RELIABILITY AND VALIDITY OF THE MEDICINE BALL THROW TEST IN CHILDREN AND ADOLESCENTS: A SYSTEMATIC REVIEW AND META-ANALYSIS

Daniel González-Devesa[1], Silvia Varela[2,3], Jose Carlos Diz-Gómez[3,4], Diego López-Amoedo[4], and Carlos Ayán-Pérez[2,3]

[1]*Research Group on Physical Activity, Education, and Health (GIAFES), Catholic University of Ávila, Ávila, Spain*
[2]*Well-Move Research Group, Galicia Sur Health Research Institute (IIS Galicia Sur), SERGAS-UVIGO, Vigo, Spain*
[3]*Department of Special Didactics, University of Vigo, Pontevedra, Spain*
[4]*Faculty of Education and Sport Sciences, University of Vigo, Pontevedra, Spain*

Systematic review
DOI 10.26582/k.57.1.11

Abstract:

The purpose of this study was to investigate the evidence regarding the reliability and validity of the medicine ball throw test when applied to children and adolescents. Systematic search was conducted in four electronic databases (Scopus, Web of Science, SPORTDiscus, and MEDLINE/PubMed) from their inception until November 2024. The meta-analysis of relative inter-session reliability synthesized data from nine studies using the intraclass correlation coefficient (n=293) and identified good test-retest reliability for the medicine ball throw test (ICC: 0.80; 95%CI [0.72-0.86]). When analysed separately, the standing chest medicine ball throw test (four studies; n=129) demonstrated fair test-retest reliability (ICC: 0.72; 95%CI [0.58-0.83]), while the seated chest medicine ball throw test (three studies; n=193) exhibited good test-retest reliability (ICC: 0.84; 95%CI [0.48-0.96]). Only three studies examined the relationship between the medicine ball throw test and other fitness measures. Various protocols are used to administer the medicine ball throw test in children and adolescents, with the seated chest medicine ball throw test proving to be particularly reliable due to its strong relative reliability. Although the medicine ball throw test demonstrates acceptable reliability, the limited investigation into its validity, particularly criterion validity, hinders definitive protocol recommendations.

Keywords: *evaluation, explosive power, field test, fitness, reproducibility*

Introduction

Monitoring and surveillance of physical fitness among children and adolescents is a priority in today's society due to its strong relationship with physical and mental health, as well as its association with lower morbidity and mortality later in life (Joensuu, et al. 2024). In this context, field-based tests are an essential tool, as they provide a low-cost and easy-to-administer method for assessing the fitness levels of children and adolescents across various settings (Castro-Pinero, et al. 2010). Among the wide variety of field-based tests and fitness batteries designed for this population, cardiorespiratory fitness and upper body strength are the most commonly assessed fitness dimensions (Marques, et al. 2021). Since these tests are not laboratory-based and are typically not considered gold-standard measures, they must demonstrate adequate psychometric properties when administered to children and adolescents before being considered reliable for identifying fitness levels.

Several systematic reviews have been conducted to identify the validity and reliability of various cardiorespiratory fitness tests administered to children and adolescents, given its importance as a health marker in this population (González-Devesa, Diz-Gómez, Sanchez-Lastra, Rodríguez & Ayán-Pérez, 2024; Lang, et al., 2018; Martínez-Lemos, Rodríguez, Diz, & Ayán-Pérez, 2024). However, there is a notable gap in systematic reviews specifically focused on the psychometric properties of field tests assessing upper body strength, despite established links between muscular fitness and lower adiposity, reduced cardiovascular and metabolic disease risk, improved academic outcomes, and higher quality of life in youth (Brazo-Sayavera, et al., 2024)

In this context, the medicine ball throw test (MBT), a cost-effective, quick measurement of upper-extremity power, emerges as a physical fitness field-based test worthy of further investigation. The MBT was first introduced as a fitness measurement tool in the early 1930s (Dunder, 1933) and has since been used for various purposes in research on children and youth across different contexts, such as physical education (Petrušič, Trajković, & Bogataj 2022), epidemiological and intervention studies (Han, Fu, Cobley, & Sanders, 2018; Kryst, Żegleń, Artymiak, Kowal, & Woronkowicz, 2023), or sports training and performance (Garcia-Carrillo, et al. 2023; Koya, Kitamura, & Takahashi, 2022). Its continued application for over nine decades highlights its longstanding presence and relevance in the field of physical fitness assessment.

Despite its widespread use, there is a lack of research that consolidates and critically evaluates the psychometric properties and practical applicability of the MBT as a tool for estimating muscular power in children and adolescents. Psychometric properties refer to the measurement qualities that determine whether an assessment tool is reliable, valid, responsive, and precise in capturing the construct it intends to measure (Ginty, 2020; Monticone, Galeoto, Berardi, & Tofani, 2021). This is crucial, as field tests designed for fitness assessment must undergo a thorough evaluation of its reliability and validity to ensure the credibility and robustness of the results (Domone, Mann, Sandercock, Wade, & Beedie, 2016). Under these circumstances, this systematic review and meta-analysis aimed to provide the best available scientific evidence of the reliability and validity of the MBT for estimating upper body muscular fitness in children and adolescents.

Material and methods

This systematic review was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Page, et al. 2021).

Search strategy

A systematic search was conducted in four electronic databases (Scopus, Web of Science, SPORTDiscus, and MEDLINE/PubMed) covering the period from their inception until November 2024. The following search terms, Boolean operators, and combinations were used: ("Throw* Test" OR "Ball Throw*" OR "Medicine Ball") AND ("Reliab*" OR "Consistency Agreement" OR "Valid*" OR "Reproducib*" OR "Measure*" OR "Accuracy" OR "Test-retest").

Eligibility criteria

Eligible studies included those that provided information on the validity and/or reliability of the MBT in children and/or adolescents. Investigations were excluded if they met any of the following criteria: a) the sample included adults, unless separate data were specifically available for children and adolescents, b) a ball-throw test was administered but a medicine ball was not used (e.g., a tennis ball or basketball was used instead), c) specific data on the validity/reliability of the MBT were not provided. To qualify for the initial screening, studies needed to be published or in-press in peer-reviewed journals (i.e., abstracts from conference proceedings, books, theses, and dissertations were not considered) and have an abstract available for review. Studies were not excluded based on the language of publication.

Study selection

The Rayyan software (QCRI, Qatar) was used to remove duplicate references before screening (Ouzzani, Hammady, Fedorowicz, & Elmagarmid, 2016). The authors (D.G.-D. and D.L.-A.) independently assessed the titles and abstracts of the identified studies to evaluate their eligibility. Following this initial review, the selected studies were examined by the mentioned authors to confirm inclusion. Any discrepancies were resolved through discussion and consensus. Full-text versions of potentially relevant studies were then obtained. When studies were unavailable, their authors were contacted directly.

Data extraction

A single researcher (D.G.-D.) extracted data from the original reports, including details on the country, year of the study, sample characteristics, MBT procedure and administration, reliability, time interval (when available), and validity. The other author (S.V.) then cross-verified the extracted data (see Table 1).

Quality appraisal

The quality evaluation of studies reporting data on the MBT reliability was conducted using a checklist assessing the precision in describing the sample (i), the time interval (ii), the results (iii), and the suitability of the statistical methods employed (iv). Each criterion was scored on a scale from 0 to 2. Based on the total scores, studies were classified into three categories: very low quality (score < 2), low quality (score 2–5), and high quality (score ≥ 6), following the methodology described by Cuenca-Garcia et al. (2022).

The methodological quality of studies addressing criterion-related validity was appraised through three specific criteria: sample size (i), description of the study population (ii), and statistical analysis (iii). Each of these was also rated from 0 to 2. Studies with a total score exceeding 5 were considered high quality, those scoring between 3 and 4 were labeled as low quality, and scores below 3 indicated very low quality. This classification approach aligns with the framework suggested by Castro-Piñeiro et al. (2010).

These quality tools had previously been used in other studies. The evaluation process was initially performed by one researcher (S.V.) and subsequently verified by the other (D.L.-A.). In cases of disagreement, the third researcher (D.G.-D.) provided the final decision.

Meta-analysis

We performed the meta-analysis calculations in Microsoft Excel with Meta-Essentials Workbooks (Workbook 5, correlational data 1.5) (Suurmond, van Rhee, & Hak, 2017). We selected the random effects model for all analysis to account for the sources of heterogeneity among different studies, and the forest plot was used to summarize the findings. Results of intraclass correlation coefficient (ICC) were interpreted using these ranges: 0.90-1 excellent, 0.80-0.89 good, 0.70-0.79 fair and < 0.69 poor reliability (Shrout & Fleiss 1979).

The I^2 was applied to assess statistical heterogeneity and inconsistency. An I^2 value of 0% indicates no observed heterogeneity, and higher values indicate greater heterogeneity. In addition to 95% confidence intervals (CI), we calculated the prediction intervals (PI). Funnel plot with Egger's test and trim-and-fill analysis were used to evaluate statistically the presence of any publication bias. We also conducted sensitivity analyses to test the robustness of the results.

