Journal of Human Sport and Exercise
E-ISSN: 1988-5202
jhse@ua.es
Universidad de Alicante
España

Mayorga Vega, Daniel; Viciana, Jesús; Cocca, Armando; Merino Marban, Rafael
Criterion-related validity of toe-touch test for estimating hamstring extensibility: A metaanalysis
Journal of Human Sport and Exercise, vol. 9, núm. 1, enero-marzo, 2014, pp. 188-200
Universidad de Alicante
Alicante, España

Original Article

Criterion-related validity of toe-touch test for estimating hamstring extensibility: A meta-analysis

DANIEL MAYORGA-VEGA[1], JESÚS VICIANA[1], ARMANDO COCCA[1], RAFAEL MERINO-MARBAN[2]

[1] *Department of Physical Education and Sport, Faculty of Sciences of Physical Activity and Sport, University of Granada, Spain*

[2] *Department of Didactics of Musical, Plastic and Body Expression, Faculty of Education Sciences, University of Malaga, Spain*

ABSTRACT

Mayorga-Vega, D., Viciana, J., Cocca, A., & Merino-Marban, R. (2014). Criterion-related validity of toe-touch test for estimating hamstring extensibility: A meta-analysis *J. Hum. Sport Exerc., 9*(1), pp. 188-200. The main purpose of the present meta-analysis was to examine the scientific literature on the criterion-related validity of the toe-touch test for estimating hamstring extensibility. For this purpose relevant studies were searched from five electronic databases dated up through September 2012. Primary outcomes of criterion-related validity were Pearson´s zero-order correlation coefficients (*r*) between the toe-touch test and hamstring extensibility criterion measure. Then, from the included studies, the Hunter-Schmidt´s psychometric meta-analysis approach was conducted to estimate population criterion-related validity of the toe-touch test. Firstly, the corrected correlation mean (r_p), unaffected by sampling error and measurement error, was calculated. 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 six studies included in the present meta-analysis, 12 correlations values were retrieved. The overall results showed that the toe-touch test have a moderate mean criterion-related validity for estimating hamstring extensibility (r_p = 0.66, 0.54-0.79). Generally, females, children and individuals with high levels of hamstring extensibility seem 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% CIs of mean correlation coefficients were overlapped, and that criterion-related validity of the toe-touch test within each category was still heterogeneous, we should be cautious with the results of the present meta-analysis. When the use of the angular tests is limited, the toe-touch test seems to be a useful alternative to estimate hamstring extensibility. **Key words**: CONCURRENT VALIDITY, RANGE OF MOTION, FLEXIBILITY, STAND AND REACH TEST, LINEAL TEST, FIELD-BASED PHYSICAL FITNESS TEST, RESEARCH SYNTHESIS, SYSTEMATIC REVIEW.

Corresponding author. Facultad de Ciencias de la Actividad Física y el Deporte, Crta. de Alfacar, s/n, 18011, Granada, Spain
E-mail: dmayorgavega@gmail.com
Submitted for publication September 2012
Accepted for publication September 2013
JOURNAL OF HUMAN SPORT & EXERCISE ISSN 1988-5202

doi:10.4100/jhse.2014.91.18

INTRODUCTION

Hamstring extensibility is a well-recognized health-related physical fitness marker that plays an important role in protecting the spine from possible risks and, therefore, allowing people to execute the normal daily living activities and social functioning (Roth-Isigkeit et al., 2005; Sato et al., 2008). For example, reduced hamstring extensibility conditions several spine disorders such as thoracic hyperkyphosis (Fisk et al., 1984), spondylolysis (Standaert & Herring, 2000), disc herniation (Harvey & Tanner, 1991), changes in lumbopelvic rhythm (López-Miñarro & Alacid, 2009) and low back pain (Sjölie, 2004). 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).

Currently there are different kinds of tests in order to examine the levels of hamstring extensibility. On one hand, the angular tests that specifically measure hip flexion with knee extended (straight leg raise test) (American Academy of Orthopaedic Surgeons, 1996), or the knee extension range with the hip flexed to 90 degrees (knee extension test) (Hartman & Looney, 2003), have been considered as the criterion measures of hamstring extensibility. However, due to the necessity of relatively sophisticated instruments, qualified technicians, and time constraints, the use of these angular tests seem to be limited in several settings (Castro-Piñero et al., 2009b). On the other hand, the lineal tests are characterized by having a simple procedure, being easy to administer, requiring minimal skills training for their application and a very affordable equipment to perform them (Castro-Piñero et al., 2009b; López Miñarro et al., 2008c). Therefore, in contrast with the angular tests, the lineal tests allow the evaluation of a large number of people in a short space of time.

