children

MDPI

Article

Reliability, Concurrent Validity, and Side-to-Side Asymmetry of a Unilateral Rotational Medicine Ball Throw Test in Adolescents

Julio Martín-Ruiz [1,*], Alejandro Leiva-Arcas [2], Amparo Aguilar-Prima [3] and Laura Ruiz-Sanchis [4]

[1] Department of Health and Functional Assessment, Catholic University of Valencia, 46900 Valencia, Spain
[2] Sport Faculty, UCAM Catholic University of Murcia, 30107 Murcia, Spain; aleiva@ucam.edu
[3] GIEPAFS Research Group, Catholic University of Valencia, 46900 Valencia, Spain; amparo.aguilar@mail.ucv.es
[4] Department of Sports Management and Physical Activity, Catholic University of Valencia, 46900 Valencia, Spain; laura.ruiz@ucv.es
* Correspondence: julio.martin@ucv.es; Tel.: +34-963637412

Highlights

What are the main findings?

- A unilateral rotational medicine ball throw showed good-to-excellent test–retest reliability in boys and moderate-to-good reliability in girls, with strong concurrent validity against the Eurofit frontal throw.
- Unlike the Eurofit frontal throw, the rotational test detected significant side-to-side (right- vs. left-directed) differences in whole-body rotational throwing performance in both sexes; although the two tests were related, they showed systematic and proportional differences.

What are the implications of the main findings?

- Physical Education teachers can use this low-cost field test, together with exploratory, sample-derived sex- and grade-specific 1–10 reference values that are not yet normative, to assess whole-body rotational throwing performance in school settings.
- The test complements, rather than replaces, the Eurofit frontal throw by providing side-to-side information that may help describe direction-specific differences in whole-body rotational performance in adolescents, although the asymmetry index was not reliable enough for individual monitoring.

Abstract

Background/Objectives: This study examined the reliability, concurrent validity, agreement, and side-to-side asymmetry of a unilateral rotational medicine ball throw test in adolescents. **Methods:** Overall, 204 students (101 boys, 103 girls; 14.1 ± 1.3 years) completed two sessions of Eurofit frontal throws and rotational throws to both sides. Reliability, measurement error, concurrent validity, Bland–Altman agreement, and asymmetry were analyzed. **Results:** Reliability was good-to-excellent in boys (ICC = 0.90) and moderate-to-good in girls (ICC = 0.70–0.82), with the right-sided throw in girls below the prespecified thresholds (SEM = 0.34–0.44 m; MDC95 = 0.94–1.21 m). Performance improved from Day 1 to Day 2, suggesting a learning effect. The rotational throw was strongly associated with the Eurofit throw (r = 0.81). Bland–Altman analysis showed systematic and proportional differences (bias 0.53 m; 95% LoA −0.68 to 1.75 m). A significant group-level right–left difference was detected (mean asymmetry 7.49%), but the asymmetry index

Academic Editor: Donatella Di Corrado

Received: 10 October 2026
Revised: 28 October 2026
Accepted: 3 November 2026
Published: 4 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

showed poor test–retest reliability (ICC = 0.22) and an inconsistent direction across sessions (κ = 0.14). **Conclusions:** The rotational throw is a related but not interchangeable complement to the Eurofit throw for whole-body rotational performance. Its asymmetry index describes group differences but is not suitable for individual monitoring. These findings are preliminary.

Keywords: adolescent; muscle strength; physical fitness; reproducibility of results; medicine ball throw; side-to-side asymmetry

1. Introduction

Physical Education (PE) develops different components of physical fitness [1], and the comprehensive assessment of all of them is usually carried out through test batteries [2] that enable teachers to assess students objectively and establish their baseline levels. Among the most relevant tests are those that assess functional capacity, such as the Functional Movement Screen (FMS) [3,4], and others that analyze strength, speed, endurance, and flexibility, such as Eurofit, one of the most widely used and validated batteries, proposed by the Council of Europe in 1986 [5,6]. The original Eurofit battery comprises nine motor tests (flamingo balance, plate tapping, sit-and-reach, standing broad jump, handgrip strength, sit-ups, bent-arm hang, 10 × 5 m shuttle run, and 20 m shuttle run), together with basic anthropometric measures, and covers balance, speed of limb movement, flexibility, explosive and static strength, muscular endurance, agility, and cardiorespiratory fitness. In Spanish school settings, it is frequently complemented with a medicine ball throw to assess upper-body explosive strength, which is the comparison test used in the present study. Such batteries should include rapid, reproducible, and validated field tests with the highest possible accuracy [2], and their design should include an appropriate selection of tests that considers the relevant variables for each educational stage [7], typically with a stronger perceptual-motor orientation until puberty and a greater emphasis on conditioning components thereafter [8].

Muscular strength is a key component of youth physical fitness and is commonly assessed using multiple field-based tests. Upper-body muscular strength, in particular, is associated with health-related outcomes in children and adolescents, such as bone health, cardiometabolic risk profile, and motor competence, and its development during childhood and adolescence is considered a central objective of long-term physical development and of health-related fitness assessment in schools [9–11]. Among strength tests, throwing tests are particularly relevant, as they reflect not only absolute strength but also the technical skill required to transfer it through proper intermuscular coordination [12]. In the Eurofit-based assessments commonly used in Spanish schools, this capacity is assessed with a bilateral standing overhead throw of a medicine ball (5 kg for boys and 3 kg for girls). Despite its widespread use and value for assessing strength [13], one of its main limitations [14] is its dependence on anthropometric factors (height and release angle of the ball), which may not reflect an individual's actual relative strength and may introduce sex bias [15,16]. Moreover, the existence of multiple variants and the influence of anthropometric factors limit its criterion validity as a general measure of individual strength [17,18].

A further limitation is that bilateral, frontal tests provide no information about side-to-side differences in performance. Unilateral measurements have been proposed to address this, partly because, in force-production tasks, the sum of unilateral performances has been reported to exceed bilateral performance (bilateral deficit) [19,20]; however, this evidence comes mainly from early neuromuscular studies and has not been confirmed in

throwing tasks. Laterality is also relevant when adapting tests to functional situations [21], and side-to-side ratios [22,23] may help describe side-to-side performance differences and identify potentially relevant functional asymmetries [24]. More recent studies in young people have linked marked inter-limb asymmetries with lower performance and, in some contexts, with a higher risk of injury, although these associations appear to be task-specific [25,26]; asymmetry in a field test should therefore not be equated with injury risk without prospective injury data. Many everyday and sporting actions are performed predominantly with one side of the body [27] and involve curvilinear or angular movements in which one side of the body acts while the other stabilizes [28]. From a developmental perspective, balanced strength development on both sides of the body is considered an important objective of youth physical development, as it supports a broad and versatile repertoire of motor skills [9,10], whereas pronounced asymmetries may limit performance in jumping, sprinting, changing direction, or throwing, and may expose the more heavily loaded side to repetitive stress [25,26]. A side-specific rotational test may therefore provide information that complements bilateral frontal tests.

Updating assessment protocols may contribute to more accurate and fairer tests [29,30] and to more specific information on students' needs after an intervention. In addition, unilateral activities have been associated with changes in motor asymmetries [31], which supports the value of tests that consider laterality.

The validation of new PE tests requires comparison with established tests of similar characteristics [32], which allows their similarities and differences to be identified. This process involves criterion, discriminant, and construct validity [32,33] and requires adequate reliability [34], so that the test can differentiate movement quality, be related to other variables such as anthropometric characteristics, and identify actual functional deficits [35]. Complementary instruments, such as perceived exertion scales, can also support test interpretation: building on Borg's scale [36], pictorial scales have been validated for resistance exercise [37] and for specific equipment and age groups [38].

Analysis of the bilateral throw has been extended, for example, by using accelerometers to increase measurement accuracy [39], or by assessing supine throws on force platforms in small samples [40] to provide alternatives to the traditional method of measuring throwing performance. However, to our knowledge, no validated unilateral throwing test that differentiates between sides and compares performance with a bilateral test has been reported. This gap is relevant for several reasons. First, the Eurofit throw is performed bilaterally and predominantly in the sagittal plane, whereas many sporting and daily actions (e.g., throwing, striking, or carrying loads) involve trunk rotation in the transverse plane and asymmetric force production [27,28]. Second, a bilateral test yields a single value that may mask side-to-side differences and therefore cannot tell teachers which side requires specific work [22,23,41]. Third, before a unilateral rotational test can be adopted in schools, it must be shown to be reliable, to reflect a construct related to that of an established field test, and to provide information beyond a simple rescaling of that test [32,34,35,42]. Comparing both protocols therefore makes it possible to determine whether the unilateral rotational throw measures related throwing performance (concurrent validity), whether both tests can be used interchangeably (agreement), and whether the new test provides additional side-specific information (side-to-side asymmetry).

Therefore, the aim of this study was to examine the reliability, validity, agreement, and side-to-side asymmetry assessment capacity of a unilateral rotational medicine ball throw test in adolescents, using the Eurofit frontal throw as a comparison field test. In this study, reliability refers to the consistency of rotational throw performance across testing days, concurrent validity to the strength of the association with the Eurofit frontal throw, and agreement to the extent to which both tests yield the same values on the same scale (i.e., metric interchangeability). It was hypothesized that the unilateral rotational throw

would show acceptable test–retest reliability (defined a priori as at least good relative reliability, ICC ≥ 0.75 [43], with an SEM below 10% of the mean score), strong concurrent validity with the Eurofit frontal throw, and the capacity to detect side-to-side asymmetries not captured by the frontal throw. It was also expected that both tests would be related but not metrically interchangeable.

2. Materials and Methods

2.1. Study Design

To examine the reliability, validity, agreement, and side-to-side asymmetry assessment capacity of a unilateral rotational throwing test in relation to the bilateral throwing test of the Eurofit battery, a cross-sectional methodological study with a test–retest design was conducted with students in Compulsory Secondary Education during the 2025–2026 school year. Participant recruitment, the familiarization session, and the two data collection sessions took place after ethical approval was granted (26 February 2026), between late February and March 2026.

Before the study began, authorization was obtained from the management of each participating school, written informed consent was obtained from the parents or legal guardians of all participants, and all students provided their assent to participate after being informed of the study procedures and of their right to withdraw at any time without consequences. The study adhered to the Declaration of Helsinki, and the research was approved by the Ethics Committee of Catholic University of Valencia (code: UCV/2025-2026/120).

2.2. Participants

The sample consisted of 204 secondary education students in Grades 1–4 (1st to 4th year of Compulsory Secondary Education) (101 boys and 103 girls; mean age 14.1 ± 1.3 years: boys 14.2 ± 1.3 years, range 12–17; girls 14.0 ± 1.3 years, range 12–17) from three secondary schools in the city of Valencia. The three schools were state-funded institutions located in urban districts, serving predominantly middle socioeconomic catchment areas. The study was offered to several secondary schools in the city, irrespective of whether they were publicly or privately owned, and was carried out in the three schools whose management teams and PE departments agreed to participate (convenience sampling). Urban or suburban areas were not specifically targeted: schools located in the city of Valencia were approached for logistical reasons, as their proximity to the research team allowed the same two evaluators to conduct all testing sessions within the regular PE timetable of each class group. No suburban or rural schools were finally included, which restricts the generalizability of the findings to comparable urban school settings (see Section 4.1). No a priori sample size calculation was performed; the sample size was determined by the number of eligible students in the participating schools. The final sample of 204 participants exceeds the minimum sample size generally recommended for reliability studies [43]; however, the sex-by-grade subgroups (18–37 students per cell) were not powered a priori, and all subgroup analyses should be regarded as exploratory. The inclusion criteria were enrollment in Grades 1–4 of Secondary Education, regular attendance at PE classes, and no engagement in physical activity on the day of the test or the previous day. Participants were reminded to refrain from strenuous physical activity on the day of each testing session and on the previous day. Sporting background was recorded with a brief questionnaire on extracurricular physical activity (participation and weekly hours of practice). Overall, 160 students reported regular extracurricular physical activity and 44 did not; however, sport type, training age, and specific throwing experience were not recorded, so habitual sport participation could only be examined as an exploratory variable

and is considered a limitation (see Section 4.1). Because the sample included students with heterogeneous sporting backgrounds, the exploratory comparison of throwing performance, asymmetry, and perceived exertion between students with and without extracurricular physical activity, and their association with weekly hours of practice, is reported in Section 3.4; a finer sub-analysis by sport discipline was not possible with the information collected. The exclusion criteria were any injury or illness that prevented safe performance of the tests and failure to attend both testing sessions. Of the 231 students initially assessed across the three schools, 5 were excluded before the analysis (3 due to an injury unrelated to the school context, 2 due to illness). Of the remaining 226 eligible students, 22 did not attend the second measurement session and were removed from the analysis, leaving a final sample of 204 students (101 boys, 103 girls) with complete data for both sessions. The reasons for non-attendance at the second session were not systematically recorded, and the Day 1 performance of the 22 non-completers (9.7% of the eligible sample) was not formally compared with that of completers, so a potential non-completion bias cannot be ruled out (see Section 4.1). Participants were nested within class groups and schools; given the small number of clusters (three schools), clustering was not modeled in the analyses. The participant flow is shown in Figure 1.