Results

Studies selection

We obtained 2417 records from the database search (Web of Science=822, Scopus=722, SportDiscus=466, and PubMed=407). After excluding duplicates, we screened the titles and abstracts of 1241 records, and subsequently, 63 articles were retrieved for the full-text assessment. Finally, 16 studies (ten identified through the main databases and six discovered via citation searching) met the inclusion criteria and were included in the systematic review (Figure 1).

Reliability

After reviewing the selected studies, reliability data were categorized into relative (inter-session, intra-session, and inter-rater) and absolute (inter-session and intra-session) reliability.

Relative inter-session reliability. A total of nine studies informed about the inter-session reliability of the MBT (Table 1). Samples ranged from 12 (Khemiri, Teboulbi, Gritli, Hachana, & Attia, 2022) up to 105 participants (Davis, et al., 2008). All investigations involved males, except for Davis et al. (2008), Jarnig, Jaunig, Kerbl, Lima, and Van Poppel (2022), Khemiri et al. (2022), and Fjørtoft, Pedersen, Sigmundsson, and Vereijken (2011),

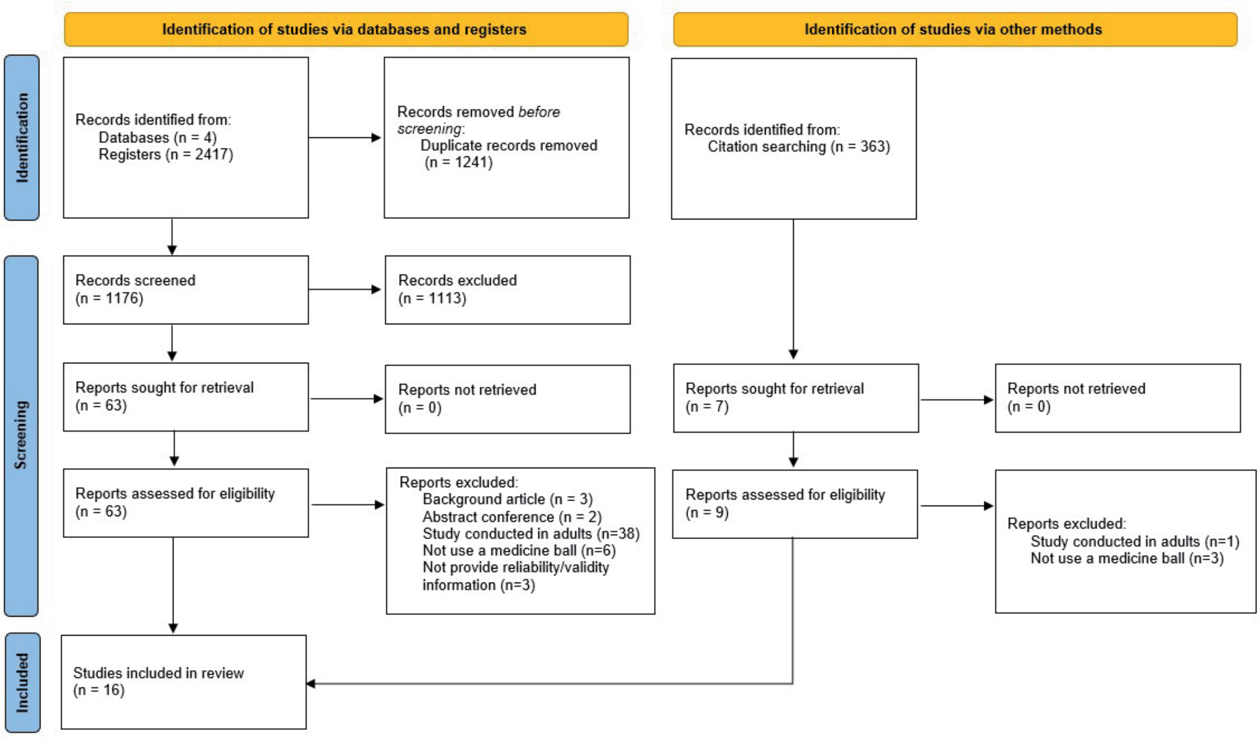

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) study flow diagram.

Table 1. Studies on the reliability and/or validity of the medicine ball test

Author (year), Country	Sample Characteristics	Test procedure and results	Reliability	Validity
Hermassi et al. (2021), Qatar	**Participants (n; sex):** 28 young handball players; 28M **Age, *years* (mean; SD):** 10.9 ± 0.72 **BMI, kg/m^2 (mean; SD):** 24.4 ± 8.01	**Test:** Seated Chest MBT **Ball weight:** 3kg **Administration:** The sitting participant had to grasp the medicine ball with both hands, and on the given signal forcefully had to push the ball from the chest. The score was calculated from the front of the sitting line to the place where the ball landed. **Distance reached, *meters* (mean; SD):** test= 4.88 ± 0.90, re-test= 4.74 ± 0.86	**Inter-session reliability** • **ICC, 95% CI:** 0.98 (0.91-1.00) • **CV (%), 95% CI:** 2.8 (2.1-4.2) **Time interval:** 14 days **Familiarization session**: All individuals were first familiarized during ad hoc preliminary sessions.	-
Davis et al. (2008), USA	**Participants (n; sex):** 105 children; 60M + 45F **Age, *years* (mean; SD):** 5.54 ± 0.5 **BMI, kg/m^2 (mean; SD):** 17.44 ± 3.17	**Test:** Seated Chest MBT **Ball weight:** 2lb **Administration:** Each student was seated on the floor with their back against the wall, holding the ball with both hands resting on their lap. Upon the tester's command ("go"), they were instructed to lift the ball to their chest and throw it forward with maximum force. Each participant completed two practice throws, followed by three recorded throws, with a 1- to 2-minute rest interval between each attempt. **Distance reached, *cm* (mean; SD):** test= 123.41 ± 41.97, re-test= 126.01 ± 39.11	Inter-session reliability • **ICC:** 0.88 **Time interval: 7** days **Familiarization session**: NR **Intra-session reliability:** • **ICC, test:** 0.93 • **ICC, re-test:** 0.94	Concurrent validity - Correlation with: ***Modified pull-up test:*** r=-0.04
Carron et al. (2024), Austria	**Participants (n; sex):** 47 young rugby players; 47M **Age, *years* (mean; SD):** 16.2 ± 1.3 **BMI, kg/m^2:** NR	**Test:** Seated Chest MBT **Ball weight:** 2kg **Administration:** Players were seated on a flat bench with their backs in contact with an adjacent wall, knees flexed to 90°, and feet in full contact with the floor. Players were not permitted to break contact with the bench or floor during each throw. **Distance reached, *meters* (mean; SD):** test= 7.06 ± 0.64, re-test= 7.07 ± 0.71	Inter-session reliability • **ICC, 95% CI:** 0.733 (0.566-0.842) • **CV (%):** 3.72 • **SEM:** 0.347 • **MDC:** 0.963 **Time interval:** 7 days **Familiarization session**: NR	-
Biggar et al. (2022), USA	**Participants (n; sex):** 113 adolescents; 56M + 57F ***12-13 years M:*** 25 ***12-13 years F:*** 31 ***13-15 years M:*** 29 ***13-15 years F:*** 25 **Age, *years* (range):** 12-15 **BMI, kg/m^2 (mean; SD):** ***12-13 years M:*** 21.5 ± 5.9 ***12-13 years F:*** 21.2 ± 5.9 ***13-15 years M:*** 22.6 ± 5.1 ***13-15 years F:*** 21.7 ± 3.2	**Test:** Seated Chest MBT **Ball weight:** 2kg **Administration:** Participants started by sitting at a 90° angle against a designated wall with their legs straight out and their head resting on the wall. Participants pushed the medicine ball in a chest-pass motion as forcefully as possible without their back or their head leaving the wall. **Distance reached, *meters* (mean; SD):** ***12-13 years M:*** 4.3 ± 0.7 ***12-13 years F:*** 3.4 ± 0.5 ***13-15 years M:*** 5.2 ± 0.8 ***13-15 years F:*** 3.7 ± 0.5	**Intra-session reliability:** • **r:** 0.85-0.97 **Familiarization session**: NR	-
Chiwaridzo et al. (2021), Zimbabwe	**Participants (n; sex):** 41 young rugby players; 41M **Age, *years* (mean; SD):** 17.5 ± 0.9 **BMI, kg/m^2 (mean; SD):** 25.9 ± 3.3	**Test:** Seated Chest MBT **Ball weight:** 2kg **Administration:** Players threw a medicine ball (dimensions=21.5cm) horizontally as far as possible while seated with the back, and legs straight. **Distance reached, *meters* (mean; SD):** test= 9.3 ± 1.3, re-test= 9.41 ± 1.3	**Inter-session reliability** • **ICC, 95% CI:** 0.89 (0.80-0.94) • **CV (%), 95% CI:** 4.48 (3.4-6.1) • **SEM, 95% CI:** 0.42 (0.2-0.67) • **MDC, 95% CI:** 1.16 (0.98-1.27) **Time interval: 7** days **Familiarization session**: Two familiarization sessions were conducted	-
Chiwaridzo et al. (2019), Zimbabwe	**Participants (n; sex):** 100 adolescents; 100M ***U16 elite rugby players:*** 41 ***U16 sub-elite rugby players:*** 30 ***U16 cricket players:*** 29 **Age, *years* (mean; SD):** 14.9 ± 0.31 **BMI, kg/m^2:** NR	**Test:** Seated Chest MBT **Ball weight:** 2kg **Administration:** Players threw a medicine ball (dimensions=21.5cm) horizontally as far as possible while seated with the back, and legs straight. **Distance reached, *meters* (mean; SD):** ***U16 elite rugby players:*** 6.97 ± 0.64 ***U16 sub-elite rugby players:*** 5.91 ± 0.86 ***U16 cricket players:*** 5.83 ± 0.86	**Subsample =** 41 U16 elite rugby players **Intra-session reliability** • **ICC:** 0.91 • **CV (%):** 1.45 • **Time interval:** 2 minutes **Familiarization session**: Two familiarization sessions were conducted	-

