

## REVIEW

# Reliability and validity of field-based tests for assessing physical fitness in gymnasts



Jorge Salse-Batán<sup>a</sup>, Silvia Varela<sup>b</sup>, Adrián García-Fresneda<sup>c,\*</sup>, Carlos Ayán<sup>b</sup>

<sup>a</sup> Institut Nacional d'Educació Física de Catalunya - Barcelona, Spain

<sup>b</sup> Universidade de Vigo, Galicia Sur Health Research Institute (IIS Galicia Sur), SERGAS-UVIGO, Departamento de Didácticas Especiais, Well-Move Research Group, Pontevedra, Spain

<sup>c</sup> TecnoCampus, University of Pompeu Fabra, Mataró, Spain

Received 1 August 2022; accepted 19 October 2022

Available online 5 November 2022

## KEYWORDS

Strength;  
Aerobic endurance;  
Anaerobic endurance;  
Flexibility;  
Evaluation and  
measurement

## Abstract

**Objective:** This systematic review aimed at analysing the reliability and validity of field-based tests for assessing physical fitness in gymnasts.

**Method:** Three electronic databases (PubMed, SPORTDiscus, and Scopus) were searched up to March 2022, in order to identify studies that assessed the psychometric properties of field-based physical fitness test among gymnastics.

**Results:** A total of 16 studies on several gymnastics modalities (artistics  $n = 11$ ; rhythmic  $n = 3$ , artistics and rhythmic  $n = 1$ ; aerobic  $n = 1$ ), were analyzed. All studies reported on reliability measured through test-retest design. Validity was reported in only four studies. Regarding specific tests, the split test (ICC = 0.998), and the handstand (ICC = 1) showed the highest test-retest reliability. The greater validity values were achieved by the split test ( $r^2 = 0.52$ ), hanging pikes test ( $r^2 = 0.86$ ), and handstand test ( $r^2 = 0.65$ ).

**Conclusion:** A great variety of both specific and non-specific physical fitness tests have been analyzed in the field of gymnastics. The side split test, the handstand test, the vertical jump test, the 20-m run test, the agility test, and the aerobic gymnast anaerobic test could be useful tools to assess flexibility, strength, balance, muscular power, speed, agility, and cardiorespiratory fitness in gymnasts. Further investigations analyzing absolute reliability and criterion validity are needed.

© 2022 FUTBOL CLUB BARCELONA and CONSELL CATALÀ DE L'ESPORT. Published by Elsevier España, S.L.U. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## Introduction

It is estimated that worldwide about 50 million people of all ages regularly perform gymnastics in a club setting.<sup>1</sup> The International Gymnastics Federation (FIG, <http://www.fig>).

\* Corresponding author at: School of Health Sciences, TecnoCampus Mataró – University of Pompeu Fabra.

E-mail address: [agarciaf@tecnocampus.cat](mailto:agarciaf@tecnocampus.cat) (A. García-Fresneda).

<https://doi.org/10.1016/j.apunsm.2022.100397>

2666-5069/© 2022 FUTBOL CLUB BARCELONA and CONSELL CATALÀ DE L'ESPORT. Published by Elsevier España, S.L.U. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

[gymnastics.com](http://gymnastics.com)) recognized a total of eight disciplines, being three of them (artistic, rhythmic and trampoline) Olympic.<sup>1</sup>

Physical fitness (PF), is strongly involved in gymnastics, since its practice requires a combination of speed, strength, endurance, agility, flexibility, balance and power.<sup>2</sup> The importance of assessing PF in gymnastics relies on the fact that it not only helps coaches and trainers to monitor the development of their athletes, but also to promote healthy, injury-free participation, as well as talent identification.<sup>2-4</sup> Consequently, coaches and trainers need meaningful, reliable, and sensitive outcome gymnastics-specific fitness tests.

Laboratory test represents the gold standard for assessing PF, however these tests are expensive and require highly trained experimenters, which compromise their feasibility and applicability in the gymnastics context. Considering these circumstances, the use of field-based PF tests is recommended, since they are easy to administer, involve minimal equipment, minimal cost, and a larger number of participants can be evaluated in a relatively short period of time.<sup>5</sup> However, the quality and weighting of the information obtained from field-based tests is conditioned by the quality of psychometric properties, especially reliability (consistency or repeatability of measurements) and validity (the capacity of the test to reflect what is has been designed to measure), which should be informed in advance.<sup>6</sup>

Information regarding the accuracy of field-based tests for assessing athletes' PF, is usually provided by systematic reviews that summarize and critically analyze their psychometric properties. This has been the case of different sports modalities such as soccer,<sup>7</sup> basketball.<sup>8</sup>

Coaches and trainers can identify which are the more accurate field-based PF tests that should be administered to their athletes, through systematic reviews that have summarized and critically analyzed their psychometric properties. However, these reviews are usually focused on the most famous and practiced sports, such as soccer,<sup>7</sup> or basketball,<sup>8</sup> while to the very best of the authors' knowledge, scant research of this kind has been carried out in lesser popular sport modalities, such as gymnastics. In the light of all this, the objective of this study is to carry out a comprehensive review of the scientific evidence about the reliability and validity of field-based tests for assessing PF in gymnasts.