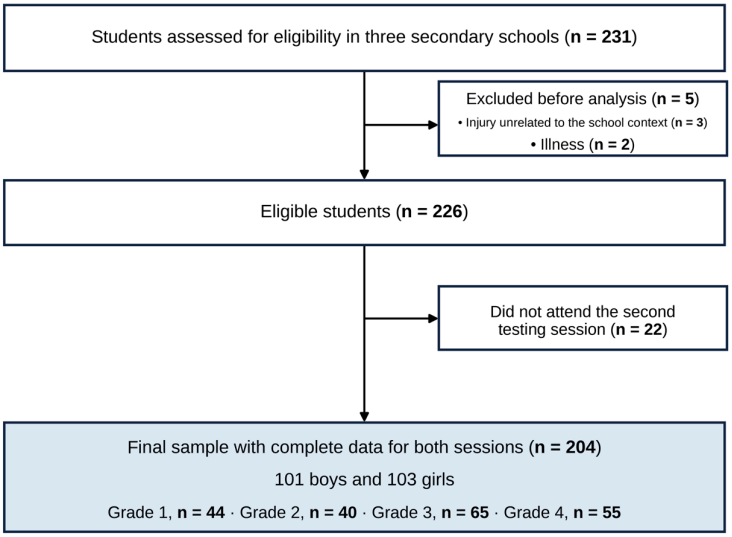


Figure 1. Participant flow diagram.

2.3. Procedure

One week before the testing sessions, participants were familiarized with both the Eurofit frontal throw and the unilateral rotational throw during a regular PE session, performing 10 Eurofit frontal throws and 10 unilateral rotational throws (five throws to the right side and five throws to the left side), i.e., 20 practice throws per participant in total, with verbal feedback on technique. Two testing sessions were conducted 72 h to one week apart, depending on the PE class schedule. The interval could not be standardized because both sessions had to be scheduled within the regular PE timetable of each class group, which differed across schools and grades, and no testing outside school hours was possible. Because the rotational throw showed a small systematic improvement from Day 1 to Day 2, this variability may have influenced the reliability estimates, as discussed in Section 4.1. In the first session, anthropometric measurements were taken, and the frontal and unilateral throws were recorded. In the second session, participants repeated both types of throws. The order of the tests was counterbalanced to avoid order effects: the six possible sequences of the three throws (Eurofit frontal, right rotational, and left rotational) were

used, and the two rows of participants always performed opposite sequences (i.e., when one row followed the sequence A–B–C, the other followed C–B–A). Two attempts were allowed for each type of throw, and the mean of the two attempts on each day was used for analysis.

The main researcher (ISAK-I) was responsible for the anthropometric assessments and recorded the following measurements: arm span, biacromial distance, height, weight, and leg length. Arm span was measured with the participant standing upright with the back against the wall (i.e., measured on the posterior side of the body), shoulders abducted at 90° and elbows extended, as the distance between the fingertips of both hands, using centimeter-graduated paper fixed to the wall. The biacromial distance was measured with the participant in the anatomical position, recording the distance (in cm) between both acromions with a flexible, non-extensible anthropometric tape (Lufkin W606PM, Crescent Lufkin, Apex Tool Group, Sparks, MD, USA). Weight was measured using a digital scale (Seca 750, Hamburg, Germany), height using a stadiometer (Seca 213, Hamburg, Germany), and leg length in the seated position, with the knee extended and the ankle in plantar flexion, by measuring the distance from the greater trochanter to the tip of the foot. The remaining measurements followed the procedures of the International Society for the Advancement of Kinanthropometry [44]. Because the leg-length definition used here is not a standard ISAK segment, it should be regarded as a pragmatic field estimate of lower-limb length; the technical error of measurement was not calculated, which limits the reproducibility of this variable (see Section 4.1).

The participants completed a standardized warm-up including mobilization, muscle activation, and core stabilization exercises: cat-camel (six repetitions) and front plank (15 s) exercises. Next, general joint mobility exercises (2 min), explosive jumps at half-knee flexion (six repetitions), and vertical medicine ball throws (10 repetitions) at low-to-moderate intensity, intended to prepare the musculature involved in throwing while avoiding fatigue, were performed. As part of the specific warm-up, at least four practice throws of each type were performed at progressively increasing, submaximal intensity, with feedback, before the test throws. Intensity was controlled through standardized verbal instructions: participants were asked to perform the successive practice throws at a progressively increasing self-perceived effort (from approximately half to near-maximal effort), without reaching maximal effort until the test throws, and the evaluators verbally corrected participants who threw at an evidently maximal intensity; intensity was therefore self-regulated and not monitored instrumentally. Once this phase was completed, the test throws were performed with two evaluators: one measured the first attempt and the other measured the second (i.e., each evaluator measured and recorded the throwing distance with a 30 m fiberglass measuring tape (PA077, PROACT, Sant Joan Despí, Spain) to the nearest 5 cm). Both evaluators held a PhD in Physical Activity and Sport Science and had prior experience in field-based physical fitness testing. Because each evaluator measured different attempts, this design did not allow inter-rater reliability to be estimated (see Section 4.1). Participants lined up in two rows and performed their first attempt in front of one of the evaluators, either the bilateral, right unilateral, or left unilateral throw. Throws were not performed consecutively: participants waited until everyone had thrown before attempting their next throw, allowing rest between attempts. At the end of the first round, the evaluators switched rows to assess the second attempts impartially (Figure 2). The waiting time built into this rotating throwing order, in which participants attempted their next throw only after the whole group had completed the previous round, was intended to provide sufficient recovery between attempts and to limit the influence of muscular fatigue on subsequent trials. The recovery period was not individually timed; however, the interval between consecutive attempts by the same participant was at least 2 min. Depending on group size and testing flow, this interval was typically approximately 2 min.

Because rest duration was not fixed, small differences in recovery between class groups, and therefore in fatigue or potentiation, cannot be ruled out (see Section 4.1).

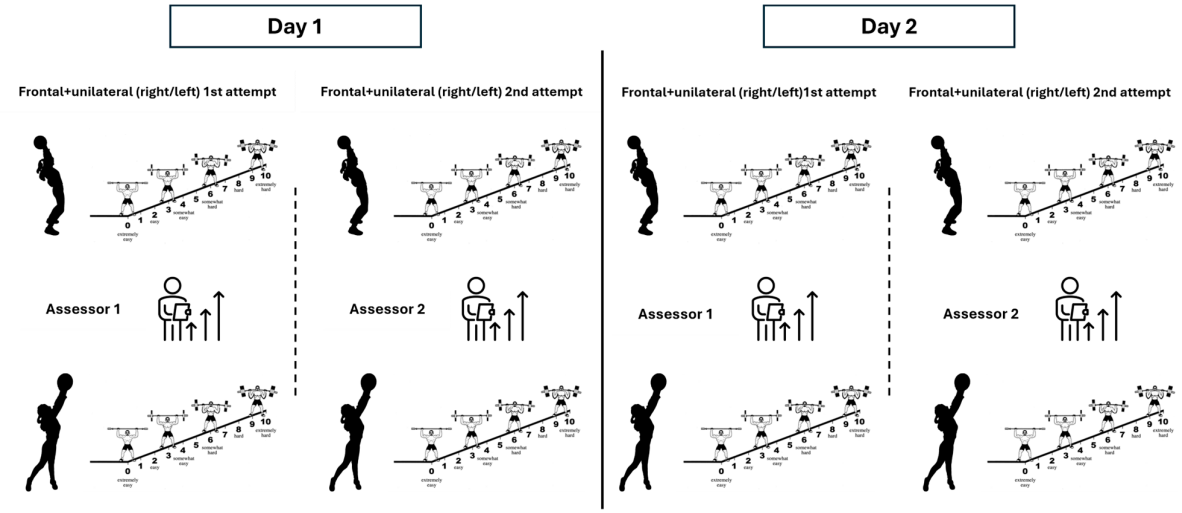


Figure 2. Overview of the testing protocol. Note: Overview of the testing protocol across the two assessment days. On each day, participants performed the Eurofit frontal throw and the rotational throw to the right and left sides, in two attempts each, with the two evaluators alternating between the first and second attempt.

2.3.1. Eurofit Frontal Medicine Ball Throw

The throw was performed while standing in a bipedal position with the feet at shoulder width (i.e., the biacromial distance; Figure 3A), facing the direction of the throw, with the medicine ball (5 kg for boys and 3 kg for girls) held in both hands. In the preparatory phase, participants brought the ball behind the head with the elbows extended, slightly flexing the knees and extending the trunk backward. They then performed a single explosive action, extending the knees and flexing the trunk and shoulders forward, and released the ball forward and upward in front of the head. Distance was measured from the line behind which the feet were placed to the nearest mark left by the ball (i.e., the point of the landing mark closest to the throwing line); the same criterion was applied in the rotational throw. The participants were instructed not to jump or cross the throw line and to finish the throw with their elbows extended and hands held high. The procedure is illustrated in Figure 3.

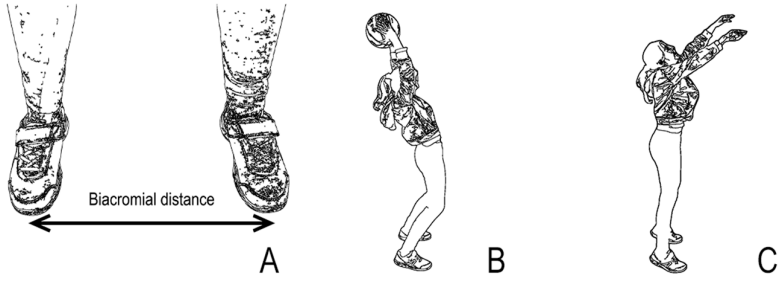


Figure 3. Eurofit frontal medicine ball throw: (**A**) distance between the feet (shoulder width, i.e., biacromial distance); (**B**) initial position; (**C**) final position.

2.3.2. Unilateral Rotational Medicine Ball Throw

Participants stood in a bipedal position, sideways to the direction of the throw, with the feet separated by 1.5 times the biacromial distance. This stance width was chosen

because it provides an easily calculated reference for the evaluators and because unilateral throws require greater stability to generate optimal throwing force [45–47]. Because this criterion was derived from studies of shot-put technique [45–47], mostly in adult or elite athletes, it should be regarded as a pragmatic, body-size-scaled reference that ensures a comparable base of support across participants, rather than as a validated optimal stance for adolescents; this issue is addressed in the Limitations. The term "unilateral" is used in this study to denote a throw directed toward one side of the body (right- or left-directed rotation), not an isolated single-limb action: the throw is executed from a bipedal stance with both hands on the ball, and both lower limbs, the pelvis, the trunk, and both upper limbs contribute to force production. Accordingly, the construct assessed by the rotational throw is whole-body rotational throwing performance (the ability to generate and transfer force to the ball through transverse-plane rotation), not isolated upper-limb muscular strength. In the same sense, side-to-side asymmetry refers in this study to a direction-specific difference in whole-body rotational throwing performance (right- vs. left-directed throws), which involves both legs, the pelvis, the trunk, and both arms; because it does not isolate the contribution of a single limb, the term "inter-limb asymmetry" is not used for the present measurements. The term "unilateral" was retained to maintain consistency with previous studies on side-to-side differences in field tests [22,23,25,26,41] and should be read as referring to the direction of the throw rather than to an isolated limb.

Participants rotated the trunk away from the throwing direction while flexing the knees and then reversed the movement through rapid joint extension to release the ball. The throw was performed on both sides (right and left) of the body, using the same medicine balls as in the Eurofit throw (5 kg for boys and 3 kg for girls). These masses were chosen to match those used in the frontal throw, so that both tests were performed with identical implements. The balls were the standard medicine balls available in the schools' PE equipment, and their diameter was not recorded. This is a relevant omission because, for a given mass, ball diameter may vary between manufacturers, and the relationship between ball diameter and hand size may affect grip, control of the ball during trunk rotation, and release conditions, particularly in younger or smaller students; the lack of control over ball diameter is therefore acknowledged as a limitation (see Section 4.1), and future studies should record ball diameter and hand dimensions. Relative to body mass, the ball represented 8.6 ± 2.0% (range 4.2–14.0%) of body mass in boys and 5.6 ± 1.1% (range 3.4–9.2%) in girls; this difference in relative load should be considered when interpreting sex differences (see Section 4.1). As in the Eurofit throw, distance was measured from the line behind which the feet were positioned to the nearest mark left by the ball, using the same criterion as in the Eurofit throw. Participants were again instructed not to jump or cross the line (or cross their feet) and to finish the throw with their hands as high as possible to optimize the release height. An example of this test is presented in Figure 4.

