Journal of Coaching and Sports Science
Volume 5, Issue 3, 1-19.
e_ISSN: 2963-1483
https://www.journal.foundae.com/index.php/jcss/index
DOI: 10.58524/jcss.v5i3.1157

Multi-domain field-test battery for performance profiling and inter-limb asymmetry screening in competitive Sanda Wushu

Novita*
Universitas Negeri Medan, INDONESIA

Ying Lei
Chengdu Sports University, CHINA

Wan Ahmad Jaafar Wan Yahaya
Universiti Sains Malaysia, MALAYSIA

Tansa Trisna Astono Putri
Universitas Negeri Medan, INDONESIA

Article Info

Article history:
Received: March 28, 2026
Revised: May 4, 2026
Accepted: July 8, 2026

Keywords:
Athlete monitoring;
Combat sports;
Field testing;
Inter-limb asymmetry;
Sanda Wushu.

Abstract

Background: Despite the complex physiological, technical, and mechanical demands of Sanda Wushu, athlete monitoring still relies largely on subjective coach observation and self-report. A feasible, objective, and coach-interpretable field-testing framework is therefore needed to detect performance changes and inter-limb asymmetry before integrating more complex wearable-sensor technologies.
Aims: To examine the feasibility, responsiveness, and descriptive asymmetry-monitoring value of a standardized multi-domain field-test battery administered across a full training squad of Sanda Wushu athletes.
Methods: An observational, field-based, repeated-measures monitoring study involved 72 nationally competing athletes (36 males and 36 females; age 18–30 years, M = 24.2, SD = 3.7). The battery was administered in July (Session 1) and October 2025 (Session 2). Valid results were benchmark-normalized to a composite Performance Index (PerfIndex, 0–1 scale; higher = better) and a categorical Score Index (ScoreIndex, 1–5 ordinal scale; higher = better). Inter-limb asymmetry (%) was computed for SL Wall Sit and Triple Hop Jump. Paired t-tests were reported with 95% CIs, Cohen's dz, Hedges' g, and a sensitivity analysis excluding four athletes with block-missing Session 1 items.
Results: PerfIndex increased significantly from 0.727 ± 0.031 to 0.755 ± 0.033, mean difference +0.029, 95% CI [0.024, 0.034], $t(71) = 11.46$, $p < .001$, dz = 1.35, Hedges' g = 1.34. ScoreIndex showed the same pattern (2.92 ± 0.25 to 3.16 ± 0.26; $t(71) = 9.41$, $p < .001$, dz = 1.11, Hedges' g = 1.10). SL Wall Sit asymmetry increased slightly but significantly (27.9% to 30.9%, $p = .023$, dz = 0.28); Triple Hop asymmetry did not change significantly (11.5% to 11.7%, $p = .718$). Asymmetry values above the descriptive 15% reference zone were observed in a substantial share of the squad at both sessions.
Conclusion: The field-test monitoring framework was feasible for squad-level implementation and detected small-to-moderate improvements in composite performance across two sessions. Inter-limb asymmetry screening also identified athletes requiring further assessment, although the findings should be interpreted as monitoring signals rather than diagnostic indicators.

To cite this article: Novita., Lei. Y., Yahaya. W.A.J., Putri. T.T.A., (2026). Multi-domain field-test battery for performance profiling and inter-limb asymmetry screening in competitive Sanda wushu. *Journal of Coaching and Sports Science, 5*(3), 1-19. https://doi.org/10.58524/jcss.v5i3.1157

INTRODUCTION

Sanda Wushu is a high-intensity striking-and-throwing combat sport that imposes sustained demands on anaerobic power, neuromuscular speed, tactical decision-making, and technical control under fatigue. Recent martial-arts injury surveillance and Sanda biomechanics research indicate that combat-sport performance and injury exposure emerge from combined technical, physiological, and

Corresponding author:
Novita, Universitas Negeri Medan, INDONESIA. ✉noviade@unimed.ac.id

mechanical demands rather than a single fitness quality (Bickley et al., 2023; Eltifi-Ghanmi et al., 2025; Wang et al., 2025). These compound demands create a persistent challenge for performance monitoring: isolating trainable signal from session-to-session noise requires objective, multi-domain measurement.

Despite increasing sophistication in combat sport conditioning science, daily monitoring in most Sanda programs continues to rely on coach observation and athlete self-report. While indispensable for qualitative feedback on technique, these methods have identifiable limitations: subtle neuromuscular fatigue, emerging inter-limb imbalances, and early deterioration in physical capacity are difficult to detect visually across consecutive sessions. Recent load-monitoring work in youth sport and female collegiate boxing indicates that internal and external load indicators may vary across athletes, sessions, and contexts, supporting objective monitoring alongside coach observation (Dudley et al., 2023; Xu et al., 2025). Standardized field-test monitoring, therefore, offers a low-cost supplement to coaching judgment rather than a replacement for it.

Wearable sensing technologies, including inertial measurement units (IMUs), surface electromyography, and heart-rate-derived measures, represent a promising frontier for monitoring in combat sports. In taekwondo, a sport that shares selected biomechanical parallels with Sanda striking, accelerometer-based pipelines have been used to classify kicking techniques, and separate validation work has examined low-cost IMU measurement of punch and kick velocity (Liu et al., 2024; Pezenka & Wirth, 2025). Taekwondo-specific field tests that integrate heart rate monitoring have also been examined to estimate aerobic power, anaerobic fitness, and agility (Taati et al., 2022). These studies support the plausibility of sensor-augmented monitoring in striking combat sports. Still, they do not validate the present Sanda battery or imply that wearable data were collected in this study (Xue et al., 2025).

A prerequisite for effective wearable-sensor integration in applied sport is an interpretable and reproducible field-monitoring pipeline: one that converts multi-test outcomes into coach-usable indices before adding more complex sensor streams. Without such a framework, raw wearable data may produce information overload rather than decision support, especially when test selection, data reduction, and reporting are not standardized (Asimakidis et al., 2024; Weakley et al., 2024; Xue et al., 2025). The present study, therefore, field-tests a monitoring framework for Sanda Wushu across a full training squad, incorporating benchmark-normalized performance indices and proxies for inter-limb asymmetry monitoring. The study is designed to be compatible with future IMU and HR/HRV integration, but is evaluated here in its field-test-only form.

Multi-domain composite scoring, in which heterogeneous tests are converted onto a common scale, is increasingly used for athlete profiling because it helps practitioners summarize flexibility, endurance, power, speed, and recovery-relevant indicators within a common interpretive frame (Asimakidis et al., 2024; Maestroni et al., 2023; Weakley et al., 2024). By converting raw test results into a continuous Performance Index (PerfIndex) and an ordinal Score Index (ScoreIndex), the present study provides coach-interpretable summary statistics intended to retain within-session and between-session comparability. This approach is conceptually consistent with composite profiling work in return-to-sport contexts, although the present Sanda-specific index should be regarded as preliminary rather than validated.

Within Sanda Wushu specifically, sport-science-driven performance profiling remains limited compared with the broader combat-sport literature. Recent martial-arts injury surveillance, Sanda-specific biomechanics work, and combat-sport core-training reviews support the practical importance of monitoring lower-limb, trunk, power, and movement-control indicators, but none directly validates the present field-test battery (Bickley et al., 2023; Eltifi-Ghanmi et al., 2025; Huang et al., 2023; Wang et al., 2025; Zhang et al., 2025). This evidence, therefore, supports the rationale for including core, lower-limb, and power-related measures in a Sanda-oriented monitoring battery, but requires cautious interpretation.

Recent studies by Sanda and others have examined kick kinetics and kinematics, punching-force prediction, taekwondo field testing, judo fitness classification, and mixed martial arts physiological testing (Ceylan et al., 2022; Eltifi-Ghanmi et al., 2025; Goncalves et al., 2024; Taati et al., 2022; Wang et al., 2025). These studies do not establish a standard for Sanda monitoring. Still, they support the broader case for multi-domain field-test batteries that assess physical capacity across flexibility, strength endurance, explosive power, speed, reaction, and aerobic-anaerobic fitness domains within a single monitoring session.

The theoretical framework is grounded in an athlete-monitoring logic that distinguishes field-test outputs, external workload, and internal physiological response. Field tests provide periodic snapshots of physical capacity and movement-related asymmetry; external load indicators describe what the athlete does; and internal load indicators describe the psychophysiological response to that work (Dudley et al., 2023; Impellizzeri et al., 2023; Xu et al., 2025). For combat sports, this distinction is essential because similar technical sessions may impose different internal stress depending on fatigue, recovery status, injury history, and athlete-specific movement efficiency. A field-test index can therefore be useful only when interpreted as one component of a broader load-response model rather than as a direct causal measure of training adaptation.

Wearable technologies can strengthen this model by providing continuous or session-level measures that field tests alone cannot capture. Recent reviews of wearable sensors in sport and combat sports emphasize that IMUs, surface electromyography, pressure sensors, and heart-rate-derived measures can support analysis of strike kinematics, limb biomechanics, fatigue-related movement features, and training-load indicators, while also warning that sensor placement, gold-standard validation, data quality, and coach-facing data reduction remain major barriers (Kovoor et al., 2024; Preatoni et al., 2022; Xue et al., 2025). In the present study, the field-test battery serves as an interpretable base layer for a future sensor-augmented system, not as a validation of such a system.

