

Review Article

Validity and Reliability of the Athletic Shoulder Test: A Brief Systematic Review and Meta-Analysis

Süleyman Ulupınar¹, İzzet İnce², CebraİL Gençođlu¹, Selim Asan¹, Salih Çabuk¹, Serhat Özbay¹

¹Faculty of Sports Sciences, Erzurum Technical University, Erzurum, Turkey;

²Faculty of Sports Sciences, Ankara Yıldırım Beyazıt University, Ankara, Turkey

Abstract

This systematic review and meta-analysis aimed to synthesize the validity and reliability of the AST and explore its measurement precision through standard error of measurement (SEM) and minimal detectable change (MDC). A systematic search of PubMed, Scopus, and Web of Science databases was conducted to identify studies evaluating the AST's psychometric properties. Studies reporting quantitative validity or reliability data (e.g., correlation coefficients, ICC) were included. A random-effects meta-analysis was performed to pool validity and reliability estimates, with heterogeneity assessed using I^2 and τ^2 . SEM and MDC values were systematically reviewed to evaluate measurement precision. Seven studies were included in the systematic review, with five studies (42 validity estimates) and six studies (35 reliability estimates) contributing to the meta-analysis. The AST demonstrated promising levels of validity (pooled $r = 0.923$) and reliability (pooled $r = 0.958$), although substantial heterogeneity was observed ($I^2 > 85\%$) and should be considered when interpreting these findings. No significant publication bias was detected. SEM ranged from 0.85 – 20.43 N, and MDC from 2.37 – 56.63 N, indicating variable measurement precision. Overall, the AST appears to be a potentially useful tool for assessing shoulder function, with relevance for clinical and research applications.

Keywords: Reliability, Shoulder function, Standard error of measurement, Validity

Introduction

The shoulder complex is crucial for daily and athletic tasks, enabling the wide range of movements needed for function and sport-specific performance^{1,2}. Its intricate structure, however, makes it susceptible to injuries and dysfunction, particularly in populations engaging in repetitive overhead motions, such as athletes in team and individual sports^{3,4}. Accurate assessment of shoulder strength and stability is therefore critical for injury prevention, rehabilitation, and performance optimization^{5,6}. The Athletic Shoulder Test (AST), a standardized assessment involving isometric force measurements in the I (180°), Y (135°), and T (90°)

positions, has been developed to evaluate shoulder function by targeting key muscle groups involved in stability and force production⁷. This test offers a practical and objective method to quantify shoulder performance, making it a valuable tool in clinical and research settings^{7,8}.

The AST specifically targets the shoulder's musculoskeletal system, engaging key muscle groups that contribute to stability and force generation⁷⁻¹³. The shoulder complex, comprising the glenohumeral, acromioclavicular, and scapulothoracic articulations, relies on the coordinated action of the rotator cuff muscles (supraspinatus, infraspinatus, teres minor, and subscapularis), deltoid, and scapular stabilizers such as the trapezius and serratus anterior^{6,14}. During the AST, the I (180°) position primarily activates the deltoid and supraspinatus, facilitating abduction, while the Y (135°) position emphasizes the rotator cuff's role in external rotation and scapular stability, with increased involvement of the infraspinatus and teres minor^{7,15,16}. The T (90°) position, in contrast, targets the subscapularis and pectoralis major, focusing on internal rotation and horizontal adduction^{7,15,16}. Measuring isometric force in these positions, the AST biomechanically evaluates

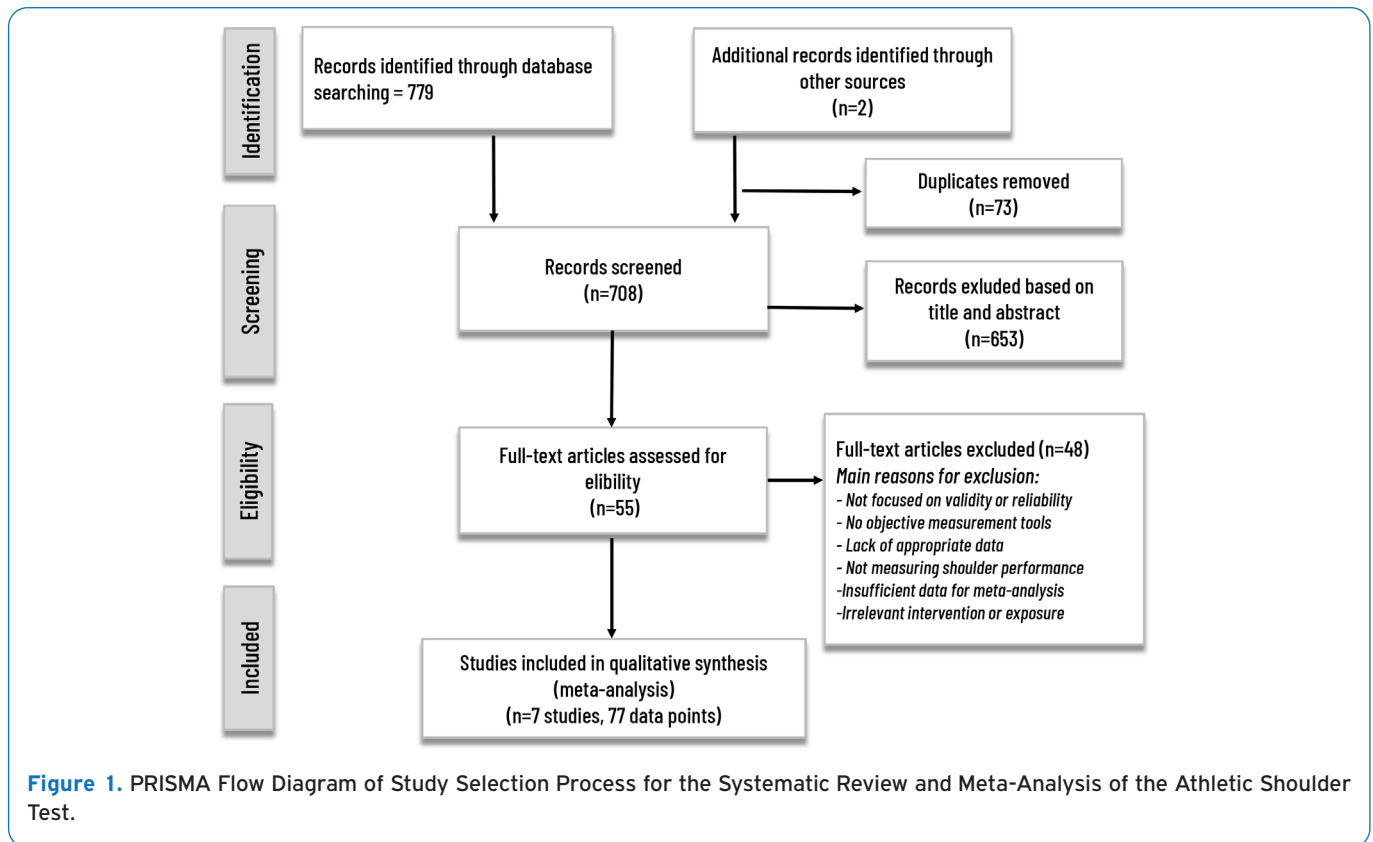
The authors have no conflict of interest.

Corresponding author: Süleyman Ulupınar, Faculty of Sports Sciences, Erzurum Technical University, Erzurum, Turkey
E-mail: suleyman.ulupinar@erzurum.edu.tr

Edited by: G. Lyritis

Accepted 9 June 2025





shoulder function, capturing the interaction of muscle activation, joint stability, and force generation^{9,11,13}.

Previous studies have investigated the psychometric properties of the AST, focusing on its validity and reliability across diverse populations, including elite and amateur athletes as well as healthy individuals⁷⁻¹³. While some studies have reported high validity, with correlation coefficients often exceeding 0.80^{11,12}, others have shown more variable results, particularly when assessing specific subgroups or modified test protocols^{8,9}. Reliability estimates have generally been strong, with intra-class correlation coefficients (ICC) frequently above 0.90^{7,10}, though inconsistencies arise due to differences in test settings, participant characteristics, and measurement tools. Such heterogeneity, along with small sample sizes, limits generalizability and underscores the need for comprehensive synthesis. A systematic review and meta-analysis are thus warranted to provide a robust evaluation of the AST's psychometric properties, addressing these gaps and offering a clearer understanding of its applicability in clinical and research contexts.