Hackett et al. (2017), Australia	**Participants (n; sex):** 190 adolescents; 130M + 60F **Age, *years* (mean; SD):** ***M:*** 12.8 ± 0.5 ***F:*** 12.8 ± 1.0 **BMI, *kg/m²*:** NR	**Test:** Seated Chest MBT **Ball weight:** 3kg **Administration:** Participants sat on an upright bench. The seat position was adjusted so that their knees were at approximately 90 degrees of flexion, with their feet flat on the floor. Participants were instructed to push a medicine ball from the center of their chest as far as possible, keeping their head, shoulders, and lower back in contact with the bench. A 10-meter tape measure was placed alongside the bench, and a researcher visually assessed the initial ground contact point of the ball for each attempt. **Distance reached, *cm* (mean; SD):** test= 339.9 ± 75.8, re-test= 348.5 ± 78.5	**Intra-session reliability** • **ICC, 95% CI:** 0.98 (0.96-0.98) • **CV (%), 95% CI:** 3.4 (2.9-4) • **ME:** 7.4 • **TE (%):** 2.2 **Time interval: <**1 day **Familiarization session**: Participants were allowed one to two practice attempts to familiarize themselves with the procedure before beginning the actual attempts.	-
Jarnig et al. (2022), Austria	**Participants (n; sex):** 821 children; 58M + 28F **Age, *years* (mean; SD):** 8.3 ± 0.7 **BMI, *kg/m²* (mean; SD):** 16.9 ± 2.9	**Test:** Standing Chest MBT **Ball weight:** 1kg **Administration:** Each child stood on a starting line holding a medicine ball with both hands, the ball touching their chest, then threw the ball with both hands as far forward as possible. **Distance reached, *cm* (mean; SD):** test= 411 ± 84, re-test= 436 ± 76	**Subsample=** 17 children **Inter-session reliability** • **ICC, 95% CI:** 0.70 (0.35-0.88) **Time interval: 7** days **Familiarization session**: NR Inter-rater reliability • **ICC, 95% CI:** 0.99 (0.99-0.99)	-
Bös et al. (2001) Germany	**Participants (n; sex):** 138; Kindergarten: 40 Elementary school: 50 Secondary school: 48 **Age, *years* (range):** 6-10 **BMI, *kg/m²*:** NR	**Test:** Standing Chest MBT (with upper body swing) **Ball weight:** 1kg **Administration:** Participants were instructed to hold a medicine ball with bent arms in front of their chest and to push the ball as far as possible from a standing position. They stood upright with their back against the wall, with the tips of their shoes touching the starting line. Swinging the upper body up to the wall was permitted. **Distance reached, *cm* (mean; SD):** Reported by age and sex	**Subsample=** 38 children **Inter-session reliability** • **r:** 0.73-0.85 **Time interval:** 10 days **Familiarization session**: NR	-
Dobbin et al. (2018), UK	**Participants (n; sex):** 50 young rugby players; 50M **Age, *years* (mean; SD):** 17.1 ± 1.1 **BMI, *kg/m²*:** NR	**Test:** Standing Chest MBT (with squat) **Ball weight:** 4kg **Administration:** Participants began standing upright with the ball above their head. They then lowered the ball toward their chest while squatting down to a self-selected depth before extending up onto their toes and pushing the ball as far as possible. Feet remained shoulder width apart, stationary, and behind a line that determined the start of the measurement. **Distance reached, *meters* (mean; SD):** ***Trial 1***= 6.4 ± 0.8 ***Trial 2***= 6.9 ± 0.7 ***Trial 3***= 6.6 ± 1	**3 separate occasions* **Inter-session reliability** • **ICC, 95% CI:** 0.74 (0.57-0.84) • **CV (%), 95% CI:** 9 (7.9-10.5) • **TE, 95% CI:** 0.5 (0.4-0.6) • **SWC:** 0.2 **Time interval:** 5-14 days **Familiarization session**: NR **Comparisons, *Trial 1 vs 2:*** • **ICC, 95% CI:** 0.71 (0.28-0.86) • **CV (%), 95% CI:** 6.9 (5.8-8.3) • **TE, 95% CI:** 0.5 (0.4-0.5) **Comparisons, *Trial 1 vs 3:*** • **ICC, 95% CI:** 0.5 (0.11-0.71) • **CV (%), 95% CI:** 14 (11.9-17.1) • **TE, 95% CI:** 0.7 (0.6-0.85) **Comparisons, *Trial 2 vs 3:*** • **ICC, 95% CI:** 0.73 (0.5-0.85) • **CV (%), 95% CI:** 10.7 (9.1-13.1) • **TE, 95% CI:** 0.5 (0.43-0.6)	-

Fjørtoft et al. (2011), Norway	**Participants (n; sex):** 195 children; 94M +101F ***5 years:*** 21 ***6 years:*** 21 ***7 years:*** 21 ***8 years:*** 21 ***9 years:*** 21 ***10 years:*** 21 ***11 years:*** 21 ***12 years:*** 21 **Age, *years* (mean; SD):** 8.3 ± 2.21 **BMI, *kg/m^2* (mean; SD):** ***5 years:*** 15.4 ± 1.6 ***6 years:*** 16.8 ± 2.01 ***7 years:*** 16.6 ± 1.46 ***8 years:*** 17.8 ± 1.84 ***9 years:*** 17.6 ± 1.58 ***10 years:*** 16.7 ± 1.89 ***11 years:*** 18.5 ± 3.22 ***12 years:*** 19.1 ± 2.38	**Test:** Standing Chest MBT **Ball weight:** 1kg **Administration:** The starting position is with the feet parallel to each other and shoulder width apart, with the ball held against the chest. **Distance reached, *meters* (mean; SD):** test= 3.34 ± 0.47, re-test= 3.44 ± 0.53	**Subsample =** 24 children **Inter-session reliability** • **ICC, 95% CI:** 0.54 (0.18-0.77) • **SEM, 95% CI:** 0.34 (0.28-0.45) **Time interval:** 7 days **Familiarization session**: NR	**Concurrent Validity - Correlation with:** ***Standing broad jump:*** r= 0.72 ***Jumping on 2 feet:*** r=0.43 ***Jumping on 1 foot:*** r=0.41 ***Throwing a tennis ball:*** r=0.80 ***Climbing wall bars:*** r=0.83 ***Shuttle run:*** r=0.72 ***Running 20 m:*** r=0.77 ***Reduced Cooper test:*** r=0.54
Luna-Villouta et al. (2022), Chile	**Participants (n; sex):** 86 young tennis players; 58M + 28F **Age, *years* (mean; SD):** 15.4 ± 0.8 **BMI, *kg/m^2*:** NR	**Test:** Standing Overhead MBT **Ball weight:** 3kg **Administration:** The players stood at a line facing the throwing direction, feet slightly apart. They brought the ball behind their head with both hands and threw it forward as far as possible without moving their feet or crossing the line for the overhead MBT. **Distance reached, *meters* (mean; SD):** ***Total sample:*** test= 7 ± 1.5, re-test= 6.9 ± 1.5 ***M:*** test= 7.7 ± 1.3, re-test= 7.6 ± 1.3 ***F***: test= 5.7 ± 1, re-test= 5.7 ± 1	**Intra-session reliability** • SEM: ***Total sample:*** 0.07; ***M:*** 0.09; ***F***: 0.03 • **MDC** ***Total sample:*** 0.16; ***M:*** 0.20 **Time interval:** < 1 day **Familiarization session**: NR	-
van den Tillaar & Marques (2011), Portugal	**Participants (n; sex):** 63 adolescents; 24M + 39F **Age, *years* (mean; SD):** 16.5 ± 1.8 **BMI, *kg/m^2*:** NR	**Test:** Standing Overhead MBT **Ball weight:** 1kg and 3kg **Administration:** Participants stood with feet parallel, holding the ball with both hands in front. They performed 2-handed overhead throws, aiming for maximum distance and speed, while keeping both feet grounded and avoiding torso or hip rotation. Non-compliant attempts were disqualified and repeated. The maximal throwing velocity was determined using a Doppler radar gun **Throwing velocity, *$m \cdot s^{-1}$*:** NR	**Intra-session reliability** • **ICC *(1-kg medicine ball):*** 0.93 • **ICC *(3-kg medicine ball):*** 0.86 **Time interval:** <1 day **Familiarization session**: Participants underwent a practice session to familiarize themselves with 2-handed overhead throws using different weighted balls.	-
Khemiri et al. (2022), Tunisia	**Participants (n; sex):** 12 young volleyball players; 12M **Age, *years* (mean; SD):** 16.5 ± 0.52 **BMI, *kg/m^2*:** 21.63 ± 1.8	**Test:** Backwards Overhead MBT **Ball weight:** 3kg **Administration:** Participants started standing, feet shoulder-width apart, facing backward. The starting line was set at heel level. Holding the medicine ball with straight arms, they performed a countermovement by bending hips and knees, lowering the ball below hip level, and then extending to throw it backward. **Distance reached, *meters* (mean; SD):** test= 11.08 ± 0.72, re-test= 11.27 ± 0.93	**Inter-session reliability** • **ICC, 95% CI:** 0.83 (0.393-0.950) • **CV (%):** 5.7 **Time interval:** 1 day **Familiarization session**: NR	-
Duncan & Hankey (2010), UK	**Participants (n; sex):** 47 adolescents; 22M + 25F **Age, *years* (mean; SD):** 12.7 ± 1.5 **BMI, *kg/m^2*:** NR	**Test:** Backwards Overhead MBT **Ball weight:** 3kg **Administration:** The exercise involved participants standing with feet shoulder-width apart, holding a medicine ball at shoulder height. They performed a countermovement by bending hips, knees, and trunk to lower the ball below hip height. Then, they extended their hips, knees, and trunk, raising the ball back to shoulder height and throwing it overhead, allowing their feet to leave the ground to minimize deceleration. **Distance reached, *meters* (mean; SD):** 5.6 ± 1.3	**Intra-session reliability** • **ICC, 95% CI:** 0.89 (0.71-0.96) • **CV (%):** 6.6 **Familiarization session**: Participants were familiarized with the tests	**Concurrent Validity- Correlation with:** ***Peak power of the CMJ:*** r=0.806 ***Peak power of the WSJ:*** r= 0.632 ***Relative power output of the CMJ:*** r= 0.79 ***Relative power output of the WSJ:*** r=0.203

