY

Daniel MAYORGA-VEGA
Employment
Research fellow, Department of Physical Education and Sport, University of Granada, Spain.
Degree
PhD student
Research interests
Measurement and evaluation, health-related physical fitness, physical education-based interventions, motivation toward physical activity.
E-mail: dmayorgavega@gmail.com

Rafael MERINO-MARBAN
Employment
Tenured Professor, Department of Didactics of Musical, Plastic and Corporal Expression, University of Malaga, Spain
Degree
PhD
Research interests
Health-related physical fitness, physical education-based interventions, kinesio tape.
E-mail: rmerino@uma.es

Jesús VICIANA
Employment
Tenured Professor, Department of Physical Education and Sport, University of Granada, Spain.
Degree
PhD
Research interests
Planning in physical education, physical education-based interventions, motivation toward physical activity.
E-mail: jviciana@ugr.es

✉ **Daniel Mayorga-Vega**
Department of Physical Education and Sport, University of Granada, Spain