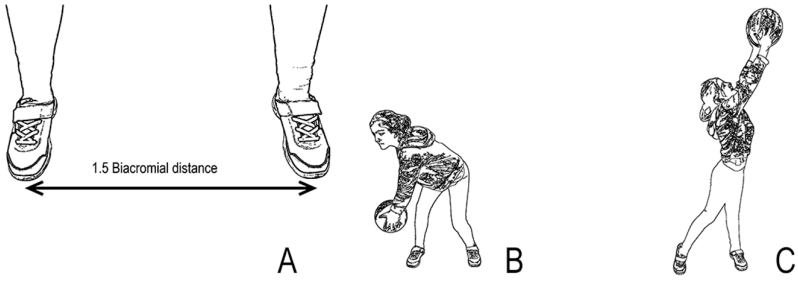


Figure 4. Unilateral rotational medicine ball throw: (**A**) distance between the feet (1.5 times the biacromial distance); (**B**) initial position; (**C**) final position.

For both tests, two attempts were made and both were recorded. Two attempts were used because of the time constraints of regular PE classes, in which each student had to complete six measured throws per session (two frontal, two right, and two left), and to limit fatigue and the cumulative load of repeated maximal throws in adolescents. The mean of both attempts, rather than the best attempt, was used because it reduces the influence of random error; whether three or more attempts would yield a more stable score is discussed in Section 4.1. Additionally, perceived exertion was recorded for each throw to examine its relationship with throw distance and type, using the OMNI Resistance Exercise Scale (OMNI-RES; 0–10) validated by Lagally and Robertson [37], which includes an illustrative, age-appropriate pictogram. Participants rated their exertion immediately after each attempt.

2.4. Statistical Analysis

The data were analyzed using descriptive and inferential statistics in R version 4.4.1 (R Foundation for Statistical Computing, Vienna, Austria; https://www.R-project.org/). The analyses were organized hierarchically according to the study hypotheses. The primary analyses addressed (i) test–retest reliability and measurement error of the rotational throw, (ii) concurrent validity against the Eurofit frontal throw, (iii) agreement between both tests, and (iv) side-to-side asymmetry in the rotational throw. The regression models, the sex and grade comparisons, the preliminary reference values, and the extracurricular-activity comparisons were considered secondary or exploratory analyses. Continuous variables are expressed as means and standard deviations. The test–retest reliability for each throw was assessed using a two-way random-effects, absolute-agreement, single-measures intraclass correlation coefficient (ICC(A,1)), in which the unit of analysis was each participant's daily performance (i.e., the mean of the two attempts on Day 1 and the mean of the two attempts on Day 2), calculated separately for each test, side, and sex. In addition, the standard error of measurement (SEM) and the minimal detectable change at the 95% confidence level (MDC95) were calculated. The SEM was calculated as $SD \times \sqrt{(1 - ICC)}$, where SD is the pooled standard deviation of the Day 1 and Day 2 scores, and the MDC95 as $SEM \times 1.96 \times \sqrt{2}$ [48]. Systematic error of measurement was additionally assessed by comparing Day 1 and Day 2 performance with paired-samples *t*-tests, reporting the mean difference (Day 2 − Day 1) and its statistical significance; this complements the SEM and MDC95, which reflect random measurement error. The magnitude of the systematic change was expressed as a paired effect size (Cohen's d_z, mean difference divided by the SD of the differences) with its 95% confidence interval. The SEM was also expressed as a percentage of the mean, which provides an index of typical error comparable to a coefficient of variation.

The concurrent validity of the unilateral rotational throw was examined using Pearson's correlations between the bilateral average of the rotational throw and the Eurofit frontal throw, which served as the comparison field test. Because the Eurofit throw is a predominantly bilateral, sagittal-plane action, whereas the rotational throw relies substantially on transverse-plane trunk and hip rotation, it was considered a comparison field test rather than a criterion standard; a strong correlation may therefore partly reflect shared dependence on general strength, power, body size, and throwing skill rather than measurement of an identical construct. Separate correlations were calculated for the right and left sides of the body. To determine whether the relationship between the two tests remained after controlling for potential confounding variables, a partial correlation adjusted for age, sex, grade, anthropometric variables, and perceived exertion was calculated.

The agreement between the unilateral rotational throw and Eurofit frontal throw was examined using the Bland–Altman method. For this purpose, the mean bias between the methods, 95% limits of agreement, and the presence of proportional bias, assessed

through the association between the mean of both methods and the difference between them, were calculated. This analysis was used to determine whether both procedures could be considered metrically interchangeable, complementing the information provided by concurrent validity correlations. The rationale for assessing numerical interchangeability was practical rather than theoretical: because PE teachers might be tempted to substitute one throw for the other, or to apply the reference values of one test to the other, it was considered necessary to document whether both tests yield equivalent distances. Given the different movement patterns involved, metric equivalence was not expected a priori, and no acceptable limits of agreement were specified in advance; the analysis was therefore descriptive, intended to quantify the magnitude and nature (systematic and proportional) of the differences between tests rather than to test a predefined equivalence hypothesis.

Although Bland–Altman analysis is often used to examine test–retest agreement of a single instrument, the method was originally proposed, and is widely used, to examine agreement between two different methods of measuring the same underlying construct [42]; it was applied here in that sense, to determine whether the rotational and Eurofit throws could be considered metrically interchangeable. ICC values were interpreted as poor (<0.50), moderate (0.50–0.75), good (0.75–0.90), or excellent (>0.90) [43], and Pearson's *r* values were interpreted as negligible (<0.30), low (0.30–0.50), moderate (0.50–0.70), high (0.70–0.90), or very high (>0.90) [49].

The influence of anthropometric variables on performance was analyzed using hierarchical linear regression analysis. In the first block, age, sex, and grade were included; in the second block, weight, height, BMI, arm span, biacromial width, and leg length were added; and in the third block, perceived exertion was incorporated. The contribution of each block was interpreted through changes in explained variance. Standardized regression coefficients (β) with 95% confidence intervals were obtained for each predictor, and multicollinearity among the anthropometric variables was examined using variance inflation factors (VIF). As a sensitivity analysis, the models were re-estimated with a reduced set of theoretically justified, non-redundant anthropometric predictors (height as an indicator of body size, BMI as an indicator of body mass relative to height, and biacromial width as an indicator of shoulder breadth). In addition, to examine directly whether body dimensions were differentially related to the two tests, the same blocks were used to predict the between-test difference (bilateral mean rotational throw minus Eurofit throw).

Side-to-side asymmetries (as defined in Section 2.3.2) were assessed by comparing the mean performances of the right and left sides using paired-samples *t*-tests. The magnitude of the difference was expressed using the effect size d_z. The relative asymmetry index was calculated as the absolute difference between the sides divided by the better side and multiplied by 100. This index was computed for each participant and then averaged; because it reflects the magnitude of side-to-side differences regardless of direction, the group mean asymmetry index can be substantially larger than the mean signed right–left difference, in which differences in opposite directions cancel out. Asymmetry analyses were performed on the total sample and stratified by sex and grade. To examine the reliability of the asymmetry index, it was also calculated separately for each testing day from that day's right- and left-side means, and its test–retest reliability was assessed with the ICC(A,1), SEM, MDC95, and Bland–Altman analysis (mean bias and 95% limits of agreement). The consistency of the direction of asymmetry (i.e., the favored throwing side) across days was examined using percentage agreement and Cohen's kappa, excluding participants with identical right- and left-side distances on either day.

To compare performance by sex, independent samples *t*-tests were used, and to analyze differences between grades, one-way analysis of variance (ANOVA) was used, with eta squared (η^2) reported as the effect size. A preliminary 1–10 reference scale was

constructed based on percentile cut-off points (deciles) of the bilateral average of the unilateral rotational throw, stratifying the cutoff points by sex and grade. As an exploratory analysis, performance was compared between students with and without extracurricular physical activity using Welch's independent-samples *t*-tests, given the unequal group sizes (160 vs. 44), and associations with weekly hours of practice were examined using Pearson's correlations. To control for multiple testing, *p* values within each family of related comparisons (sex differences by grade, side-to-side comparisons by sex and grade, and the exploratory extracurricular comparisons) were adjusted using the Holm–Bonferroni procedure [50]. The level of statistical significance was set at $p < 0.05$.

3. Results

3.1. Description of the Sample and Performance by Sex and Grade

A total of 204 secondary education students were analyzed in this study. The mean performance in the unilateral rotational throw increased with grade level and differed between the sexes. Boys demonstrated significantly higher rotational performance than girls ($p = 0.003$; $d = 0.420$), and significant differences were also observed by grade level ($p < 0.001$; $\eta^2 = 0.19$ for the rotational throw and 0.25 for the Eurofit throw) (Table 1; Figure 5). After Holm–Bonferroni adjustment, sex differences in the rotational throw remained significant only in Grade 3 (adjusted $p = 0.012$), and those in the Eurofit throw only in Grade 1 (adjusted $p = 0.032$). Because boys and girls threw balls of different mass (5 vs. 3 kg) and differ in body mass, the relative load was not equivalent between sexes; sex comparisons should therefore be interpreted with caution. The anthropometric characteristics of the participants were as follows (boys vs. girls, mean ± SD): height 168.4 ± 9.6 vs. 159.6 ± 7.2 cm; body mass 61.2 ± 13.5 vs. 55.1 ± 10.2 kg; BMI 21.5 ± 3.9 vs. 21.6 ± 4.0 kg·m^{-2}; arm span 170.2 ± 11.5 vs. 159.1 ± 7.6 cm; biacromial width 39.4 ± 3.5 vs. 37.4 ± 2.7 cm; and leg length 101.5 ± 6.2 vs. 97.6 ± 5.2 cm.

Table 1. Throwing performance and side-to-side asymmetry by sex and grade.

Sex	Grade	*n*	Eurofit Mean (SD)	Rotational Mean (SD)	Asymmetry % (SD)	Sex Diff. *p* (Eurofit)	Sex Diff. *p* (Rotational)
Total	All	204	3.73 (0.84)	4.26 (1.06)	7.49 (5.91)	—	—
Boys	1	24	2.78 (0.66)	3.33 (1.06)	7.78 (4.96)	0.008 **	0.265
Girls	1	20	3.34 (0.66)	3.63 (0.60)	7.68 (5.62)		
Boys	2	18	3.48 (0.58)	4.57 (1.09)	9.33 (5.98)	0.373	0.048 *
Girls	2	22	3.64 (0.59)	4.01 (0.62)	8.15 (7.02)		
Boys	3	28	3.97 (0.94)	4.78 (1.26)	8.26 (6.33)	0.452	0.003 **
Girls	3	37	3.84 (0.40)	4.05 (0.64)	7.68 (5.60)		
Boys	4	31	4.41 (0.92)	5.05 (1.05)	6.50 (6.51)	0.038 *	0.018 *
Girls	4	24	3.92 (0.77)	4.41 (0.83)	5.14 (4.90)		

Note: Values are presented as mean (standard deviation, SD). Eurofit mean refers to the average distance obtained in the frontal medicine ball throw. The rotational mean refers to the bilateral average of unilateral rotational throws. Asymmetry was calculated as the absolute difference between the right and left sides, divided by the better side and multiplied by 100. Sex differences (Boys vs. Girls) within each grade were assessed with independent-samples *t*-tests; the *p* value for each grade is shown in the Boys row. Overall (pooled) sex effect: Eurofit $p = 0.877$ (ns); Rotational $p = 0.003$ ($d = 0.420$). Grade effect (one-way ANOVA): Eurofit $F = 22.33$, $p < 0.001$; Rotational $F = 15.36$, $p < 0.001$. * $p < 0.05$; ** $p < 0.01$. Effect size of the grade effect: Eurofit $\eta^2 = 0.25$; Rotational $\eta^2 = 0.19$. After Holm–Bonferroni adjustment within each test, sex differences remained significant only for the Eurofit throw in Grade 1 (adjusted $p = 0.032$) and for the rotational throw in Grade 3 (adjusted $p = 0.012$).

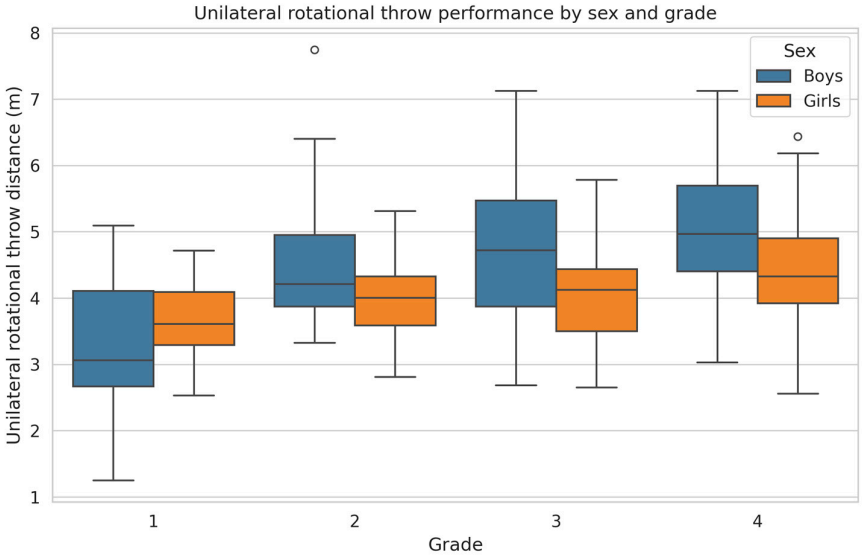


Figure 5. Distribution of bilateral mean distances in the unilateral rotational throw for boys and girls in each grade (Grades 1–4 of secondary education). Note: Boxplots show the distribution of the bilateral mean rotational throw distance according to sex and grade. The horizontal line within each box represents the median, the boxes represent the interquartile range, and the whiskers represent the distribution of observed values. Boys and girls were analyzed separately to illustrate sex- and grade-related differences in rotational throwing performance.