Duncan et al. (2005), UK	**Participants (n; sex):** 28 adolescents rugby players; 28M **Age, *years* (mean; SD):** 15.1 ± 0.5 **BMI, *kg/m²*:** NR	**Test:** Backwards Overhead MBT **Ball weight:** 3kg **Administration:** The exercise involved participants standing with feet shoulder-width apart, holding a medicine ball at shoulder height. They performed a countermovement by bending hips, knees, and trunk to lower the ball below hip height. Then, they extended their hips, knees, and trunk, raising the ball back to shoulder height and throwing it overhead, allowing their feet to leave the ground to minimize deceleration. **Distance reached, *meters* (mean; SD):** ***Trial 1***= 9.7 ± 1.4 ***Trial 2***= 10.1 ± 1.7 ***Trial 3***= 9.6 ± 1.5 ***Trial 4***= 10 ± 1.5 ***Trial 5***= 10.5 ± 1.4 ***Trial 6***= 10.6 ± 1.4	**Intra-session reliability** • **ICC, *Trial 5 vs 6*:** 0.983 • **LOA (%), *Trial 5 vs 6*:** 3.3 • **TE, *Trial 1 vs 2*:** 0.84 • **TE, *Trial 2 vs 3*:** 0.66 • **TE, *Trial 3 vs 4*:** 0.48 • **TE, *Trial 4 vs 5*:** 0.31 • **TE, *Trial 5 vs 6*:** 0.10 **Familiarization session**: NR	-

Note. BMI: body mass index; CI: confidence interval; CMJ: counter movement jump; CV: coefficient of variation; F: female; ICC: intraclass correlation coefficient; LOA: limits of agreement; M: male; MBT: medicine ball throw; MDC: minimal detectable change; ME: measurement error; NR: not reported; SD: standard deviation; SDC: smallest detectable change; SEM: standard error of measurements; SWC: smallest worthwhile change; TE: Typical Error; WSJ: weighted (7kg bar) squat jump.

which involved mixed samples. However, Bös et al. (2001) did not provide information regarding participants' sex. Participants' age ranged from 5 to 18 years, with three investigations including preschoolers (Bös, et al., 2001; Davis, et al., 2008; Fjørtoft, et al., 2011).

A wide variety of protocols were identified. Four studies provided data on the seated chest MBT. Of these, three utilized a seated position on the floor (Chiwaridzo, Ferguson, & Smits-Engelsman, 2019; Davis, et al., 2008; Hermassi, van den Tillaar, Bragazzi, & Schwesig, 2021), while the remaining study involved participants seated on a bench (Carron, Scanlan, & Doering, 2024). The weights of the medicine balls varied, ranging from 0.90 kg (2 lb) (Davis, et al., 2008) to 3 kg (Hermassi, et al. 2021). Similarly, four studies provided information on the standing chest MBT. In two of these studies, participants were not allowed to use any upper body swing (Fjørtoft, et al., 2011; Jarnig, et al., 2022), one investigation permitted upper body swing (Bös, et al., 2001), and the remaining research allowed participants to perform a squat to generate additional force (Dobbin, Hunwicks, Highton, & Twist, 2018). The weights of the medicine balls used in these studies ranged from 1 kg (Bös, et al., 2001; Fjørtoft, et al., 2011; Jarnig, et al., 2022) to 4 kg (Dobbin, et al., 2018). Finally, one investigation provided information on the backwards overhead MBT (Khemiri, et al., 2022) using a 3 kg medicine ball.

Time interval was reported in all studies, with a duration between five (Dobbin, et al. 2018) to 14 days (Dobbin, et al., 2018; Hermassi, et al., 2021). Information regarding familiarization sessions was only provided in two out of the nine studies (Chiwaridzo, et al., 2021; Hermassi, et al., 2021). Seven studies evaluated the consistency of test results using intraclass correlation coefficients, which ranged from 0.54 to 0.98. The remaining study reported Pearson's correlation coefficient values, which ranged from 0.73 to 0.85 (Bös, et al., 2001).

The pooled data from nine samples (n= 293) indicated a good test-retest reliability for the MBT (ICC*:* 0.80; 95% CI [0.72-0.86]) (Figure 2), without evidence of publication bias (Egger 1.22 [0.45; 2], p=0.74; trim and fill= 0). When the data were meta-analyzed separately, the standing chest MBT (four studies; n=129) demonstrated a fair test-retest reliability (ICC: 0.72; 95% CI [0.58–0.83]) while the seated chest MBT (three studies; n=193), showed good test-retest reliability (ICC: 0.84; 95% CI [0.48–0.96]). Additionally, the standing version exhibited higher heterogeneity (I^2=63%) than the seated version (I^2=0%) (Figure 3).

Relative intra-session reliability. Seven studies assessed the intra-session reliability of the MBT (Table 1), with sample sizes ranging from 28 participants (Duncan, Al-Nakeeb, & Nevill, 2005) to 190 participants (Hackett, Davies, Ibel, Cobley, & Sanders, 2017). Most studies involved mixed samples, except for Chiwaridzo et al. (2019) and Duncan et al. (2005), which included only male participants. While adolescents formed the primary focus of these studies, Davis et al. (2008) uniquely included children aged 5-6 years.