The present study aims to address these limitations by conducting a systematic review and meta-analysis to synthesize the validity and reliability of the Athletic Shoulder Test across existing literature. By pooling data

from diverse studies, this meta-analysis seeks to provide a more precise estimate of the AST's psychometric properties, accounting for variability in study designs and populations. Furthermore, it aims to explore potential sources of heterogeneity, such as differences in participant characteristics and test protocols, to better understand the factors influencing the test's performance. Ultimately, this study offers evidence-based insights into the AST's clinical and research use—supporting its role in injury prevention, rehabilitation, and performance assessment across populations.

Methods

Study Selection and Inclusion Criteria

A systematic literature search was conducted to identify studies evaluating the psychometric properties of the AST (Figure 1). Databases including PubMed, Scopus, and Web of Science were searched from inception to March 2025, using a combination of keywords such as "Athletic Shoulder Test," "validity," "reliability," "shoulder assessment," and "isometric strength." Additional studies were identified through manual searches of reference lists and relevant journals. Studies were included if they met the following criteria: (1) evaluated the validity or reliability of the AST

Table 1. Characteristics and Psychometric Outcomes of Studies Included in the Systematic Review and Meta-Analysis of the AST.

Study (Authors and Year)	Participant Characteristics (N, Age, Gender, Level)	Test Setting	Validity Outputs	Reliability Outputs
Ashworth et al. ⁷	18; Age: 22.4 ± 4.6 years; Gender: M/F: 18/O; Sport: Team sports; Level: Elite	Positions: I (180°), Y (135°), T (90°); Tools: Force plate	No quantitative data reported	Reliability: Dominant: ICC: I: 0.97, Y: 0.96, T: 0.98, SEM: I: 8.6 N, Y: 8.2 N, T: 4.8 N, MDC90: I: 17.1 N, Y: 22.7 N, T: 13.3 N; Non-Dominant: ICC: I: 0.95, Y: 0.94, T: 0.96, SEM: I: 10.8 N, Y: 9.4 N, T: 6.3 N, MDC90: I: 21.3 N, Y: 25.9 N, T: 17.5 N
İnce ¹²	12; Age: 16.13 ± 2.21 years; Gender: M/F: 12/O; Sport: Individual sports; Level: Amateur	Positions: I (180°), Y (135°), T (90°); Tools: Force plate; Condition: Prone position, Bilateral	R ² : I: 0.995, Y: 0.986, T: 0.994	Reliability: ICC: I: 0.997 (0.994–0.999), Y: 0.996 (0.995–0.998), T: 0.999 (0.997–0.999), SEM: I: 1.82 N, Y: 1.20 N, T: 0.85 N, MDC: I: 5.07 N, Y: 3.34 N, T: 2.37 N
Morrison et al. ¹¹	20; Age: 25.15 ± 3.27 years; Gender: M/F: 20/O; Sport: Team sports; Level: Amateur	Positions: I (180°), Y (135°), T (90°); Tools: Force plate, Sphygmomanometer; Condition: Prone position	r: I: 0.770 (0.64–0.9), Y: 0.817 (0.66–0.93), T: 0.764 (0.64–0.9)	Reliability: I: SEM: 20.43 N, MDC90: 56.63 N; Y: SEM: 12.26 N, MDC90: 33.99 N; T: SEM: 13.29 N, MDC90: 36.83 N
Olds et al. ¹³	27; Age: 37.0 ± 12.0 years; Gender: M/F: 10/17; Sport: Not specified; Level: Healthy	Positions: I (180°), Y (135°), T (90°); Tools: Force plate, Handheld dynamometer; Condition: Prone position	ICC (Force, kg): Dominant: I: 0.86 (0.72–0.94), Y: 0.87 (0.73–0.94), T: 0.87 (0.74–0.94); Non-Dominant: I: 0.89 (0.78–0.95), Y: 0.92 (0.82–0.96), T: 0.84 (0.69–0.93); ICC (Torque, N·m): Dominant: I: 0.89 (0.76–0.95), Y: 0.89 (0.77–0.95), T: 0.89 (0.76–0.95); Non-Dominant: I: 0.92 (0.82–0.96), Y: 0.94 (0.87–0.97), T: 0.87 (0.73–0.94)	Reliability (Force, HHD): Dominant: ICC: I: 0.93 (0.88–0.97), Y: 0.94 (0.89–0.97), T: 0.95 (0.91–0.98), SEM: I: 9.03 N, Y: 7.06 N, T: 6.08 N, MDC95: I: 24.43 N, Y: 19.52 N, T: 16.78 N; Non-Dominant: ICC: I: 0.93 (0.86–0.96), Y: 0.80 (0.66–0.89), T: 0.93 (0.88–0.97), SEM: I: 8.53 N, Y: 13.73 N, T: 6.47 N, MDC95: I: 23.74 N, Y: 37.96 N, T: 17.95 N; (Force, Force plate): Dominant: ICC: I: 0.92 (0.86–0.96), Y: 0.91 (0.84–0.96)
Paliouras et al. ⁹	33; Age: 12.6 ± 2.0 years; Gender: M/F: 17/16; Sport: Individual sports; Level: Amateur	Positions: I (180°), Y (135°), T (90°); Tools: Force plate, Handheld dynamometer; Condition: Includes SSASPT	ICC (SSASPT): Dominant: I: 0.808, Y: 0.856, T: 0.824; Non-Dominant: I: 0.755, Y: 0.823, T: 0.758; ICC (ER): Dominant: I: 0.605, Y: 0.742, T: 0.712; Non-Dominant: I: 0.584, Y: 0.631, T: 0.645; ICC (IR): Dominant: I: 0.772, Y: 0.867, T: 0.820; Non-Dominant: I: 0.771, Y: 0.752, T: 0.693	Test-Retest: Dominant: ICC: I: 0.949 (0.898–0.975), Y: 0.963 (0.925–0.981), T: 0.924 (0.845–0.962), SEM: I: 6.47 N, Y: 4.12 N, T: 5.29 N, MDC: I: 18.04 N, Y: 11.53 N, T: 14.71 N; Non-Dominant: ICC: I: 0.970 (0.939–0.985), Y: 0.967 (0.934–0.984), T: 0.959 (0.934–0.984), SEM: I: 4.71 N, Y: 3.53 N, T: 3.33 N, MDC: I: 13.04 N, Y: 9.71 N, T: 9.32 N
Schellekens et al. ¹⁰	42; Age: 28.66 ± 9.56 years; Gender: M/F: 28/14; Sport: Individual sports; Level: Mixed	Positions: I (180°), Y (135°), T (90°); Tools: Force plate; Condition: Clinic setting	No quantitative data reported	Intra-Rater (Asymptomatic, Dominant): ICC: I: 0.93 (0.84–0.97), Y: 0.95 (0.87–0.98), T: 0.97 (0.94–0.99), SEM: I: 9.31 N, Y: 7.87 N, T: 4.69 N, MDC95: I: 25.81 N, Y: 21.81 N, T: 13.00 N; (Symptomatic, Dominant): ICC: I: 0.96 (0.82–0.99), Y: 0.91 (0.80–0.96), T: 0.93 (0.83–0.97), SEM: I: 7.54 N, Y: 7.38 N, T: 6.77 N, MDC95: I: 20.90 N, Y: 20.46 N, T: 18.77 N
Tooth et al. ⁸	20; Age: 22.1 ± 2.1 years; Gender: M/F: 20/O; Sport: Individual sports; Level: Amateur	Positions: I (180°), Y (135°), T (90°); Tools: Force plate, Handheld dynamometer	ICC: I: 0.952 (0.884–0.981), Y: 0.926 (0.582–0.970), T: 0.864 (0.688–0.944); Bias: I: -0.23 Nm, Y: -0.65 Nm, T: 2.36 Nm	Reliability: I: ICC: 0.810 (AST), 0.923 (M-AST), SEM: 10.44 Nm (AST), 6.78 Nm (M-AST), MDC95: 28.93 Nm (AST), 18.78 Nm (M-AST); Y: ICC: 0.743 (AST), 0.643 (M-AST), SEM: 9.07 Nm (AST), 11.60 Nm (M-AST), MDC95: 25.15 Nm (AST), 32.17 Nm (M-AST); T: ICC: 0.715 (AST), 0.826 (M-AST), SEM: 8.59 Nm (AST), 6.42 Nm (M-AST), MDC95: 23.82 Nm (AST), 17.80 Nm (M-AST)