3.2. Reliability and Concurrent Validity of the Unilateral Rotational Throw

The unilateral rotational throw demonstrated good-to-excellent test–retest reliability in boys for both the right and left sides (ICC = 0.900 and 0.902), whereas reliability in girls was good for the left side (ICC = 0.819) and moderate for the right side (ICC = 0.696) (Table 2). Day 2 performance was systematically higher than Day 1 performance for the rotational throw, with small-to-moderate paired effect sizes (boys: d_z = 0.26 [95% CI 0.06 to 0.46] for the right side and 0.41 [0.21 to 0.61] for the left side; girls: d_z = 0.34 [0.14 to 0.54] and 0.52 [0.31 to 0.72], respectively), whereas no systematic change was observed for the Eurofit throw (boys: d_z = −0.19 [−0.39 to 0.01]; girls: d_z = 0.03 [−0.16 to 0.23]). The SEM ranged from 8.5% to 10.5% of the mean for the rotational throw and from 7.9% to 8.1% for the Eurofit throw (Table 2). Right- and left-side distances on each testing day are also shown in Table 2. Accordingly, the a priori reliability thresholds (ICC ≥ 0.75 and SEM < 10% of the mean) were met for both sides in boys and for the left side in girls, but not for the right side in girls (ICC = 0.696; SEM = 10.5%).

Table 2. Test–retest reliability and measurement error of the throwing tests, by sex.

Test	Sex	*n*	Day 1, m	Day 2, m	Mean Diff. (SD), m	*p*	ICC (95% CI)	SEM, m (%)	MDC95, m (%)
Eurofit frontal	Boys	101	3.78	3.69	−0.08 (0.43)	0.053	0.914 (0.87–0.94)	0.30 (8.1%)	0.84 (22.6%)
Eurofit frontal	Girls	103	3.71	3.72	+0.01 (0.42)	0.747	0.797 (0.71–0.86)	0.29 (7.9%)	0.82 (21.9%)
Rotational (right)	Boys	101	4.49	4.64	+0.15 (0.60)	0.011	0.900 (0.85–0.93)	0.44 (9.5%)	1.21 (26.4%)
Rotational (right)	Girls	103	4.01	4.21	+0.20 (0.59)	<0.001	0.696 (0.56–0.79)	0.43 (10.5%)	1.20 (29.2%)
Rotational (left)	Boys	101	4.28	4.51	+0.22 (0.54)	<0.001	0.902 (0.83–0.94)	0.41 (9.3%)	1.14 (25.8%)
Rotational (left)	Girls	103	3.87	4.09	+0.22 (0.43)	<0.001	0.819 (0.66–0.89)	0.34 (8.5%)	0.94 (23.6%)

Note: Reliability was assessed using the intraclass correlation coefficient (two-way random, absolute agreement, single measures, ICC(A,1)) between the Day 1 and Day 2 mean distances, calculated separately by sex given the difference in reliability observed between boys and girls. SEM represents

the standard error of measurement and MDC95 represents the minimum detectable change at the 95% confidence level; the percentages in parentheses express these values relative to the group mean. Systematic error was assessed with a paired-samples *t*-test comparing Day 1 and Day 2 distances (mean difference, Day 2 − Day 1, with SD in parentheses; exact *p* values are shown); a significant positive value indicates greater distances on Day 2, consistent with a learning effect. ICC: intraclass correlation coefficient; SEM: standard error of measurement; MDC95: minimum detectable change at 95% confidence.

The bilateral mean rotational throw was strongly associated with the Eurofit frontal throw (r = 0.813; 95% CI, 0.76 to 0.85; $p < 0.001$). Associations were also strong when the right (r = 0.805; $p < 0.001$) and left (r = 0.789; $p < 0.001$) sides were analyzed separately. After controlling for age, sex, grade, anthropometric variables, and perceived exertion, the association between the bilateral rotational throw and the Eurofit frontal throw remained strong (partial r = 0.742; $p < 0.001$) (Table 3; Figure 6).

Table 3. Concurrent validity of the rotational throw against the Eurofit frontal throw.

Comparison	*n*	*r*	95% CI	*p*
Rotational bilateral mean vs. Eurofit	204	0.813	0.76 to 0.85	<0.001
Right rotational vs. Eurofit	204	0.805	0.75 to 0.85	<0.001
Left rotational vs. Eurofit	204	0.789	0.73 to 0.84	<0.001
Asymmetry index vs. Eurofit	204	−0.037	−0.17 to 0.10	0.598
Rotational perceived exertion vs. Eurofit	204	0.174	0.04 to 0.30	0.013

Note: Concurrent validity was assessed using Pearson's correlation coefficients between the Eurofit frontal throw performance and rotational throw outcomes. The bilateral rotational mean was calculated as the average of the right and left performances. CI: 95% confidence interval. RPE: rating of perceived exertion.

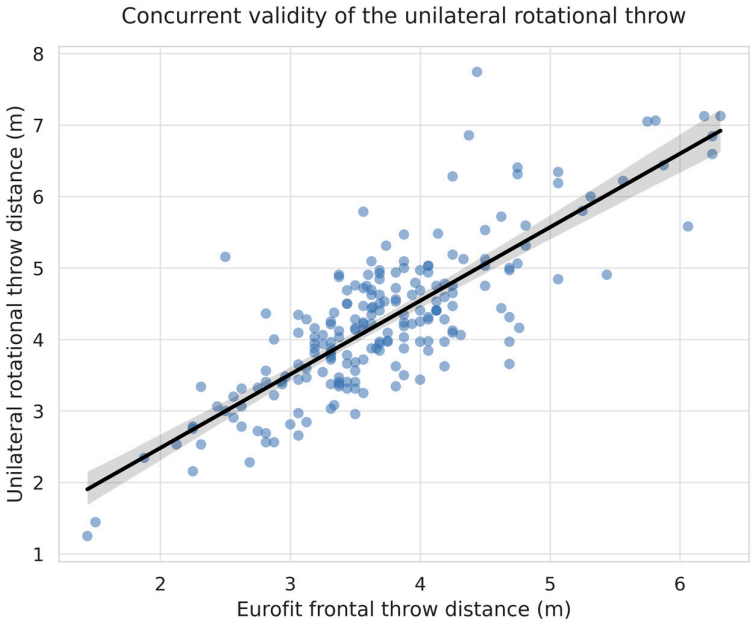


Figure 6. Association between the Eurofit frontal throw and the bilateral mean of the unilateral rotational throw (concurrent validity against a comparison field test). Note: Scatter plot showing the association between the mean Eurofit frontal throw distance and bilateral mean rotational throw distance. The line represents the linear regression fit between the two tests.

3.3. Agreement Between the Unilateral Rotational Throw and the Eurofit Frontal Throw: Bland–Altman Analysis

Bland–Altman analysis was incorporated at this stage because it complements the concurrent validity correlations. While the correlation showed that both tests were strongly related, the Bland–Altman analysis made it possible to assess whether the results were metrically interchangeable.

The comparison between the bilateral average of the rotational throw and Eurofit showed a mean bias of 0.533 m, indicating that the rotational throw achieved greater distances than the Eurofit throw. The 95% limits of agreement ranged from −0.682 to 1.748 m. The mean difference between the two methods was significant ($p < 0.001$), and a significant proportional bias was observed ($p < 0.001$), indicating that the difference between the two procedures varied according to the performance level (difference = −0.50 + 0.26 × mean of both methods); that is, the rotational throw exceeded the Eurofit throw to a greater extent in higher-performing students (Table 4; Figure 7).

Table 4. Bland–Altman agreement analysis between the rotational throw and the Eurofit frontal throw.

Comparison	*n*	Mean Bias	SD Diff	Lower LoA	Upper LoA	Bias *p*	Proportional Bias Slope	Proportional Bias *p*
Rotational bilateral mean vs. Eurofit	204	0.533	0.620	−0.682	1.748	<0.001	0.26	<0.001

Note: Bland–Altman analysis was performed using the bilateral mean of the rotational throw and mean Eurofit frontal throw distance. The mean bias represents the average difference between the methods, calculated as the rotational throw minus the Eurofit throw. The limits of agreement were calculated as the mean bias ± 1.96 standard deviation (SD) of the differences. Proportional bias was examined by regressing the difference between the methods on the mean of both methods. LoA = limits of agreement. Regression equation for proportional bias: difference = −0.50 + 0.26 × mean of both methods.

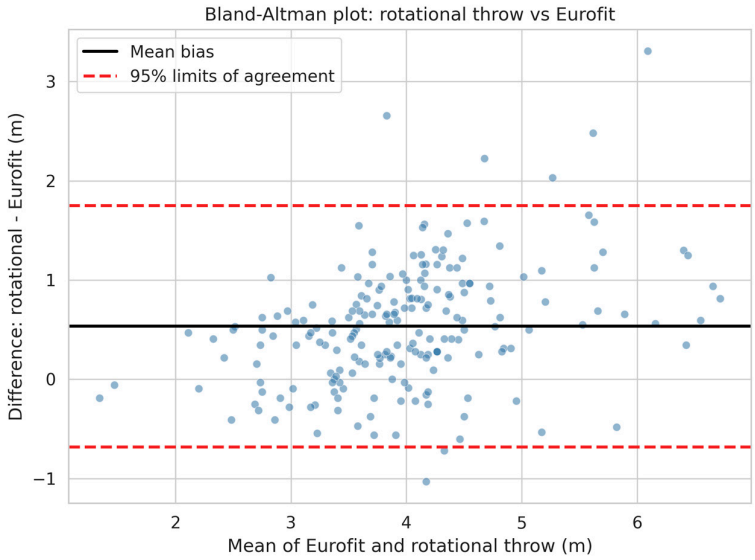


Figure 7. Agreement between the bilateral mean of the unilateral rotational throw and the Eurofit frontal throw (Bland–Altman analysis). Note: Bland–Altman plot showing agreement between the bilateral mean rotational throw and the Eurofit frontal throw. The solid line represents the mean bias, and the dashed lines represent the 95% limits of agreement (LoA). The differences were calculated as the rotational throw minus the Eurofit throw.

Thus, the rotational throw yielded systematically higher values than the Eurofit throw, and the between-method difference increased with performance level.

3.4. Influence of Anthropometric Variables

Anthropometric variables contributed significantly to performance in both throwing tests. For the bilateral mean rotational throw, the model with age, sex, and grade explained 22.3% of the variance; when weight, height, BMI, arm span, biacromial width, and leg length were added, the explained variance increased to 38.2%. This increase was similar to that observed for the Eurofit throw, where the explained variance increased from 24.8% to 40.1% (Table 5).

Table 5. Contribution of anthropometric variables and perceived exertion to throwing performance.

Outcome	*n*	R^2 Age-Sex-Grade	R^2 + Anthropometry	ΔR^2 Anthropometry	R^2 + RPE	ΔR^2 RPE
Eurofit throw	204	0.248	0.401	0.152	0.412	0.011
Rotational throw (bilateral mean)	204	0.223	0.382	0.159	0.383	0.001
Rotational throw (right)	204	0.212	0.372	0.160	0.372	0.000
Rotational throw (left)	204	0.217	0.365	0.148	0.367	0.002
Asymmetry index (%)	204	0.020	0.082	0.062	0.083	0.000

Note: Hierarchical regression models were used to examine the contribution of anthropometric variables and perceived exertion to throwing performance. The first block included age, sex, and grade. The second block included the anthropometric variables. The third block included perceived exertion. ΔR^2 represents the additional explained variance after adding each block of predictors. RPE: rating of perceived exertion. Grade was entered as an ordinal variable (1–4); RPE was entered as the mean rating of the corresponding throws. Standardized coefficients and VIFs for the anthropometric models are reported in Appendix A (Table A1).