The injury-risk component is framed cautiously. Inter-limb asymmetry is commonly used as a screening signal, but threshold-based interpretation remains contested because asymmetry is task-specific, population-specific, metric-dependent, and variable over time (Bishop et al., 2022; D'Hondt et al., 2024; Keogh et al., 2023). Prospective and review evidence suggests that lower-limb asymmetry may be associated with injury risk in some sporting contexts, but the evidence does not justify deterministic prediction from a single asymmetry value (Fort-Vanmeerhaeghe et al., 2022; Guan et al., 2022; Guan et al., 2023; Jordan & Bishop, 2024). Accordingly, this study uses asymmetry as a descriptive monitoring proxy that can flag athletes for follow-up assessment, not as a validated injury-prediction tool. Finally, because prospective injury outcomes were not tracked, the asymmetry-related findings are interpreted as squad-level monitoring signals rather than confirmatory evidence of injury-risk reduction or elevation (Teresi et al., 2022).

The main research question is: Can a structured field-test monitoring framework be implemented feasibly and show statistically detectable responsiveness in composite performance indices and descriptive inter-limb asymmetry indicators across two monitoring sessions in a squad-level sample of Sanda Wushu athletes?. The study addressed two objectives. Objective 1 was to examine whether benchmark-normalized performance indices (PerfIndex and ScoreIndex) showed statistically detectable directional changes from Session 1 to Session 2 across the training squad. Objective 2 was to describe whether inter-limb asymmetry indicators could identify athlete-specific and squad-level monitoring signals that may inform future protocol design and threshold development. The study followed a convenience (squad-level) sampling framework, and participants were classified as nationally trained athletes using the participant-caliber taxonomy proposed by McKay et al. (2022).

METHOD

Study Design

This study employed an observational, field-based, repeated-measures (pre-post) monitoring design to evaluate a standardized field-test battery for monitoring performance and asymmetry in a squad of Sanda Wushu athletes. Two monitoring sessions were conducted: Session 1 (July 2025) served as the initial point, and Session 2 (October 2025) as a follow-up approximately 3 months later. Both sessions were embedded within the athletes' active training mesocycle, and no intervention, random allocation, control group, or wearable-sensor data collection was implemented. Reporting emphasized transparent presentation of design, setting, participants, variables, data sources, bias considerations, study size, quantitative variables, and statistical methods. This reporting logic follows contemporary recommendations for observational field-monitoring studies in applied sport science. A completed reporting checklist with item numbers, manuscript locations, and page references is provided as Supplementary S3, submitted alongside this manuscript for editorial and peer-review verification.

Participants

Ethical approval for this study was granted by the Research Ethics Committee, Faculty of Sport Science, Universitas Negeri Medan, Indonesia (Ethical Clearance No.: 0234/UN33.8/SP2H/PM/2025; Approval date: 10 June 2025). All procedures were conducted in accordance with the ethical principles of the Declaration of Helsinki (2013 revision) and applicable national research ethics guidelines. Written informed consent was obtained from each participant before the first testing session, following full disclosure of study procedures, associated risks, the voluntary nature of participation, and the right to withdraw at any time without penalty. Participants were assigned de-identified codes (SIM-001 to SIM-072) for data handling and are not individually identifiable in this manuscript or its supplementary materials.

The final sample comprised 72 nationally competing Sanda Wushu athletes (36 males, 36 females; age 18–30, M = 24.2 years, SD = 3.7). Anthropometric characteristics were: height 166.9 ± 8.6 cm (male: 172.1 ± 8.0 cm; female: 161.7 ± 6.0 cm); body mass 61.4 ± 10.0 kg (male: 65.7 ± 9.9 kg; female: 57.1 ± 7.7 kg); and BMI 22.1 ± 3.6 kg/m^2 (male: 23.1 ± 3.2 kg/m^2; female: 21.1 ± 3.6 kg/m^2). For four athletes (5.6% of the sample), between two and three of the 15 Session 1 items could not be validly completed because of scheduling and equipment-availability constraints on the Session 1 testing day (e.g., an item requiring equipment that was temporarily unavailable to a testing lane); these entries were coded as missing with a zero indicator flag for audit purposes and excluded from both the numerator and denominator of the affected athlete's PerfIndex and ScoreIndex composites, rather than treated as zero performance. All four athletes completed the full 15-item battery at Session 2.

Testing Setting and Equipment

Testing was conducted in the indoor sports hall of Universitas Negeri Medan (Medan, North Sumatra, Indonesia) on a hard, rubber-coated sports floor suitable for sprint, agility, and hop testing. Both sessions were administered between 08:00 and 12:00 WIB (Western Indonesian Time) to minimize diurnal variation in physiological performance. The same lead examiner conducted both sessions, assisted by two trained observers. Testing procedures were standardized: each session began with an identical 10-minute dynamic warm-up (5 min light jogging, 5 min dynamic stretching targeting the major muscle groups). Tests were administered in a fixed sequence from least to most fatiguing, with standardized rest intervals between items (2 min for upper-body tests, 3 min for lower-body and aerobic/anaerobic items). Timing was recorded using a digital stopwatch (Casio HS-70W, precision 0.01 s) for sprint and agility tests; hop distances were measured with a non-elastic measuring tape (precision 0.5 cm); V-Sit and Reach were measured with a standard sit-and-reach box (precision 0.5 cm). Ambient temperature during both sessions ranged between 26°C and 30°C. No participants reported illness, acute muscle soreness, or deviations from normal training in the 48 hours preceding each session.

Field-Test Battery and Instrumentation

The 15-item battery was selected to represent performance domains relevant to Sanda Wushu, including flexibility for kicking range, abdominal and upper-limb endurance for repeated striking and clinch activity, unilateral lower-limb endurance and hop performance for limb-disparity monitoring, agility and sprinting for attack-defense transitions, throwing power for upper-body force expression, reaction time for response speed, core stability for postural control, and aerobic/anaerobic tests for repeated high-intensity efforts. The battery covered: flexibility (V-Sit and Reach), abdominal endurance (2-min Sit-Up), arm endurance (1-min Push-Up), bilateral leg endurance (SL Wall Sit - right/left), agility (8 x 5m Shuttle Run), bilateral leg power (Triple Hop Jump - right/left), arm power (MB Overhead Throw), reaction time (Ruler Drop), speed (20m Sprint), core stability (12-Level Core Stability Test), aerobic endurance (Multi-Stage Fitness Test), anaerobic capacity (300m Run), and anaerobic endurance/fatigue index (RAST Test).

Each test produced: (i) a continuous ratio score (pct_ratio, 0-1, capped at 1.0); (ii) a 5-level ordinal score; and (iii) a zero_indicator flag. For tests in which lower raw values indicated better performance, scores were normalized to align with each benchmark's directionality, ensuring that higher pct_ratio and ordinal scores always indicated better performance. Athlete-session composites were then computed as follows: PerfIndex = mean of valid pct_ratio items, and ScoreIndex = mean of valid ordinal scores. For the four athletes with block-missing Session 1 items, indices were based on 12-13 valid items; the affected items were excluded from both the numerator and denominator and documented separately via the zero_indicator flag. Complete benchmark thresholds, directionality rules, and scoring syntax are provided in Supplementary S1-S2 to support reproducibility.

Benchmark norms for all 15 test items were derived from sex-specific internal federation-standard benchmarks supplied by the research team and reproduced in Supplementary Tables S1a and S1b. Sex-specific norms were applied separately for male and female participants. Each benchmark represents the minimum performance value for the highest categorical tier (Score = 5) among national-level competitors in the available benchmark file. Because these norms are not peer-reviewed journal norms and were established approximately 2-3 years before the present study, they are treated as operational benchmarks rather than validated national reference values. This point is identified as a limitation. PerfIndex values exceeding 1.0 (i.e., athlete outperforms the benchmark) were capped at 1.0 to prevent inflation of composite indices.

Injury-Risk Proxy Metrics

Asymmetry (%) = |Right - Left| / max (Right, Left) x 100, applied to the SL Wall Sit (right vs. left) and Triple Hop Jump (right vs. left). Higher values indicate greater inter-limb disparity. The 10-15% range is used only as a descriptive reference zone because asymmetry interpretation remains task-specific and cannot be treated as a universal diagnostic threshold (D'Hondt et al., 2024; Fort-Vanmeerhaeghe et al., 2022; Guan et al., 2022; Guan et al., 2023; Jordan & Bishop, 2024; Keogh et al., 2023).

Bias Control and Study Size

To reduce procedural bias, both sessions followed the same test sequence, field-testing procedures, scoring rules, and benchmark-normalization logic. The principal risks of bias were selection bias due to convenience (squad-level) sampling, performance variability associated with the active training mesocycle, and block-missing Session 1 data for four athletes. These risks were addressed through transparent reporting, individual-level trajectories, zero_indicator documentation, and a sensitivity analysis that excluded the affected athletes, rather than through causal inference. With 72 paired observations, the study was adequately sized for formal inferential testing of the primary composite outcomes, although the single-group, non-randomized, non-controlled design still precludes causal attribution of the observed changes to training adaptation (Teresi et al., 2022; Totton et al., 2023).