## Methods

A systematic review about the reliability and/or validity of field-based tests was used to assess the fitness level in gymnasts was carried out. This systematic review was conducted following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines.<sup>9</sup>

### Search strategy

Three electronic databases (PubMed, SPORTDiscus, and Scopus) were searched from inception to March 2022. The literature search was conducted by one researcher. The following keywords, Boolean operators, and combinations were used: ["Gymnastics" OR "Rhythm gymnastics" OR "Artistic gymnastics"] AND ["Physical fitness" OR "Physical performance" OR "Strength" OR "Muscular strength" OR "Endurance" OR

"Aerobic endurance" OR "Flexibility" OR "Anaerobic" OR "Aerobic endurance"] AND ["Evaluation" OR "Measurement"]. To be included in the review, studies were required to meet the following criteria: (i) provided information about the reliability and/or validity of at least one field-based PF test in gymnastics, (ii) published in English, Spanish or Portuguese and (iii) in a peer-reviewed journal. Investigations that reported data on the psychometric properties of field-based PF test without describing the methodological approach used for identifying reliability or validity, were excluded.

### Study selection

One author screened the titles and abstracts identified during the search. When the information provided suggested that the study met the selection criteria, a full-text copy was examined. Doubts about inclusion were discussed with a third author until a consensus was reached.

### Data extraction

All included studies were reviewed by one author. Information on participants' characteristics (n, age and gymnastics specialty), gymnastics fitness test/s performed, and values related to their reliability and/or validity (the method to identify them and the type of statistical analysis and its coefficients) was extracted. Two expert authors identified specific and non-specific tests. Discrepancies were resolved by a second author. The bibliography in all selected studies was analyzed in search of new evidence.

## Results

A total of 677 studies were found after different search strategies. Authors obtained 56 studies after removal duplicate results and those not related to the main aim. After reading full texts, a total of 16 studies about the psychometric properties of field-based physical fitness tests in gymnasts were selected for the further analysis (Table 1).

### Design and samples

Regarding specialty, artistic specialty was reported in eleven studies.<sup>2,4,10-18</sup> On the other hand, three investigations reported rhythmic specialty.<sup>19-21</sup> Only one study indicated aerobic specialty.<sup>22</sup> Another record included rhythmic and artistic specialties.<sup>23</sup>

All studies reported on reliability measured through test-retest design. Validity was reported in only four studies.<sup>2,15,18,22</sup>

### Relative reliability

A total of 15 investigations indicated test-retest reliability,<sup>2,4,10-14,16-23</sup> while only two studies added interrater reliability.<sup>15,19</sup>

A total of 12 studies reported on relative reliability for different physical fitness-related subtests.<sup>2,4,11-15,17-20,23</sup> Four studies reported on reliability data for one physical fitness test: flexibility,<sup>10,16</sup> cardiorespiratory fitness,<sup>22</sup> and

**Table 1** Studies included in the final selection.

First Author, Year	Sample (n, age, specialty)	Test (Specific/ Non-specific)	PF Component	Reliability (time interval) and validity
Batista, 2019	<i>n</i> = 157 Age = 13.87 ± 1.97 Specialty: Rhythmic	Leg up with help of the hand (Specific)	Flexibility	<b>Intra-examiner reliability</b> (10 days): Kendall Coefficient of Concordance = 0.83 – 0.97 <b>Inter-examiner reliability:</b> ICC = 0.84 – 0.97 <b>Validity:</b> NR
		Leg up without help of the hand (Specific)	Flexibility	
		Rotation of the upper limbs (Specific)	Flexibility	
		Trunk lift (Specific)	Flexibility	
		Forward stand-and-reach (Non-specific)	Flexibility	
Dallas, 2014	<i>n</i> = 18 Age = 21.83 ± 1.76 Specialty: Artistic	Sit and reach (Non-specific)	Flexibility	<b>Test-retest reliability</b> (at least five days): ICC = 0.992 <b>Validity:</b> NR
		SJ (Non-specific)	Muscular power	
		CMJ (Non-specific)	Muscular power	
Dallas, 2014	<i>n</i> = 34 Age = 9.22 ± 1.34 Specialty: Artistic	Sit and reach (Non-specific)	Flexibility	<b>Test-retest reliability</b> (at least five days): ICC = 0.996 <b>Validity:</b> NR <b>Test-retest reliability</b> (at least five days): ICC = 0.995 <b>Validity:</b> NR <b>Test-retest reliability</b> (same day): ICC = 0.994 <b>Validity:</b> NR
Kaldas, 2017	<i>n</i> = 90 Age = 8 - 18 Specialty: Artistic	The rope climb test (Non-specific)	Strength	<b>Inter-rater reliability:</b> ICC = 0.92 <b>Construct validity</b> (Test score vs. Competitive level): <i>r</i> <sup>2</sup> = 0.82 <b>Inter-rater reliability:</b> ICC = 0.98 <b>Construct validity</b> (Test score vs. Competitive level): <i>r</i> <sup>2</sup> = 0.88 <b>Inter-rater reliability:</b> ICC = 0.94 <b>Construct validity</b> (Test score vs. Competitive level): <i>r</i> <sup>2</sup> = 0.86 <b>Inter-rater reliability:</b> ICC = 0.99 <b>Construct validity</b> (Test score vs. Competitive level): <i>r</i> <sup>2</sup> = 0.01 <b>Inter-rater reliability:</b> ICC = 0.95 <b>Construct validity</b> (Test score vs. Competitive level): <i>r</i> <sup>2</sup> = 0.96 <b>Inter-rater reliability:</b> ICC = 0.98 <b>Construct validity</b> (Test score vs. Competitive level): <i>r</i> <sup>2</sup> = 0.80 <b>Inter-rater reliability:</b> ICC = 0.97 <b>Construct validity</b> (Test score vs. Competitive level): <i>r</i> <sup>2</sup> = 0.52 <b>Inter-rater reliability:</b> ICC = 0.91 <b>Construct validity</b> (Test score vs. Competitive level): <i>r</i> <sup>2</sup> = 0.91
		The jump test (Non-specific)	Muscular power	
		The hanging pikes test (Specific)	Strength	
		The shoulder flexibility test (Non-specific)	Flexibility	
		The agility test (Non-specific)	Agility	
		The over-grip pull-up test (Non-specific)	Strength	
		The splits test (Specific)	Flexibility	
		The push-up test (Non-specific)	Strength	