The classic sit-and-reach test (also called traditional or standard sit-and-reach), originally designed by Wells & Dillon (1952), is probably the most widely used lineal test of flexibility in exercise science laboratories, physical education classes, and commercial fitness centres (Cepero et al., 2011; Holt et al., 1999; Mirzaei et al., 2011). Two years later, Kraus & Hirschland (1954) designed the toe-touch (TT) test in which the individuals were assessed standing instead of sitting on the floor like in the classic sit-and-reach test. The most common assumption when interpreting the results of TT test is that individuals with better scores possess a higher level of hamstring extensibility than those with lower scores (Muyor et al., 2012; Sainz de Baranda et al., 2006).

Nevertheless, the primary studies examining the criterion-related validity of the TT test for estimating hamstring extensibility have shown inconclusive results (López-Miñarro et al., 2010c; Rodríguez-García et al., 2008). Each primary study that is published about criterion-related validity of the TT test only constitutes as a single piece of a constantly growing body of evidence (Cooper et al., 2009). To clarify the often conflicting results found in the scientific literature, researchers have to conduct meta-analyses (Cooper et al., 2009; Hunter & Schmidt, 2004; Lipsey & Wilson, 2001). Therefore, the meta-analyses remain a useful tool for the evaluation of evidence (Flather et al., 1997).

In this line, recently Mayorga-Vega et al. (2014) carried out a meta-analysis about the criterion-related validity of the sit-and-reach tests. Beyond the simple but important function of describing and summarizing the scientific findings of this research area, the main contribution of the above mentioned meta-analysis was to estimate as accurately as possible the population parameters. Therefore, the results of the Mayorga-Vega´s et al. (2014) meta-analysis let us generalize the research findings, as well as test hypotheses that may have never been tested in primary studies.

Unfortunately, to our knowledge there are no meta-analyses addressing the criterion-related validity of the TT test. Consequently, the main purpose of the present study was to examine the scientific literature on criterion-related validity of the TT test for estimating hamstring extensibility. More specifically, the objectives of this meta-analysis were: (a) to describe and summarize the up-to-date scientific findings of criterion-related validity of the TT test for estimating hamstring extensibility; (b) to estimate the overall population mean of the criterion-related validity coefficients of the TT test for estimating hamstring 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 the TT test.

MATERIAL AND METHODS

All the methodological procedure followed in the present study was based on the Mayorga-Vega´s et al. (2014) meta-analysis. See the methods section of the mentioned manuscript for more detailed information.

Search strategy

The following five electronic databases were searched from their inception through September 2012: SportDiscus, Scopus, Medline, Pubmed, and Web of Science. The search terms used were based on expressions related to the TT test (toe touch, stand and reach, finger to floor distance, fingertip floor distance). The terms of TT test were combined together with the Boolean operator "OR". Since the keywords consisted of more than one word they 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 the TT test for estimating hamstring extensibility were: (a) studies with apparently healthy participants who did not present any injury, physical and/or mental disabilities; (b) studies with the TT test that yielded the values of the maximum reach of the fingertips, and (c) studies in which hamstring extensibility criterion measurements used are widely accepted in the scientific literature (i.e., the angular tests straight leg raise or knee extension). In addition to papers, master/doctoral dissertations and conference proceedings were also accepted. No language or publication date restrictions were imposed.

Coding study features

For this meta-analysis, data were collected from studies that reported relationships between the TT test and a hamstring extensibility criterion measure 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), criterion-related validity result (Pearson´s *r* correlation coefficient value), reliability of the TT test (intraclass correlation coefficient value), reliability of the hamstring extensibility criterion measure (intraclass correlation coefficient value), and the level of hamstring extensibility (1 = low level of hamstring extensibility, < 80º in the straight leg raise test; 2 = high level of hamstring extensibility, < 80º in the straight leg raise test) (Kendall et al., 2005).

Since identification of the study features is usually explicitly stated in each of the primary articles, use of more than one rater was deemed unnecessary. For a study to be included in this meta-analysis, sample size, a widely accepted hamstring extensibility criterion measure and Pearson's *r* value were considered to be critical. In the event that the authors mixed subgroups of a study feature (e.g., males mixed with

females) or failed to identify a study feature (e.g., reliability scores) the data was omitted. When in the same study data were expressed for both legs separately, the average value of the coefficients was 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 the TT test, which represents the strength of association between the estimate of TT test and the hamstring extensibility criterion measure. If a single study reported more than on *r* value from different subsamples (e.g., males and females), we assumed each *r* value form different subsamples to be independent from each other and included them in a single meta-analysis (Lipsey & 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 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). According to Cohen´s guidelines (1992), the correlation coefficient was interpreted as small when $r < 0.30$. Secondly, according to Light and Pillemer´s graphic method (1984), the scatter plots of correlations coefficients against sample size for the TT test related to hamstring extensibility were analyzed. Finally, with the objective of quantifying the outcomes of the scatter plots, as suggested by Begg & Mazumdar (1994), a Spearman´s rank order correlation between *r* values and sample size was calculated.