In bivariate terms, arm span ($r = 0.569$ and 0.483 for the rotational and Eurofit throws, respectively), height ($r = 0.564$ and 0.476), and biacromial width ($r = 0.498$ and 0.495) showed the strongest associations with performance, whereas BMI showed only weak associations ($r = 0.123$ and 0.137). In the full model, however, the high collinearity among anthropometric variables (VIF = 33.7, 67.1, and 92.2 for height, BMI, and weight, respectively) meant that biacromial width was the only individually significant anthropometric predictor, and only for the Eurofit throw ($p = 0.018$; standardized $\beta = 0.20$, 95% CI 0.04 to 0.37); no single anthropometric variable was individually significant for the rotational throw, although the block as a whole added significant explained variance ($p < 0.001$). Standardized coefficients, 95% CIs, and VIFs for all predictors are reported in Appendix A (Table A1). The increase in explained variance after adding the anthropometric block was of similar magnitude in both protocols ($\Delta R^2 = 0.159$ for the rotational throw and 0.153 for the Eurofit throw).

Perceived exertion did not add relevant explanatory variance to the model (Table 5).

In the sensitivity analysis with a reduced set of non-redundant anthropometric predictors (height, BMI, and biacromial width; all VIF ≤ 5.1, anthropometric VIF ≤ 2.3), the increase in explained variance was very similar to that of the full models (Eurofit: R^2 from 0.248 to 0.398, $\Delta R^2 = 0.150$; bilateral rotational throw: R^2 from 0.223 to 0.382, $\Delta R^2 = 0.159$). Height was the strongest anthropometric predictor in both tests (Eurofit: $\beta = 0.36$, 95% CI 0.20 to 0.52; rotational: $\beta = 0.42$, 95% CI 0.26 to 0.58; both $p < 0.001$), whereas biacromial width was significant only for the Eurofit throw ($\beta = 0.20$, 95% CI 0.04 to 0.37, $p = 0.016$) and BMI was not significant in either test. When the between-test difference (rotational − Eurofit) was used as the dependent variable, the anthropometric block added only 3.2% of explained variance ($p = 0.063$), although height was individually associated with a

larger difference (β = 0.24, 95% CI 0.05 to 0.43, $p = 0.012$), indicating that taller students tended to obtain proportionally greater distances in the rotational throw than in the Eurofit throw.

In an exploratory analysis, extracurricular practice was also related to throwing performance. Students who engaged in extracurricular activities achieved greater distances in both the Eurofit frontal throw and the unilateral rotational throw. The difference was significant for Eurofit ($p = 0.002$, $d = 0.495$) and for the rotational throw ($p < 0.001$, $d = 0.722$). Likewise, weekly hours of extracurricular practice were positively associated with performance in Eurofit ($r = 0.369$, $p < 0.001$) and in the rotational throw ($r = 0.406$, $p < 0.001$). In contrast, extracurricular practice was not associated with the asymmetry index ($p = 0.726$) or perceived exertion ($p = 0.494$). The significant differences remained so after Holm–Bonferroni adjustment (adjusted $p \leq 0.006$). Because sport type, training age, and specific throwing experience were not recorded, a finer sub-analysis by sport discipline could not be performed; therefore, the possible influence of the heterogeneous sporting background of the sample on the reliability and asymmetry estimates cannot be fully excluded (see Section 4.1).

3.5. Side-to-Side Asymmetries

The unilateral rotational throw allowed the detection of significant side-to-side asymmetries. In the total sample, the right side achieved a greater mean distance than the left side, with a mean difference of 0.150 m and a mean relative asymmetry of 7.49%. The paired-samples *t*-test was significant ($p < 0.001$), and the effect size was small to moderate ($d_z = 0.356$). Descriptive statistics for the right and left sides are reported by sex and grade in Table 6.

Asymmetries were significant in both boys and girls. By grade, the differences were significant in Grades 2 and 3 but not in Grades 1 or 4 (Table 6; Figure 8). Applying the Holm–Bonferroni adjustment to the six subgroup comparisons did not change this pattern. Notably, in Grade 1 the mean signed right–left difference was small (0.051 m), whereas the mean asymmetry index was 7.73%. This is because the index is calculated for each participant using absolute values, so individual asymmetries favoring the right and the left sides do not cancel out, whereas they do in the signed mean difference. Thus, many Grade 1 students showed relevant side-to-side differences, but without a consistent direction at the group level. Limb dominance (handedness) was not assessed in this study; therefore, these asymmetry results should be interpreted in absolute right-versus-left terms rather than dominant-versus-non-dominant terms.

When the asymmetry index was calculated separately for each testing day, its mean value was 10.23% (SD 7.97) on Day 1 and 8.60% (SD 6.56) on Day 2 (mean change −1.63 percentage points, $p = 0.011$). Its test–retest reliability was poor (ICC = 0.22; boys 0.19, girls 0.25), with an SEM of 6.4 percentage points, an MDC95 of 17.9 percentage points, and 95% limits of agreement from −19.5 to 16.2 percentage points. The direction of asymmetry was also weakly consistent: among the 170 participants with a side-to-side difference on both days, the same side was favoured in both sessions in 60.0% of cases (Cohen's κ = 0.14; boys 62.5%, κ = 0.19; girls 57.3%, κ = 0.09).

Table 6. Side-to-side asymmetry in the rotational throw by sex and grade.

Group	*n*	Right Mean (SD)	Left Mean (SD)	Right–Left (SD)	Asymmetry % (SD)	*p*	d_z
Overall	204	4.334 (1.096)	4.184 (1.070)	0.150 (0.422)	7.49 (5.91)	<0.001	0.356
Boys	101	4.564 (1.342)	4.396 (1.283)	0.168 (0.468)	7.80 (6.02)	<0.001	0.360
Girls	103	4.109 (0.723)	3.977 (0.761)	0.132 (0.372)	7.19 (5.81)	<0.001	0.355
Grade 1	44	3.489 (0.888)	3.438 (0.922)	0.051 (0.341)	7.73 (5.21)	0.323	0.151

Grade 2	40	4.405 (0.957)	4.117 (0.889)	0.288 (0.423)	8.68 (6.51)	<0.001	0.681
Grade 3	65	4.456 (1.044)	4.273 (1.038)	0.183 (0.424)	7.93 (5.88)	<0.001	0.430
Grade 4	55	4.815 (1.048)	4.726 (1.008)	0.090 (0.454)	5.91 (5.85)	0.148	0.198

Note: Right and left values represent the mean rotational throw distances. The right–left difference was calculated as the right-side performance minus the left-side performance. Side-to-side asymmetry was calculated as the absolute difference between the sides divided by the better side and multiplied by 100. Paired-samples *t*-tests were used to compare right and left side performances. d_z represents the standardized mean difference for paired comparisons. The asymmetry index was computed for each participant and then averaged; therefore, it can be large even when the mean signed right–left difference is small (e.g., Grade 1), because individual differences favoring opposite sides cancel out in the signed mean but not in the absolute index. Subgroup significance was unchanged after Holm–Bonferroni adjustment. Values are presented as mean (standard deviation, SD).

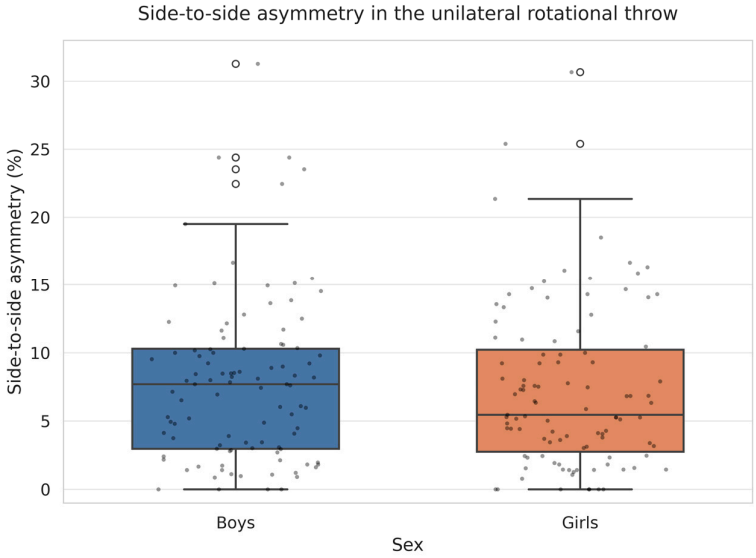


Figure 8. Distribution of side-to-side asymmetry percentages (right vs. left side) in the unilateral rotational throw in boys and girls. Note: Boxplots and individual data points show side-to-side asymmetry percentages in the rotational throw by sex. Asymmetry was calculated as the absolute difference between right- and left-side performance divided by the better side and multiplied by 100. The figure illustrates the distribution and variability of asymmetry values in boys and girls.

3.6. Preliminary 1–10 Reference Scale

Preliminary sex- and grade-specific reference values were generated to support practical interpretation of the unilateral rotational throw in adolescent field settings. Cut-off points were calculated separately for each sex-by-grade cell. Given the small number of participants per sex-by-grade cell (18–37), each decile represents only two to four students; these thresholds should therefore be considered preliminary percentile-based reference values that require validation in a larger, independent, and representative sample before being used as norms.

Each student's score was obtained by placing their mean bilateral distance within the interval corresponding to their sex and grade level. The bilateral average is more stable than the better side alone and preserves the functional information of performance on both sides (Table 7).

Table 7. Preliminary 1–10 reference scale for the rotational throw by sex and grade.

Sex	Grade	1	2	3	4	5	6	7	8	9	10
Boys	1	<2.19	2.19–2.46	2.46–2.77	2.77–2.93	2.93–3.06	3.06–3.86	3.86–4.00	4.00–4.36	4.36–4.61	≥4.61
Boys	2	<3.58	3.58–3.83	3.83–3.98	3.98–4.14	4.14–4.22	4.22–4.57	4.57–4.90	4.90–5.01	5.01–5.53	≥5.53
Boys	3	<3.26	3.26–3.63	3.63–4.17	4.17–4.53	4.53–4.72	4.72–4.82	4.82–5.03	5.03–5.62	5.62–6.91	≥6.91
Boys	4	<3.72	3.72–4.25	4.25–4.53	4.53–4.72	4.72–4.97	4.97–5.13	5.13–5.58	5.58–6.00	6.00–6.34	≥6.34
Girls	1	<2.73	2.73–3.24	3.24–3.36	3.36–3.41	3.41–3.61	3.61–3.86	3.86–4.07	4.07–4.12	4.12–4.28	≥4.28
Girls	2	<3.23	3.23–3.55	3.55–3.69	3.69–3.97	3.97–4.01	4.01–4.13	4.13–4.22	4.22–4.52	4.52–4.78	≥4.78
Girls	3	<3.33	3.33–3.44	3.44–3.62	3.62–3.91	3.91–4.12	4.12–4.26	4.26–4.41	4.41–4.47	4.47–4.69	≥4.69
Girls	4	<3.55	3.55–3.88	3.88–3.97	3.97–4.26	4.26–4.33	4.33–4.56	4.56–4.77	4.77–4.92	4.92–5.04	≥5.04

Note: The scoring scale was developed using preliminary percentile-based (decile) thresholds from the bilateral mean of the rotational throw. Scores range from 1 to 10, with higher scores indicating better performance. The cut-off values are presented in meters and stratified by sex and grade to support educational interpretation in secondary school physical education classes. Each interval includes its lower limit and excludes its upper limit. Sample size per cell: boys n = 24, 18, 28, and 31; girls n = 20, 22, 37, and 24 (Grades 1–4, respectively). These values are exploratory and sample-specific: they were derived from small sex-by-grade subgroups (n = 18–37) of a convenience sample, are not normative, and should be validated in an independent sample before any educational use.

3.7. Usefulness Compared with the Eurofit Frontal Throw

The Bland–Altman analysis showed systematic and proportional bias between the rotational and Eurofit throws (Section 3.3), whereas the rotational throw additionally provided side-specific performance values and an asymmetry index.

The absence of an association between the asymmetry index and Eurofit performance (r = −0.037; p = 0.598) (Table 3) shows that Eurofit performance was not related to the magnitude of side-to-side differences in the rotational throw.

4. Discussion

The main finding of the present study was that the unilateral rotational throw showed acceptable test–retest reliability (good-to-excellent in boys and moderate-to-good in girls) and strong concurrent validity with the Eurofit frontal throw in secondary school students. However, the Bland–Altman analysis showed that the rotational throw yielded higher values and that the difference between the two methods was not constant across the entire performance range. Therefore, the new test should not be considered a direct substitute for Eurofit in metric terms. Because no acceptable limits of agreement were defined a priori, this conclusion describes the observed systematic and proportional differences between the procedures; it does not constitute a formal equivalence or non-equivalence decision against predefined practical criteria. Rather, the rotational throw should be understood as a functionally related test that expands the available information by incorporating unilateral analysis. This interpretation aligns with the methodological literature on method comparison, which clearly distinguishes between association and agreement [42,51].