In terms of test procedures, three investigations evaluated the seated chest MBT with participants seated on the floor (Biggar, Larson, & Debeliso, 2022; Chiwaridzo, et al., 2019; Davis, et al., 2008), while one study assessed the seated chest MBT with participants seated on a bench (Hackett, et al., 2017). The weight used in these studies ranged from 0.9 kg (2 lbs) to 3 kg. Additionally, one study reported data on the standing overhead MBT (van den Tillar & Marques 2011) with medicine balls weighing 1 kg and 3 kg. Finally, two other studies (Duncan, et al., 2005; Duncan & Hankey, 2010) evaluated the intra-session reliability of the back-

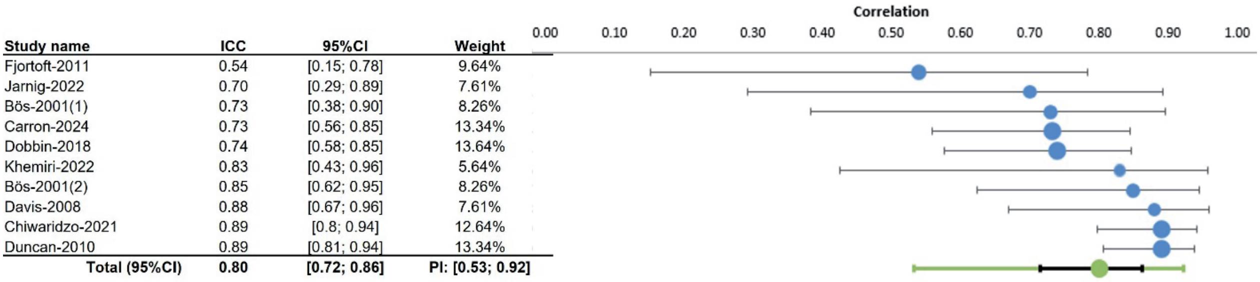

Figure 2. Forest plot for the test-retest reliability. A total of 281 participants in nine different samples. Random effects model. I^2=57%. ICC: intraclass correlation coefficient, CI: confidence interval, PI: prediction interval.

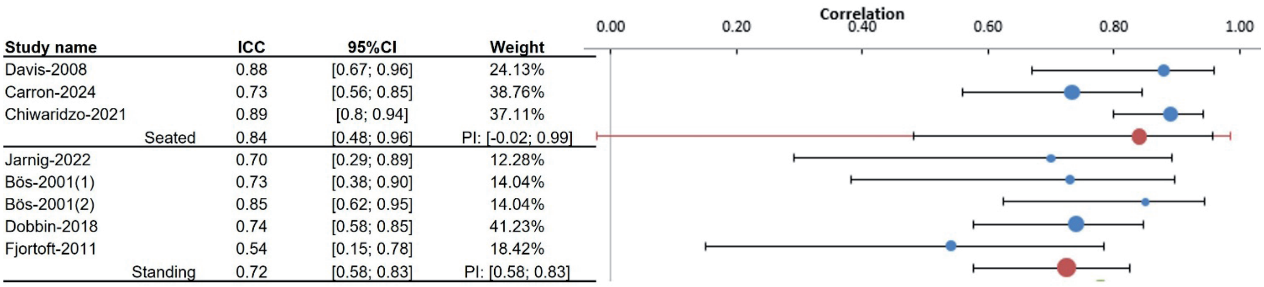

Figure 3. Forest plot for the test-retest reliability in subgroups of different samples according to throwing method. A total of 234 participants in eight studies. Random effects model. I^2=63% for seated and I^2=0% for standing. ICC: intraclass correlation coefficient, CI: confidence interval, PI: prediction interval.

wards overhead MBT using a 3 kg medicine ball.

Five studies employed ICCs to measure consistency, reporting values between 0.86 and 0.98. One study, Biggar et al. (2022), used Pearson's correlation coefficients, which ranged from 0.85 to 0.97.

Relative inter-rater reliability. One research analysed the relative inter-rater reliability of the MBT (Jarnig, et al., 2022) in a mixed sub-sample of 17 children. The study utilized the standing chest MBT with a 1 kg medicine ball. Inter-rater reliability was exceptionally high (ICC: 0.99; 95% CI [0.99–0.99]).

Absolute inter-session reliability. Six out of the 16 studies included in this review reported data on the absolute inter-session reliability of the MBT. Hermassi et al. (2021) evaluated the seated chest MBT performed on the floor with a 3 kg medicine ball and observed a coefficient of variation (CV) of 2.8% (95% CI: 2.1–4.2). Similarly, Chiwaridzo et al. (2021) analysed the same protocol with a 2 kg medicine ball and reported a CV of 4.48% (95% CI: 3.4–6.1), standard error of measurement (SEM) of 0.42 (95% CI: 0.2–0.67), and minimum detectable change (MDC) of 1.16 (95% CI: 0.98–1.27). Carron et al. (2024) examined the seated chest MBT performed on a bench and found a CV of 3.72%, SEM of 0.347, and MDC of 0.963.

For the standing chest MBT, Fjørtoft et al. (2011) observed a SEM of 0.34 (95% CI: 0.28–0.45) when the test was performed with 1 kg medicine ball. Dobbin et al. (2018) investigated the standing chest MBT with squat assistance using a 4 kg medicine ball and reported a CV of 9% (95% CI: 7.9–10.5), a technical error (TE) of 0.5 (95% CI: 0.4–0.6), and a smallest worthwhile change (SWC) of 0.2.

The only study that provided data on the backwards overhead MBT was conducted by Khemiri et al. (2022), who reported a CV of 5.7%.

Absolute intra-session reliability. Five investigations reported data on the absolute intra-session reliability of the MBT. Chiwaridzo et al. (2019) investigated the seated chest MBT performed on the floor using a 2 kg medicine ball and found a CV of 1.45%. Hackett et al. (2017) analysed the same test performed on a bench with a 3 kg medicine ball. Their findings included a CV of 3.4% (95% CI: 2.9–4), a measurement error (ME) of 7.4, and a TE of 2.2%.

Two additional studies focused on the backwards overhead MBT protocol, both using a 3 kg medicine ball. Duncan and Hankey (2010) observed a CV of 6.6%, while Duncan et al. (2005) examined reliability across repeated trials (Trial 5 vs. Trial 6) and reported limits of agreement (LOA) of 3.3% and a TE of 0.10.

Only one study analysed the absolute intra-session reliability of the standing overhead MBT (Luna-Villouta, et al., 2022) using a 3 kg medicine ball. The SEM was 0.07 for the total sample, 0.09 for males, and 0.03 for females. The MDC was 0.16 for the total sample and 0.20 for males (data for females were not reported).

Validity

A total of three studies provided information on the validity of the MBT (Table 1). Sample size ranged from 47 (Duncan & Hankey, 2010) to 195 (Fjørtoft, et al., 2011) participants and the three studies involved mixed samples (male and female). According to the studies' initial sample, the participants' age ranged from 5 (Davis, et al., 2008; Fjørtoft, et al., 2011) to 14 years old (Duncan & Hankey, 2010). Results by age group were not specifically reported. The association between the MBT and various fitness measures was reported with validity coefficients ranked from highest to lowest. Fjørtoft et al. (2011) observed the strongest correlations for the standing chest MBT with climbing wall bars (r=0.83), throwing a tennis ball (r=0.80), standing broad jump and shuttle run (both r=0.72), running 20 m (r=0.77), and the reduced Cooper test (r=0.54). Duncan and Hankey (2010) found strong correlations for the backwards overhead MBT with the peak power of the CMJ (r=0.806) and relative power output of the CMJ (r=0.79), while weaker correlations were reported for the peak power (r=0.632) and relative power output (r=0.203) of the WSJ. Finally, Davis et al. (2008) identified a weak correlation with the modified pull-up test (r=0.04).

Quality appraisal

Regarding reliability, six out of the 16 analyzed studies were classified as high quality (Carron, et al., 2024; Chiwaridzo, et al., 2021; Davis, et al., 2008; Hackett, et al., 2017; Hermassi, et al., 2021; Khemiri, et al., 2022), whereas the remaining studies were rated as low quality (Table 2). The three investigations that provided data on the MBT validity were deemed to be of low (Duncan & Hankey, 2010) or very low quality (Davis, et al., 2008; Fjørtoft, et al., 2011) (Table 3).

Table 2. Quality assessment criteria for reliability studies (n=16)

Study		Description of the participants	Time interval	Description of results	Statistical analysis	Total score
Hermassi et al. (2021)	Inter-session	1	2	1	2	6
Davis et al. (2008)	Inter-session	1	2	1	2	6
	Intra-session	1	2	0	2	5
Carron et al. (2024)	Inter-session	1	2	2	2	7
Biggar et al. (2022)	Intra-session	1	2	0	0	3
Chiwaridzo et al. (2021)	Inter-session	1	2	2	2	7
Chiwaridzo et al. (2019)	Intra-session	1	2	0	2	5
Hackett et al. (2017)	Intra-session	2	1	1	2	6
Luna-Villouta et al. (2022)	Intra-session	1	1	1	2	5
van den Tillaar & Marques (2011)	Intra-session	1	1	1	2	5
Khemiri et al. (2022)	Inter-session	1	2	1	2	6
Jarnig et al. (2022)	Inter-session	0	2	1	2	5
	Inter-rated	0	2	1	2	5
Bös et al. (2001)	Inter-session	0	2	1	0	3
Dobbin et al. (2018)	Inter-session	1	1	1	2	5
Fjørtoft et al. (2011)	Inter-session	0	2	1	2	5
Duncan & Hankey (2010)	Intra-session	1	2	0	2	5
Duncan et al. (2005)	Intra-session	0	2	1	2	5

Note. ≥6: high quality; 2-5: low quality; <2: very low quality.