Computation of correlations: The Hunter-Schmidt´s psychometric meta-analysis approach was conducted to obtain the population estimates of the criterion-related validity of the TT test (Hunter & 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 the TT 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 the TT and criterion measure tests 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 the TT test (r_p), three different criteria were simultaneously examined: (a) the 95% credibly interval (95% CV) is relatively large or includes the value zero; (b) the percentage of variance accounted for by statistical

artefacts is less than 75% of the observed variance in r_p; and (c) the *Q* homogeneity statistic is statistically significant ($p < 0.05$).

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 the TT 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 and high average level).

RESULTS AND DISCUSSION

Study description

Figure 1 shows a flow chart of the study selection process. Of the 473 literature search results, 17 potentially relevant publications were identified and retrieved for a more detailed evaluation. Finally, due to duplication issues, of the eight studies that met the inclusion criteria, only six studies were included in the present meta-analysis.

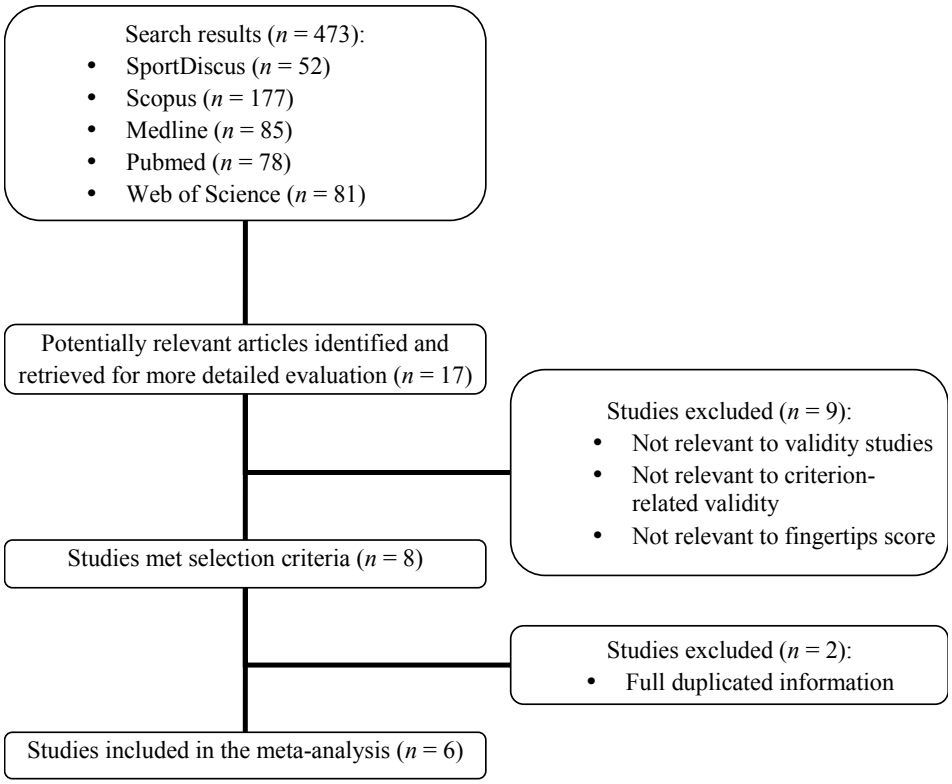

Figure 1. *Flow chart of studies selection process*

Table 1 presents a summary of the retrieved studies of criterion-related validity of the TT test for estimating hamstring extensibility. A total of 15 criterion-related validity coefficients (*r*) for the TT test were retrieved, ranging from low ($r = 0.25$) to high values ($r = 0.92$). A total sample of 1,307 participants (794 males and 513 females) was retrieved. The average age of participants ranged from 13.3 (± 0.6) to 65.3 (± 9.1) years old. Six studies examined the adults and two the children.