**Table 1** (Continued)

First Author, Year	Sample (n, age, specialty)	Test (Specific/ Non-specific)	PF Component	Reliability (time interval) and validity
Kritikou, 2017	n = 46 Age = 9.9 ± 1.3 Specialty: Rhythmic	The 20-yard sprint test (Non-specific)	Speed	<b>Inter-rater reliability:</b> ICC = 0.68 <b>Construct validity</b> (Test score vs. Competitive level): r <sup>2</sup> = 0.92
		The handstand test (Specific)	Strength and balance	<b>Inter-rater reliability:</b> ICC = 1.00 <b>Construct validity</b> (Test score vs. Competitive level): r <sup>2</sup> = 0.65
		Rhythmic gymnastics specific test (Specific)	Balance	<b>Test-retest reliability</b> (same day): ICC = 0.85 <b>Validity:</b> NR
		Shoulder flexion test (Non-specific)	Mobility	<b>Test-retest reliability</b> (same day): ICC = 0.94 <b>Validity:</b> NR
		Shoulder extension test (Non-specific)	Mobility	<b>Test-retest reliability</b> (same day): ICC = 0.97 <b>Validity:</b> NR
		Sit and reach test (Non-specific)	Flexibility	<b>Test-retest reliability</b> (same day): ICC = 0.92 <b>Validity:</b> NR
		Straight leg raise (Non-specific)	Flexibility	<b>Test-retest reliability</b> (same day): ICC = 0.91 <b>Validity:</b> NR
		Sideways leg extension (Non-specific)	Flexibility	<b>Test-retest reliability</b> (same day): ICC = 0.96 <b>Validity:</b> NR
		Bridge test – Body hyperextension test (Specific)	Mobility	<b>Test-retest reliability</b> (same day): ICC = 0.89 <b>Validity:</b> NR
		Push-up test (Non-specific)	Strength	<b>Test-retest reliability</b> (same day): ICC = 0.91 <b>Validity:</b> NR
		Sit-up test (Non-specific)	Strength	<b>Test-retest reliability</b> (same day): ICC = 0.91 <b>Validity:</b> NR
		Back extension test (Non-specific)	Strength	<b>Test-retest reliability</b> (same day): ICC = 0.91 <b>Validity:</b> NR
		CMJ (Non-specific)	Muscular power	<b>Test-retest reliability</b> (same day): ICC = 0.96 <b>Validity:</b> NR
		DJ (30 cm) (Non-specific)	Muscular power	<b>Test-retest reliability</b> (same day): ICC = 0.95 <b>Validity:</b> NR
León-Prados, 2011	n = 11 Age = 21.5 ± 3.5 Specialty: Artistic	20-m shuttle run test (Non-specific)	Cardiorespiratory fitness	<b>Test-retest reliability</b> (same day): ICC = 0.91 <b>Validity:</b> NR
		Rope climb test (5 m) (Non-specific)	Strength	<b>Test-retest reliability</b> (one day): ICC = 0.91 <b>Validity:</b> NR
Marina, 2013	n = 50 Age = 8.84 ± 0.62 Specialty: Artistic	Repetitions to failure. Straddle handstand on parallel bars (Specific)	Strength	<b>Test-retest reliability</b> (NR): ICC = 0.99 <b>Validity:</b> NR
		SJ (Non-specific)	Muscular power	<b>Test-retest reliability</b> (one week): ICC = 0.93 <b>Validity:</b> NR
		CMJ (Non-specific)	Muscular power	<b>Test-retest reliability</b> (one week): ICC = 0.95 <b>Validity:</b> NR
		CMJA (Non-specific)	Muscular power	<b>Test-retest reliability</b> (one week): ICC = 0.96 <b>Validity:</b> NR
		DJ (from 40 and 60 cm drop height) (Non-specific)	Muscular power	<b>Test-retest reliability</b> (one week) (DJ40): ICC = 0.94 <b>Validity:</b> NR <b>Test-retest reliability</b> (one week) (DJ60): ICC = 0.95 <b>Validity:</b> NR

Table 1 (Continued)