Notes: AST: Athletic Shoulder Test. A standardized assessment designed to evaluate shoulder strength and function, involving isometric force measurements in the I (180°), Y (135°), and T (90°) positions. SSASPT: Seated Single-Arm Shot-Put Test. A functional test performed in a seated position using a single-arm shot-put throw to assess shoulder strength and performance (Paliouras et al., 2025). ER: External Rotation. Refers to the outward rotation of the shoulder, typically measured to evaluate rotator cuff function. IR: Internal Rotation. Refers to the inward rotation of the shoulder, used to assess shoulder stability and strength. M-AST: Modified Athletic Shoulder Test (hypothesized definition). Reported in Tooth et al. (2022) as a variant of the AST; specific modifications (e.g., protocol, position, or equipment) are not explicitly defined in the study and require clarification. ICC: Intra-class Correlation Coefficient. A statistical measure of reliability, ranging from 0 to 1, with values ≥0.75 generally indicating high reliability. Confidence intervals (e.g., 0.884–0.981) are provided where reported. SEM: Standard Error of Measurement. Represents the standard deviation of measurement error; lower values indicate greater precision (units: N or Nm). MDC: Minimal Detectable Change. The smallest change in a measurement that can be considered a true difference beyond measurement error; reported as MDC90 (90% confidence level) or MDC95 (95% confidence level) (units: N or Nm). N: Newton. Unit of force. Nm: Newton-meter. Unit of torque. Dominant/Non-Dominant: Indicates whether the test was performed on the dominant (typically preferred) or non-dominant arm. Force Plate: A device used to measure ground reaction forces, commonly employed in biomechanical assessments. Handheld Dynamometer (HHD): A portable instrument for measuring muscle strength, expressed in force (N) or torque (Nm). Sphygmomanometer: A device traditionally used to measure blood pressure, adapted in Morrison et al. (2021) to assess shoulder strength in a prone position.

in measuring shoulder strength or function, (2) reported quantitative outcomes such as correlation coefficients (e.g., ICC, r , R^2), standard error of measurement (SEM), or minimal detectable change (MDC), (3) involved human participants of any age, gender, or activity level, and (4) were published in English. Studies were excluded if they lacked sufficient quantitative data, focused on versions of another shoulder assessment without clear documentation, or were not peer-reviewed (e.g., conference abstracts, theses). The initial search yielded 781 records, which were screened for eligibility based on titles and abstracts. After removing duplicates and applying the inclusion criteria, 55 studies were included in the full text review, of which 7 provided data suitable for meta-analysis of validity and reliability. The selection process adhered to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines.

Data Extraction

Data extraction was performed systematically from the included studies to capture relevant information for both the meta-analysis and systematic review components (Table 1). Two independent reviewers extracted the following data: (1) study characteristics, including author names, publication year, and sample size; (2) participant characteristics, such as age, gender, sport type, and activity level (e.g., elite, amateur, healthy); (3) test settings, including AST positions (I, Y, T), measurement tools (e.g., force plate, handheld dynamometer), and conditions (e.g., prone position, dominant/non-dominant arm); (4) validity outcomes, such as correlation coefficients (e.g., ICC, r , R^2) with confidence intervals where available; (5) reliability outcomes, including intra-class correlation coefficients (ICC), standard error of measurement (SEM), and minimal detectable change (MDC) across different positions and subgroups; and (6) additional methodological details, such as test-retest intervals or modifications to the AST protocol. In this review, validity was defined as the extent to which AST scores reflect true shoulder function, and it was categorized into two primary types: (a) concurrent validity—where AST results were compared to functional performance tests such as the shot put test; and (b) criterion validity—where AST was evaluated against gold-standard force measurement systems, such as force plates. Reliability was defined as the consistency of AST scores across repeated trials or raters, most commonly assessed via ICC. These definitions were used to classify and interpret the extracted psychometric outcomes. For the meta-analysis, validity and reliability correlation coefficients were prioritized, resulting in 42 validity estimates and 107 reliability estimates (35 correlation coefficients, 36 SEM, and 36 MDC). Discrepancies between reviewers were resolved through discussion or consultation with a third reviewer. When multiple subgroups (e.g., dominant vs. non-dominant arms) were reported within a study, data were extracted separately for each subgroup to ensure comprehensive representation. Missing or unclear

data were addressed by contacting study authors where feasible; otherwise, such data were excluded from the meta-analysis but retained for the systematic review if relevant.

Meta-Analysis

The meta-analysis was conducted to synthesize the validity and reliability estimates of the Athletic Shoulder Test across included studies. Data were extracted from a structured dataset containing study-specific information, including sample sizes (N), correlation coefficients (r) for validity and reliability, and subgroup details. Separate analyses were performed for validity and reliability using Python (version 3.x) within the PyCharm integrated development environment, with the following packages: Pandas (version 2.2.3) for data management, NumPy (version 1.24.3) for numerical computations, and Statsmodels (version 0.14.4) for statistical modeling.

Correlation coefficients (r) were transformed into Fisher's Z scores to normalize their distribution, following established methods¹⁷. For studies with multiple subgroups, a single effect size per study was calculated by computing a weighted average of Z scores, with weights based on sample size adjusted for degrees of freedom. Variance for each Z score was derived accordingly. A random-effects model, employing the DerSimonian-Laird (DL) method, was used to estimate the pooled effect size (Z), which was then back-transformed to a pooled correlation coefficient (r) using the inverse Fisher's Z transformation^{18,19}. The 95% confidence intervals (CI) were calculated based on the standard error of the pooled Z estimate.

For validity, Pearson's correlation coefficients (r) were primarily used, while for reliability, intra-class correlation coefficients (ICC) were extracted or calculated. Where necessary, ICC values were converted or interpreted as equivalent to r for transformation and pooling purposes. To aid interpretation, correlation strength was categorized according to Portney and Watkins (2009) as follows: r or ICC < 0.50 = poor, 0.50 – 0.74 = moderate, 0.75 – 0.89 = good, and ≥ 0.90 = excellent²⁰. These classifications were used descriptively to enhance clarity and should not be considered absolute clinical thresholds.

While SEM and MDC values were also extracted and descriptively summarized, they were not included in the meta-analytic synthesis due to inconsistencies in units (e.g., Newton vs. Newton-meter), varying confidence intervals (90% vs. 95%), and non-standardized reporting formats across studies. In contrast, validity and reliability metrics—reported as correlation coefficients or ICC—were statistically compatible and could be transformed using Fisher's Z method, allowing for appropriate quantitative pooling under the assumptions of a random-effects model.

Heterogeneity across studies was quantified using τ^2 (between-study variance) and I^2 (percentage of total variation due to heterogeneity), both derived from the random-effects model. Publication bias was assessed by generating funnel plots, plotting effect sizes (r) against

their standard errors in the Fisher Z scale, and conducting Egger's regression test, with significance set at $p < 0.05^{21}$. Forest plots were used to visualize individual and pooled effect sizes with 95% CIs, while funnel plots assessed symmetry indicative of publication bias. Visualizations were created using Matplotlib (version 3.x).

Systematic Review of SEM and MDC

The standard error of measurement (SEM) and minimal detectable change (MDC) values reported in the included studies were systematically reviewed to provide insights into the measurement precision and sensitivity of the AST. These metrics were extracted for each study across the I (180°), Y (135°), and T (90°) positions, capturing data for both dominant and non-dominant arms where specified. SEM and MDC values were recorded in their reported units (Newtons [N] or Newton-meters [Nm]) and confidence levels (MDC90 or MDC95) to ensure accuracy. Studies that provided SEM and MDC for modified versions of the AST (e.g., M-AST) or specific conditions (e.g., asymptomatic vs. symptomatic shoulders) were also included, with distinctions noted accordingly. No meta-analytic pooling was performed for these metrics due to heterogeneity in units, confidence levels, and reporting formats across studies. Instead, the data were summarized descriptively to highlight the range of measurement error and detectable change, facilitating a qualitative comparison of the AST's reliability across different populations and test settings.

Quality Assessment

To ensure a rigorous evaluation of the methodological quality of the studies included in this systematic review and meta-analysis, a structured quality assessment was conducted using the Downs and Black Checklist. This tool, widely recognized for its applicability to both randomized and non-randomized studies in health sciences, was selected due to its comprehensive coverage of key methodological domains relevant to psychometric research, including reporting, external validity, internal validity (bias and confounding), and power²². The checklist comprises 27 items, scored on a scale from 0 to 32, with higher scores indicating greater methodological rigor.