The test–retest reliability (ICC = 0.70–0.90; Table 2) indicates acceptable stability across measurement days in most conditions, although it was lower in girls, particularly for the right side, which may be related to the systematic Day 1–Day 2 improvement discussed below. Importantly, reliability was not uniform across testing conditions: the right-sided throw in girls (ICC = 0.696; SEM = 10.5% of the mean) did not meet the prespecified thresholds (ICC ≥ 0.75 and SEM < 10%), whereas both sides in boys and the left side in girls did. The conclusion regarding reliability should therefore be condition-specific: the protocol can be considered reliable in boys and for the left side in girls, whereas for the

right side in girls its reliability was only moderate and should be confirmed in further studies, ideally after additional familiarization. The inclusion of the SEM and MDC95 for the rotational throw (SEM = 0.34–0.44 m; MDC95 = 0.94–1.21 m, i.e., 23.6–29.2% of the mean; Table 2) allows for the interpretation of individual change beyond measurement error, a fundamental aspect if the test is used to monitor students or to assess the effect of instructional units aimed at developing strength and coordination. The relative reliability observed is comparable to that reported in a recent meta-analysis of medicine ball throw tests in children and adolescents (pooled ICC = 0.80; 95% CI 0.72–0.86), in which standing chest throws showed only fair reliability (ICC = 0.72) [17]. Likewise, a distance-based rotational medicine ball throw has shown moderate test–retest reliability in adults (ICC = 0.72), whereas ball velocity was more reliable [52], and rotational medicine ball tosses have shown good-to-excellent reliability in adults (ICC = 0.87–0.91) [53]. Thus, the values obtained in boys are at the upper end of those reported for field throwing tests, whereas those obtained in girls fall within the range typically reported for standing throws. The relative SEM of the rotational throw (8.5–10.5%) was slightly higher than that of the Eurofit throw (7.9–8.1%), which is consistent with the greater coordinative demand and the residual learning effect of a novel task. The systematic improvement from Day 1 to Day 2 (0.15–0.22 m; d_z = 0.26–0.52), which occurred despite the familiarization session, has two implications. First, because the absolute-agreement ICC and the SEM incorporate systematic as well as random error, part of the measurement error reported here reflects learning rather than instability of the instrument; the present estimates should therefore be regarded as those applicable when the test is administered after a single familiarization session, and reliability might be higher once the task has been fully learned, although this remains to be demonstrated. Second, for individual performance monitoring, a difference between a first and a second assessment that is smaller than, or similar to, the group learning effect should not be interpreted as a true improvement; a baseline obtained after at least two familiarization sessions, together with the MDC95, is suggested before interpreting individual change.

The strong association observed between the unilateral rotational throw and the Eurofit frontal throw suggests that both procedures share common performance components related to throwing ability. This result provides preliminary evidence of concurrent validity for the proposed protocol. However, because the Eurofit frontal throw was used as a comparison field test rather than as a biomechanical criterion standard, a strong correlation indicates that both tests measure related, but not necessarily identical, constructs; it may partly reflect shared dependence on general strength, power, body size, and throwing skill. The present findings should therefore not be interpreted as comprehensive construct validation, which would require criterion measures (e.g., ball velocity or rotational power derived from kinematic or kinetic analyses) and evidence of convergent and discriminant validity. Moreover, association should not be confused with agreement, because a high correlation may be mistakenly interpreted as equivalence between methods. In the present study, the correlation supported concurrent validity, whereas the Bland–Altman analysis delineated the practical use of the test as a related but independent tool with its own scale [42,51].

The bias identified by the Bland–Altman analysis is also consistent with the fact that the two protocols do not share the same kinematic chain: the Eurofit throw is executed predominantly in the sagittal plane with limited trunk rotation, whereas the rotational throw relies substantially on trunk and hip rotation to generate and transfer force to the ball, engaging the core musculature to a greater extent. This mechanical difference offers a plausible explanation for why the two tests, despite being strongly correlated, show systematic and proportional differences rather than metric equivalence.

These results are consistent with the literature on medicine ball throw tests, which has shown that this type of task can provide reliable and useful field measurements for assessing upper body strength and power in children and adolescents. For example, previous studies have reported high levels of reliability in medicine ball throws and highlighted their feasibility in educational and field assessment contexts [11,54]. The specific contribution of the present study is that the protocol is not limited to a frontal or bilateral throw but instead introduces a unilateral rotational execution, which provides additional information on performance by side.

Sex and grade differences, along with the influence of anthropometric variables, highlight the need to interpret performance using specific reference values. Height, arm span, and biacromial width were associated with performance, which is consistent with the mechanical nature of throwing actions and with previous studies that have shown relationships between performance in throwing tests, age, and body variables such as height and weight [54]. The increase in explained variance attributable to the anthropometric block was similar in the rotational and Eurofit throws (ΔR^2 = 0.159 vs. 0.152), although a similar block-level ΔR^2 does not by itself demonstrate that both tests are equally influenced by body dimensions. Moreover, the substantial multicollinearity among body weight, BMI, and height (VIF > 10; Appendix A) precludes a meaningful interpretation of individual coefficients and, consequently, of which specific body dimensions drive performance in each test. A sensitivity analysis with a reduced set of non-redundant predictors (height, BMI, and biacromial width), free of problematic collinearity, yielded virtually the same block-level increases in explained variance (ΔR^2 = 0.150 vs. 0.159) and identified height as the main anthropometric predictor in both tests. When the between-test difference was modeled directly, the anthropometric block did not add significant explained variance, but height was associated with a larger rotational-minus-Eurofit difference, suggesting that the rotational throw may be slightly more favorable to taller students. Accordingly, the rotational throw can be interpreted as a test influenced by physical development, and possibly slightly more dependent on height than the Eurofit throw; this should be considered when comparing students of different stature. In addition, because the balls weighed 5 kg for boys and 3 kg for girls and body mass differs between sexes, the relative load was not equivalent: the ball represented on average 8.6% of body mass in boys but only 5.6% in girls, so boys threw a proportionally heavier implement. This means that the sex difference observed in the rotational throw (d = 0.42) was obtained under a heavier relative load for boys and may underestimate the sex difference that would be observed with an equivalent relative load, whereas the absence of a pooled sex difference in the Eurofit throw should be interpreted in the same light. Consequently, the present design does not allow differences attributable to the loading protocol to be separated from those attributable to biological, anthropometric, maturational, or skill-related factors, and sex-related differences should be interpreted descriptively. Expressing ball mass relative to body mass, using allometric scaling, or using a common relative load would help clarify this issue in future studies.

BMI showed only weak bivariate associations with performance and was not an independent predictor in either test, which suggests that overall body size and segment lengths, rather than body mass relative to height, are the anthropometric factors most relevant to throwing distance. Because standing throw protocols may favor taller participants, future studies could explore whether seated or kneeling variants of the medicine ball throw reduce the influence of standing height on the score, although this possibility was not examined in the present study.

In addition to the differences associated with sex, grade level, and anthropometric characteristics, extracurricular participation was positively related to performance. Students who engaged in physical or sports activities outside school hours achieved greater

distances, especially in the unilateral rotational throw. This result suggests that the test may be sensitive not only to body profile but also to accumulated motor experience. This interpretation is consistent with motor development models that propose a dynamic relationship between motor competence, physical practice, and physical condition during childhood and adolescence [55]. However, extracurricular practice was not associated with side-to-side asymmetry, indicating that throwing farther does not necessarily imply a more balanced bilateral profile.

The detection of group-level side-to-side differences is one of the potential advantages of the unilateral rotational throw. The fact that the asymmetry index was not associated with Eurofit performance suggests that the frontal test does not capture side-to-side performance differences. At the group level, this information may help teachers plan balanced practice on both sides of the body, especially at ages when laterality, coordination, and motor experience are still being consolidated. However, the right–left differences detected should not be interpreted as necessarily undesirable imbalances: asymmetry is task-specific, may be directionally consistent with sport practice or handedness, and is not inherently pathological. The mean asymmetry of 7.49% lies below the 10–15% thresholds often used in practice to flag asymmetries, although such thresholds are task-specific and lack strong empirical support [25,41]; because handedness was not measured, this descriptive, sample-derived value should not be interpreted as a threshold of functional impairment or injury risk. Furthermore, statistical significance should be distinguished from practical relevance: the mean right–left difference (0.15 m) was statistically significant in a sample of 204 students but smaller than the SEM of the test (0.34–0.44 m). Moreover, the test–retest analysis of the asymmetry index itself showed poor reliability (ICC = 0.22), a large MDC95 (17.9 percentage points), and weak consistency in the favored throwing side across sessions (60% agreement; $\kappa = 0.14$). This is consistent with the well-known low reliability of difference scores derived from two correlated measures: although right- and left-side distances were reliable, much of the between-subject variance they share cancels out when the difference is calculated, leaving a large proportion of measurement error. Therefore, the asymmetry index is useful for describing side-to-side differences at the group level, where random errors tend to average out, but it is not sufficiently reliable to identify asymmetric individuals, to classify the direction of an individual's asymmetry, or to monitor individual changes in asymmetry. Averaging several attempts or sessions might improve its reliability, but this remains to be tested. The literature on inter-limb asymmetries indicates that the magnitude and direction of the differences between sides should be interpreted specifically for each task, avoiding generalizations across different tests [26,41].

Several mechanisms related to coordination and laterality may explain why the asymmetry index was unrelated to Eurofit performance. The Eurofit throw is a bilateral, symmetrical action in which both upper limbs and both sides of the trunk act simultaneously and in the same direction; in such tasks, the stronger or more skilled side can compensate for the weaker one, and performance depends mainly on the overall capacity of the kinetic chain in the sagittal plane [19,20,41]. In contrast, the unilateral rotational throw requires one side of the body to act as the main force generator while the contralateral side stabilizes and decelerates the movement, so performance depends on the side-specific sequencing of the hips, trunk, and shoulders in the transverse plane [27,28,46]. Moreover, during adolescence, lateral preference, inter-limb coordination, and the stabilization of motor patterns are still being consolidated [9,10,55], and prior motor experience is often biased toward the preferred side (e.g., in throwing or racket sports) [21,31]. Consequently, a student may achieve a high bilateral score while displaying a marked difference between sides, or vice versa. This suggests that bilateral output and lateral balance are distinct qualities that require specific tasks for their assessment, although this interpretation

should be confirmed in studies that assess limb dominance and include kinematic analyses.

From an applied perspective, the exploratory 1–10 reference scale may facilitate the transfer of results to the school context. However, because it was derived from small subgroups (18–37 students per cell) of a convenience sample and has not been validated in an independent sample, it is explicitly sample-derived and exploratory: it should not be presented or used as an established normative reference or as an educational assessment standard (e.g., for grading) until it has been validated in a larger, independent, and representative sample. Stratification by sex and grade may allow a fairer interpretation of performance, avoiding direct comparison of students with different maturational and anthropometric characteristics. However, stratification by grade does not resolve maturational heterogeneity within each grade: adolescents of the same chronological age may differ markedly in pubertal timing, and biological maturation may therefore explain part of the observed sex and grade differences, which should not be interpreted solely as age-related or developmental effects [56]. Additionally, the use of the bilateral average as the main indicator combines measurement stability with functional representation of both sides, while the asymmetry index provides complementary, group-level information for planning curricular units. This approach is consistent with the need for fitness and motor competence tests in schoolchildren to be valid, reliable, feasible, and useful for decision-making in educational contexts [11,57].

Perceived exertion ratings indicated a low-to-moderate load and did not add relevant explanatory variance to performance. This suggests that the rotational test can be administered without causing a high perceived demand on students. From an educational perspective, this finding suggests that the test could be integrated into regular physical education sessions without imposing an excessive load. Nevertheless, perceived exertion should be interpreted as a complementary indicator of tolerance and not as a primary determinant of performance.

From a practical standpoint, several recommendations can be offered to PE teachers and youth sport coaches; some derive directly from the present data, whereas others are proposals that require further validation. First (proposal), because a single familiarization session did not prevent a systematic Day 1–Day 2 improvement, at least two familiarization sessions are suggested before formal testing, although whether this removes the learning effect has not been tested. Second (based on the present protocol), the test should be administered after a standardized warm-up, with two attempts per side, using the mean of both attempts as the score. Third (based on the present data), individual changes over time should be interpreted as real only when they exceed both the MDC95 (approximately 1 m for the rotational throw) and the systematic learning effect observed between the first two assessments (0.15–0.22 m); smaller changes may be attributable to measurement error. It should be noted that an MDC95 of 0.94–1.21 m is large relative to the changes that may be expected within a school term, which restricts the usefulness of the test for detecting small individual improvements over short periods; moreover, the MDC95 reflects measurement precision only and should not be confused with a practically or educationally meaningful change, which has not been established for this test. Consequently, at present the test is better suited to describing group performance and identifying large individual changes than to monitoring small individual improvements. Fourth, the bilateral mean can be used together with the preliminary sex- and grade-specific reference values to position students relative to their peers, bearing in mind that these values are not yet normative. Fifth (based on the present data), because the asymmetry index showed poor test–retest reliability, it should be used only to describe side-to-side differences at the group or class level, for example, to justify balanced practice of rotational tasks to both sides, and not to classify or monitor individual students. Finally, rotational

throw performance should not be used interchangeably with Eurofit frontal throw performance, because the two tests yield different scales. These recommendations should be read in the light of the study's limitations (see Section 4.1): the evidence comes from a convenience sample of three urban, state-funded schools; inter-rater reliability was not assessed; and the retest interval and rest periods were not fully standardized. In practice, this means that the test is best used by a single trained evaluator to monitor students within a given school setting, keeping the familiarization, warm-up, rest intervals, and retest interval as constant as possible, and that comparisons between schools, evaluators, or populations that differ from the one studied should be made with caution until these aspects have been examined.