Statistical Analysis

Paired-samples t-tests (IBM SPSS v.26; alpha = 0.05) compared Session 2 vs. Session 1 for PerfIndex, ScoreIndex, SL Wall Sit asymmetry, and Triple Hop asymmetry. Mean differences were supplemented with 95% confidence intervals, Cohen's dz (= mean paired difference / SD of paired differences), and Hedges' g corrected for small-sample bias using J(df) = 1 - 3/(4df - 1). Normality of the paired difference scores was formally assessed using the Shapiro-Wilk test; neither PerfIndex nor ScoreIndex differences deviated significantly from normality (PerfIndex: W = 0.988, p = .702; ScoreIndex: W = 0.973, p = .132), supporting the use of parametric paired tests. Missing data were handled at the item level before composite calculation: block-missing Session 1 values for four athletes were excluded from valid-item averaging and retained as zero_indicator documentation. A sensitivity analysis excluding these four athletes (n = 68) compared full-sample and missing-data-adjusted estimates. Wilcoxon signed-rank tests were additionally computed as non-parametric alternatives for the primary outcomes and were consistent in direction and significance with the parametric results. This strategy follows recent guidance that sport-science studies should report estimation (effect sizes, confidence intervals) alongside significance testing rather than significance testing alone (Abt et al., 2025; Hecksteden et al., 2022; Mesquida et al., 2022).

RESULTS AND DISCUSSION

Result

Data Completeness

Of 2,160 possible observations (72 athletes x 15 tests x 2 sessions), Session 2 yielded 1,080 valid entries (100%). Session 1 yielded 1,071 valid entries (99.2%), with 9 block-missing entries across 4 athletes documented via zero_indicator flags. These entries were excluded from valid-item composite scoring rather than treated as true zero performance. Shapiro-Wilk tests confirmed that the paired difference scores for PerfIndex ($W = 0.988$, $p = .702$) and ScoreIndex ($W = 0.973$, $p = .132$) did not deviate significantly from normality, supporting the primary use of paired-samples t-tests. Wilcoxon signed-rank tests were additionally reported for all primary outcomes to ensure transparency regarding distributional assumptions.

Primary Outcomes: PerfIndex and ScoreIndex

Full-sample analyses ($N = 72$) showed statistically significant increases in composite performance indices. PerfIndex increased from 0.727 +/- 0.031 (S1) to 0.755 +/- 0.033 (S2), mean difference +0.029 (95% CI [0.024, 0.034]), $t(71) = 11.46$, $p < .001$, $d_z = 1.35$, Hedges' $g = 1.34$. ScoreIndex increased from 2.92 +/- 0.25 to 3.16 +/- 0.26, mean difference +0.240 (95% CI [0.190, 0.291]), $t(71) = 9.41$, $p < .001$, $d_z = 1.11$, Hedges' $g = 1.10$. Of the 72 athletes, 64 (88.9%) showed a positive PerfIndex change and 63 (87.5%) showed a positive ScoreIndex change; no athlete showed zero change. Individual and group-level distributions are displayed in Figure 1 (PerfIndex) and Figure 2 (ScoreIndex). The Wilcoxon signed-rank test corroborated both parametric results (PerfIndex: $W = 79$, $p < .001$; ScoreIndex: $W = 115$, $p < .001$), confirming directional and statistical consistency across the two approaches.

Sex-disaggregated descriptive inspection (36 males, 36 females) was not analyzed inferentially given the study's primary focus on the whole-squad comparison. Male athletes showed a Session 1 PerfIndex of 0.732 +/- 0.031, increasing to 0.763 +/- 0.033 at Session 2; female athletes showed 0.721 +/- 0.030 at Session 1, increasing to 0.747 +/- 0.032 at Session 2. The corresponding ScoreIndex values were 2.96 +/- 0.24 to 3.24 +/- 0.25 (male) and 2.88 +/- 0.25 to 3.09 +/- 0.26 (female). Both sexes, therefore, showed a directionally and numerically similar improvement pattern; these values are reported to improve transparency and should not be interpreted as confirmatory evidence of a sex-based difference, since no formal interaction test was conducted.

Table 1 summarizes the sex-stratified and whole-squad composite indices so that the reader can inspect both the direction of change and the sensitivity of the estimate to the four athletes with block-missing Session 1 data before the formal effect-size summary is presented.

Table 1. Sex-Stratified and Whole-Squad Performance Indices Across Two Monitoring Sessions (N = 72).

Group	Sex	PerfIndex S1	PerfIndex S2	ScoreIndex S1	ScoreIndex S2	PerfIndex Delta
Male (n = 36)	M	0.732 ± 0.031	0.763 ± 0.033	2.96 ± 0.24	3.24 ± 0.25	+0.031
Female (n = 36)	F	0.721 ± 0.030	0.747 ± 0.032	2.88 ± 0.25	3.09 ± 0.26	+0.026
Total (N = 72)	—	0.727 ± 0.031	0.755 ± 0.033	2.92 ± 0.25	3.16 ± 0.26	+0.029
Sensitivity (n = 68) [a]	—	0.724 ± 0.028	0.754 ± 0.033	2.90 ± 0.22	3.15 ± 0.26	+0.030
t(71); p	—	—	—	—	—	11.46; <.001
95% CI	—	—	—	—	—	[0.024, 0.034]
Cohen's dz; Hedges' g	—	—	—	—	—	1.35; 1.34
Shapiro-Wilk p (ΔPerfIndex)	—	—	—	—	—	0.702

Note. PerfIndex: 0-1 (higher = better); ScoreIndex: 1-5 ordinal (higher = better). Inferential statistics (t, 95% CI, dz, Hedges' g) apply to the whole-squad PerfIndex delta (N = 72). The sensitivity row excludes the four athletes with missing block items in Session 1 (n = 68). Full outcome-by-outcome inferential statistics, including ScoreIndex and asymmetry proxies, are reported in Table 2.

Individual PerfIndex changes ranged from -0.025 to +0.092 across the sample (Figure 1); only eight of 72 athletes (11.1%) showed a decline between sessions, and none showed a decline exceeding 0.03 on the 0-1 scale. Figure 1 presents an estimation-focused display of PerfIndex, combining session-specific means and 95% confidence intervals with the paired mean-change estimate. This format emphasizes the magnitude of effects and precision without implying access to distributional information beyond the reported data.

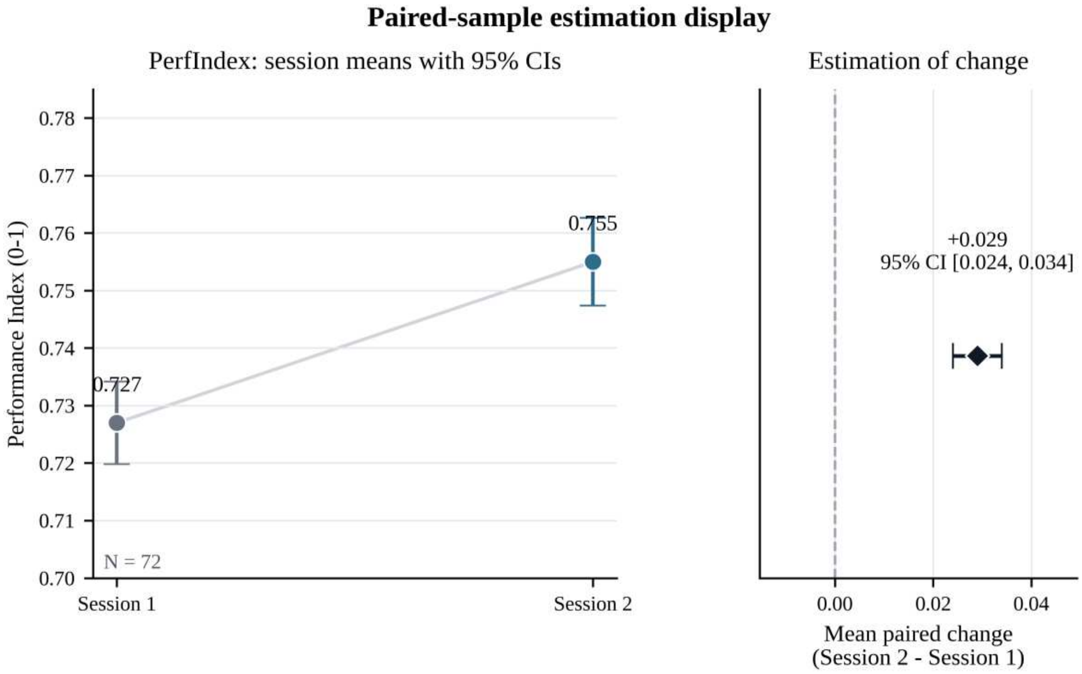

Figure 1. PerfIndex session means and paired-change estimation display (N = 72). Points represent means; vertical bars represent 95% confidence intervals for session means; the right panel shows the mean paired change and its 95% confidence interval.

As shown in Figure 1, the Session 2 mean was higher than the Session 1 mean, and the 95% confidence interval for the paired change excluded zero. The figure is intentionally estimation-centered; causal attribution remains inappropriate because the study used a single-group observational design. Figure 2 applies the same estimation framework to ScoreIndex, allowing direct inspection of the session means, their uncertainty, and the magnitude and precision of the paired change.