Table 1. *Summary of studies of criterion-related validity of the toe-touch test for estimating hamstring extensibility*

Reference	Sample	*n*	Age (years)	Criterion measure	Criterion-related validity (*r*)	
					♂	♀
Ayala et al. (2011)	Professional futsal players	♂=55 ♀=48	26.0 ± 4.5 23.0 ± 5.3	PSLR	0.25	0.92*
Ayala et al. (2012)	Recreationally active university students	♂=156 ♀=87	21.3 ± 2.5 20.7 ± 1.6	PSLR	0.70*	
López Miñarro et al. (2008a)	Canoeists	♂=44 ♀=22	13.3 ± 0.6	PSLR	0.73*-0.66*	0.78*-0.85*
López Miñarro et al. (2008b)	Canoeists	?=66	13.3 ± 0.6	PSLR	**0.73*-0.73***	
López-Miñarro et al. (2010b)	University students	♂=130 ♀=110	22.9 ± 3.2 23.2 ± 4.5	PSLR	**0.57*-0.62***	**0.72*-0.75***
López-Miñarro et al. (2011)	Older women: Low, moderate and high flexibility	♀=36	65.3 ± 9.1	PSLR		0.48*-0.46*
		♀=35				0.61*-0.59*
		♀=35				0.78*-0.76*
López-Miñarro and Rodríguez-García (2010c)	Recreationally active university students: Low and normal flexibility	♂=120	22.9 ± 3.6	PSLR	0.28*-0.40*	
		♂=120			0.60*-0.55*	
Rodríguez-García et al. (2008)	Fit sports activities practitioners	♂=125 ♀=118	22.9 ± 3.2 23.2 ± 4.5	PSLR	0.57*-0.62*	0.72*-0.75*

Note. This table includes all the studies that met selection criteria, however, full information of two studies was not included in the meta-analysis due to duplication issues (in bold); ♂, males; ♀, females; ?, information unavailable; PSLR, Passive straight leg raise test; 0.xx-0.xx, Pearson´s *r* for the left and right leg, respectively.
* Pearson´s *r* statistically significant at $p < 0.05$

Publication bias

Due to two studies having fully duplicated information, these *r* coefficients values were not analyzed in the present meta-analyses despite the fact that these studies met the selection criteria. 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, López-Miñarro´s et al. (2010b) study information (for both males and females) was not computed because the same data had been published previously in other journal (although the sample size was slightly different, it was considered a duplication issue because all the other information was equal) (Rodríguez-García et al., 2008). Pearson´s *r* correlation values of selected studies that were excluded for the meta-analysis are indicated (in bold) in Table 1.

Subsequently, several exploratory analyses were conducted to detect the presence of publication bias. Firstly, the file drawer analysis based on the effect size was calculated. The results of the file drawer analysis showed that 14 unlocated studies averaging null results (r = 0) would have to exist to bring the mean r_p down to 0.29. This is a large number of "lost" studies, especially if we aware of the percentage of unlocated/located studies (117%). Hence, we concluded that it was unlikely that there would be this particular number of "lost" studies.

Then, the Figure 2 shows the scatter plot of sample size against criterion-related validity coefficients of the TT test for estimating hamstring extensibility. Based on the statistical significance of the studies, if there would be publication bias in the scatter plot the small-sample studies reporting small r values would be disproportionately absent because they are studies that would fail to attain statistical significance. Therefore, according to this graphic method, the figure suggested that in the present study there was an absence of publication bias. Finally, in line with the graphic method, the results of Spearman´s rank order correlation between r values and sample size did not show a statistically significant correlation for estimating hamstring extensibility (r = - 0.35, p = 0.263). Similarly, 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.

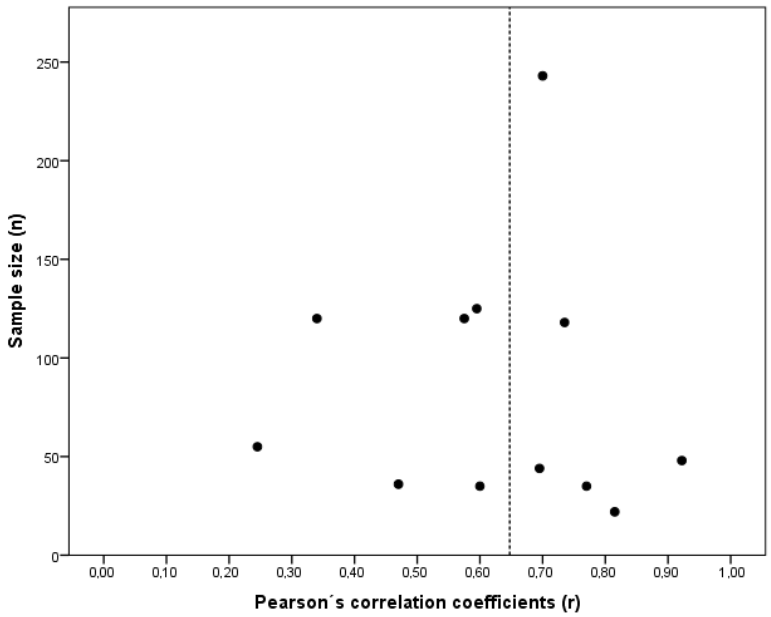

Figure 2. *Scatter plot of sample size and criterion-related validity coefficients (r) of the toe-touch test for estimating hamstring extensibility. Dashed line represents the median value of the validity coefficients.*

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 weighted mean of r corrected for sampling error only (r_c), the weighted mean of r corrected for both sampling error and measurement error (r_p), as well as the 95% CI of the criterion-related validity correlation coefficients (r_p) for the TT test for estimating hamstring extensibility. In addition, to detect the presence of moderator effects which may affect the criterion-related validity of the TT test, the 95% CV, the percentage of variance accounted for by statistical artefacts, and the *Q* homogeneity statistic were calculated.