4.1. Limitations

This study has several limitations. First, the sample consisted of secondary education students from three state-funded, urban secondary schools in the city of Valencia serving predominantly middle socioeconomic catchment areas; therefore, the proposed reference values should be interpreted as exploratory, sample-derived values requiring confirmation in larger and more diverse samples, and the findings cannot be generalized to suburban or rural schools or to other socioeconomic contexts. Second, although participants completed a familiarization session (10 frontal and 10 rotational practice throws, five per side) during a PE class one week before testing and specific warm-up throws immediately before testing, the rotational throw—unlike the well-established Eurofit throw—showed a small but statistically significant increase from Day 1 to Day 2 (mean difference 0.15–0.22 m, $p < 0.001$ to 0.011), suggesting a residual learning effect for the novel protocol and that a single familiarization session may not have been sufficient. Third, the interval between sessions was not homogeneous across participants (72 h to one week) and was shorter than the 1–2-week interval typically recommended for test–retest designs, which may have contributed to this learning effect. Fourth, the two-evaluator protocol used (one evaluator scoring the first attempt, the other the second) does not allow a valid estimate of inter-rater reliability, since this requires both raters to independently score the same attempts; inter-rater reliability could therefore not be assessed [58]. Fifth, although both evaluators held a PhD in Physical Activity and Sport Science and had prior field-testing experience, a more detailed record of their specific training on this protocol was not kept. Sixth, limb dominance (handedness) was not assessed, so side-to-side asymmetry is reported in absolute right-versus-left terms rather than dominant-versus-non-dominant terms. Seventh, an exploratory association was observed between extracurricular physical activity and throwing performance, but this variable was not part of the primary hypotheses and its type, dose, and specialization were not systematically characterized. In addition, biological maturation (e.g., maturity offset or peak height velocity) was not assessed [56], although it may influence throwing performance and asymmetry beyond chronological age and grade; ball mass relative to body mass differed between sexes, which limits direct sex comparisons; the stance width of 1.5 times the biacromial distance was adapted from the shot-put literature and has not been specifically validated in adolescents; the 1–10 reference values were derived from small sex-by-grade subgroups (18–37 participants) and should therefore be considered preliminary rather than normative [6]; the rest interval between attempts was not fixed; only two attempts per throw were recorded, and three or more attempts might provide a more stable score; the rotational technique was standardized through instructions, demonstration, and feedback, but movement execution was not controlled biomechanically; throwing distance was not expressed in a relative or allometrically scaled form; sport participation was characterized only by participation and weekly hours of practice; leg length was measured with a non-standard definition and its technical error of measurement was not calculated; the diameter of the medicine balls, and

therefore its relationship with hand size and grip, was not recorded or standardized; the submaximal intensity of the warm-up throws was self-regulated and not monitored instrumentally; no a priori sample size calculation was performed; the strong multicollinearity among body weight, BMI, and height in the regression models (VIF > 10) precludes interpretation of individual anthropometric coefficients, so the relative influence of body dimensions on each test cannot be determined; the reliability of the right-sided throw in girls did not meet the prespecified thresholds, so the reliability conclusions are condition-specific; and, despite the Holm–Bonferroni adjustment, the large number of sex-, grade-, side-, and activity-based comparisons increases the risk of Type I error, so these subgroup analyses should be considered exploratory. Furthermore, the asymmetry index showed poor test–retest reliability and the favored throwing direction was only weakly consistent across sessions, so the asymmetry results apply to the group level only; the nesting of students within class groups and schools was not modeled, given the small number of clusters; assessors were not blinded to participants' Day 1 performance; the medicine-ball dimensions were not standardized; and the MDC95 values (0.94–1.21 m) indicate that relatively large changes are required to distinguish genuine individual improvement from measurement variability, while a practically meaningful change has not yet been defined for this test. Finally, the Eurofit throw was used as a concurrent comparison test, not a biomechanical criterion standard; future studies should incorporate kinematic, kinetic, and electromyographic measurements, multiple familiarization sessions, a fixed retest interval, a fully crossed inter-rater design, and an assessment of limb dominance.

Several of these limitations have direct practical implications and deserve further comment. First, regarding generalizability, the three participating schools were those whose management teams and PE departments agreed to take part; these schools may have been more motivated or more familiar with fitness testing than non-participating schools, and the sample included only urban, state-funded schools with predominantly middle socioeconomic catchment areas. The reliability estimates and reference values should therefore be applied only to similar school populations until they are replicated in other settings. Second, regarding inter-rater reliability, each evaluator scored a different attempt, so systematic differences between evaluators, if present, would have been incorporated into the measurement error rather than quantified separately; the present SEM and MDC95 values are therefore applicable to the situation in which trained evaluators follow the same protocol, and a fully crossed design, in which both evaluators independently score the same attempts, is needed before results obtained by different evaluators can be compared. Third, regarding standardization, neither the retest interval (72 h to one week) nor the rest between attempts (at least 2 min, determined by the rotation of the rows) was fixed. Longer intervals may have allowed more consolidation of learning, and variable rest periods may have introduced additional within-session variability; both factors may therefore have contributed to the systematic Day 1–Day 2 improvement and to the measurement error reported. In practice, teachers should keep these conditions as constant as possible between assessments. Finally, 22 of the 226 eligible students (9.7%) did not attend the second session and were excluded from the analysis. Because Day 1 data were not retained for these students, their characteristics and performance could not be compared with those of completers; non-completion bias therefore cannot be excluded, and the findings may not be fully representative of students with lower attendance.

4.2. Future Research

Future research should analyze the role of motor experience in greater detail, differentiating not only between students with and without extracurricular practice but also according to sports type, frequency and volume of practice, sports specialization, lateral dominance, and specific experience in throwing tasks, among other variables. This

information would make it possible to determine whether rotational performance and side-to-side asymmetries respond differently depending on students' physical practice profiles. Future studies should also validate the preliminary reference values in larger, independent, and representative samples, examine the influence of biological maturation, and explore seated or kneeling variants of the test to reduce the influence of standing height.

5. Conclusions

The unilateral rotational medicine ball throw showed good-to-excellent relative reliability in boys (ICC 0.90), good reliability for the left side in girls (ICC 0.82), and only moderate reliability for the right side in girls (ICC 0.70), the latter condition failing to meet the prespecified thresholds; absolute reliability was acceptable in boys and for the left side in girls (SEM 8.5–9.5% of the mean) but marginal for the right side in girls (SEM 10.5%). The test showed a strong association with the Eurofit frontal throw (r = 0.81, partial r = 0.74 after adjustment), which provides preliminary evidence of concurrent validity against a comparison field test but not comprehensive construct validation. Bland–Altman analysis revealed systematic and proportional differences between the two tests, which may reflect mechanical differences between the protocols; because no acceptable limits of agreement were prespecified, this finding describes the nature of the differences rather than a formal equivalence decision. At the block level, anthropometric variables explained a similar proportion of variance in both tests, and a sensitivity analysis with non-redundant predictors confirmed this pattern, although height was associated with a slightly larger rotational-minus-Eurofit difference. Overall, within the limits of this sample, the rotational throw appears to be a promising complementary field test of whole-body rotational throwing performance that adds side-specific performance information and group-level, direction-specific asymmetry information not available from the Eurofit throw; the asymmetry values reported are descriptive and should not be interpreted as thresholds of functional impairment or injury risk. However, given the systematic Day 1–Day 2 improvement, the moderate reliability observed in girls for the right side, and the use of a comparison field test rather than a criterion standard, further validation against biomechanical measures and in independent samples is required. At present, therefore, the test provides preliminary measurement evidence rather than comprehensive validation for individual monitoring, and the asymmetry index showed poor test–retest reliability (ICC = 0.22) and an inconsistent direction across sessions, so it is suitable only for describing group-level side-to-side differences and not for individual monitoring. The proposed reference values should be considered preliminary until validated in larger samples.

Author Contributions: Conceptualization, J.M.-R.; methodology, J.M.-R. and L.R.-S.; formal analysis, J.M.-R. and L.R.-S.; investigation, J.M.-R., A.L.-A., A.A.-P. and L.R.-S.; data curation, J.M.-R.; writing—original draft preparation, J.M.-R.; writing—review and editing, J.M.-R. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of the Catholic University of Valencia (protocol code UCV/2025-2026/120, date of approval 26 February 2026). All recruitment, familiarization, and data collection procedures were conducted after ethical approval had been obtained, between late February and March 2026.

Informed Consent Statement: Written informed consent was obtained from the parents or legal guardians of all participants involved in the study, and all students provided assent before participation.

Data Availability Statement: The datasets used and analyzed during the current study are available from the corresponding author upon reasonable request due to privacy and ethical restrictions.

Acknowledgments: The authors are very grateful to the schools, teachers, and students who contributed to this study.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ANOVA	Analysis of variance
BMI	Body mass index
CI	Confidence interval
FMS	Functional Movement Screen
ICC	Intraclass correlation coefficient
ISAK	International Society for the Advancement of Kinanthropometry
LoA	Limits of agreement
MDC95	Minimal detectable change at the 95% confidence level
PE	Physical Education
PES	Perceived exertion scale
RPE	Rating of perceived exertion
SD	Standard deviation
SEM	Standard error of measurement

Appendix A

Table A1. Standardized regression coefficients (β), 95% confidence intervals (CI), and variance inflation factors (VIF) for the full hierarchical models (age, sex, and grade plus anthropometric variables).

Predictor	VIF	Eurofit Throw: β (95% CI); *p*	Rotational Throw (Bilateral Mean): β (95% CI); *p*
Age	5.3	−0.05 (−0.31 to 0.20); 0.674	0.06 (−0.19 to 0.32); 0.637
Sex (0 = girls, 1 = boys)	1.4	−0.21 (−0.34 to −0.08); 0.002	−0.04 (−0.17 to 0.10); 0.582
Grade	5.3	0.31 (0.06 to 0.57); 0.015	0.12 (−0.14 to 0.38); 0.361
Weight	92.2	0.02 (−1.03 to 1.07); 0.967	0.14 (−0.93 to 1.20); 0.802
Height	33.7	0.25 (−0.38 to 0.89); 0.433	0.33 (−0.31 to 0.98); 0.313
BMI	67.1	0.03 (−0.87 to 0.92); 0.953	−0.05 (−0.96 to 0.86); 0.917
Arm span	2.5	0.00 (−0.17 to 0.18); 0.989	0.02 (−0.15 to 0.20); 0.785
Biacromial width	2.3	0.20 (0.04 to 0.37); 0.018	0.13 (−0.04 to 0.30); 0.123
Leg length	3.2	0.10 (−0.10 to 0.29); 0.335	0.00 (−0.20 to 0.20); 0.987

Note: All variables were standardized before entering the model. VIF values above 10 for weight, BMI, and height indicate strong collinearity among these variables, which explains why none of them reached individual significance despite their bivariate associations with performance.

References

1. Bass, R.W.; Brown, D.D.; Laurson, K.R.; Coleman, M.M. Physical fitness and academic performance in middle school students. *Acta Paediatr.* **2013**, *102*, 832–837.
2. Cuadrado, G.; Morante, J.C.; Redondo, J.C.; Zarzuela, R. *Valoración de la Condición Física de la Población Escolar Mediante la Batería Eurofit*; Wanceulen: Sevilla, Spain, 2009.