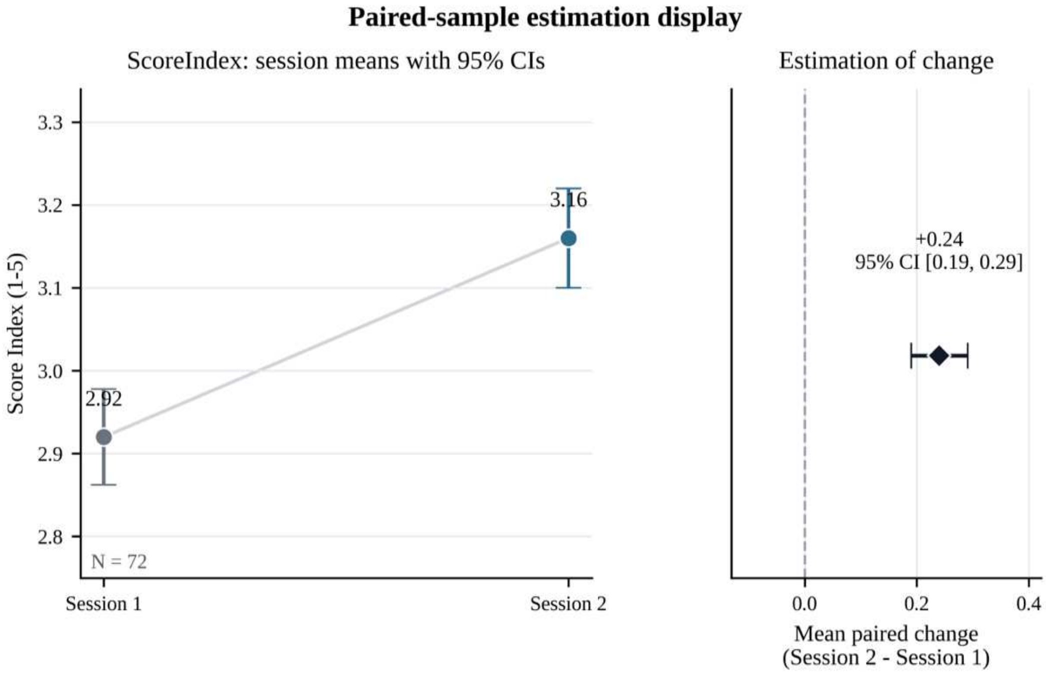

Figure 2. ScoreIndex session means and paired-change estimation display (N = 72). Points represent means; vertical bars represent 95% confidence intervals for session means; the right panel shows the mean paired change and its 95% confidence interval.

Figure 2 shows that ScoreIndex increased in the same direction as PerfIndex. The paired-change estimate was positive, and its 95% confidence interval excluded zero, consistent with the reported paired t-test and Wilcoxon signed-rank results.

Sensitivity Analysis (n = 68, Excluding Athletes with Block-Missing Session 1 Data)

Excluding the four athletes with block-missing Session 1 items, PerfIndex changed from 0.724 +/- 0.028 (S1) to 0.754 +/- 0.033 (S2), mean difference +0.030 (95% CI [0.025, 0.035]), $t(67) = 11.66$, $p < .001$, $d_z = 1.41$, Hedges' $g = 1.40$. The sensitivity estimate is closely consistent with the full-sample result (full-sample mean difference +0.029 vs. sensitivity +0.030), indicating that the four affected athletes did not materially distort the whole-squad estimate. This consistency supports treating the zero_indicator-flagged items as a minor, non-differential source of missingness rather than a threat to the primary composite finding. The full set of full-sample and sensitivity estimates, together with the asymmetry-proxy inferential statistics, is presented in Table 2. Table 2 reports the formal sensitivity and effect-size estimates for all primary and secondary outcomes, enabling the reader to compare the full-sample result with the missing-data-adjusted analysis.

Table 2. Formal Sensitivity and Effect-Size Summary for Full-Sample and Missing-Data-Adjusted Analyses.

Outcome	Sample	Mean difference	95% CI	t(df)	p	dz	Hedges' g
PerfIndex	Full sample (N=72)	+0.029	[0.024, 0.034]	11.46 (71)	<.001	1.35	1.34
ScoreIndex	Full sample (N=72)	+0.240	[0.190, 0.291]	9.41 (71)	<.001	1.11	1.10
PerfIndex	Excl. block-missing (n=68)	+0.030	[0.025, 0.035]	11.66 (67)	<.001	1.41	1.40
SL Wall Sit asymmetry	Full sample (n=71)	+3.00 pp	[0.43, 5.56]	2.33 (70)	.023	0.28	0.27
Triple Hop asymmetry	Full sample (n=72)	+0.21 pp	[-0.93, 1.35]	0.36 (71)	.718	0.04	0.04

Note. pp = percentage points. Hedges' g was calculated using $J(df) = 1 - 3/(4df - 1)$. All estimates are based on two-tailed paired-samples t-tests; alpha = 0.05.

As shown in Table 2, both the full-sample and the missing-data-adjusted analyses converge on the same conclusion: PerfIndex and ScoreIndex increased significantly with large standardized effect sizes, while SL Wall Sit asymmetry showed a small but significant increase and Triple Hop asymmetry showed no significant change.

Secondary Outcomes: Asymmetry Proxies

SL Wall Sit asymmetry (n = 71; one athlete missing a Session 1 side due to block-missing data) increased from 27.9 +/- 19.6% at S1 to 30.9 +/- 20.1% at S2, mean difference +3.00 percentage points (95% CI [0.43, 5.56]), t(70) = 2.33, p = .023, dz = 0.28, Hedges' g = 0.27. This small but statistically significant increase, combined with the very large baseline dispersion, indicates that Wall Sit asymmetry is highly heterogeneous across the squad rather than a stable trait. Triple Hop asymmetry (n = 72) changed from 11.5 +/- 8.5% at S1 to 11.7 +/- 8.1% at S2, mean difference +0.21 percentage points (95% CI [-0.93, 1.35]), t(71) = 0.36, p = .718, dz = 0.04, Hedges' g = 0.04, indicating no meaningful change. Table 3 and Figure 3 summarize these group-level and threshold-based asymmetry patterns.

Table 3 presents group-level asymmetry statistics, along with the proportion of athletes exceeding the descriptive 10% and 15% reference zones at each session, so that the reader can assess how widely individual disparities were distributed across the squad rather than relying on the group mean alone.

Table 3. Asymmetry-Based Inter-Limb Monitoring Proxies Across Sessions (N = 71-72).

Test	n	Metric	S1 (%)	S2 (%)	t(df)	p
SL Wall Sit	71	Group Mean ± SD	27.90 ± 19.65	30.90 ± 20.10		
SL Wall Sit	71	Athletes > 10% (%)	81.7 (58/71)	83.1 (59/71)		
SL Wall Sit	71	Athletes > 15% (%)	67.6 (48/71)	74.6 (53/71)		
SL Wall Sit	71	Highest individual [b]	84.7	68.5		
SL Wall Sit	71	Group comparison			2.33 (70)	.023
Triple Hop	72	Group Mean ± SD	11.45 ± 8.47	11.66 ± 8.08		
Triple Hop	72	Athletes > 15% (%)	31.9 (23/72)	27.8 (20/72)		
Triple Hop	72	Highest individual [c]	35.0	23.9		
Triple Hop	72	Group comparison			0.36 (71)	.718

Note. [a] n varies slightly by test due to one block-missing Session 1 entry for SL Wall Sit. [b] Highest individual SL Wall Sit asymmetry: SIM-051 (84.7% S1, 68.5% S2). [c] Highest individual Triple Hop asymmetry: SIM-008 (35.0% S1, 23.9% S2). The high SD for SL Wall Sit at both sessions (~20 percentage points) reflects genuine between-athlete heterogeneity rather than a single outlier, given that 48/71 athletes (67.6%) exceeded the descriptive 15% zone at S1. Asymmetry = |Right - Left| / max(Right, Left) x 100.

Because a majority of the squad exceeded the descriptive 10-15% Wall Sit reference zone at both sessions, the group means alone understate how common elevated asymmetry was; the threshold-prevalence rows in Table 3 are therefore more informative for coaching triage than the mean and SD in isolation. Figure 3 integrates three complementary views of asymmetry: group means with 95% confidence intervals, paired mean-change estimates with 95% confidence intervals, and the proportion of athletes exceeding the descriptive reference zones. This avoids treating a single summary statistic as sufficient for screening interpretation.