Each included study⁷⁻¹³ was independently evaluated by two reviewers with expertise in sports science and exercise physiology. The assessment process involved a detailed review of the full text of each article, focusing on the clarity and completeness of reported information, the representativeness of participants and test settings, the robustness of measurement protocols, the control of confounding factors, and the adequacy of statistical power. The checklist was applied as follows:

Reporting (0–11 points): An additional point was available for reporting adverse events (Item 11), though this was rarely applicable in reliability and validity studies.

External Validity (0–3 points): This section evaluated

the generalizability of findings, focusing on the representativeness of participants to the target population (Item 11), the test environment to real-world settings (Item 12), and the personnel conducting the assessments (Item 13).

Internal Validity—Bias (0–7 points): This domain examined potential biases in study execution, including blinding (Item 14, marked as not applicable [N/A] for reliability studies), absence of data dredging (Item 15), consistency of test timing (Item 16), protocol adherence (Item 17), appropriateness of statistical tests (Item 18), reliability of measurement tools (Item 19), and suitability of the participant pool (Item 20).

Internal Validity—Confounding (0–6 points): This section assessed control of confounding factors, with Items 21–24 (randomization and group comparability) marked as N/A for non-comparative reliability studies. Control of confounders (Item 25) and consistency of participation rates (Item 26) were scored based on reported methodological controls.

Power (0–5 points): The power domain was adapted to a 0–5 scale, as per previous studies in exercise physiology, to reflect sample size adequacy and power analysis reporting (Item 27). Scores were assigned as follows: 0 (no power analysis, $N < 20$), 1 (no analysis, $N = 20–30$, limited adequacy), 2 (a priori analysis, $N = 20–40$), 3 (a priori or post-hoc analysis, $N > 40$ or justified by prior studies), 4 (robust analysis, $N > 50$), and 5 (comprehensive analysis with effect size and power $> 80\%$).

For each item, a score of 1 was awarded if the criterion was fully met, 0 if unmet or unclear, and N/A where inapplicable (e.g., blinding in reliability studies). Specific scoring adjustments were made to accommodate the psychometric focus: for instance, Item 14 (blinding) was consistently marked N/A, as blinding is not feasible in test-retest reliability designs. Discrepancies between reviewers were resolved through discussion, and consensus scores were recorded. Total scores for each study were calculated by summing the applicable items (maximum 32, adjusted for N/A items), providing a quantitative measure of methodological quality. Quality assessment results were not used to exclude studies from the meta-analysis, given that all included articles were sourced from peer-reviewed journals indexed in PubMed, Scopus, and Web of Science, ensuring a baseline level of scientific integrity. Instead, these scores served to contextualize the reliability and validity findings, identify potential sources of heterogeneity, and inform the interpretation of pooled estimates.

Results

Validity Analysis

The meta-analysis of the Athletic Shoulder Test (AST) validity included five studies^{8,9,11-13}, encompassing 42 validity estimates derived from sample sizes ranging from 12 to 33 participants (Figure 2). A random-effects model

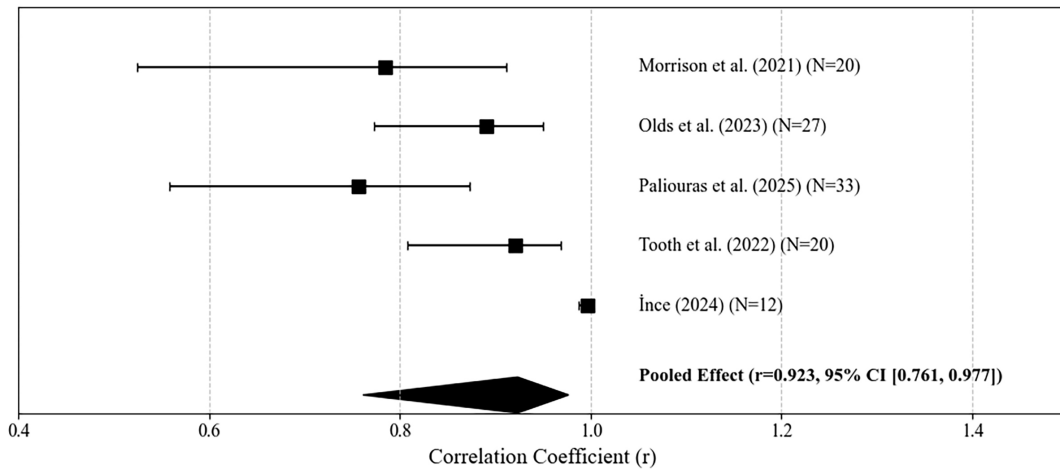


Figure 2. Forest Plot of Validity Correlation Coefficients for the Athletic Shoulder Test.

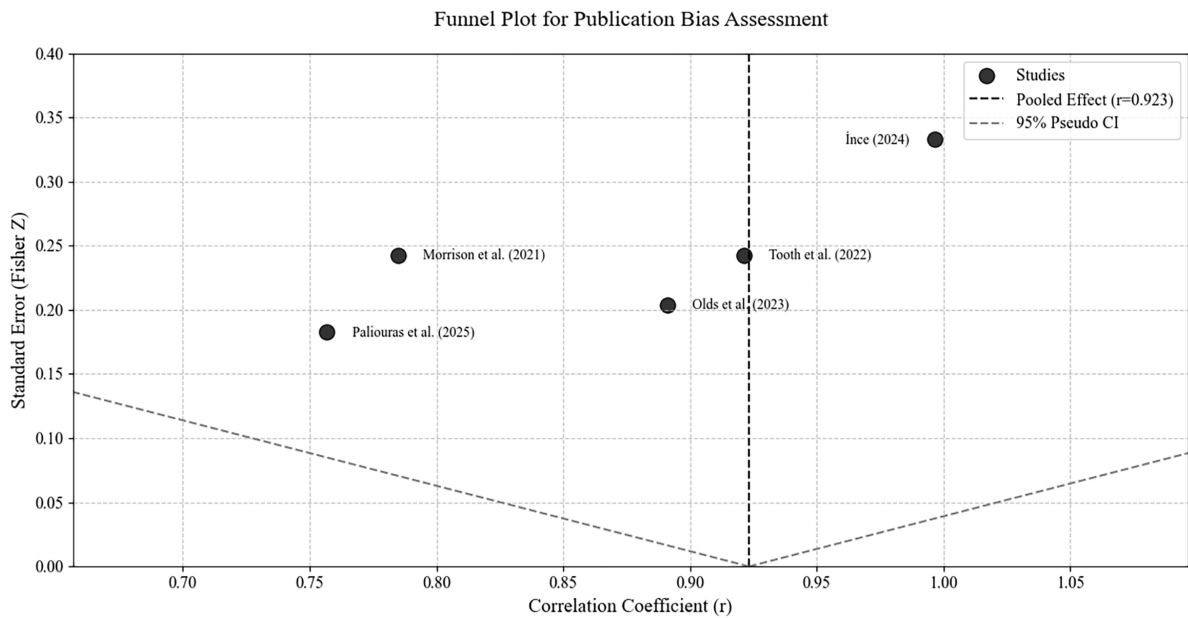


Figure 3. Funnel Plot for Assessing Publication Bias in Validity Studies of the Athletic Shoulder Test.

was employed, yielding a pooled correlation coefficient of $r = 0.923$ (95% CI: 0.761–0.977), with a z-value of 5.158 and $p < 0.001$. This result indicates a statistically significant and strong validity of the AST in assessing shoulder strength and function. Individual study effect sizes varied widely, ranging from $r = 0.757^9$ to $r = 0.997^{12}$, reflecting differences in validation approaches and measurement contexts.