Table 2. *Results of the meta-analysis for the criterion-related validity correlation coefficients of the toe-touch test for estimating hamstring extensibility*

	K	*n*	*N*	r_c	r_p	95% CI[a]	95% CV[b]	% of variance[c]	*Q* statistic
Overall	6	12	1,001	0.61	0.66	0.54-0.79	0.32-1.00	10.96	109.50*
Sex of participants[d]									
Males	4	5	464	0.49	0.52	0.37-0.67	0.25-0.79	23.49	21.29*
Females	4	6	294	0.73	0.77	0.66-0.89	0.53-1.00	17.68	33.94*
Age of participants									
Children	1	2	66	0.74	0.78	0.65-0.92	0.78-0.78	100.00	1.51
Adults	5	10	935	0.60	0.66	0.54-0.77	0.30-1.00	9.82	101.88*
Level of hamstring extensibility									
Low	5	6	428	0.52	0.55	0.38-0.71	0.15-0.94	14.79	40.57*
High	5	6	573	0.68	0.75	0.66-0.84	0.61-0.89	28.13	21.33*

Note. *K*, number of studies; *n*, number of *r*s; *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 toe-touch test. [d] Because a study mixed males and females, the overall *n* for this category is lower. * $p < 0.05$

The overall result showed that the TT test had a moderate mean correlation coefficient of criterion-related validity for estimating hamstring extensibility (r_p = 0.66, 0.54-0.79) in which the 95% CI did not include the value zero. However, we should be extremely cautious because the low numbers of *r* values over the present meta-analysis were supported. In line with the present meta-analysis, recently Mayorga-Vega et al. (2014) carried out a meta-analysis about the criterion-related validity of the sit-and-reach tests for estimating hamstring and lumbar extensibility. These authors found that overall the classic sit-and-reach tests also had a moderate criterion-related validity for estimating hamstring extensibility (r_p = 0.67, 0.55-0.80). Additionally, because 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 large, we concluded that the results could be affected by moderator effects. Therefore, follow-up moderator analyses were conducted using predefined moderators as it was hypothesized in the present study.

The classic sit-and-reach test is probably the most widely used lineal measure of flexibility in exercise science laboratories, physical education classes, and commercial fitness centres (Cepero et al., 2011; Holt et al., 1999; Mirzaei et al., 2011). In this line, Castro-Piñero et al. (2009a) carried out a systematic review of the criterion-related validity of field-based fitness battery tests worldwide for youth. These authors found that 91% (10 of 11) of battery tests that included the flexibility assessment proposed the classic sit-and-reach test. However, in none of these batteries the TT test was proposed. However, according to the

results of the present meta-analysis, if the purpose is to assess hamstring extensibility, it seems that the use of one test over the other is not justified.

Moderator analyses

Table 2 reports the results of moderator analyses to examine the effects of the sex of the participants (i.e., male and female), the age of the participants (i.e., children and adults), and the level of hamstring extensibility (i.e., low average level, < 80º, and high average level, ≥ 80º) on overall criterion-related validity correlation coefficient of the TT test for estimating hamstring extensibility.

Sex of participants: The results of the present study showed that the TT test had a moderate-low mean correlation coefficient of criterion-related validity for estimating hamstring extensibility for males (r_p = 0.52, 0.37-0.67) and moderate-high for females (r_p = 0.77, 0.66-0.89) in which the 95% CIs did not include the value zero. Additionally, aside from this considerably greater value for females compared to males, the 95% CIs of mean correlation coefficients were hardly overlapped. However, according to moderator analyses criteria, all the criteria were met in both males and females, indicating that the criterion-related validity of the TT test separately for sex was still heterogeneous. Finally, because a study grouped males and females together, in Table 3 the overall *n* for the sex of participants is lower.