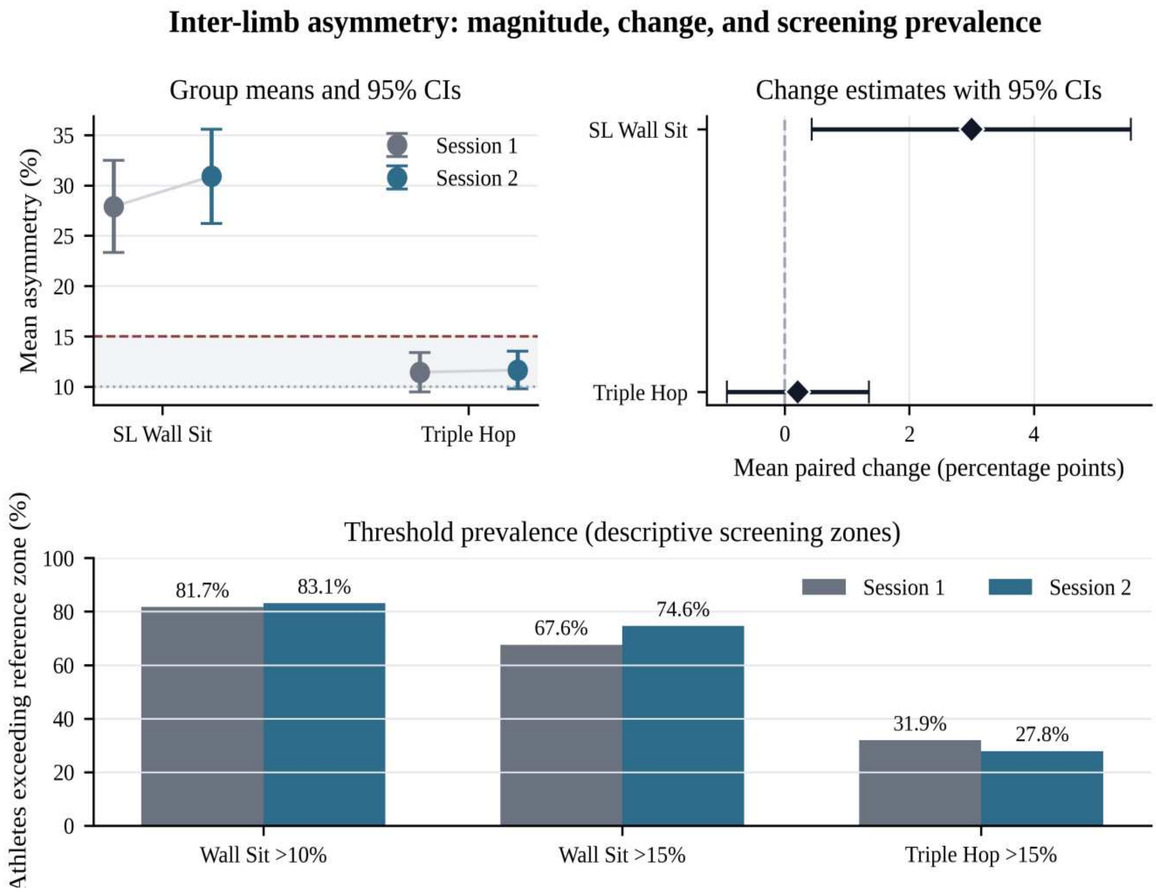

Figure 3. Inter-limb asymmetry magnitude, paired change, and descriptive threshold prevalence. Reference zones are presented for screening description only and must not be interpreted as diagnostic cut-points.

Figure 3 shows that SL Wall Sit asymmetry was higher and more prevalent than Triple Hop asymmetry above the descriptive reference zones. The paired-change interval for SL Wall Sit excluded zero, whereas the interval for Triple Hop crossed zero. Threshold prevalence is therefore presented alongside, rather than in place of, the continuous estimates.

Item-Level Test Outcomes

Table 4 reports paired-samples results for each of the 15 individual test items (pct_ratio scale, 0-1), so that the composite-level PerfIndex and ScoreIndex findings can be traced back to their constituent tests and the unusually large composite effect sizes reported above can be compared against more typical single-test effect sizes.

Table 4. Item-Level Paired Comparisons for All 15 Field-Test Items (pct_ratio Scale, 0–1).

Test	n	S1 Mean ± SD	S2 Mean ± SD	t(df)	p	dz
1. V-Sit and Reach	71	0.618 ± 0.192	0.664 ± 0.216	3.92 (70)	<.001	0.47
2. 2-min Sit-Up	71	0.740 ± 0.145	0.771 ± 0.159	3.62 (70)	<.001	0.43
3. 1-min Push-Up	72	0.642 ± 0.165	0.672 ± 0.184	3.31 (71)	.001	0.39
4. SL Wall Sit – Right	71	0.638 ± 0.168	0.669 ± 0.189	3.64 (70)	<.001	0.43
5. SL Wall Sit – Left	72	0.584 ± 0.170	0.601 ± 0.190	1.94 (71)	.057	0.23
6. 8x5m Shuttle Run	72	0.886 ± 0.057	0.895 ± 0.062	2.61 (71)	.011	0.31

Test	n	S1 Mean ± SD	S2 Mean ± SD	t(df)	p	dz
7. Triple Hop Jump – Right	72	0.847 ± 0.078	0.866 ± 0.081	4.07 (71)	<.001	0.48
8. Triple Hop Jump – Left	72	0.814 ± 0.086	0.832 ± 0.090	4.23 (71)	<.001	0.50
9. MB Overhead Throw	71	0.773 ± 0.111	0.801 ± 0.124	4.45 (70)	<.001	0.53
10. Ruler Drop (reaction)	72	0.715 ± 0.138	0.746 ± 0.148	4.69 (71)	<.001	0.55
11. 20m Sprint	71	0.897 ± 0.051	0.912 ± 0.059	4.27 (70)	<.001	0.51
12. 12-Level Core Stability	71	0.547 ± 0.153	0.626 ± 0.176	8.19 (70)	<.001	0.97
13. Multi-Stage Fitness Test	72	0.688 ± 0.121	0.725 ± 0.129	5.25 (71)	<.001	0.62
14. 300m Run	71	0.833 ± 0.072	0.852 ± 0.080	4.13 (70)	<.001	0.49
15. RAST Fatigue Index	70	0.670 ± 0.158	0.703 ± 0.167	3.38 (69)	.001	0.40

Note. All comparisons are two-tailed paired-samples t-tests (Session 2 vs. Session 1); alpha = 0.05. n varies slightly across items due to missing entries in the block for four athletes in Session 1. dz = Cohen's dz. The 12-Level Core Stability item shows the largest single-item standardized effect (dz = 0.97) in the battery; SL Wall Sit – Left is the only item that did not reach significance at alpha = 0.05 (p = .057).

Figure 4 summarizes the standardized paired changes for all 15 field-test items in a single forest plot. The display permits simultaneous comparison of effect magnitude and uncertainty and makes clear that the composite-level effects should not be interpreted as uniform across the constituent tests.

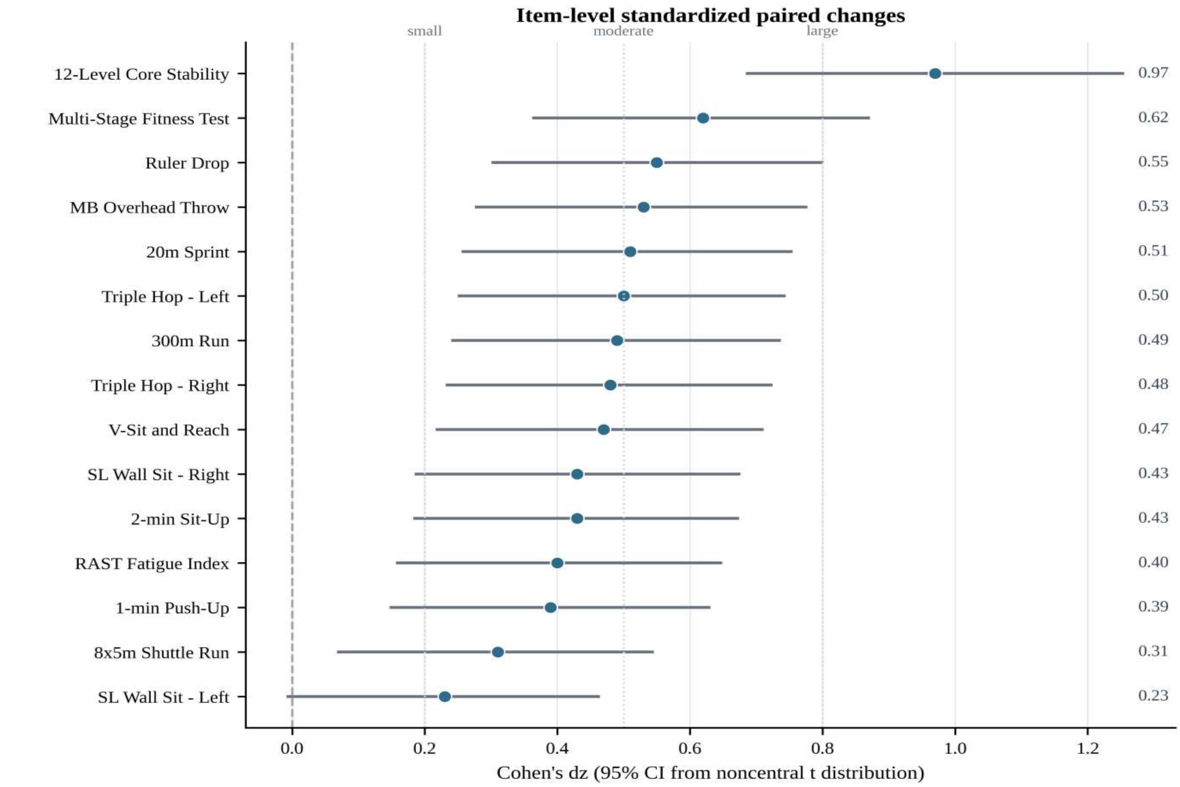

Figure 4. Forest plot of item-level Cohen's dz estimates with 95% confidence intervals obtained by inverting the noncentral t distribution from the reported paired t statistics and item-specific sample sizes.

The forest plot confirms that the largest standardized change occurred for the 12-Level Core Stability Test, whereas the SL Wall Sit - Left estimate was the smallest. Most item-level estimates were small to moderate, supporting the manuscript's caution that the larger composite effect partly reflects variance reduction from averaging multiple items. Fourteen of the 15 items showed a statistically significant Session 1-to-Session 2 improvement, with item-level dz values ranging from 0.23 (SL Wall Sit – Left, non-significant) to 0.97 (Core Stability). These single-item effect sizes are, as expected, smaller and more heterogeneous than the composite PerfIndex effect size (dz = 1.35), confirming that the large composite effect reflects the reduced variance of an averaged multi-item index rather than a uniformly large effect on every underlying physical quality.