Validity estimates were based on comparisons of the AST with various reference standards, including force plates as the presumed gold standard, handheld dynamometers, sphygmomanometers, and alternative functional tests such as the Seated Single-Arm Shot-Put Test (SSASPT). For instance, İnce (2024) reported exceptionally high validity ($R^2: 0.986\text{--}0.995$) using bilateral force plate

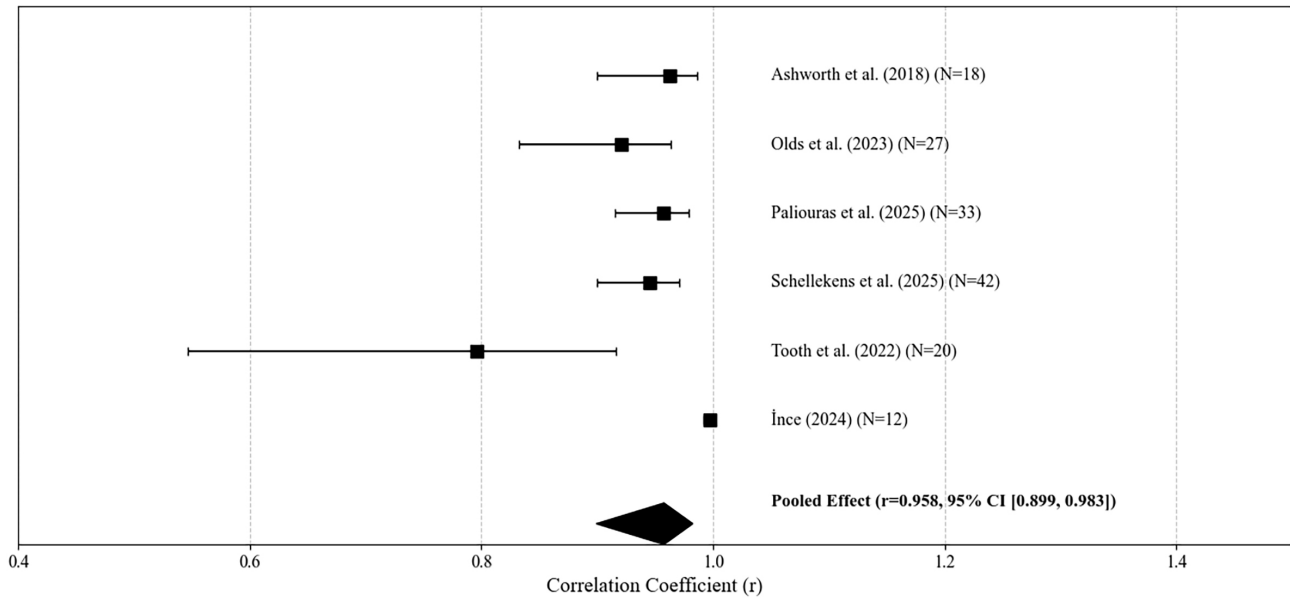


Figure 4. Forest Plot of Reliability Correlation Coefficients for the Athletic Shoulder Test.

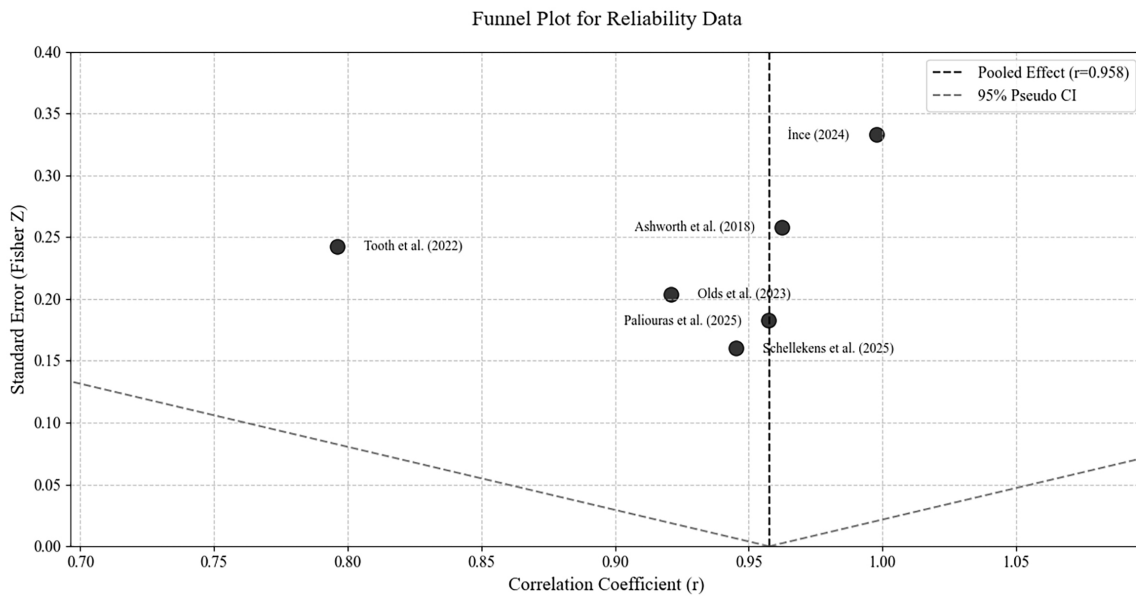


Figure 5. Funnel Plot for Assessing Publication Bias in Reliability Studies of the Athletic Shoulder Test.

measurements in amateur athletes, while Morrison et al. (2021) found moderate correlations ($r: 0.764-0.817$) when comparing sphygmomanometer outputs to force plate data in a prone position. Similarly, Olds et al. (2023) and Tooth et al. (2022) validated the AST against force plates and

handheld dynamometers, reporting ICC values ranging from 0.84 to 0.952, whereas Paliouras et al. (2025) included comparisons with SSASPT and rotator cuff strength (external/internal rotation), yielding ICCs of 0.584–0.867. These diverse reference standards underscore the AST’s

Table 2. Methodological Quality Assessment Using the Downs and Black Checklist.

Study	①	②	③	④	⑤	⑥	Summary of Scoring Justification
Ashworth et al. ⁷	9	2	6	3	0	20	Clear aims, methods, and findings; lacks p-values and adverse events reporting. Participants are representative, but personnel details are missing. No power analysis.
İnce ¹²	10	2	6	4	3	25	Detailed protocol and findings; lacks adverse events and personnel info. Strong power analysis based on prior study; confounders well controlled.
Morrison et al. ¹¹	9	2	5	3	1	20	Clear objectives and findings; timing unclear, adverse events not reported. Participant group is representative, but assessor data missing. Limited power estimation.
Olds et al. ¹³	10	2	6	3	2	23	Protocol and outcomes are clearly reported; lacks adverse event data and assessor clarity. Power analysis conducted; confounding control is limited.
Paliouras et al. ⁹	10	2	6	4	2	24	Transparent timing and outcomes; lacks adverse events and assessor info. Moderate control of confounders; power analysis adequately reported.
Schellekens et al. ¹⁰	10	2	6	4	2	24	Clear hypotheses and analysis; lacks details on adverse events and assessor experience. Confounding variables reasonably addressed; appropriate power analysis.
Tooth et al. ⁸	10	2	6	3	1	22	Hypotheses, findings, and statistics well-detailed; timing incomplete, adverse events not reported. Assessor characteristics not described. No power analysis.

Note: ① Reporting (0–11) | ② External Validity (0–3) | ③ Internal Validity – Bias (0–7) | ④ Internal Validity – Confounding (0–6) | ⑤ Power (0–5) | ⑥ Total Score (0–32).

criterion validity across different tools and settings, though the variability in measurement precision and protocols likely influenced the pooled estimate.

Heterogeneity across studies was substantial, with $\text{Tau}^2 = 0.428$ and $I^2 = 88.9\%$, indicating considerable between-study variation in validity estimates. This heterogeneity may be attributed to differences in participant characteristics (e.g., age: 12.6–37.0 years; athletic level: elite vs. amateur), testing conditions (e.g., prone vs. clinic settings), and the sensitivity of measurement tools (e.g., force plate vs. sphygmomanometer). Publication bias was assessed using a funnel plot and Egger's regression test, which yielded an intercept of 12.232 with $p = 0.256$, suggesting no significant evidence of publication bias (Figure 3). Collectively, these findings affirm the AST's strong validity for evaluating shoulder function, albeit with context-specific considerations due to the observed heterogeneity.

Reliability Analysis

The reliability meta-analysis incorporated six studies^{7-10,12,13}, comprising 35 reliability estimates with sample sizes ranging from 12 to 42 participants (Figure 4). The pooled correlation coefficient was $r = 0.958$ (95% CI: 0.899–0.983), with a z-value of 8.36 and $p < 0.001$, demonstrating high and statistically significant reliability of the AST. Individual study effect sizes varied from $r = 0.796^8$ to $r = 0.998^{12}$. Heterogeneity was notable, with $\text{Tau}^2 = 0.262$ and $I^2 = 84.9\%$, indicating substantial between-

study variability. Publication bias was evaluated using a funnel plot, and Egger's regression test yielded an intercept of 4.757 with $p = 0.437$, suggesting no significant evidence of publication bias (Figure 5).