The results of the present meta-analysis suggest that the sex of participants affects the criterion-related validity of the TT test for estimating hamstring extensibility. Therefore, it seems that the use of the TT test is more appropriate among females than among males. In this line, most studies either examined only one sex or both but without grouping them probably because they intuited that the sex of participants was a feature than would affect the relationship between the TT test and criterion measure (Ayala et al., 2011; López Miñarro et al., 2008a; López-Miñarro et al., 2011; López-Miñarro et al., 2010a; López-Miñarro et al., 2010b; Rodríguez-García et al., 2008). Similarly, Mayorga-Vega et al. (2014) found that the classic sit-and-reach showed a trend to be more valid for females (r_p = 0.70, 0.58-0.82) than for males (r_p = 0.64, 0.50-0.78). However, in contrast with the present meta-analysis, in the Mayorga-Vega´s et al. (2014) study these differences were really small and the 95% CIs were overlapped.

Age of participants: The results of the present meta-analysis showed that the TT test had a moderate-high mean correlation coefficient of criterion-related validity for estimating hamstring extensibility for children (r_p = 0.78, 0.65-0.92) and moderate for adults (r_p = 0.66, 0.54-0.77) in which the 95% CIs did not include the value zero. These results suggested that the criterion-related validity of the TT test is greater for children than for adults. However, the 95% CIs of mean correlation coefficients were overlapped. Furthermore, we should be extremely cautious because the analyses for children were supported only over two *r* values. In this line, future studies with children about the criterion-related validity of the TT test are required. In this line, Mayorga-Vega et al. (2014) did not find out differences in the results of the classic sit-and-reach between age categories. Finally, according to moderator analyses criteria, the three criteria were met among adults (for children clearly were not met because these analyses had only two *r* values), indicating that the criterion-related validity of the TT test separately for age were still heterogeneous.

Level of hamstring extensibility: The results of this study showed that the TT test had a moderate-low 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 = 0.55, 0.38-0.71) and moderate-high for participants with a high level of hamstring extensibility (≥ 80º in the average score of the straight leg raise test) (r_p = 0.75, 0.66-0.84) in which the 95% CIs did not include the value zero. The results of the present meta-analysis suggested that the criterion-related validity of the TT test is larger for participants with high

level of hamstring extensibility than those with low hamstring extensibility. However, we have to be aware that the 95% CIs of mean correlation coefficients were slightly overlapped. Additionally, according to moderator analyses criteria, at least two of the three criteria were met in both categories, indicating that the criterion-related validity of the TT test separately for level of hamstring extensibility were still heterogeneous.

Similarly, in line with the results of the present meta-analysis, previous primary studies carried out with young adults (López-Miñarro & Rodríguez-García, 2010b) and elderly women (López-Miñarro et al., 2011) found that the level of hamstring extensibility influenced the criterion-related validity of the TT test. 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. Although Mayorga-Vega et al. (2014) found the same trend in their meta-analysis with the classic sit-and-reach test, the 95% CIs were overlapped and they also pointed out this methodological limitation.

STRENGTHS AND LIMITATIONS

An extensive revision of the general strengths and limitations of the meta-analyses, as well as specifically in the meta-analysis of the criterion-related validity of the field-based flexibility tests, has been previously published (Mayorga-Vega et al., 2014). Briefly, regarding the strengths of the present meta-analysis, we followed several measures to avoid (or at least to reduce) publication bias. Firstly, to avoid availability bias, we conducted a wide literature search thought several databases without limiting any kind of manuscript (i.e., articles, master/doctoral dissertations, and conference proceedings), language (i.e., English and non-English language) or publication date. Secondly, in the present meta-analysis all the studies by the same authors were thoroughly cross-referenced with each other in order to avoid duplicated information. Lastly, several exploratory analyses were also conducted to detect the potential presence of publication bias.

Another strength of the present meta-analysis is related to the statistical approach used. In the present study, the Hunter-Schmidt´s psychometric meta-analysis approach (2004) was conducted in order to obtain the population estimates of criterion-related validity of the TT test. Since this method estimates the population correlation by individually correcting the observed correlations due to various artefacts such as sampling error and measurement error, it has been considered one of the best meta-analyses approaches.

On the other hand, there were some limitations that should be considered when examining the results of the present meta-analysis. The main limitations 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, because a partially hierarchical breakdown had to be used, quite misleading results due to confounding and interaction effects might be produced. Therefore, the results of the present study should be considered with caution.

Another limitation of the present meta-analysis is related to the criterion measures used in the included studies. Although all the previous studies found considered the angular tests measured by goniometers as the criterion measures, nowadays some studies have suggested that the criterion measures of hamstring extensibility must be reexamined and readjusted (Cardoso et al., 2007; Hartman & Looney, 2003). Finally, coding some study features was problematic due to different reasons. For instance, because in the present meta-analysis the level of hamstring extensibility was classified based on the average scores, we are

aware that several individuals with low hamstring extensibility could be classified as high flexibility and vice versa. Additionally, 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.