First Author, Year	Sample (n, age, specialty)	Test (Specific/ Non-specific)	PF Component	Reliability (time interval) and validity
Mkaouer, 2018	<i>n</i> = 51 Age = 11.03 ± 0.95 Specialty: Artistic	20 m run (Non-specific)	Speed	<b>Test-retest reliability</b> (same day): ICC = 0.996 <b>Validity:</b> NR
		4 m rope climb (Non-specific)	Strength	<b>Test-retest reliability</b> (same day): ICC = 0.999 <b>Validity:</b> NR
		Vertical jump test (Non-specific)	Muscular power	<b>Test-retest reliability</b> (same day): ICC = 0.999 <b>Validity:</b> NR
		Broad jump (Non-specific)	Muscular power	<b>Test-retest reliability</b> (same day): ICC = 0.995 <b>Validity:</b> NR
		Flexion legs upon upper body (Specific)	Strength	<b>Test-retest reliability</b> (same day): ICC = 0.994 <b>Validity:</b> NR
		Extension legs upon upper body (Specific)	Strength	<b>Test-retest reliability</b> (same day): ICC = 0.975 <b>Validity:</b> NR
		Pull up (Specific)	Strength	<b>Test-retest reliability</b> (same day): ICC = 0.999 <b>Validity:</b> NR
		Dips (Specific)	Strength	<b>Test-retest reliability</b> (same day): ICC = 0.993 <b>Validity:</b> NR
		Straddle lift to handstand (Specific)	Strength	<b>Test-retest reliability</b> (same day): ICC = 0.992 <b>Validity:</b> NR
		Double legs circle (Specific)	Strength	<b>Test-retest reliability</b> (same day): ICC = 0.99 <b>Validity:</b> NR
		V lever (Specific)	Strength	<b>Test-retest reliability</b> (same day): ICC = 0.999 <b>Validity:</b> NR
		Tucked top planchet (Specific)	Strength	<b>Test-retest reliability</b> (same day): ICC = 0.999 <b>Validity:</b> NR
		Back hang scale (Specific)	Strength	<b>Test-retest reliability</b> (same day): ICC = 0.999 <b>Validity:</b> NR
		Side split sit (Specific)	Flexibility	<b>Test-retest reliability</b> (same day): ICC = 0.998 <b>Validity:</b> NR
		Right split sit (Specific)	Flexibility	<b>Test-retest reliability</b> (same day): ICC = 0.992 <b>Validity:</b> NR
		Left split sit (Specific)	Flexibility	<b>Test-retest reliability</b> (same day): ICC = 0.985 <b>Validity:</b> NR
Bridge (Specific)	Flexibility	<b>Test-retest reliability</b> (same day): ICC = 0.985 <b>Validity:</b> NR		
Body bent (Non-specific)	Flexibility	<b>Test-retest reliability</b> (same day): ICC = 0.989 <b>Validity:</b> NR		
Leg lift forwards (Specific)	Flexibility	<b>Test-retest reliability</b> (same day): ICC = 0.968 <b>Validity:</b> NR		
Active shoulder flexibility (Non-specific)	Flexibility	<b>Test-retest reliability</b> (same day): ICC = 0.996 <b>Validity:</b> NR		
Robles, 2015	<i>n</i> = 10 (Reliability) <i>n</i> = 8 (Validity) Age = 17.7 ± 2.4 (Reliability); 18.2 ± 2.4 (Validity) Specialty: Aerobic	Specific anaerobic field test (Specific)	Anaerobic performance	<b>Test-retest reliability</b> (three days): ICC = 0.97 <b>Concurrent validity:</b> Lower-body power – SAGAT time trial: <i>r</i> = 0.69 – 0.73; <i>p</i> = 0.02 – 0.03 Upper-body power – SAGAT time trial: <i>r</i> = 0.69 – 0.70; <i>p</i> = 0.03
Russo, 2020	<i>n</i> = 45 Age = 10 – 12 Specialty: Both	Sit and reach test (Non-specific)	Flexibility	<b>Test-retest reliability</b> (one week): ICC > 0.870 <b>Validity:</b> NR

**Table 1** (Continued)

First Author, Year	Sample (n, age, specialty)	Test (Specific/ Non-specific)	PF Component	Reliability (time interval) and validity
Sands, 2008	n = 10 Age = 10.7 ± 0.99 Specialty: Artistic	Abduction test of the hips (Non-specific)	Mobility	<b>Test-retest reliability</b> (same day): Cronbach's alpha = 0.93 – 0.99 <b>Validity:</b> NR
		Wand test (Non-specific)	Mobility	
		Flamingo test (Non-specific)	Balance	
		Standstill long jump (Non-specific)	Strength	
		10 m test (Non-specific)	Speed	
		Cooper's test (Non-specific)	Cardiorespiratory fitness	
		Forward split test (Specific)	Flexibility	
		Santana, 2019	n = 51 Age = 13.76 ± 1.53 Specialty: Rhythmic	
Siatras, 2010	n = 32 Age = 9.9 ± 1.1 Specialty: Artistic			Shoulder flexors test (0° - 180°) (Non-specific)
		Shoulder extensors test (0° - 180°) (Non-specific)	Strength	<b>Intra-rater reliability</b> (one to seven days): ICC = 0.73 – 0.88 <b>Validity:</b> NR
		Hip flexors test (-30° - 60°) (Non-specific)	Strength	<b>Intra-rater reliability</b> (one to seven days): ICC = 0.83 – 0.88 <b>Validity:</b> NR
		Hip extensors test (-90°) (Non-specific)	Strength	<b>Intra-rater reliability</b> (one to seven days): ICC = 0.85 – 0.88 <b>Validity:</b> NR
Sleeper, 2012	n = 50 (Reliability) n = 105 (Validity) Age = 6 - 18 Specialty: Artistic	The rope climb test (Non-specific)	Strength	<b>Test-retest reliability</b> (one week): ICC = 0.80 <b>Construct validity</b> (Test score vs. Competitive level): r <sup>2</sup> = 0.27
		The jump test (Non-specific)	Muscular power	<b>Test-retest reliability</b> (one week): ICC = 0.83 <b>Construct validity</b> (Test score vs. Competitive level) r <sup>2</sup> = 0.42
		The hanging pikes test (Specific)	Strength	<b>Test-retest reliability</b> (one week): ICC = 0.88 <b>Construct validity</b> (Test score vs. Competitive level): r <sup>2</sup> = 0.27
		The shoulder flexibility test (Non-specific)	Flexibility	<b>Test-retest reliability</b> (one week): ICC = 0.92 <b>Construct validity</b> (Test score vs. Competitive level): r <sup>2</sup> = 0.05
		The agility test (Non-specific)	Agility	<b>Test-retest reliability</b> (one week): ICC = 0.86 <b>Construct validity</b> (Test score vs. Competitive level): r <sup>2</sup> = 0.47
		The over-grip pull-up test (Specific)	Strength	<b>Test-retest reliability</b> (one week): ICC = 0.89 <b>Construct validity</b> (Test score vs. Competitive level): r <sup>2</sup> = 0.20
		The splits test (Specific)	Flexibility	<b>Test-retest reliability</b> (one week): ICC = 0.91 <b>Construct validity</b> (Test score vs. Competitive level): r <sup>2</sup> = 0.09
		The push-up test (Non-specific)	Strength	<b>Test-retest reliability</b> (one week): ICC = 0.84 <b>Construct validity</b> (Test score vs. Competitive level): r <sup>2</sup> = 0.25