Standard Error of Measurement (SEM) and Minimal Detectable Change (MDC)

The systematic review of the AST included SEM and MDC values reported across seven studies that provided reliability data⁷⁻¹³. These metrics were evaluated in the I (180°), Y (135°), and T (90°) positions, with measurements reflecting both dominant and non-dominant arms where specified, and expressed in Newtons (N) or Newton-meters (Nm) depending on the study (Table 1).

SEM values, indicating measurement precision, varied widely across studies. The lowest SEM was reported by İnce (2024) at 0.85 N (T position), 1.20 N (Y), and 1.82 N (I), suggesting high precision in a small sample of amateur athletes ($N = 12$). In contrast, Morrison et al. (2021) reported the highest SEM values at 20.43 N (I), 12.26 N (Y), and 13.29 N (T) in an amateur team sport cohort ($N = 20$), indicating greater measurement variability. Other studies showed intermediate SEM ranges: Ashworth et al. (2018) reported 4.8–10.8 N across positions and arms ($N = 18$, elite athletes); Paliouras et al. (2025) ranged from 3.33–6.47 N ($N = 33$, amateur athletes); Schellekens et al. (2025) from 4.69–9.31 N ($N = 42$, mixed levels); Tooth et al. (2022) from 6.42–11.60 Nm for AST and Modified AST

(M-AST) variants (N = 20, amateur athletes); and Olds et al. (2023) from 6.08–13.73 N across dominant and non-dominant arms using handheld dynamometer (HHD) (N = 27, healthy individuals).

MDC values, representing the smallest detectable change beyond measurement error, also exhibited variability. İnce (2024) reported the lowest MDC values at 2.37 N (T), 3.34 N (Y), and 5.07 N (I) at an unspecified confidence level, reflecting high sensitivity to change. Conversely, Morrison et al. (2021) reported the highest MDC90 values at 56.63 N (I), 33.99 N (Y), and 36.83 N (T), suggesting a larger threshold for detecting true differences. Other studies included: Ashworth et al. (2018) with MDC90 ranging from 13.3–25.9 N; Paliouras et al. (2025) with MDC (unspecified level) from 9.32–18.04 N; Schellekens et al. (2025) with MDC95 from 13.00–25.81 N; Tooth et al. (2022) with MDC95 from 17.80–32.17 Nm; and Olds et al. (2023) with MDC95 from 16.78–37.96 N. Differences in units (N vs. Nm) and confidence levels (MDC90 vs. MDC95) across studies precluded direct comparisons without standardization.

Methodological Quality Assessment

A methodological quality assessment was conducted for the seven included studies using the Downs and Black checklist, which evaluates five key domains: reporting, external validity, internal validity (bias and confounding), and statistical power (Table 2). Total scores ranged from 20 to 25 out of a possible 32 points, indicating moderate to high methodological quality across studies.

Reporting quality was generally strong. Most studies clearly described their aims, protocols, and statistical findings. However, only five studies^{8-10,12,13} achieved a score of 10/11, with the primary limitation being the absence of adverse events reporting. Ashworth et al. and Morrison et al. received slightly lower reporting scores (9/11) due to missing p-values and unclear timepoint descriptions.

External validity scores were consistent across all studies (2/3). While study populations (e.g., elite and amateur athletes) and testing environments were deemed representative, none of the studies provided detailed information about the evaluators or test administrators, which limited full external generalizability.

Internal validity – bias scores ranged from 5 to 6 out of 7. Most studies demonstrated appropriate use of reliable measurement instruments and standardized procedures. However, Morrison et al. (2021) scored lower due to timing inconsistencies and potential risks associated with manual measurement tools.

Internal validity – confounding showed more variation (3–4/6). While basic demographic variables such as sex and age were typically controlled, potential confounding factors—such as fatigue, biological maturation, or learning effects—were only partially addressed in most studies. Paliouras et al. (2025), Schellekens et al. (2025), and İnce (2024) provided relatively stronger control by employing standardized test protocols and stratified analyses.

Statistical power was the most heterogeneous domain, with scores ranging from 0 to 3 out of 5. Three studies (Ashworth, Morrison, and Tooth) did not report a priori power analyses or justify sample sizes adequately. In contrast, Olds et al. (2023), Paliouras et al. (2025), and Schellekens et al. (2025) provided sufficient power estimates based on sample calculations. İnce (2024) achieved the highest score in this domain by referencing a previous large-scale study to justify statistical power.

In terms of total scores, Ashworth and Morrison received the lowest ratings (20/32), primarily due to the absence of power analysis and partial reporting. Tooth scored slightly higher (22/32) due to improved reporting, despite similar limitations in power estimation. Olds et al. (2023) (23/32) benefitted from a priori sample size justification. Paliouras et al. (2025) and Schellekens et al. (2025) each scored 24/32 by combining robust reporting, moderate confounder control, and acceptable power estimates. The highest score (25/32) was observed in İnce (2024), which demonstrated excellent statistical rigor and comprehensive methodological reporting; however, generalizability remained limited due to the small sample size (N = 12).

Discussion

This systematic review and meta-analysis provide robust evidence supporting the psychometric properties of the AST as a reliable and valid tool for assessing shoulder function. Pooled correlation coefficients indicated high validity ($r = 0.923$) and even stronger reliability ($r = 0.958$), both statistically significant ($p < 0.001$). These outcomes were consistent across elite athletes, amateurs, and healthy individuals. These findings indicate that the AST effectively measures shoulder strength and stability, with consistent performance across repeated assessments. The absence of significant publication bias, as confirmed by Egger's regression tests ($p = 0.256$ for validity, $p = 0.437$ for reliability), further strengthens the confidence in these results. Collectively, this meta-analysis underscores the AST's potential as a standardized method for evaluating shoulder function in both clinical and research settings, offering a solid foundation for its broader application.

The findings of this meta-analysis align closely with individual studies in the literature, reinforcing the AST's strong psychometric performance while providing a more precise estimate through data synthesis. For instance, Ashworth et al. (2018) reported high reliability with ICC values ranging from 0.94 to 0.98 across I, Y, and T positions in elite athletes, consistent with our pooled reliability estimate of $r = 0.958$. Similarly, İnce (2024) demonstrated exceptional validity ($R^2 > 0.99$) in a small cohort of amateur athletes, which supports our pooled validity of $r = 0.923$. However, some studies, such as Tooth et al. (2022), reported more variable validity (ICC: 0.643–0.952) and reliability (ICC: 0.643–0.923) outcomes, particularly when comparing the standard AST with its modified version (M-AST). Such

discrepancies underscore the value of meta-analysis in reducing the influence of study-specific factors (e.g., sample size, protocol) and producing more generalizable results. By synthesizing data across diverse populations and settings, this study offers a clearer and more robust understanding of the AST's psychometric properties compared to individual reports, thereby enhancing its evidence base for practical use. In particular, the AST demonstrates high applicability in populations with elevated shoulder demands, such as overhead athletes (e.g., baseball, tennis, swimming), where repetitive shoulder elevation places significant stress on stabilizing musculature. Its capacity to detect asymmetries and monitor strength progression makes it an effective pre-season screening tool for injury risk stratification. Furthermore, in rehabilitation contexts, the AST offers a simple and reproducible means of tracking improvements in shoulder function over time, supporting clinical decision-making during return-to-play phases.

Despite the strong psychometric performance of the AST, the meta-analysis revealed substantial heterogeneity, with I^2 values of 88.9% for validity and 84.9% for reliability, indicating significant variability across studies. This heterogeneity likely stems from several factors, including differences in participant characteristics, such as age (ranging from 12.6 to 37.0 years), activity level (elite vs. amateur), and sport type (team vs. individual sports), which may influence shoulder strength and test performance^{1,4}. Additionally, variations in test protocols, such as the use of different measurement tools (e.g., force plate, handheld dynamometer, sphygmomanometer) and testing conditions (e.g., prone position, clinic vs. field settings), likely contributed to the observed inconsistencies^{23,24}. For example, studies like Morrison et al. (2021) adapted a sphygmomanometer for measurement, potentially introducing variability compared to force plate-based assessments in Ashworth et al. (2018). Although a random-effects model was employed to statistically account for between-study variability, it does not eliminate the underlying methodological and clinical heterogeneity. Therefore, the pooled validity and reliability coefficients should be interpreted with caution—as indicative estimates rather than definitive conclusions. These findings highlight the importance of considering contextual factors when using the AST, paving the way for a more nuanced understanding of its application. To provide a clearer synthesis of these heterogeneous influences, we conceptually classified the major sources of variation in AST implementation into three domains: (1) measurement device (e.g., force plate vs. handheld dynamometer), (2) test setting and posture (e.g., prone vs. upright), and (3) participant characteristics (e.g., age, sport, training level). Each of these domains can uniquely affect force output, joint stability, and neuromuscular control during the test, thereby contributing to the variability observed in psychometric outcomes. This multifactorial heterogeneity is visualized in Figure 6, which may serve as a reference framework for future studies aiming to standardize AST protocols across different contexts.