3. Hincapié, C.A.; Tomlinson, G.A.; Hapuarachchi, M.; Stankovic, T.; Hirsch, S.; Carnegie, D.; Richards, D.; Frost, D.; Beach, T.A. Functional Movement Screen Task Scores and Joint Range-of-motion: A Construct Validity Study. *Int. J. Sports Med.* **2022**, *43*, 648–656.
4. O'Brien, W.; Khodaverdi, Z.; Bolger, L.; Tarantino, G.; Philpott, C.; Neville, R.D. The Assessment of Functional Movement in Children and Adolescents: A Systematic Review and Meta-Analysis. *Sports Med.* **2022**, *52*, 37–53.
5. Lundgren, U.P. Nuevos desafíos para los profesores y para la formación del profesorado: [Informe presentado a la Conferencia Permanente de Ministros Europeos de Educación del Consejo de Europa (15ª Sesión, Helsinki, 5-7 de mayo de 1987)]. *Rev. Educ.* **1988**, *285*, 291–328.
6. Tomkinson, G.R.; Carver, K.D.; Atkinson, F.; Daniell, N.D.; Lewis, L.K.; Fitzgerald, J.S.; Lang, J.J.; Ortega, F.B. European normative values for physical fitness in children and adolescents aged 9–17 years: Results from 2 779 165 Eurofit performances representing 30 countries. *Br. J. Sports Med.* **2018**, *52*, 1445–1456.
7. Hernández, J. Relación entre diferentes pruebas de campo: Fuerza, potencia y velocidad. *Pensar Mov. Rev. Cienc. Ejerc. Salud* **2003**, *3*, 1–10.
8. Bompa, T. *Periodización. Teoría y Metodología del Entrenamiento*, 1st ed.; Hispano Europea: Barcelona, Spain, 2016.
9. Lloyd, R.S.; Oliver, J.L. The Youth Physical Development Model: A new approach to long-term athletic development. *Strength Cond. J.* **2012**, *34*, 61–72.
10. Lloyd, R.S.; Faigenbaum, A.D.; Stone, M.H.; Oliver, J.L.; Jeffreys, I.; Moody, J.A.; Brewer, C.; Pierce, K.C.; McCambridge, T.M.; Howard, R.; et al. Position statement on youth resistance training: The 2014 International Consensus. *Br. J. Sports Med.* **2014**, *48*, 498–505.
11. Marques, A.; Henriques-Neto, D.; Peralta, M.; Martins, J.; Gomes, F.; Popovic, S.; Masanovic, B.; Demetriou, Y.; Schlund, A.; Ihle, A. Field-Based Health-Related Physical Fitness Tests in Children and Adolescents: A Systematic Review. *Front. Pediatr.* **2021**, *9*, 640028.
12. Bernate, J.; Fonseca, I.; Babativa, H. Revisión sistemática de las estrategias didácticas en la Educación Física para el desarrollo de habilidades motrices. *Cienc. Deporte* **2023**, *8*, 16–31.
13. Borms, D.; Maenhout, A.; Cools, A.M. Upper quadrant field tests and isokinetic upper limb strength in overhead athletes. *J. Athl. Train.* **2016**, *51*, 789–796.
14. Martínez López, E.J. Aplicación de la prueba de lanzamiento de balón medicinal, abdominales superiores y salto horizontal a pies juntos. Resultados y análisis estadístico en Educación Secundaria. *Rev. Int. Med. Cienc. Act. Física Deporte* **2003**, *3*, 223–241.
15. Jarnig, G.; van Poppel, M.N.M.; Kerbl, R. Model Calculation of the Influence of Body Height on Performance in Medicine Ball Throw. *Paediatr. Paedol.* **2025**, *60*, 308–311.
16. Podstawski, R.; Markowski, P.; Clark, C. Sex-mediated differences and correlations between the anthropometric characteristics and motor abilities of university students. *J. Phys. Educ. Sport.* **2020**, *20*, 86–96.
17. González-Devesa, D.; Varela, S.; Diz-Gómez, J.C.; López-Amoedo, D.; Ayán-Pérez, C. Reliability and validity of the medicine ball throw test in children and adolescents: A systematic review and meta-analysis. *Kinesiology* **2025**, *57*, 122–135.
18. Erikoğlu, Ö.; Güzel, N.; Pense, M.; Örer, G.E. Comparison of physical fitness parameters with EUROFIT test battery of male adolescent soccer players and sedentary counterparts. *Int. J. Sport Cult. Sci.* **2015**, *3*, 43–52.
19. Cannon, R.J.; Cafarelli, E. Neuromuscular adaptations to training. *J. Appl. Physiol.* **1987**, *63*, 2396–2402.
20. Howard, J.D.; Enoka, R.M. Enhancement of maximal force by contralateral-limb stimulation. *J. Biomech.* **1987**, *20*, 908.
21. Mayolas, C. Un nuevo test de valoración de la lateralidad para los profesionales de la educación física. *Apunt. Educ. Física Deport.* **2003**, *71*, 14–22.
22. Oxendale, C.; Smith, G. Association between 2D landing biomechanics, isokinetic muscle strength and asymmetry in females using novel, task specific metrics based on ACL injury mechanisms. *PLoS ONE* **2025**, *20*, e0326882.
23. Turati, M.; Benedettini, E.; Sugimoto, D.; Crippa, M.; Alessandro, C.; Bacchin, V.; Piatti, M.; Albanese, F.; Accadbled, F.; Rigamonti, L.; et al. Quadriceps and hamstring muscles strength differences in adolescent and adult recreational athletes 6 months after autograft bone-patellar-tendon-bone anterior cruciate ligament reconstruction: A retrospective study. *Knee* **2025**, *54*, 9–18.
24. Díaz-López, N.; Monge-Pereira, E.; Jodra-Centeno, E.; Molina-Rueda, F.; Miangolarra-Page, J.C. Uso del reconocimiento de la lateralidad a través de imaginería motora implícita para la mejora del control postural y el equilibrio en pacientes con ictus subagudo: Un estudio controlado aleatorizado. *Rev. Neurol.* **2022**, *74*, 375.
25. Maloney, S.J. The relationship between asymmetry and athletic performance: A critical review. *J. Strength Cond. Res.* **2019**, *33*, 2579–2593.

26. Bishop, C.; Turner, A.; Read, P. Effects of inter-limb asymmetries on physical and sports performance: A systematic review. *J. Sports Sci.* **2018**, *36*, 1135–1144.
27. Fang, Q.; Xia, Y.; Zhang, X.; Huang, F. Asymmetry of interlimb transfer: Pedagogical innovations in physical education. *Front. Psychol.* **2022**, *13*, 1029888.
28. Guillén del Castillo, M.; Linares Girela, D. *Bases Biológicas y Fisiológicas del Movimiento Humano*; Médica Panamericana: Madrid, Spain, 2002.
29. Lijewski, M.; Burdukiewicz, A.; Pietraszewska, J.; Andrzejewska, J.; Stachoń, A. Asymmetry of Muscle Mass Distribution and Grip Strength in Professional Handball Players. *Int. J. Environ. Res. Public Health* **2021**, *18*, 1913.
30. Stern, D.; Gonzalo-Skok, O.; Loturco, I.; Turner, A.; Bishop, C. A Comparison of Bilateral vs. Unilateral-Biased Strength and Power Training Interventions on Measures of Physical Performance in Elite Youth Soccer Players. *J. Strength Cond. Res.* **2020**, *34*, 2105–2111.
31. Chapelle, L.; Bishop, C.; D'Hondt, J.; D'Hondt, E.; Clarys, P. Morphological and functional asymmetry in elite youth tennis players compared to sex- and age-matched controls. *J. Sports Sci.* **2022**, *40*, 1618–1628.
32. López-Barrancos, S. *Fiabilidad y Validez de un Protocolo de Evaluación de la Condición Física Relacionada Con la Salud (COFISA) en Escolares [tesis de Maestría]*; Universidad de Murcia: Murcia, Spain, 2008.
33. Guijarro-Romero, S.; Mayorga-Vega, D.; Casado-Robles, C.; Viciana, J. Desarrollo y validación de una prueba escrita objetiva de elección múltiple para evaluar el conocimiento del entorno para el acondicionamiento físico (CENAFI) en escolares (Development and validation of an objective written test of multiple-choice to assess the knowledge about the environment for physical conditioning (CENAFI) in schoolchildren). *Retos* **2024**, *51*, 426–441.
34. Hendrik; Karo Karo, A.A.P.; Sari, L.P. Validity and Reliability of a Physical Fitness Test Instrument for Physical Education Students. *ACPES J. Phys. Educ. Sport Health* **2025**, *5*, 96–102. https://doi.org/10.15294/ajpesh.v5i2.33951.
35. Vehrs, P.R.; Uvacsek, M.; Johnson, A.W. Assessment of Dysfunctional Movements and Asymmetries in Children and Adolescents Using the Functional Movement Screen-A Narrative Review. *Int. J. Environ. Res. Public Health* **2021**, *18*, 12501.
36. Borg, E.; Kaijser, L. A comparison between three rating scales for perceived exertion and two different work tests. *Scand. J. Med. Sci. Sports* **2006**, *16*, 57–69.
37. Lagally, K.M.; Robertson, R.J. Construct validity of the OMNI resistance exercise scale. *J. Strength Cond. Res.* **2006**, *20*, 252–256.
38. Colado, J.C.; García-Massó, X.; Triplett, N.T.; Flandez, J.; Borreani, S.; Tella, V. Concurrent Validation of the OMNI-Resistance Exercise Scale of Perceived Exertion with Thera-Band Resistance Bands. *J. Strength Cond. Res.* **2012**, *26*, 3018–3024.
39. Bishop, C.; Read, P.; Lake, J.; Chavda, S.; Turner, A. Interlimb Asymmetries: Understanding How to Calculate Differences From Bilateral and Unilateral Tests. *Strength Cond. J.* **2018**, *40*, 1–6.
40. Weir, J.P. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *J. Strength Cond. Res.* **2005**, *19*, 231–240.
41. Roe, G.; Shaw, W.; Darrall-Jones, J.; Phibbs, P.J.; Read, D.; Weakley, J.J.; Till, K.; Jones, B. Reliability and Validity of a Medicine Ball–Contained Accelerometer for Measuring Upper-Body Neuromuscular Performance. *J. Strength Cond. Res.* **2018**, *32*, 1915–1918.
42. Sayers, M.G.L.; Bishop, S. Reliability of a New Medicine Ball Throw Power Test. *J. Appl. Biomech.* **2017**, *33*, 311–315.
43. Koo, T.K.; Li, M.Y. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J. Chiropr. Med.* **2016**, *15*, 155–163.
44. Stewart, A.; Marfell-Jones, M.; Olds, T.; de Ridder, H. *International Standards for Anthropometric Assessment*; International Society for the Advancement of Kinanthropometry: Lower Hutt, New Zealand, 2011.
45. Jiao, J.; Yang, T.; Ma, Y.; Wang, H.; Chen, J.; Xu, H.; Yang, G. Kinematic Characteristics of the Non-Throwing Arm During the Completion Phase of the Glide Shot Put in Elite Female Athletes: A Case Study. *Appl. Sci.* **2025**, *15*, 1397.
46. Kato, T.; Maeda, K.; Mizushima, J.; Maeda, A. Investigation of a Theoretical Model for the Rotational Shot Put Technique. *J. Appl. Biomech.* **2024**, *40*, 528–536.
47. Almadhkhori, H.; Pavlović, R.; Skrypchenko, I.; Rafahiya, B.; Singh, R.R.M. Predictive value of kinematic indicators for shot put result and selection of novice athletes. *Health Sport Rehabil.* **2021**, *7*, 35–45.
48. Bland, J.M.; Altman, D.G. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* **1986**, *1*, 307–310.
49. Mukaka, M.M. Statistics corner: A guide to appropriate use of correlation coefficient in medical research. *Malawi Med. J.* **2012**, *24*, 69–71.
50. Holm, S. A simple sequentially rejective multiple test procedure. *Scand. J. Stat.* **1979**, *6*, 65–70.

51. Davis, K.L.; Kang, M.; Boswell, B.B.; DuBose, K.D.; Altman, S.R.; Binkley, H.M. Validity and reliability of the medicine ball throw for kindergarten children. *J. Strength Cond. Res.* **2008**, *22*, 1958–1963.
52. Atkinson, G.; Nevill, A.M. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Med.* **1998**, *26*, 217–238.
53. Hardy, S.G.J.; Stelzer-Hiller, O.W.; Edwards, K.M.; Freeston, J. Criterion validity and reliability of a new medicine ball rotational power test. *J. Strength Cond. Res.* **2025**, *39*, e429–e435.
54. Sell, M.A.; Abt, J.P.; Sell, T.C.; Keenan, K.A.; Allison, K.F.; Lovalekar, M.T.; Lephart, S.M. Reliability and validity of medicine ball toss tests as clinical measures of core strength. *Isokinet. Exerc. Sci.* **2015**, *23*, 151–160.
55. Stodden, D.F.; Goodway, J.D.; Langendorfer, S.J.; Roberton, M.A.; Rudisill, M.E.; Garcia, C.; Garcia, L.E. A developmental perspective on the role of motor skill competence in physical activity: An emergent relationship. *Quest* **2008**, *60*, 290–306.
56. Mirwald, R.L.; Baxter-Jones, A.D.G.; Bailey, D.A.; Beunen, G.P. An assessment of maturity from anthropometric measurements. *Med. Sci. Sports Exerc.* **2002**, *34*, 689–694.
57. Barnett, L.M.; Jerebine, A.; Keegan, R.; Watson-Mackie, K.; Arundell, L.; Ridgers, N.D.; Salmon, J.; Dudley, D. Validity, Reliability, and Feasibility of Physical Literacy Assessments Designed for School Children: A Systematic Review. *Sports Med.* **2023**, *53*, 1905–1929.
58. Shrout, P.E.; Fleiss, J.L. Intraclass correlations: Uses in assessing rater reliability. *Psychol. Bull.* **1979**, *86*, 420–428.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.