Table 3. Quality assessment criteria for criterion-related validity studies (n=3)

Study	Number of study subjects	Description of the study population	Statistical analysis	Total score
Davis et al. (2008)	2	0	0	2
Fjørtoft et al. (2011)	1	0	0	1
Duncan & Hankey (2010)	1	1	1	3

Note. >5: high quality; 3-4: low quality; <3: very low quality.

Discussion and conclusions

Researchers and practitioners must accurately measure youth muscular strength with results supported by evidence of reliability and validity. This review presents findings on both psychometric dimensions of the MBT when administered to children and adolescents, while also evaluating its accuracy and practical utility.

According to Shrout and Fleiss (1979), relative reliability is classified as excellent for ICC values between 0.90 and 1.00, good for 0.80–0.89, fair for 0.70–0.79, and poor for values below 0.69. We identified a substantial number of studies investigating the inter-session test-retest reliability of the MBT, with pooled data demonstrating good agreement. These results are consistent with findings from previous systematic reviews on upper-body physical performance tests (ICC>0.70) (Barbosa, Calixtre, Fialho, Locks, & Kamonseki, 2024), and ball throwing velocity and distance tests (ICC ≥ 0.76) (Paraskevopoulos, Pamboris, Plakoutsis, & Papandreou, 2023). Other well-known upper-body field-based muscular strength tests such as the handgrip test or the basketball throw test yielded higher coefficients (ICC>0.90) when administered to children (Belhaidas, Dahoune, Eather, & Oukebdane, 2021; Fernandez Santos, Ruiz, Gonzalez-Montesinos, & Castro-Piñero, 2016). In contrast, lower inter-session test-retest reliability coefficients have been reported for commonly used upper-body strength assessments included in the Fitnessgram® battery, such as the flexed arm hang test (k: 0.60) and the push-up test (k: 0.48) (Morrow, Martin, & Jackson, 2010).

Regarding other muscular power field-based tests available for young populations, the existing scientific evidence primarily focuses on the lower body extremities. In this regard, inter-session test-retest reliability was generally higher than that observed for the MBT, with the standing long jump test showing an ICC of 0.95 when administered to children and adolescents (Fernandez-Santos, Ruiz, Cohen, Gonzalez-Montesinos, & Castro-Piñero, 2015), and the Sargent jump test reporting ICC values ranging from 0.84 to 0.95 when performed by preschoolers (Ayán-Pérez, Cancela-Carral, Lago-Ballesteros, & Martínez-Lemos, 2017).

An interesting finding of this research is that the position in which the MBT is performed appears to influence its inter-session test-retest reliability, with meta-analyzed data revealing higher coefficient values for the seated chest tests using a 2 kg ball (n=3) compared to the standing chest tests (1 kg: n=3; 4 kg: n=1). The inter-session test-retest reliability of the seated chest MBT has been reported to be good for young adults (ICC: 0.77) (Ferreira et al., 2021), high for older adults (ICC>0.90) (Harris, et al., 2011) and very high for university students (ICC: 0.97) (Harasin, Dizdar, & Marković, 2006). Regarding the standing MBT, the literature is limited, with studies reporting good values for the overhead MBT among university students (r: 0.84) (Rosni, Abas, & Mohamad, 2014) and high reliability coefficients for the backward overhead MBT in volleyball players (r>0.9) (Stockbrugger & Haennel, 2001). In any event, our findings are in line with previous research, suggesting a higher reliability for the seated vs. the standing MBT (van den Tillaar & Marques, 2013). Based on these findings and considering that the seated chest SMBT demonstrated good predictive ability for muscular strength and power in adolescents (Hackett, et al., 2017), this test appears to be the most suitable option currently available for use.

The MBT is a quick and easy test to administer. However, a familiarization process is essential for achieving stable scores (Duncan, et al., 2005). Notably, very few of the reviewed investigations reported including a familiarization phase, despite its importance as a crucial factor in ensuring the reliability of fitness field-based tests (González-Devesa, et al., 2024). This omission is particularly relevant for practitioners, as failing to incorporate a familiarization session may compromise the accuracy of test results by introducing avoidable variability due to a lack of prior exposure to the task. In applied settings, especially with children and adolescents unfamiliar with maximal throwing tasks, this step becomes essential to reduce learning effects and improve measurement consistency. On the other hand, several studies informed on the intra-session reliability of the MBT, which was considered to be high. This is a noteworthy finding, as this type of reliability reflects the random variability, inaccuracy of the measurement itself, or variations in the performance of the subject and the tested phenomenon (Bauer, Gröger, Rupprecht, & Gaßmann, 2008). The intra-session reliability values observed in this review align with results from prior research conducted with college students performing various medicine ball throw distance tests (ICC: 0.88-0.97) (van den Tillaar & Marques, 2009, 2013).

In contrast, only one of the reviewed studies reported data on inter-rater reliability, which was deemed very high. The scarcity of research on inter-rater reliability for upper-body physical performance and throwing distance tests has been previously noted (Barbosa, et al., 2024; Paraskevopoulos, et al., 2023), highlighting a potential area for future investigation.

Absolute reliability is a critical psychometric property in sports science, as it aids in predicting the magnitude of a "real" change following an intervention (Atkinson & Nevill, 1998). Several of the reviewed investigations explored this aspect, with their findings corroborating those of previous studies. For instance, Ferreira et al.

(2021) reported a SEM of 13 cm and a MDC of 36 cm when analyzing inter-session reliability of the seated MBT among young adults. Similarly, Borms and Cools (2018) found a SEM of 11 cm and an MDC of 30 cm when investigating the intra-session reliability of the overhead MBT performed by athletes. Regarding the CV, the results from the reviewed studies were comparable to those reported for undergraduate students performing the seated MBT (intra-session CV%: 3.2–4.7%) (Beckham, et al., 2019) and the standing overhead MBT (inter-session CV%: 7.82%) (Rosni, et al., 2014). Additionally, Johnson, Maurya, Sisneros, Ford, & Palmer, (2024) observed a CV of approximately 9% when analyzing the inter-session reliability of the supine MBT performed by young women. All together, these findings support selecting the MBT for assessing the effectiveness of strategies aimed at improving upper-body muscular strength among children and adolescents.

Physical fitness field-based tests must exhibit strong validity to ensure they accurately measure the intended variable. Ideally, these tests should demonstrate robust criterion-related validity, which reflects the degree to which the test correlates with the criterion measure, often referred to as the gold standard (Currell & Jeukendrup, 2008). In this sense, the available scientific evidence on the validity of the MBT remains limited. Indeed, only a few studies addressing the validity of this test were identified, and these lacked methodological rigor. Furthermore, these investigations primarily focused on convergent validity, comparing MBT scores with those from other fitness tests. The reported validity coefficients varied widely, underscoring the necessity for further rigorous research into this psychometric property.

In this sense, other investigations have tried to explore the criterion-related validity of the MBT through the use of different strategies. For instance, Leite et al. (2016) compared upper limb muscle power measured using the MBT with performance in the bench press exercise, utilizing a Myotest® accelerometer attached to the bench press bar to estimate muscle power in a sample of rugby players. In this line, Harris et al. (2011) proposed using an explosive push-up test on a floor-mounted force platform as a criterion measure for assessing the validity of the MBT among older adults. In this line, Roe et al. (2018) introduced the innovative approach of embedding an accelerometer in a medicine ball to establish a criterion measure for monitoring upper-body neuromuscular performance.

From a practical standpoint, the seated chest MBT is recommended for use in school and general physical education contexts, especially with younger children or in large group settings, due to its simplicity and high inter-session reliability. For adolescents involved in competitive or performance-focused programs, standing or overhead variations may offer a more sport-specific alternative, although standardization and familiarization remain essential. Therefore, professionals should select the protocol that best aligns with their assessment objectives, available resources, and participant characteristics.

This review represents the first systematic evaluation and meta-analysis of the psychometric properties of the MBT when administered to children and adolescents. Despite its originality, several limitations must be acknowledged. Firstly, there was a lack of consensus on the MBT protocols investigated, and in some cases, the statistical approaches were insufficiently detailed. Secondly, the large heterogeneity in procedures, including body position, ball weight, number of attempts, and the presence or absence of familiarization protocols, should be noted, as it may significantly influence inter-session reliability and complicates comparisons across studies. Thirdly, the small number of validity studies identified precluded the possibility of conducting a meta-analysis on this outcome. Fourthly, most studies did not report inter-rater reproducibility data, limiting the assessment of consistency between evaluators. Lastly, although a comprehensive search was undertaken, the omission of specific grey literature databases may have limited the scope of studies included.

In conclusion, various protocols exist for administering the MBT to children and adolescents, with the seated chest MBT emerging as a particularly reliable option due to its good relative reliability. While different investigations report acceptable absolute reliability values for the MBT, the limited research on its validity, especially criterion validity, hinders a conclusive determination of the most appropriate protocol for assessing upper-body muscular strength in this population. Therefore, findings related to the validity should be interpreted with caution, and further research in this area is strongly recommended.