Table 1 (Continued)

First Author, Year	Sample (n, age, specialty)	Test (Specific/ Non-specific)	PF Component	Reliability (time interval) and validity
Sleeper, 2016	n = 30 (Reliability) n = 83 (Validity) Age = 11.1 ± 2.9 Specialty: Artistic	The 20-yard sprint test (Non-specific)	Speed	<b>Test-retest reliability</b> (one week): ICC = 0.85 <b>Construct validity</b> (Test score vs. Competitive level): r <sup>2</sup> = 0.47
		The handstand test (Specific)	Strength and balance	<b>Test-retest reliability</b> (one week): ICC = 0.92 <b>Construct validity</b> (Test score vs. Competitive level): r <sup>2</sup> = 0.38
		Splits test (Specific)	Flexibility	<b>Test-retest reliability</b> (one week): ICC = 0.97 <b>Construct validity</b> (Test score vs. Competitive level): r <sup>2</sup> = 0.004
		Shoulder flexibility test (Non-specific)	Flexibility	<b>Test-retest reliability</b> (one week): ICC = 0.90 <b>Construct validity</b> (Test score vs. Competitive level): r <sup>2</sup> = 0.01
		Hanging pikes test (Specific)	Strength	<b>Test-retest reliability</b> (one week): ICC = 0.83 <b>Construct validity</b> (Test score vs. Competitive level): r <sup>2</sup> = 0.29
		Overgrip pull-ups test (Specific)	Strength	<b>Test-retest reliability</b> (one week): ICC = 0.89 <b>Construct validity</b> (Test score vs. Competitive level): r <sup>2</sup> = 0.51
		Handstand push-up test (Specific)	Strength and balance	<b>Test-retest reliability</b> (one week): ICC = 0.96 <b>Construct validity</b> (Test score vs. Competitive level): r <sup>2</sup> = 0.24
		Rings hold test (Specific)	Strength	<b>Test-retest reliability</b> (one week): ICC = 0.75 <b>Construct validity</b> (Test score vs. Competitive level): r <sup>2</sup> = 0.36
		Agility test (Non-specific)	Agility	<b>Test-retest reliability</b> (one week): ICC = 0.95 <b>Construct validity</b> (Test score vs. Competitive level): r <sup>2</sup> = 0.55
		Vertical jump test (Non-specific)	Muscular power	<b>Test-retest reliability</b> (one week): ICC = 0.91 <b>Construct validity</b> (Test score vs. Competitive level): r <sup>2</sup> = 0.64
		Star excursion balance test (Non-specific)	Balance	<b>Test-retest reliability</b> (one week): ICC = 0.77 <b>Construct validity</b> (Test score vs. Competitive level): r <sup>2</sup> = 0.10
Torrado, 2011	n = 36 Age = 7 - 12 Specialty: Artistic	SJ (Non-specific)	Muscular power	<b>Test-retest reliability</b> (two days): ICC = 0.703 Pearson's correlation coefficient = 0.595 <b>Validity</b> : NR
		CMJ (Non-specific)	Muscular power	<b>Test-retest reliability</b> (two days): ICC = 0.767 Pearson's correlation coefficient = 0.624 <b>Validity</b> : NR
		CMJA (Non-specific)	Muscular power	<b>Test-retest reliability</b> (two days): ICC = 0.829

**Table 1** (Continued)

First Author, Year	Sample (n, age, specialty)	Test (Specific/ Non-specific)	PF Component	Reliability (time interval) and validity
		DJ (from 40 and 60 cm drop height) (Non-specific)	Muscular power	Pearson's correlation coefficient = 0.713 Validity: NR Test-retest reliability (two days) (DJ40): ICC = 0.776 Pearson's correlation coefficient = 0.658 Validity: NR Test-retest reliability (two days) (DJ60): ICC = 0.769 Pearson's correlation coefficient = 0.625 Validity: NR

CMJ: Countermovement jump; CMJA: Countermovement jump with arm swing; DJ: Drop jump; ICC: Intraclass correlation coefficient; KTK: Körperkoordinationstest für Kinder; NR: Not reported; SJ: Squat jump.

coordination.<sup>21</sup> Time interval between test and retest varied from same day<sup>4,10,16,20</sup> to ten days.<sup>19</sup>

### Flexibility

Flexibility tests were the most frequent assessment in 10 studies.<sup>2,4,11,15,16,18–20,23,24</sup> Regarding to the most reliable test, the side split test was the most reliable specific assessment (ICC = 0.998), while between non-specific tests, active shoulder flexibility was the most reliable (ICC = 0.996).<sup>4</sup>

### Strength

A total of eight records analyzed strength.<sup>2,4,12,15,17,18,20,23</sup> The handstand test (ICC = 1.00)<sup>15</sup> and 4-m rope climb (ICC = 0.999)<sup>4</sup> obtained the higher reliable values in the specific and the non-specific tests, respectively.