The systematic review of SEM and MDC values further highlights the variability in the AST's measurement precision across studies, with implications for its clinical utility. SEM values ranged from as low as 0.85 N in Ince (2024) to as high as 20.43 N in Morrison et al. (2021), reflecting a wide spectrum of measurement error that appears to be influenced by test settings and equipment. Similarly, MDC values varied significantly, from 2.37 N (Ince, 2024) to 56.63 N (Morrison et al., 2021), indicating that the threshold for detecting meaningful changes in shoulder strength differs substantially across studies. This variability likely stems from differences in measurement tools (e.g., force plates vs. handheld dynamometers) and participant factors like age or training status, which can affect muscle function and test outcomes. In clinical practice, these findings suggest that practitioners should interpret AST results in the context of the specific SEM and MDC values relevant to their testing conditions, ensuring that observed changes exceed the measurement error to confirm true improvements or declines in shoulder function. This underscores the need for careful consideration of test standardization to enhance the AST's reliability in detecting clinically significant changes.

The high validity and reliability of the AST, as demonstrated in this meta-analysis, position it as a valuable tool for both clinical and research applications. In clinical settings, the AST can be effectively utilized to assess shoulder strength and stability, particularly in populations at risk of shoulder dysfunction, such as overhead athletes or individuals undergoing rehabilitation following shoulder injuries^{25,26}. Its ability to provide consistent measurements across repeated assessments ($r = 0.958$) makes it a reliable method for monitoring progress during rehabilitation, enabling clinicians to track improvements in shoulder function over time and adjust treatment plans accordingly. In research contexts, the AST's strong psychometric properties support its use as a standardized outcome measure in studies investigating shoulder biomechanics, injury prevention strategies, or the effectiveness of therapeutic interventions^{7-9,12}. The AST's objective and reproducible nature facilitates cross-study comparisons and strengthens the generalizability of findings^{7,10,12}. These practical applications highlight the test's potential to bridge the gap between clinical practice and scientific inquiry, contributing to improved shoulder health outcomes.

When compared to other shoulder assessment tools, the AST presents unique advantages in terms of standardization, muscle-specific evaluation, and ecological validity^{7,10,23}. For instance, the Seated Single-Arm Shot-Put Test emphasizes explosive power and functional range but lacks positional standardization and is sensitive to trunk compensation^{27,28}. Handheld dynamometers offer portability but are more susceptible to tester-induced variability and lack control over joint alignment, particularly in high-force outputs^{29,30}. In contrast, isokinetic dynamometry provides detailed torque-velocity profiles under dynamic conditions, yet remains cost-prohibitive and

less feasible in field-based or high-throughput settings^{31,32}. The AST offers a middle ground—quantifying isometric strength across biomechanically controlled positions (I, Y, T) while minimizing equipment demands and allowing reproducible testing with reduced operator dependency^{7,9}. This balance makes the AST a compelling option for both clinical and sport-specific screening environments.

Beyond its psychometric strengths, the AST offers valuable insights into the physiological and biomechanical underpinnings of shoulder function through its focus on isometric muscle activation^{23,33}. Specifically, the test engages high-threshold motor units and challenges the neuromuscular system under static load, providing indirect indicators of muscle fatigue resistance, scapular stabilization, and force-generating capacity. During the test, the sustained isometric contractions in I, Y, and T positions induce localized muscle fatigue, particularly in the rotator cuff and deltoid muscles, which can be observed through increased electromyographic activity and metabolic stress^{4,34,35}. This fatigue response, coupled with the test's ability to isolate specific movement planes, provides a window into the neuromuscular control mechanisms governing shoulder stability, such as the co-activation of scapular stabilizers (e.g., trapezius, serratus anterior) to maintain proper joint alignment under load^{35,36}. Furthermore, the biomechanical demands of each position—such as the high abduction torque in the I position or the rotational stress in the Y and T positions—reflect the differential loading patterns on the glenohumeral joint and surrounding soft tissues, offering a functional assessment of the shoulder's capacity to withstand dynamic stresses encountered in sports or daily activities^{6,37}. Capturing these physiological and biomechanical dynamics, the AST helps assess shoulder performance while revealing underlying mechanisms—supporting more targeted interventions to optimize function and prevent injuries.

Strengths and Limitations

This study has several notable strengths that enhance its contribution to the literature on shoulder assessment. First, the systematic review and meta-analysis were conducted using a rigorous methodology, adhering to PRISMA guidelines and sourcing studies from high-quality, peer-reviewed databases such as PubMed, Scopus, and Web of Science. This ensured a baseline level of methodological rigor in the included studies. Second, the meta-analysis provided a robust synthesis of the Athletic Shoulder Test's (AST) validity and reliability, offering more precise estimates than individual studies by pooling data across diverse populations. Additionally, the systematic review of SEM and MDC values offered valuable insights into the test's measurement precision, addressing an often-overlooked aspect of psychometric evaluation. Finally, the absence of significant publication bias, as confirmed by Egger's regression tests, strengthens the confidence in the findings.

However, several limitations must be acknowledged. A

primary limitation is the limited number of studies available for the meta-analysis of validity, with only five studies providing quantitative data. Moreover, these studies exhibited heterogeneity in their approach to validity assessment: some compared the AST with force platforms versus more cost-effective devices (e.g., handheld dynamometers, sphygmomanometers), while others evaluated the AST against alternative shoulder tests, such as the Seated Single-Arm Shot-Put Test⁹. This variability in validation methods may have contributed to the high heterogeneity observed ($I^2 = 88.9\%$) and suggests that the literature on AST's validity is still maturing, necessitating further independent studies to establish a more comprehensive evidence base. Another limitation is the lack of a universally accepted, standardized test for evaluating the complex shoulder joint and its associated musculature. While the AST demonstrates strong psychometric properties and may serve as a pioneering tool in this domain, the field has yet to reach a consensus on a gold-standard assessment, which limits the ability to benchmark the AST against a definitive reference. Additionally, the relatively small sample sizes in some included studies (e.g., $N = 12$ in Ince, 2024) and the variability in participant characteristics (e.g., age ranging from 12.6 to 37.0 years) may affect the generalizability of the findings. These limitations highlight the need for cautious interpretation of the results and underscore the importance of continued research to refine the AST's role in shoulder assessment. Additionally, the current synthesis did not incorporate GRADE-based quality appraisal, which may limit the stratified evaluation of evidence strength across outcomes. Future meta-analyses should consider integrating GRADE criteria to enhance interpretability for clinical guidelines. Moreover, the absence of longitudinal studies restricts our ability to assess the AST's predictive validity, particularly in terms of injury risk stratification or return-to-play outcomes. Addressing these gaps would further solidify the AST's utility in both prospective monitoring and decision-making contexts.

Conclusions

This systematic review and meta-analysis confirm that the AST is a highly valid ($r = 0.923$, 95% CI: 0.761–0.977) and reliable ($r = 0.958$, 95% CI: 0.899–0.983) tool for assessing shoulder function across diverse populations, including elite and amateur athletes as well as healthy individuals. The test's strong psychometric properties, coupled with its ability to provide insights into the physiological and biomechanical mechanisms of shoulder performance, position it as a valuable method for both clinical practice and research. Clinicians can confidently use the AST to evaluate shoulder strength and monitor rehabilitation progress, while researchers can adopt it as a standardized measure to investigate shoulder-related outcomes. However, the observed heterogeneity in validity and reliability estimates, as well as the variability in SEM and

MDC values, underscores the need for careful interpretation of results in the context of specific test conditions and populations.

To further advance the utility of the AST, future studies should focus on addressing the identified limitations. Larger and more diverse cohorts are needed to enhance the generalizability of findings, particularly across different age groups, activity levels, and clinical populations, such as those with shoulder pathologies. Additionally, standardizing test protocols, including the use of consistent measurement tools (e.g., force plates) and testing conditions (e.g., prone position), could reduce heterogeneity and improve the comparability of results across studies. Independent validations of the AST against a broader range of shoulder assessment tools would also strengthen its evidence base, potentially establishing it as a leading method in the absence of a universally accepted gold standard for shoulder evaluation. By pursuing these research directions, the AST can be further refined, solidifying its role as a cornerstone in the assessment and management of shoulder function.