Discussion

Performance Monitoring Sensitivity and Statistical Power

The primary finding - a statistically robust and large composite performance increase (PerfIndex: dz = 1.35, Hedges' g = 1.34, $p < .001$; ScoreIndex: dz = 1.11, $p < .001$) across a squad of 72 athletes - should still be contextualized carefully. The standardized effect sizes are unusually large for a repeated field-testing interval, which most plausibly reflects the low between-athlete variability of an averaged 13-15-item composite rather than an unusually large true training effect on any single physical quality; composite indices formed by averaging many items tend to have smaller standard deviations than their constituent items, which mechanically inflates dz and Hedges' g relative to single-test effect sizes (compare Table 4 item-level results). The 95% confidence interval for the mean change in PerfIndex excludes zero, indicating a statistically significant group-level signal. Still, the single-group, non-randomized, pre-post design cannot separate training-related adaptation from maturation, seasonal/competition-cycle effects, familiarization with the test battery, or regression to the mean (Abt et al., 2025; Mesquida et al., 2022; Teresi et al., 2022).

This tension between statistical robustness and causal ambiguity is not unique to the present study. Recent methodological work in sport science emphasizes that even well-powered observational designs require transparent estimation and explicit acknowledgment of confounds, rather than treating statistical significance as equivalent to intervention effectiveness (Abt et al., 2025; Hecksteden et al., 2022; Mesquida et al., 2022). For the present framework, the most defensible interpretation is therefore that the monitoring workflow produced complete, reproducible composite estimates and a statistically dependable directional signal across the training interval. In contrast, the attribution of that signal to a specific training stimulus remains an open question for a future controlled design.

Within this context, the Performance Index approach of the present study aligns conceptually, but not methodologically identically, with the Total Score of Athleticism (TSA) framework and broader composite profiling approaches (Asimakidis et al., 2024; Maestroni et al., 2023; Weakley et al., 2024). TSA relies on standardized profiling logic, whereas the present PerfIndex uses benchmark-normalized capped ratios. The manuscript, therefore, treats PerfIndex and ScoreIndex as coach-facing monitoring indices rather than fully validated performance constructs. The directional consistency of the indices across 88.9% of athletes (PerfIndex) suggests plausible workflow responsiveness, but reliability, smallest worthwhile change, and minimal detectable change must still be established before individual-level change-detection claims can be made with confidence. A formal participant calibration step, following the classification taxonomy outlined by McKay et al. (2022), would also improve comparability across future Sanda monitoring studies by anchoring participant descriptors to standardized performance tiers rather than to self-reported competitive levels. Because the present study did not include a dedicated test-retest reliability session, SWC and MDC values for PerfIndex and ScoreIndex cannot be estimated from the current data. They must be derived from a separate reliability study under stable conditions before individual-athlete change-detection thresholds can be implemented.

A brief sex-disaggregated descriptive note is warranted, though no formal interaction test was conducted. Male athletes (n = 36) showed a slightly higher mean Session 1 PerfIndex (0.732) than female athletes (n = 36; 0.721), and this small gap persisted at Session 2 (0.763 vs. 0.747). Both sexes improved by a similar absolute margin (male +0.031; female +0.026), suggesting a broadly parallel response pattern rather than a sex-specific monitoring signal. Because the benchmark norms are operational rather than independently validated, and because no formal sex × session interaction was tested, this pattern should not be interpreted as confirmatory evidence of a sex-based performance difference. Future iterations of the battery should retain sex-stratified benchmark norms and, where sample size allows, formally test the sex x session interaction.

Individual-Level Variation and Missing-Data Handling

Beyond the group-level trend, individual PerfIndex trajectories varied meaningfully (range: -0.025 to +0.092), as expected in heterogeneous squad training under normal seasonal variation. The four athletes with block-missing Session 1 items (5.6% of the sample) illustrate an applied data-management scenario that any field-monitoring pipeline must handle transparently: two or three domains per affected athlete were not validly testable on the Session 1 testing day, and these missing values were flagged with zero_indicator rather than treated as true zero performance. The sensitivity analysis excluding these athletes (Table 2) produced estimates that closely matched those from the full sample, indicating that this modest, non-systematic missingness did not meaningfully bias the whole-squad composite estimates.

This experience supports two practical recommendations for future implementations of the battery. First, testing-day logistics (equipment availability across lanes, athlete flow) should be scheduled with sufficient redundancy to minimize block-missing entries, since even a small proportion of affected athletes requires an explicit sensitivity analysis to rule out bias. Second, the zero_indicator flag should be treated as a mandatory field in any coach-facing dashboard built on this pipeline, so that missing-item composites are never visually confused with genuinely low performance. Future iterations should also log the specific reason for each missing entry (e.g., equipment, scheduling, injury, illness) as a structured field, which was not consistently captured in the present dataset and would strengthen future missing-data audits.

Asymmetry Indicators: Screening Thresholds and Squad-Level Prevalence

The asymmetry-based proxies showed a statistically significant, though numerically small, pre-post increase for SL Wall Sit ($dz = 0.28$, $p = .023$) and no significant change for Triple Hop ($dz = 0.04$, $p = .718$). The more striking finding is descriptive rather than inferential: 48 of 71 athletes (67.6%) exceeded the descriptive 15% Wall Sit reference zone at Session 1, and 53 of 71 (74.6%) did so at Session 2. This prevalence is far higher than the reference zone's framing as an exceptional flag would suggest, indicating either that the commonly cited 10-15% range is not well calibrated to this squad and this specific unilateral wall-sit protocol, or that elevated Wall Sit asymmetry is simply common in this population. Because asymmetry thresholds are task-specific and not strongly validated as universal cut-points, this pattern should be interpreted as evidence for developing squad-specific and test-specific reference distributions rather than applying an imported threshold uncritically (D'Hondt et al., 2024; Guan et al., 2022; Keogh et al., 2023).

The highest individual Wall Sit asymmetry (SIM-051: 84.7% at S1, decreasing to 68.5% at S2) illustrates that, even after accounting for high squad-wide prevalence, some individuals sit well outside the group distribution and would still warrant follow-up assessment. From a coaching perspective, a two-tier interpretation is therefore proposed: squad-relative percentile flags (e.g., top decile) for routine triage, combined with the descriptive 10-15% zone only as a secondary, coarser check. Neither should be treated as a binary pass/fail decision. This more cautious interpretation is consistent with evidence showing that asymmetry should be considered alongside sport, test type, athlete history, and longitudinal variation rather than interpreted through a single universal threshold. Prospective and review evidence suggest that larger lower-limb asymmetries can be associated with injury incidence in some athletic settings, but effect sizes and transferability vary across tasks and populations (Bishop et al., 2022; D'Hondt et al., 2024; Fort-Vanmeerhaeghe et al., 2022; Guan et al., 2022; Guan et al., 2023; Keogh et al., 2023; Sliwowski et al., 2024). Individual outlier values in the present squad should therefore be treated as monitoring concerns warranting reassessment rather than as confirmed injury-risk classifications.

Khoygani and Esmaeili (2025) further reported that normalizing hop performance to the body-height ratio may outperform the limb symmetry index for predicting second ACL injury, supporting the incorporation of body-height-adjusted metrics alongside percentage asymmetry in future protocol iterations. The present Triple Hop protocol, which computes asymmetry between the right and left limbs, can readily be extended to incorporate this normalization with no additional testing burden, and could be evaluated against the squad-specific percentile approach proposed above.

Core Stability and Striking Performance

The Core Stability (12-Level Core Stability Test) item showed the largest item-level standardized effect in the entire battery (pct_ratio: 0.547 to 0.626, dz = 0.97, $p < .001$; see Table 4), consistent with the likely importance of postural control in Sanda preparation. This finding should be framed as a plausible performance-relevant signal rather than evidence that technical execution improved, since the present study did not measure kick velocity, strike accuracy, throwing quality, or technical degradation under fatigue. Recent systematic reviews suggest that core stability or core strength training can benefit balance, strength, and selected sport-specific outcomes in combat-sport athletes (Huang et al., 2023; Zhang et al., 2025). Future studies should pair the field battery with IMU-derived kick kinematics, punch-force indicators, or video-based skill metrics to directly test this link.

Training Load Management and HRV in Combat Sports

The current monitoring framework quantifies periodic performance outcomes but does not capture training load or recovery status between sessions. This is a central limitation because changes in field-test scores cannot be attributed to a specific training dose without external and internal load data. In workload theory, external load reflects the work completed, while internal load captures the athlete-specific physiological or psychophysiological response (Impellizzeri et al., 2023). Reviews and combat-sport case-study evidence also show that single load metrics often have limited or context-dependent relationships with injury, physical qualities, or session response, supporting a multivariate approach to future monitoring (Dudley et al., 2023; Xu et al., 2025). Heart rate variability (HRV) should be considered a plausible adjunct to, rather than a stand-alone replacement for, field-test monitoring. Recent strength-and-conditioning guidance describes HRV as a potentially useful recovery-monitoring tool when measurement timing, interpretation, and athlete context are controlled (Addleman et al., 2024). Nakamura et al. (2022) further reported reliable nocturnal HRV indices and sensitivity to maximal endurance-exercise perturbation, supporting the methodological relevance of timing and repeated measurement. For the present Sanda framework, HRV integration should therefore be treated as a future empirical step that requires standardized measurement timing, repeated observations, and validation against training-load and recovery markers.