### Muscular power

Eight investigations used muscular power assessments.<sup>2,4,11,13–15,18,20</sup> The non-specific vertical jump test was the most reliable evaluation in muscular power (ICC = 0.999).<sup>4</sup>

### Balance

A total of five studies used balance assessments.<sup>2,15,18,20,23</sup> The handstand test was also obtained the best value in the specific balance tests (ICC = 1.00)<sup>15</sup> and regarding to non-specific balance tests, the flamingo test accomplished the higher result (ICC = 0.870).<sup>23</sup>

### Speed

Four studies analyzed speed performance.<sup>2,4,15,23</sup> The 20-m run obtained the best value in the non-specific speed assessments (ICC = 0.996).<sup>4</sup>

### Agility

Three studies reported agility assessments.<sup>2,15,18</sup> The non-specific agility test accomplished the most reliable value (ICC = 0.95).<sup>15,18</sup>

### Cardiorespiratory fitness

Three studies analyzed cardiorespiratory fitness.<sup>20,22,23</sup> Specific aerobic gymnast anaerobic test was the most reliable test (ICC = 0.97)<sup>22</sup> and 20-m shuttle run test obtained the higher value in the non-specific cardiorespiratory fitness evaluations (ICC = 0.91).<sup>20</sup>

### Mobility

Mobility assessments were reported in two studies.<sup>20,23</sup> Shoulder extension test obtained the best result in non-specific mobility tests (ICC = 0.97).

### Coordination

Only one study reported coordination assessment.<sup>21</sup> In this study, throwing the ball and reversing forward specific test obtained 0.799 in the Pearson's correlation coefficient.

### Absolute reliability

Standard error of measurement (SEM) and minimum detectable change (MDC) were calculated to assess absolute reliability in only four studies<sup>13,15,17,22</sup>. Indicated that SEM with 95% confidence intervals was 0.96 to 0.99 in the overall reliability analysis of the battery total score. Another study reported that SEM for all muscle groups varied from  $\pm 0.4$  to  $\pm 1.0$  kg at different joint angular positions in four muscle groups (shoulder flexors and extensors, hip flexors and extensors).<sup>17</sup> MDC 95% resulting from the ICC was 0.12 s in the specific anaerobic field test for aerobic gymnastics.<sup>13,22</sup> Reported that SEM (flight time) varied from 5.84 ms to 10.26 ms in the CMJ and DJ40, respectively. SEM (estimated mechanical power) varied between 1.84 w/kg (DJ60) and 2.25 w/kg (DJ40). MDC % (flight time) was also indicated and ranged between 3.3 (CMJA) and 5.9 (DJ40). MDC % (estimated mechanical power) ranged from 12.5 (DJ60) to 13.9 (DJ40).

### Validity

Three studies analyzed construct validity using a simple regression analysis between total test scores and the gymnasts' competition level.<sup>2,15,18</sup> Only one used criterion validity evaluating whether the specific anaerobic field test for aerobic gymnastics correlated with the Wingate test.<sup>22</sup>

### Flexibility

Three studies reported flexibility assessments.<sup>2,15,18</sup> Regarding to validity values, the specific splits test was the most valid test ( $r^2 = 0.52$ )<sup>15</sup> and the shoulder flexibility test obtained the higher value among non-specific tests ( $r^2 = 0.05$ ).<sup>2</sup>

**Strength**

Strength assessments were used in three records.<sup>2,15,18</sup> The specific hanging pikes test accomplished the higher value in strength assessments ( $r^2 = 0.86$ ), while the non-specific push-up test obtained the higher value ( $r^2 = 0.91$ ).<sup>15</sup>

**Muscular power**

Three investigations reported muscular power evaluations.<sup>2,15,18</sup> The non-specific jump test was the most valid test in muscular power assessments ( $r^2 = 0.88$ ).<sup>15</sup>

**Balance**

Balance evaluations were analyzed in three investigations.<sup>2,15,18</sup> The handstand test obtained the best value in balance evaluations ( $r^2 = 0.65$ ).<sup>15</sup>

**Speed**

Two records reported speed evaluations.<sup>2,15</sup> The 20-yard sprint test was the most valid test ( $r^2 = 0.92$ ).<sup>15</sup>

**Agility**

Three investigations reported agility assessments.<sup>2,15,18</sup> The agility test showed the higher value ( $r^2 = 0.96$ ).<sup>15</sup>

**Cardiorespiratory fitness**

Cardiorespiratory fitness validity test was reported in one study.<sup>22</sup> This study indicated that validity ranged from 0.69 to 0.73.

**Discussion**

The main goal of this review was to show the scientific evidence about the reliable and validity data of field-based tests to assess PF in gymnastics. Thus, coaches could use these findings to know and control physical condition in gymnasts, that it could be helpful to improve the performance of athletes.