CONCLUSIONS

Overall the TT test has a moderate mean correlation coefficient of criterion-related validity for estimating hamstring extensibility. The results of the present meta-analysis show that the TT test has the same population estimated criterion-related validity for estimating hamstring extensibility than the widely used lineal test classic sit-and-reach. Regarding the three potential moderators examined (sex of participants, age of participants, and level of hamstring extensibility), generally females, children, and individuals with high levels of hamstring extensibility seem 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% CIs of mean correlation coefficients were overlapped, and that criterion-related validity of the TT test within each category was still heterogeneous, we should be cautious with the results of the present meta-analysis.

When angular tests such as the straight leg raise or knee extension tests cannot be used, the TT test seems to be a useful alternative to estimate hamstring extensibility. Nevertheless, as in the application of any field-based fitness test, evaluators must be aware that the results of the TT test is 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 the TT test for estimating hamstring extensibility, especially among populations such as children, and go deeply into other related aspects such as the influence of the level of hamstring extensibility.

ACKNOWLEDGEMENTS

We thank Anna Szczesniak for the English revision. The author Daniel Mayorga-Vega is supported by a research grant from the Spanish Ministry of Education, Culture and Sport (AP2010-5905).

REFERENCES

1. American Academy of Orthopaedic Surgeons (1996). *Joint motion: Method of measuring and recording*. Livingstone: Edinburgh.
2. Ayala, F., Sainz de Baranda, P., De Ste Croix, M., & Santonja, F. (2011). Criterion-related validity of four clinical tests used to measure hamstring flexibility in professional futsal players. *Phys Ther Sport, 12*(4), pp.175-181.
3. Ayala, F., Sainz de Baranda, P., De Ste Croix, M., & 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. *Phys Ther Sport*, *13*(4), pp.219-226.
4. Begg, C. B., & Mazumdar, M. (1994). Operating characteristics of a rank order correlation for publication bias. *Biometrics*, *50*(4), pp.1088-1101.
5. Cardoso, J. R., Azevedo, N. C. T., Cassano, C. S., Kawano, M. M., & 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. *Rev Bras Fisioterapia*, *11*(2), pp.133-138.

6. Castro-Piñero, J., Artero, E. G., España-Romero, V., Ortega, F. B., Sjöström, M., & Ruiz, J. R. (2009a). Criterion-related validity of field-based fitness tests in youth: A systematic review. *Brit J Sport Med*, *44*(13), pp.934-943.
7. Castro-Piñero, J., Chillón, P., Ortega, F. B., Montesinos, J. L., Sjöström, M., & 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. *Int J Sport Med*, *30*(9), pp.658-662.
8. Cepero, M., López, R., Suárez-Llorca, C., Andreu-Cabrera, E., & Rojas, F. J. (2011). Fitness test profiles in children aged 8-12 years old in Granada (Spain). *J Hum Sport Exerc, 6*(1), pp.135-145.
9. Cohen, J. A. (1992). Power primer. *Psychological Bulletin*, *112*(1), pp.155-159.
10. Cooper, H., Hedges, L. V., & Valentine, J. C. (2009). *The handbook of research synthesis and meta-analysis* (2nd ed.). New York: Sage.
11. Erkula, G., Demirkan, F., Kilic, B. A., & Kiter, E. (2002). Hamstring shortening in healthy adults. *J Back Musculoskelet*, *16*(2), pp.77-81.
12. Fisk, J. W., Baigent, M. L., & Hill, P. D. (1984). Scheuermann´s disease. Clinical and radiological survey of 17 and 18 year olds. *Am J Phys Med*, *63*(1), pp.18-30.
13. Flather, M. D., Farkouh, M. E., Pogue, J. M., & Yusuf, S. (1997). Strengths and limitations of meta-analysis: Larger studies may be more reliable. *Control Clin Trials*, *18*(6), pp.568-579.
14. Hartman, J. G., & Looney, M. (2003). Norm-referenced and criterion-referenced reliability and validity of the back-saver sit-and-reach. *Meas Phys Educ Exer Sci*, *7*(2), pp.71-87.
15. Harvey, J., & Tanner, S. (1991). Low back pain in young athletes: A practical approach. *Sports Med*, *12*(6), pp.394-406.
16. Holt, L. E., Pelma, T. W., & Burke, D. G. (1999). Modifications to the standard sit-and-reach flexibility protocol. *J Athl Training*, *34*(1), pp.43-47.
17. Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.). Newbury Park: Sage.
18. Jones, C. J., Rikli, R. E., Max, J., & Noffal, G. (1998). The reliability and validity of a chair sit-and-reach test as a measure of hamstring flexibility in older adult. *Res Q Exercise Sport*, *69*(4), pp.338-343.
19. Kendall, F. P., McCreary, E. K., Provance, P. G., Rodgers, M. M., & Romani, W. A. (2005). *Muscles: Testing and function with posture and pain* (5th ed.). Baltimore: Lippincott, Williams, & Wilkins.
20. Kraus, H., & Hirschland, R. P. (1954). Minimum muscular fitness tests in school children. *The Res Quarterly, 25*(2), pp.178-188.
21. Light, R. J., & Pillemer, D. B. (1984). *Summing up: The science of reviewing research*. Cambridge: Harvard University Press.
22. Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Newbury Park: Sage.
23. López Miñarro, P. A., Ferragut Fiol, C., Alacid Cárceles, F., Yuste Lucas, J. L., & 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*(157), pp.24-29.
24. López Miñarro, P. A., Rodríguez García, P. L., Yuste, J. L., Alacid, F., Ferragut, C., & García Ibarra, A. (2008b). Validity of the lumbo-sacral position in bending as measure of hamstring muscle extensibility on young athletes. *Arch Med Depor, 25*(124), pp.103-110.
25. López Miñarro, P. A., Sainz de Baranda Andújar, P., Yuste Lucas, J. L., & 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*(8), pp.87-92.
26. López-Miñarro, P. A., & Alacid, F. (2009). Influence of hamstring muscle extensibility on spinal curvatures in young athletes. *Sci Sport*, *25*, pp.188-193.