References

Atkinson, G., & Nevill, A.M. (1998). Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Medicine*, *26*(4), 217-238. doi: 10.2165/00007256-199826040-00002

Ayán-Pérez, C., Cancela-Carral, J.M., Lago-Ballesteros, J., & Martínez-Lemos, I. (2017). Reliability of Sargent jump test in 4- to 5-year-old children. *Perceptual and Motor Skills*, *124*(1), 39-57. doi: 10.1177/0031512516676174

Barbosa, G.M., Calixtre, L.B., Fialho, H.R.F., Locks, F., & Kamonseki, D.H. (2024). Measurement properties of upper extremity physical performance tests in athletes: A systematic review. *Brazilian Journal of Physical Therapy*, *28*(1), 100575. doi: 10.1016/j.bjpt.2023.100575

Bauer, C., Gröger, I., Rupprecht, R., & Gaßmann, K.G. (2008). Intrasession reliability of force platform parameters in community-dwelling older adults. *Archives of Physical Medicine*, *89*(10), 1977-1982. doi: 10.1016/j.apmr.2008.02.033

Beckham, G., Lish, S., Keebler, L., Longaker, C., Disney, C., DeBeliso, M., & Adams, K.J. (2019). The reliability of the seated medicine ball throw for distance. *Journal of Physical Activity Research*, *4*(2), 131-136. doi: 10.12691/jpar-4-2-9

Belhaidas, M.B., Dahoune, O., Eather, N., & Oukebdane, M.A. (2021). Objectivity, reliability, and validity of the basketball throw test as a health-related measure of upper-body muscular strength in a sample of Algerian primary school children. *Measurement in Physical Education and Exercise Science*, *27*(1), 1-9. doi: 10.1080/1091367X.2021.2021205

Biggar, C., Larson, A., & DeBeliso, M. (2022). Establishing normative reference values for the Utah seated medicine ball throw protocol in adolescents. *The Sport Journal / BioRxiv*. https://doi.org/10.1101/2021.12.14.472637

Borms, D., & Cools, A. (2018). Upper-extremity functional performance tests: Reference values for overhead athletes. *International Journal of Sports Medicine*, *39*(6), 433-441. https://doi.org/10.1055/a-0573-1388

Bös, K., Opper, E., Woll, A., Liebisch, R., Breithecker, D., & Kremer, B. (2001). Das Karlsruher Testsystem für Kinder (KATS-K)—Testmanual. *Haltung und Bewegung*, *21*(4), 4-66.

Brazo-Sayavera, J., Silva, D.R., Lang, J.J., Tomkinson, G.R., Agostinis-Sobrinho, C., …, & Dos Santos, L. (2024). Physical fitness surveillance and monitoring systems inventory for children and adolescents: A scoping review with a global perspective. *Sports Medicine*, *54*(7), 1755-1769. doi: 10.1007/s40279-024-02038-9

Carron, M.A., Scanlan, A.T., & Doering, T.M. (2024). The retest reliability of common tests to assess physical qualities in adolescent rugby league players. *Pediatric Exercise Science*, ahead of print. doi: 10.1123/pes.2024-0031

Castro-Pinero, J., Artero, E.G., Espana-Romero, V., Ortega, F.B., Sjostrom, M., Suni, J., & Ruiz, J.R. (2010). Criterion-related validity of field-based fitness tests in youth: A systematic review. *British Journal of Sports Medicine*, *44*(13), 934-943. https://doi.org/10.1136/bjsm.2009.058321

Chiwaridzo, M., Ferguson, G.D., & Smits-Engelsman, B.C.M. (2019). Anthropometric, physiological characteristics and rugby-specific game skills discriminating Zimbabwean under-16 male adolescent rugby players by level of competition. *BMJ Open Sport and Exercise Medicine*, *5*(1), 1-7. doi: 10.1136/bmjsem-2019-000576

Chiwaridzo, M., Tadyanemhandu, C., Mkumbuzi, N.S., Dambi, J.M., Ferguson, G.D., & Smits-Engelsman, B.C. (2021). Absolute and relative reliability of scrum test battery components assembled for schoolboy rugby players playing competitive rugby in low-resource settings: A pragmatic in-season test-retest approach. *South African Journal of Sports Medicine*, *33*(1), 1-7. doi: 10.17159/2078-516X/2021/v33i1a12220

Cuenca-Garcia, M., Marin-Jimenez, N., Perez-Bey, A., Sánchez-Oliva, D., Camiletti-Moiron, D., Alvarez-Gallardo, I.C., Ortega, F.B., & Castro-Piñero, J. (2022). Reliability of field-based fitness tests in adults: A systematic review. *Sports Medicine*, *52*(8), 1961-1979. doi: 10.1007/s40279-021-01635-2

Currell, K., & Jeukendrup, A. (2008). Validity, reliability and sensitivity of measures of sporting performance. *Sports Medicine*, *38*(4), 297–316. doi: 10.2165/00007256-200838040-00003

Davis, K.L., Kang, M., Boswell, B.B., Dubose, K.D., Altman, S.R., & Binkle, H.M. (2008). Validity and reliability of the medicine ball throw for kindergarten children. *Journal of Strength and Conditioning Research*, *22*(6), 1958-1963. doi: 10.1519/JSC.0b013e3181821b20

Dobbin, N., Hunwicks, R., Highton, J., & Twist, C. (2018). A reliable testing battery for assessing physical qualities of elite academy rugby league players. *Journal of Strength and Conditioning Research*, *32*(11), 3232-3238. doi: 10.1519/JSC.0000000000002280

Domone, S., Mann, S., Sandercock, G., Wade, M., & Beedie, C. (2016). A method by which to assess the scalability of field-based fitness tests of cardiorespiratory fitness among schoolchildren. *Sports Medicine*, *46*(12), 1819-1831. doi: 10.1007/s40279-016-0553-6

Duncan, M.J., Al-Nakeeb, Y., & Nevill, A.M. (2005). Influence of familiarization on a backward, overhead medicine ball explosive power test. *Research in Sports Medicine*, *13*(4), 345-352. doi: 10.1080/15438620500359950

Duncan, M.J., & Hankey, J. (2010). Concurrent validity of the backwards overhead medicine ball throw as a test of explosive power in adolescents. *Medicina Sportiva*, *14*(3), 103-107. doi: 10.2478/v10036-010-0019-0

Dunder, V.C. (1933). A multiple strength index of general motor ability. *Research Quarterly. American Physical Education Association*, *4*(3), 132-142.

Fernandez-Santos, J.R., Ruiz, J.R., Cohen, D.D., Gonzalez-Montesinos, J.L., & Castro-Piñero, J. (2015). Reliability and validity of tests to assess lower-body muscular power in children. *Journal of Strength and Conditioning Research*, *29*(8), 2277-2285. https://doi.org/10.1519/JSC.0000000000000864

Fernandez Santos, J.R., Ruiz, J.R., Gonzalez-Montesinos, J.L., & Castro-Piñero, J. (2016). Reliability and validity of field-based tests to assess upper-body muscular strength in children aged 6-12 years. *Pediatric Exercise Science*, *28*(2), 331-340. https://doi.org/10.1123/pes.2014-0196

Ferreira, L.G.R., de Oliveira, A.S., do Carmo, N.D., Santos Bueno, G.A., Lemos, T.V., Matheus, J.P.C., & de Souza Júnior, J.R. (2021). Reliability and validity of the one arm hop test and seated medicine ball throw test in young adults: A cross-sectional study. *Journal of Bodywork and Movement Therapies*, *28*, 26-33. https://doi.org/10.1016/j.jbmt.2021.07.018

Fjørtoft, I., Pedersen, A.V., Sigmundsson, H., & Vereijken, B. (2011). Measuring physical fitness in children who are 5 to 12 years old with a test battery that is functional and easy to administer. *Physical Therapy*, *91*(7), 1087-1095. doi: 10.2522/ptj.20090350

Garcia-Carrillo, E., Ramirez-Campillo, R., Thapa, R.K., Afonso, J., Granacher, U, & Izquierdo, M. (2023). Effects of upper-body plyometric training on physical fitness in healthy youth and young adult participants: A systematic review with meta-analysis. *Sports Medicine—Open*, *9*(1). doi: 10.1186/s40798-023-00631-2

Ginty, A.T. (2020). Psychometric properties. In *Encyclopedia of Behavioral Medicine*. Springer.

González-Devesa, D., Diz-Gómez, J.C., Sanchez-Lastra, M.A., Rodríguez, A.O., & Ayán-Pérez, C. (2024). Criterion-related validity and reliability of the six-minute run walk test among children: Findings from a systematic review and meta-analysis. *Measurement in Physical Education and Exercise Science*, 29(2), 160-173. doi: 10.1080/1091367X.2024.2426150

Hackett, D.A., Davies, T.B., Ibel, D., Cobley, S., & Sanders, R. (2017). Predictive Ability of the medicine ball chest throw and vertical jump tests for determining muscular strength and power in adolescents. *Measurement in Physical Education and Exercise Science*, *22*(1), 79-87. doi: 10.1080/1091367X.2017.1385462

Han, A., Fu, A., Cobley, S., & Sanders, R.H. (2018). Effectiveness of exercise intervention on improving fundamental movement skills and motor coordination in overweight/obese children and adolescents: A systematic review. *Journal of Science and Medicine in Sport*, *21*(1), 89-102. doi: 10.1016/j.jsams.2017.07.001

Harasin, D., Dizdar, D., & Marković, G. (2006). High reliability of tests of maximum throwing performance. *Journal of Human Movement Studies*, *51*(1), 63-76.