On the one hand, it was reported that reliability values were high in most tests ( $ICC > 0.9$ ). Anyway, it is important to clarify that protocols were not adequate in several records. According to this, most studies included in this review used a time between trials too short ( $< 2$  days). In most of them, the familiarization session was not included. Thus, it exists scientific evidence which indicates that subjects should be familiarized with the performance protocol by at least one trial before measurement commences.<sup>6</sup> Furthermore, the time between assessments could affect the test-retest reliability and it could not be too short or too long.<sup>25</sup> Although it was shown that the optimal time interval will vary depending on the construct being measured, on the stability of the construct over time and on the target population, two weeks seem the most frequently recommended interval.<sup>26</sup> It seems relevant to highlight that various studies included in this review used one week interval to control fatigue or learning effects while trying to avoid enough passage of time to allow a true change in a gymnast's overall fitness.<sup>2</sup>

Regarding to physical abilities, flexibility, strength, power, balance, speed, agility, and cardiorespiratory fitness play an important role in the success of a competitive gymnast.<sup>2</sup> Considering this, it has shown that the most reliable tests were the side split test (flexibility), the handstand test

(strength and balance), the vertical jump test (muscular power), the 20-m run test (speed), the agility test (agility), and the aerobic gymnast anaerobic test (cardiorespiratory fitness). Thus, three of them (the vertical jump test, the 20-m run test and the agility test) were non-specific tests. So, it is important to consider that specific tests are useful tool to assess specific performance because the non-specific tests do not always correlate well with gymnasts' performance, and this could suppose a weakness.<sup>27</sup> Furthermore, some of these tests showed high reliability values in other populations. For example, the 20-m run test showed a high reliability among elite youth female soccer players ( $ICC = 0.96$ ).<sup>28</sup> Regarding jump assessments, the vertical jump test also showed a high reliability values in men and women physically active ( $ICC = 0.87 - 0.94$ ).<sup>29</sup>

It is relevant to highlight that only two studies indicated inter-rater reliability data.<sup>15</sup> only reported inter-rater reliability data, while<sup>19</sup> showed intra- and inter-rater reliability values. Thus, authors affirmed that only one rater administered the measurements supposed a limitation in one study.<sup>17</sup>

Regarding to reliability types, relative reliability is the degree to which individuals maintain their position in a sample with repeated measurements,<sup>30</sup> being reported in most studies included in this review. Nevertheless, only four records indicated absolute reliability data through SEM and MDC. Absolute reliability refers to the degree to which repeated measurements vary for individuals. In addition, it helps predict the magnitude of a real change in individual athletes and could be employed to estimate statistical power for a repeated-measures experiment.<sup>30</sup> According to this, it seems necessary that future investigations also show absolute reliability data with the main goal of knowing how repeated assessments vary among subjects.

On the other hand, validity was reported in only four studies. Three of them used construct validity, contrasting total test scores with gymnasts' competition level. Nonetheless, only one study used criterion validity evaluating the correlation between the specific anaerobic field test for aerobic gymnastics and a gold standard assessment (the Wingate test). This supposes a weakness because criterion validity allows for an objective measure of validity.<sup>6</sup> The lack of studies that analyze the validity of different tests used in gymnasts hinders the implementation of these tests to assess and control athletes' performance.

**Limitations**

There are some limitations that need to be considered when interpreting the findings of this review. First, only four studies analyzed absolute reliability, and since this helps predict the magnitude of a true change. Moreover, criterion validity is the only objective measure of validity and only one record reported criterion validity data. Thus, these results should be interpreted with caution.

In addition, most studies did not report a correct explanation about tests. Furthermore, most tests were non-specific assessments, so this should be interpreted considering this aspect. It is important the details of assessments to understand and apply these tests in the best way. Finally, language restrictions in the search process and the non-inclusion of grey literature might have affected these results.

## Conclusions

There are several tests to assess gymnasts' fitness. The side split test, the handstand test, the vertical jump test, the 20-m run test, the agility test, and the aerobic gymnast anaerobic test could be useful tools to assess flexibility, strength, balance, muscular power, speed, agility, and cardiorespiratory fitness in gymnasts. Further investigations aimed at analyze absolute reliability and criterion validity are needed.

## Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## Conflicts of interest

The authors declare that they don't have any conflict of interests.