Authors' Contributions

Süleyman Ulupinar: Conceptualization, methodology, formal analysis, data curation, original draft preparation, writing—review and editing, supervision, and project administration. The primary author and guarantor of the study. İzzet İnce: Literature search, data interpretation, critical revision of the manuscript for intellectual content, and validation. Contributed significantly to the theoretical framework. Cebrail Gençoğlu: Data extraction, statistical support, visualization, and quality assessment of included studies. Selim Asan: Screening of studies, risk-of-bias evaluation, and manuscript proofreading. Salih Çabuk: Reference standardization, technical editing, and compliance checks. Serhat Özbay: Supplementary material preparation and cross-checking data accuracy.

Acknowledgments

The authors acknowledge the use of artificial intelligence tools for language editing during the preparation of this manuscript.

References

- Vocelle AR, Weidig G, Bush TR. Shoulder structure and function: The impact of osteoarthritis and rehabilitation strategies. *J Hand Ther.* 2022;35(3):377–87.
- Baffour FI, Ahmed AK, Gopireddy DR, Hsieh J, Halaweish AF, et al. Ultra-high-resolution imaging of the shoulder and pelvis using photon-counting-detector CT: a feasibility study in patients. *Eur Radiol.* 2022;32(10):7079–86.
- Moiroux-Sahraoui A, Le Garrec S, Martin L, Ucay R, Dauty M, et al. Prevention of overhead shoulder injuries in throwing athletes: a systematic review. *Diagnostics (Basel).* 2024;14(21):2415.
- Escalante G, Longhurst C, Gorman S, Carroll T. Progressive exercise strategies to mitigate shoulder injuries among weight-training participants. *Strength Cond J.* 2021;43(1):72–85.
- Escalante G. Exercise modification strategies to prevent and train around shoulder pain. *Strength Cond J.* 2017;39(3):74–86.
- McKenzie AK, Swanson BT, Chalmers PN. Glenohumeral extension and the dip: considerations for the strength and conditioning professional. *Strength Cond J.* 2021;43(1):93–100.
- Ashworth B, Hogben P, Singh N, Tulloch L, Cohen DD. The Athletic Shoulder (ASH) test: reliability of a novel upperbody isometric strength test in elite rugby players. *BMJ Open Sport Exerc Med.* 2018;4(1):e000365.
- Tooth C, Thomas C, Le Roux A, Mandrusiak A, King E. The modified-athletic shoulder test: reliability and validity of a new on-field assessment tool. *Phys Ther Sport.* 2022;58:8–15.
- Paliouras A, Evaggelidis G, Tsapralis K, Tsopani D, Dallas G, et al. Psychometric properties of the Athletic Shoulder Test in adolescent tennis players. *J Clin Med.* 2025;14(4):1146.
- Schellekens M, Gouttebauge V, van der Kruk E, Stubbe JH. Reliability of the Athletic Shoulder Test in asymptomatic and symptomatic overhead racquet athletes. *Phys Ther Sport.* 2025;72:86–94.
- Morrison G, Ashworth B, Taylor-Kaveney T. The validity of the sphygmomanometer for shoulder strength assessment in amateur rugby union players. *Phys Ther Sport.* 2021;47:59–65.
- İnce İ. Validation of software developed for force measurements with Nintendo Wii Balance Board: reliability and validity study. *Online J Recreat Sports.* 2024;13(2):184–90.
- Olds M, McLaine S, Magni N. Validity and reliability of the Kinvent handheld dynamometer in the Athletic Shoulder Test. *J Sport Rehabil.* 2023;32(7):764–72.
- Williams MD, Edwards TB, Walch G. Understanding the importance of the teres minor for shoulder function: functional anatomy and pathology. *J Am Acad Orthop Surg.* 2018;26(5):150–61.
- Tsuruie M, Ellenbecker TS. A comparison of teres minor and infraspinatus muscle activation in the prone position. *JSES Int.* 2022;6(1):116–22.
- Montgomery S, Suri M. Physical examination of the shoulder: the basics and specific tests. In: *Musculoskeletal Examination of the Shoulder.* Boca Raton: CRC Press; 2024. p. 2–22.
- van Aert RC. Meta-analyzing partial correlation coefficients using Fisher's z transformation. *Res Synth Methods.* 2023;14(5):768–73.
- Kontopantelis E, Reeves D. Performance of statistical methods for meta-analysis when true study effects are non-normally distributed: a comparison between DerSimonian–Laird and restricted maximum likelihood. *Stat Methods Med Res.* 2012;21(6):657–59.
- Jackson D, White IR, Thompson SG. Extending DerSimonian and Laird's methodology to perform multivariate random effects meta-analyses. *Stat Med.* 2010;29(12):1282–97.
- Portney LG, Watkins MP. Foundations of clinical

- research: applications to practice. 3rd ed. Upper Saddle River (NJ): Pearson/Prentice Hall; 2009.
21. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ*. 1997;315(7109):629–34.
 22. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *J Epidemiol Community Health*. 1998;52(6):377–84.
 23. Królikowska A, Klich S, Maszczyk A, Zając A. Reliability and validity of the Athletic Shoulder (ASH) test performed using portable isometric-based strength training device. *Biology (Basel)*. 2022;11(4):577.
 24. Lis R, Klich S, Stastny P, Czaplicki A. The relationship between various jump tests and baseball pitching performance: a brief review. *Strength Cond J*. 2024;46(5):520–33.
 25. Trunt A, Fisher BT, MacFadden LN. Athletic Shoulder Test differences exist bilaterally in healthy pitchers. *Int J Sports Phys Ther*. 2022;17(4):715.
 26. Cools AM, Borms D, Cagnie B, Johansson FR, Maenhout A. The challenge of the sporting shoulder: from injury prevention through sport-specific rehabilitation toward return to play. *Ann Phys Rehabil Med*. 2021;64(4):101384.
 27. Pinheiro JS, Silva FM, De Medeiros MM, De Oliveira ES, Ferreira PH. Seated single-arm shot-put test to measure the functional performance of the upper limbs in exercise practitioners with chronic shoulder pain: a reliability study. *J Chiropr Med*. 2020;19(3):153–8.
 28. Riemann BL, Kline K, Bolin M, Featherston S. A bilateral comparison of the underlying mechanics contributing to the Seated Single-Arm Shot-Put functional performance test. *J Athl Train*. 2018;53(10):976–82.
 29. Stark T, Walker B, Phillips JK, Fejer R, Beck R. Hand-held dynamometry correlation with the gold standard isokinetic dynamometry: a systematic review. *PM R*. 2011;3(5):472–79.
 30. Chamorro C, Armijo-Olivo S, De la Fuente C, Fuentes J, Javier G. Absolute reliability and concurrent validity of hand held dynamometry and isokinetic dynamometry in the hip, knee and ankle joint: systematic review and meta-analysis. *Open Med (Wars)*. 2017;12(1):359–75.
 31. Schindler IFS, Matykiewicz P, Halz M, Grabara M. A systematic review of isokinetic muscle strength in a healthy population with special reference to age and gender. *Sports Health*. 2023;15(3):328–32.
 32. İnce İ, Ulupinar S, Özbay S. Body composition, isokinetic knee extensor strength and balance as predictors of competition performance in junior weightlifters. *Isokinet Exerc Sci*. 2020;28(2):215–22.
 33. Secchi LLB, De Lima C, Simões R, Zampar AC, Bezerra da Silva E, et al. Is the isometric strength of the shoulder associated with functional performance tests in overhead athletes? *Phys Ther Sport*. 2022;55:131–8.
 34. Glousman R. Electromyographic analysis and its role in the athletic shoulder. *Clin Orthop Relat Res*. 1993;(288):27–34.
 35. Minning S, Laubach LL, Porter VL. EMG analysis of shoulder muscle fatigue during resisted isometric shoulder elevation. *J Electromyogr Kinesiol*. 2007;17(2):153–9.
 36. Klich S, Krymski T, Kawczyński A, Peltonen J. Electromyographic evaluation of the shoulder muscle after a fatiguing isokinetic protocol in recreational overhead athletes. *Int J Environ Res Public Health*. 2021;18(5):2516.
 37. Tennent TD, Beach WR, Meyers JF. A review of the special tests associated with shoulder examination: part I: the rotator cuff tests. *Am J Sports Med*. 2003;31(1):154–60.