Field Testing Validity in Striking Combat Sports

The multi-domain field-test approach of this study is consistent with emerging practice in combat-sport performance science. Evidence from mixed martial arts testing indicates that submaximal and supramaximal protocols can yield distinct physiological profiles, supporting the need to match test intensity to the monitoring question (Goncalves et al., 2024). The present 15-item battery, covering aerobic endurance, anaerobic capacity, neuromuscular power, reaction time, flexibility, and core stability, addresses multiple domains relevant to Sanda. Still, it should be interpreted as a feasibility-tested battery rather than a validated performance model. In taekwondo, Taati et al. (2022) examined a sport-specific field test for estimating aerobic power, anaerobic fitness, and agility, demonstrating the practical value of combat-sport-specific testing. In judo, Ceylan et al. (2022) examined which performance tests best define Special Judo Fitness Test classification in elite athletes, supporting the broader principle that combat-sport batteries should include lower-limb power, aerobic fitness, and sport-relevant repeated-effort indicators. Recent Sanda-specific work on kick mechanics and punching force prediction further supports the relevance of biomechanical and force-related domains, while also indicating that Sanda-specific validation remains incomplete (Eltifi-Ghanmi et al., 2025; Wang et al., 2025). Together, these studies support the rationale for a multi-domain monitoring framework without implying that the present battery has already achieved criterion validity.

Wearable Sensor Integration: Future Directions

The field-test framework developed here is explicitly designed as a foundation for future wearable-sensor integration, not as evidence that sensor-augmented monitoring has already been validated in Sanda Wushu. Recent reviews suggest that sport and combat-sport wearable sensors can support analysis of limb biomechanics, impact-related variables, and fatigue-related movement features, but they also stress the need for sensor validation, standardized protocols, and coach-facing interpretation (Kovoor et al., 2024; Preatoni et al., 2022; Xue et al., 2025). Thus, IMU and HR/HRV integration should be treated as the next empirical phase, not as a conclusion from the present field-test-only study. In a future sensor-augmented version of the present framework, the PerfIndex could be supplemented by: (a) IMU-derived external load, such as session movement intensity, strike counts, and peak acceleration events; (b) HR/HRV-based internal load and recovery status; and (c) strike kinematic variables, such as velocity, acceleration, and inter-strike consistency. These additions could help evaluate fatigue accumulation, technique degradation, and individual readiness, but they would require validation before being used for injury-risk decisions. Practical sensor integration should begin with validated IMU placement and sampling protocols for punch and kick mechanics, standardization of HR/HRV measurement, and comparison against accepted criterion measures before field deployment (Liu et al., 2024; Pezenka & Wirth, 2025; Wang et al., 2025; Xue et al., 2025).

Limitations

This study has the following limitations. First, and most importantly, the single-group, non-randomized, pre-post observational design precludes separation of training-related adaptation from maturation, seasonal/competition-cycle effects, familiarization with the test battery, recovery status, or regression to the mean; the statistically significant group-level change should not be read as proof of a specific training effect. Second, convenience (squad-level) sampling limits generalizability beyond similar training contexts and Sanda squads. Third, training load and recovery status between sessions were not quantified, preventing attribution of performance changes to a specific training dose. Fourth, no prospective injury surveillance was implemented, preventing validation of asymmetry thresholds against actual injury outcomes. Fifth, benchmark norms were internally developed and operational rather than independently peer-reviewed or nationally validated; if these do not represent current national-level Sanda capability, PerfIndex compression may occur in the upper range. Sixth, no wearable sensor data were collected; the study tested only the field-test pipeline, not sensor-augmented monitoring. Seventh, four athletes (5.6%) had block-missing Session 1 items. However, the sensitivity analysis suggested minimal impact on the whole-squad estimates; the specific cause of each missing entry was not captured as a structured field. Eighth, reliability, minimal detectable change, and smallest worthwhile change were not estimated, so individual-athlete change-detection claims remain preliminary despite the group-level result being statistically robust. Ninth, published test-retest reliability data for the SL Wall Sit and Triple Hop Jump tests in Sanda or comparable combat-sport populations are limited; without a known typical error or coefficient of variation for these tests in this context, the observed group-level Wall Sit asymmetry increase ($dz = 0.28$) and individual-level changes cannot be confidently attributed to genuine neuromuscular adaptation rather than measurement variation. Tenth, the unusually large standardized effect sizes for the composite indices ($dz > 1.0$) likely reflect, at least in part, the reduced variance of a multi-item average rather than an unusually large true effect on any single physical quality, and should be interpreted alongside the more modest item-level effect sizes reported in Table 4.

Sugesstions

Future work should: (1) implement a controlled or multi-squad comparison design (e.g., a non-training-matched comparison group, or a longer time series with more than two monitoring points) to permit causal attribution of the observed changes; (2) extend to at least 4-8 monitoring sessions to model individual trajectories and within-athlete variability across a full training cycle; (3) integrate IMU-derived load and strike kinematics (Liu et al., 2024; Pezenka & Wirth, 2025; Xue et al., 2025) and HR/HRV-based readiness monitoring (Addleman et al., 2024; Nakamura et al., 2022); (4) implement prospective injury surveillance to validate asymmetry interpretation (Guan et al., 2022; Guan et al., 2023); (5) update and externally validate benchmark norms using current national-level Sanda athlete data, ideally with input from the national federation; (6) incorporate body-height-adjusted hop metrics alongside percentage asymmetry (Khoygani & Esmaeili, 2025) and develop squad-specific asymmetry percentile norms; (7) estimate test-retest reliability, minimal detectable change, and smallest worthwhile change through a dedicated reliability sub-study; (8) capture structured reasons for any missing test entries; and (9) evaluate coach usability through structured qualitative or mixed-method procedures.

CONCLUSION

This study evaluated a field-test-based monitoring framework across a squad of 72 Sanda Wushu athletes. Composite performance indices increased significantly from Session 1 to Session 2, with large standardized effect sizes for both the full sample (PerfIndex: $dz = 1.35$, Hedges' $g = 1.34$, $p < .001$) and the missing-data-adjusted sensitivity sample ($n = 68$: $dz = 1.41$, $p < .001$). SL Wall Sit asymmetry showed a small but statistically significant increase, while Triple Hop asymmetry showed no significant change. These findings support the feasibility and statistical sensitivity of the field-test pipeline at squad scale. Still, the single-group observational design means they should be read as a dependable monitoring signal rather than confirmatory evidence of training effectiveness. The benchmark-normalization pipeline, composite PerfIndex, ScoreIndex, and inter-limb asymmetry calculations provide a reproducible, squad-scale monitoring workflow for coach-facing use in Sanda Wushu, given the benchmark thresholds, scoring syntax, and missing-data rules documented in the Supplementary Materials. Asymmetry monitoring showed that elevated Wall Sit disparity was common across the squad rather than confined to isolated cases, indicating a need for squad-specific reference distributions alongside individual-level follow-up for outlier athletes. The next empirical step is a controlled or multi-cycle repeated-measures study that combines this field-test base layer with IMU-derived load data, HR/HRV readiness monitoring, formal reliability estimates, prospective injury surveillance, and coach-usability evaluation.

AUTHOR CONTRIBUTION STATEMENT

Novita (N) conceptualized the study, supervised the research workflow, and finalized the manuscript. Ying Lei (YL) conducted data processing and statistical analysis (including paired t-tests and sensitivity analysis) and prepared all tables and figures. Wan Ahmad Jaafar Wan Yahaya (WA) implemented field-testing procedures, managed data collection, and curated the dataset. Tansa Trisna Astono Putri (TT) contributed to instrumentation planning, interpretation of monitoring indicators, literature synthesis, and critical revision of the manuscript. All authors reviewed and approved the final manuscript.

AI DISCLOSURE STATEMENT

The authors used QuillBot (an AI-powered paraphrasing tool) to edit selected sentences to improve the academic register. This tool was used exclusively for language clarity and did not contribute to any data collection, analysis, interpretation, or scientific content. All content was thoroughly reviewed and edited by the authors, who take full responsibility for the accuracy, integrity, and originality of the publication.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest — financial, institutional, or personal — that could have influenced this study's conduct, analysis, manuscript preparation, or publication. No specific funding from commercial, government, or non-profit agencies was received for this research.

REFERENCES

Abt, G., Boreham, C., Davison, G., Jackson, R., Jobson, S., Wallace, E., & Williams, A. M. (2025). Sample size estimation revisited. *Journal of Sports Sciences*, 43(21), 2511-2516. https://doi.org/10.1080/02640414.2025.2499403

Addleman, J. S., Lackey, J. R., DeBlauw, J., & Hajduczok, A. (2024). Heart rate variability applications in strength and conditioning: A narrative review. *Journal of Functional Morphology and Kinesiology*, 9(2), 93. https://doi.org/10.3390/jfmk9020093

Asimakidis, N. D., Bishop, C. J., Beato, M., Mukandi, I. N., Kelly, A. L., Weldon, A., & Turner, A. N. (2024). A survey into the current fitness testing practices of elite male soccer practitioners: From assessment to communicating results. *Frontiers in Physiology*, 15, 1376047. https://doi.org/10.3389/fphys.2024.1376047

Bickley, R. J., Hazim, N. Y., Sy, J. W., Nute, D. W., & Lafferty, P. M. (2023). An epidemiological study of martial arts injuries in patients presenting to US emergency rooms. *Injury*, 54(12), 111089. https://doi.org/10.1016/j.injury.2023.111089

Bishop, C., Read, P., Bromley, T., Brazier, J., Jarvis, P., Chavda, S., & Turner, A. (2022). The association between interlimb asymmetry and athletic performance tasks: A season-long study in elite academy soccer players. *Journal of Strength and Conditioning Research*, 36(3), 787-795. https://doi.org/10.1519/JSC.0000000000003526

Ceylan, B., Simenko, J., & Balci, S. S. (2022). Which performance tests best define the Special Judo Fitness Test classification in elite judo athletes? *Journal of Functional Morphology and Kinesiology*, 7(4), 101. https://doi.org/10.3390/jfmk7040101