## References

- Edouard P, Steffen K, Junge A, Leglise M, Soligard T, Engebretsen L. Gymnastics injury incidence during the 2008, 2012 and 2016 Olympic Games: analysis of prospectively collected surveillance data from 963 registered gymnasts during Olympic Games. *Br J Sports Med.* 2018;52(7):475–81.
- Sleeper MD, Kenyon LK, Casey E. Measuring fitness in female gymnasts: the gymnastics functional measurement tool. *Int J Sports Phys Ther.* 2012;7(2):124–38.
- Bergeron MF, Mountjoy M, Armstrong N, Chia M, Côté J, Emery CA, et al. International Olympic Committee consensus statement on youth athletic development. *Br J Sports Med.* 2015;49(13):843–51.
- Mkaouer B, Hammoudi-Nassib S, Amara S, Chaabène H. Evaluating the physical and basic gymnastics skills assessment for talent identification in men's artistic gymnastics proposed by the International Gymnastics Federation. *Biol Sport.* 2018;35(4):383–92.
- Ruiz JR, Castro-Piñero J, España-Romero V, Artero EG, Ortega FB, Cuenca MAM, et al. Field-based fitness assessment in young people: the ALPHA health-related fitness test battery for children and adolescents. *Br J Sports Med.* 2011;45(6):518–24.
- Currell K, Jeukendrup A. Validity, reliability and sensitivity of measures of sporting performance LK -. *Sports Medicine TA.* 2008;38(4):297–316. <https://rug.on.worldcat.org/oclc/367041412>.
- Paul DJ, Nassiss GP. Physical fitness testing in youth soccer: issues and considerations regarding reliability, validity and sensitivity. *Pediatr Exerc Sci.* 2015;27(3):301–13.
- Mancha-Triguero D, García-Rubio J, Calleja-González J, Ibáñez SJ. Physical fitness in basketball players: a systematic review. *J Sports Med Phys Fitness.* 2019;59(9):1513–25.
- Moher D, Liberati A, Tetzlaff J, Altman DG, Antes G, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med.* 2009;6(7).
- Dallas G, Kirialanis P, Mellos V. The acute effect of whole body vibration training on flexibility and explosive strength of young gymnasts. *Biol Sport.* 2014;31(3):233–7.
- Dallas G, Smirniotou A, Tsiganos G, Tsopani D, Di Cagno A, Tsolakis C. Acute effect of different stretching methods on flexibility and jumping performance in competitive artistic gymnasts. *J Sports Med Phys Fitness.* 2014;54(6):683–90.
- León-Prados JA, Gómez-Píriz PT, González-Badillo JJ. Relación entre test físicos específicos y rendimiento en gimnastas de elite. *RICYDE.* 2011;7(22):58–71.
- Marina M, Torrado P. Does gymnastics practice improve vertical jump reliability from the age of 8 to 10 years? *J Sports Sci.* 2013;31(11):1177–86.
- Pineda PT, Evrad MM. Fiabilitat dels tests de salt vertical de les gimnastes prepuberals. *Apunts Med Esport.* 2012;47(175):91–7.
- Kaldas J, Bisson C, Hogue AC, Apinis C, Berbiche D, Gaudreault N. Construct validity and inter-rater reliability of the Gymnastic Functional Measurement Tool in the classification of female competitive gymnasts in Canada. *Phys Ther Sport.* 2017;28:9–14.
- Sands WA, McNeal JR, Stone MH, Haff GG, Kinser AM. Effect of vibration on forward split flexibility and pain perception in young male gymnasts. *Int J Sports Physiol Perform.* 2008;3(4):469–81.
- Siatras TA, Douka ID, Milosic DC. Feasibility and reproducibility of muscular strength measures in gymnastics-specific body positions using hand-held dynamometry. *Isokinet Exercise Sci.* 2010;18(4):223–34.
- Sleeper MD, Kenyon LK, Elliott JM, Cheng MS. Measuring sport-specific physical abilities in male gymnasts: the Men'S Gymnastics Functional Measurement Tool. *Int J Sports Phys Ther.* 2016;11(7):1082–100.
- Batista A, Garganta R, Ávila-Carvalho L. Flexibility and functional asymmetry in rhythmic gymnastics. *Athens J Sports.* 2019;6(2):77–94.
- Kritikou M, Donti O, Bogdanis GC, Donti A, Theodorakou K. Correlates of artistry performance scores in preadolescent rhythmic gymnasts. *Sci Gymnast J.* 2017;9(2):165–77.
- Vernetta Santana M, Montosa Mirón I, Gutiérrez Sánchez Á. Validación y fiabilidad de un test para evaluar la coordinación óculo manual y agilidad en gimnasia rítmica. *Sportis Sci J School Sport Physical Educ Psychomotr.* 2019;5(2):174–89.
- Alves CRR, Borelli MTC, Paineli V de S, Azevedo R de A, Borelli CCG, Lancha Junior AH, et al. Development of a specific anaerobic field test for aerobic gymnastics. *PLoS One.* 2015;10(4):e0123115.
- Russo L, Palermi S, Dhahbi W, Kalinski SD, Bragazzi NL, Padulo J. Selected components of physical fitness in rhythmic and artistic youth gymnast. *Sport Sci Health.* 2020.
- Despina T, George D, George T, Sotiris P, Alessandra DC, George K, et al. Short-term effect of whole-body vibration training on balance, flexibility and lower limb explosive strength in elite rhythmic gymnasts. *Hum Mov Sci.* 2014;33:149–58.
- Allen M, Yen W. Introduction to measurement theory. [Internet]. Waveland Press. Monterey (CA): Brooks/Cole; 1979 [cited 2022 Aug 1]. 310. Available from: <https://www.waveland.com/browse.php?t=367&r=s|Introduction%20to%20measurement%20theory>
- Streiner DL, Norman GR. Biases in responding. *Health Measurement Scales.* Oxford University Press; 2008. p. 103–34.
- McGuigan M. Evaluating athletic capacities. In: Joyce D, Lewindon D, eds. *High-Performance Training for Sports, Human kinetics;* 2014.
- Bishop C, Read P, McCubbine J, Turner A. Vertical and horizontal asymmetries are related to slower sprinting and jump performance in elite youth female soccer players. *J Strength Condition Res.* 2021;35(1):56–63.
- Moir G, Shastri P, Connaboy C. Intersession reliability of vertical jump height in women and men. *J Strength Condition Res.* 2008;22(6):1779–84.
- Atkinson G, Nevill AM. Statistical methods for assessing measurement error (Reliability) in variables relevant to sports medicine. *Sports Med.* 1998;26(4):217–38.