Harris, C., Wattles, A.P., DeBeliso, M., Sevene-Adams, P.G., Berning, J.M., & Adams, K.J. (2011). The seated medicine ball throw as a test of upper body power in older adults. *Journal of Strength and Conditioning Research*, *25*(8), 2344-2348. doi: https://doi.org/10.1519/JSC.0b013e3181ecd27b

Hermassi, S., van den Tillaar, R., Bragazzi, N.L., & Schwesig, R. (2021). The associations between physical performance and anthropometric characteristics in obese and non-obese schoolchild handball players. *Frontiers in Physiology*, *11*, 1-12. doi: 10.3389/fphys.2020.580991

Jarnig, G., Jaunig, J., Kerbl, R., Lima, R.A., & Van Poppel, M.N.M. (2022). A novel monitoring system (AUT FIT) for anthropometrics and physical fitness in primary school children in Austria: A cross-sectional pilot study. *Sports*, *10*(1), 4. https://doi.org/10.3390/sports10010004

Joensuu, L., Csányi, T., Huhtiniemi, M., Kälbi, K., Magalhães, J., Milanović, I., Morrison, S.A., Ortega, F.B., Sardinha, L.B., Starc, G., Tammelin, T.H., & Jurak, G. (2024). How to design and establish a national school-based physical fitness monitoring and surveillance system for children and adolescents: A 10-step approach recommended by the FitBack network. *Scandinavian Journal of Medicine and Science in Sports*, *34*(3), e14593. doi: 10.1111/sms.14593

Johnson, E.B., Maurya, P.S., Sisneros, K.P., Ford, B.R., & Palmer, T.B. (2024). Force production measurements during a supine medicine ball throw: A reliability and correlation study. *Journal of Musculoskeletal Neuronal Interactions* 24(2), 120-126.

Khemiri, A., Teboulbi, A., Gritli, A., Hachana, Y., & Attia, A. (2022). Reliability of a medicine ball explosive power test in young volleyball players. *Ukraïns'kij Žurnal Medicini, Bìologìï ta Sportu*, *7*(3), 282-287. doi: 10.26693/jmbs07.03.282

Koya, N., Kitamura, T., & Takahashi, H. (2022). Prediction of service performance based on physical strength in elite junior tennis players. *Frontiers in Psychology*, *13*, 898224. doi: 10.3389/fpsyg.2022.898224

Kryst, Ł., Żegleń, M., Artymiak, P., Kowal, M., & Woronkowicz, A. (2023). Analysis of secular trends in physical fitness of children and adolescents (8-18 years) from Kraków (Poland) between 2010 and 2020. *American Journal of Human Biology*, *35*(3), e23829. https://doi.org/10.1002/ajhb.23829

Lang, J.J., Belanger, K., Poitras, V., Janssen, I., Tomkinson, G.R., & Tremblay, M.S. (2018). Systematic review of the relationship between 20 m shuttle run performance and health indicators among children and youth. *Journal of Science and Medicine in Sport*, *21*(4), 383-397. doi: 10.1016/j.jsams.2017.08.002

Leite, M.A.F.D.J., Sasaki, J.E., Lourenço, C.L.M., Zanetti, H.R., Cruz, L.G., Mota, G.R.D., & Mendes, E.L. (2016). Medicine ball throw test predicts arm power in rugby sevens players. *Revista Brasileira de Cineantropometria & Desempenho Humano*, *18*(2), 166-176. https://doi.org/10.5007/1980-0037.2016v18n2p166

Luna-Villouta, P., Leao Ribeiro, I., Faúndez-Casanova, C., Vásquez-Gómez, J., Flores-Rivera, C., & Vargas Vitoria, R. (2022). Intra-rater and intraday test-retest reliability for physical performance tests in young Chilean tennis players. *Cultura, Ciencia y Deporte*, *17*(53), 89-97. doi: 10.12800/ccd.v17i53.1853

Marques, A., Henriques-Neto, D., Peralta, M., Martins, J., Gomes, F., Popovic, S., Masanovic, B., Demetriou, Y., Schlund, A., & Ihle, A. (2021). Field-based health-related physical fitness tests in children and adolescents: A systematic review. *Frontiers in Pediatrics*, *9*, 640028. doi: 10.3389/fped.2021.640028

Martínez-Lemos, I., Rodríguez, A., Diz, J., & Ayán-Pérez, C. (2024). Reliability and criterion-related validity of the Cooper test in pre-adolescents and adolescents: A systematic review and meta-analysis." *Journal of Sports Sciences*, *42*(3), 222-236. https://doi.org/10.1080/02640414.2024.2326352

Monticone, M., Galeoto, G., Berardi, A., & Tofani, M. (2021). Psychometric properties of assessment tools. In G. Galeoto, A. Berardi, M. Tofani, & M. Auxiliadora Marquez (Eds.), *Measuring spinal cord injury* (pp. 7-15). Springer.

Morrow, J.R., Martin, S.B., & Jackson, A.W. (2010). Reliability and validity of the FITNESSGRAM®: Quality of teacher-collected health-related fitness surveillance data. *Research Quarterly for Exercise and Sport*, *81*(Supp 3), S24-S30. doi: 10.1080/02701367.2010.10599691

Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan—A web and mobile app for systematic reviews. *Systematic Reviews*, *5*(1), 210. doi: 10.1186/s13643-016-0384-4

Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., …, & Moher, D. (2021). The PRISMA 2020 statement: An Updated Guideline for Reporting Systematic Reviews. *The British Medical Journal*, *372*, n71. doi: 10.1136/bmj.n71

Paraskevopoulos, E., Pamboris, G.M., Plakoutsis, G., & Papandreou, M. (2023). Reliability and measurement error of tests used for the assessment of throwing performance in overhead athletes: A systematic review. *Journal of Bodywork and Movement Therapies*, *35*, 284-297. https://doi.org/10.1016/j.jbmt.2023.04.042

Petrušič, T., Trajković, N., & Bogataj, Š. (2022). Twelve-week game-based school intervention improves physical fitness in 12-14-year-old girls. *Frontiers in Public Health*, *10*, 8-13. doi: 10.3389/fpubh.2022.831424

Roe, G., Shaw, W., Darrall-Jones, J., Phibbs, P.J., Read, D., Weakley, J.J., Till, K., & Jones, B. (2018). Reliability and validity of a medicine ball-contained accelerometer for measuring upper-body neuromuscular performance. *Journal of Strength and Conditioning Research*, *32*(7), 1915-1918. https://doi.org/10.1519/JSC.0000000000002470

Rosni, M.H.N.B.M., Abas, N.G., & Mohamad, N.I. (2014). Reliability of overhead medicine ball throw test as a muscular power assessment tool. In *Proceedings of the Seminar Penyelidikan Kebangsaan*, Perak, Malaysia.

Shrout, P.E., & Fleiss, J.L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *86*(2), 420-428. doi: 10.1037/0033-2909.86.2.420

Stockbrugger, B.A., & Haennel, R.G. (2001). Validity and reliability of a medicine ball explosive power test. *Journal of Strength and Conditioning Research*, *15*(4), 431-438.

Suurmond, R., van Rhee, H., & Hak, T. (2017). Introduction, comparison, and validation of meta-essentials: A free and simple tool for meta-analysis. *Research Synthesis Methods*, *8*(4), 537-553. doi: 10.1002/jrsm.1260

van den Tillaar, R., & Marques, M.C. (2009). Effect of two different training programs with the same workload on soccer overhead throwing velocity. *International Journal of Sports Physiology and Performance*, *4*(4), 474-484. doi: 10.1123/ijspp.4.4.474

van den Tillar, R., & Marques, M.C. (2011). A comparison of three training programs with the same workload on overhead throwing velocity with different weighted balls. *Journal of Strength and Conditioning Research*, *25*(8), 2316-2321. doi: 10.1519/JSC.0b013e3181f159d6

van den Tillaar, R., & Marques, M.C. (2013). Reliability of seated and standing throwing velocity using differently weighted medicine balls. *Journal of Strength and Conditioning Research*, *27*(5),1234-1238. https://doi.org/10.1519/JSC.0b013e3182654a09

Submitted: April 13, 2025
Accepted: May 18, 2025
Published Online First: June 30, 2025

Correspondence to:
Prof. Daniel González-Devesa, Ph.D.
Catholic University of Ávila, Faculty of Humanities and Education
C/ Canteros, 05005, Ávila, Spain
Email: danidevesa4@gmail.com