D'Hondt, J., Chapelle, L., Bishop, C., Aerenhouts, D., De Pauw, K., Clarys, P., & D'Hondt, E. (2024). Association between inter-limb asymmetry and determinants of middle- and long-distance running performance in healthy populations: A systematic review. *Sports Medicine - Open*, 10, 127. https://doi.org/10.1186/s40798-024-00790-w

Dudley, C., Johnston, R., Jones, B., Till, K., Westbrook, H., & Weakley, J. (2023). Methods of monitoring internal and external loads and their relationships with physical qualities, injury, or illness in adolescent athletes: A systematic review and best-evidence synthesis. *Sports Medicine*, 53, 1559-1593. https://doi.org/10.1007/s40279-023-01844-x

Eltifi-Ghanmi, S., Amara, S., & Mkaouer, B. (2025). Kinetic and kinematic analysis of three kicks in Sanda Wushu. *AIMS Biophysics*, 12(2), 174-196. https://doi.org/10.3934/biophy.2025011

Fort-Vanmeerhaeghe, A., Mila-Villarroel, R., Pujol-Marzo, M., Arboix-Alio, J., & Bishop, C. (2022). Higher vertical jumping asymmetries and lower physical performance are indicators of increased injury incidence in youth team-sport athletes. *Journal of Strength and Conditioning Research*, 36(8), 2204-2211. https://doi.org/10.1519/JSC.0000000000003828

Goncalves, A. F., Miarka, B., Mauricio, C. A., Teixeira, R. P. A., Brito, C. J., Valenzuela Perez, D. I., Slimani, M., Znazen, H., Bragazzi, N. L., & Reis, V. M. (2024). Enhancing performance: Unveiling the physiological impact of submaximal and supramaximal tests on mixed martial arts athletes in the -61 kg and -66 kg weight divisions. *Frontiers in Physiology*, 14, 1257639. https://doi.org/10.3389/fphys.2023.1257639

Guan, Y., Bredin, S. S. D., Taunton, J., Jiang, Q., Wu, N., & Warburton, D. E. R. (2022). Association between inter-limb asymmetries in lower-limb functional performance and sport injury: A systematic review of prospective cohort studies. *Journal of Clinical Medicine*, 11(2), 360. https://doi.org/10.3390/jcm11020360

Guan, Y., Bredin, S. S. D., Taunton, J., Jiang, Q., Wu, N., & Warburton, D. E. R. (2023). Predicting the risk of injuries through assessments of asymmetric lower limb functional performance: A prospective study of 415 youth Taekwondo athletes. *Orthopaedic Journal of Sports Medicine*, 11(9), 23259671231185586. https://doi.org/10.1177/23259671231185586

Hecksteden, A., Forster, S., Egger, F., Buder, F., Kellner, R., & Meyer, T. (2022). Dwarfs on the shoulders of giants: Bayesian analysis with informative priors in elite sports research and decision making. *Frontiers in Sports and Active Living*, 4, 793603. https://doi.org/10.3389/fspor.2022.793603

Huang, Y., Li, Y., Su, S., Wang, D., Huang, C., Bi, N., & Cao, Z. (2023). Effects of core stability training on balance, strength, and sports performance of combat sport athletes: A systematic review. *Frontiers in Psychology*, 14, 1021080. https://doi.org/10.3389/fpsyg.2023.1021080

Impellizzeri, F. M., Shrier, I., McLaren, S. J., Coutts, A. J., McCall, A., Slattery, K., Jeffries, M., & Kalkhoven, J. T. (2023). Understanding training load as exposure and dose. *Sports Medicine*, 53(7), 1383-1392. https://doi.org/10.1007/s40279-022-01833-8

Jordan, M. J., & Bishop, C. (2024). Testing limb symmetry and asymmetry after anterior cruciate ligament injury: Four considerations to increase its utility. *Strength & Conditioning Journal*, 46(4), 406-414. https://doi.org/10.1519/SSC.0000000000000821

Keogh, J. W. L., Winwood, P. W., Rathbone, E. N., & King, N. A. (2023). Monitoring lower limb biomechanical asymmetry and psychological measures in athletic populations: A scoping review. *Scandinavian Journal of Medicine & Science in Sports*, 33(8), 1289-1307. https://doi.org/10.1111/sms.14460

Khoygani, M. R., & Esmaeili, H. (2025). Hop distance to body height ratio outperforms limb symmetry index in predicting second anterior cruciate ligament injury. *Knee Surgery, Sports Traumatology, Arthroscopy*, 33(3), 878-887. https://doi.org/10.1002/ksa.12405

Kovoor, M., Durairaj, M., Karyakarte, M. S., Zair Hussain, M., Ashraf, M., & Maguluri, L. P. (2024). Sensor-enhanced wearables and automated analytics for injury prevention in sports. *Measurement: Sensors*, 32, 101054. https://doi.org/10.1016/j.measen.2024.101054

Liu, Z., Yang, M., Li, K., & Qin, X. (2024). Recognition of TaeKwonDo kicking techniques based on accelerometer sensors. *Heliyon*, 10(12), e32475. https://doi.org/10.1016/j.heliyon.2024.e32475

Maestroni, L., Turner, A., Papadopoulos, K., Sideris, V., & Read, P. (2023). Total score of athleticism: Profiling strength and power characteristics in professional soccer players after anterior cruciate ligament reconstruction to assess readiness to return to sport. *American Journal of Sports Medicine*, 51(12), 3121-3130. https://doi.org/10.1177/03635465231194778

McKay, A. K., Stellingwerf, T., Smith, E. S., Martin, D. T., Mujika, I., Goosey-Tolfrey, V. L., Sheppard, J., & Burke, L. M. (2022). Defining training and performance caliber: A participant classification framework. *International Journal of Sports Physiology and Performance*, 17(2), 317-331. https://doi.org/10.1123/ijspp.2021-0451

Mesquida, C., Murphy, J., Lakens, D., & Warne, J. (2022). Replication concerns in sports and exercise science: A narrative review of selected methodological issues in the field. *Royal Society Open Science*, 9(12), 220946. https://doi.org/10.1098/rsos.220946

Nakamura, M., Miura, K., Premkumar, P., Nakagawa, T., Suzuki, S., & Maffiuletti, N. A. (2022). Reliability and sensitivity of nocturnal heart rate and heart-rate variability in monitoring individual responses to training load. *International Journal of Sports Physiology and Performance*, 17(8), 1296-1302. https://doi.org/10.1123/ijspp.2022-0145

Pezenka, L., & Wirth, K. (2025). Reliability of a low-cost inertial measurement unit (IMU) to measure punch and kick velocity. *Sensors*, 25(2), 307. https://doi.org/10.3390/s25020307

Preatoni, E., Bergamini, E., Fantozzi, S., Giraud, Q., Orejel Bustos, M., Vannozzi, G., & Camomilla, V. (2022). The use of wearable sensors for preventing, assessing, and informing recovery from sport-related musculoskeletal injuries: A systematic scoping review. *Sensors*, 22(9), 3225. https://doi.org/10.3390/s22093225

Sliwowski, R., Paillard, T., Bojkowski, L., Dudzinski, W., Patek, M., & Marynowicz, J. (2024). Intra- and inter-limb strength imbalance and asymmetry in soccer: A comparison of elite senior and junior players. *PLOS ONE*, 19(4), e0302474. https://doi.org/10.1371/journal.pone.0302474

Taati, B., Arazi, H., Bridge, C. A., & Franchini, E. (2022). A new taekwondo-specific field test for estimating aerobic power, anaerobic fitness, and agility performance. *PLOS ONE*, 17(3), e0264910. https://doi.org/10.1371/journal.pone.0264910

Teresi, J. A., Yu, X., Stewart, A. L., & Hays, R. D. (2022). Guidelines for designing and evaluating feasibility pilot studies. *Medical Care*, 60(1), 95-103. https://doi.org/10.1097/MLR.0000000000001664

Totton, N., Lin, J., Julious, S., Chowdhury, M., & Brand, A. (2023). A review of sample sizes for UK pilot and feasibility studies on the ISRCTN registry from 2013 to 2020. *Pilot and Feasibility Studies*, 9, 188. https://doi.org/10.1186/s40814-023-01416-w

Wang, G., Li, H., & Huang, L. (2025). Predicting the punching force in Wushu Sanda after neuromuscular electrical stimulation by employing the KAN neural network combined with neuromuscular electricity. *Sensors*, 25(19), 5979. https://doi.org/10.3390/s25195979

Weakley, J., Black, G., McLaren, S., Scantlebury, S., Suchomel, T. J., McMahon, E., Watts, D., & Read, D. B. (2024). Testing and profiling athletes: Recommendations for test selection, implementation, and maximizing information. *Strength & Conditioning Journal*, 46(2), 159-179. https://doi.org/10.1519/SSC.0000000000000784

Xue, H., Han, C., & Zhu, D. (2025). Limb biomechanics in combat sports: Insights from wearable sensor technology. *Frontiers in Bioengineering and Biotechnology*, 13, 1663592. https://doi.org/10.3389/fbioe.2025.1663592

Xu, Q., Wang, Z., Liu, L., Mao, R., Xi, C., & Shi, G. (2025). The relationship between training and competition load characteristics of female boxers: A case study based on collegiate boxers. *Frontiers in Psychology*, 16, 1517102. https://doi.org/10.3389/fpsyg.2025.1517102

Zhang, S., Huang, W., Soh, K. G., Luo, S., Li, L., & Wang, X. (2025). Effects of core strength training on the technical skill performance of striking combat sport players: A systematic review. *PeerJ*, 13, e19615. https://doi.org/10.7717/peerj.19615