27. López-Miñarro, P. A., Alacid, F., Muyor, J. M., & López, F. J. (2010a). Criterion-related validity of the modified sit-and-reach test as hamstring muscle extensibility measure in young adults. *Kronos*, *9*(17), pp.39-46.
28. López-Miñarro, P. A., García Ibarra, A., & Rodríguez García, P. L. (2010b). Comparison between sit-and-reach tests for measuring hamstring muscle extensibility. *Apunts*, *99*(1), pp.56-64.
29. López-Miñarro, P.A., Muyor, J. M., & Alacid, F. (2011). Validity of sit-and-reach tests as measures of hamstring extensibility in older women. *Rev Int Med Ciencias Act Fís Deport, 11*(43), pp.564-572.
30. López-Miñarro, P. A., & Rodríguez-García, P. L. (2010c). Hamstring muscle extensibility influences the criterion-related validity of sit-and-reach and toe-touch tests. *J Strength Cond Res*, *24*(4), pp.1013-1018.
31. Mayorga-Vega, D., Merino-Marban, R., & Viciana, J. (2014). Criterion-related validity of sit-and-reach tests for estimating hamstring and lumbar extensibility: A meta-analysis. *J Sport Sci Med*. *J Sport Sci Med*, *13*(1), pp.1-14.
32. Mirzaei, B., Curby, D. C., Barbas, I., & Lotfi, N. (2011). Anthropometric and physical fitness traits of four-time World Greco-Roman wrestling champion in relation to national norms: A case study. *J Hum Sport Exerc, 6*(2), pp.406-413.
33. Muyor, J. M., López-Miñarro, P. A. & Casimiro, A.J. (2012). Effect of stretching program in an industrial workplace on hamstring flexibility and sagittal spinal posture of adult women workers: A randomized controlled trial. *J Back Musculoskelet*, *25*(3), pp.161-169.
34. Rodríguez-García, P. L., López-Miñarro, P. A., Yuste, J. L., & 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. *Med Sport*, *61*(1), pp.11-20.
35. Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychol Bull*, *86*(3), pp.683-641.
36. Roth-Isigkeit, A., Thyen U., Stöven, H., Schwarzenberger, J., & Schmucker, P. (2005). Pain among children and adolescents: Restrictions in daily living and triggering factors. *Pediatrics*, *115*(2), pp.e152-162.
37. Sainz de Baranda, P., Rodríguez, P. L., Santonja, F. M., López, P. A., Andújar, P., Ferrer, V., & Pastor, A. (2006). Effects of hamstring stretching exercises on the toe-touch test in elementary schoolchildren. *J Hum Mov Stud, 51*(4), pp.277-289.
38. Sato, T., Ito, T., Hirano, T., Morita, O., Kikuchi, R., Endo, N., & Tanabe, N. (2008). Low back pain in childhood and adolescence: A cross sectional study in Niigata City. *Eur Spine J*, *17*(11), pp.1441-1447.
39. Sjölie, A. N. (2004). Low back pain adolescents are associated with poor hip mobility and high body mass index. *Scand J Med Sci Sports*, *14*(3), pp.168-175.
40. Standaert, C. J., & Herring, S. A. (2000). Spondylolysis: A critical review. *Br J Sports Med*, *34*(6), 415-422.
41. Wells, K. F., & Dillon, E. K. (1952). The sit-and-reach. A test of back and leg flexibility. *The Res Quarterly*, *23*, pp.115-118.