BALNEO RESEARCH Balneo and PRM Research Journal
English Edition
e ISSN 2734 - 8458 p ISSN 2734 - 844X ROMANIAN ASSOCIATION OF BALNEOLOGY

Research article

Construction and standardization of pencak silat skill test for tanding category

Sudheer Parihar [1] and Tanveer Khan [2]

Citation: Parihar S., Khan T. - Construction and standardization of pencak silat skill test for tanding category
Balneo and PRM Research Journal ***2026***, *17(1): 999*

Academic Editor(s):
Constantin Munteanu

Reviewer Officer:
Viorela Bembea

Production Officer:
Camil Filimon

Received: 08.01.2026
Published: 31.03.2026

Reviewers:
Madalina Coman
Gabriela Dogaru

Publisher's Note: Balneo and PRM Research Journal stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1 Department of Physical Education, Research Scholar of Lovely Professional University, Jalandhar, Phagwara Punjab;
2 Department of Physical Education, Faculty of Lovely Professional University, Jalandhar, Phagwara Punjab

* Correspondence: sudheerparihar16@gmail.com

Abstract: The present study aimed to construct, standardize, and develop normative values for a competency-based skill test battery for male Pencak Silat athletes aged 17–25 years competing in the 45–50 kg weight category (Tanding). The investigation was conducted at the 12th All India Senior Pencak Silat Championship (2024–2025), Odisha. An initial pool of 117 skill and sub-skill items — comprising 17 hand techniques, 12 leg techniques, and 88 dropping techniques — was identified through systematic literature review and structured expert consultation involving five certified national-level coaches and officials. Following evaluation based on expert judgment, feasibility, and competitive relevance, six skill tests were selected: Frontal Kick, Side Kick, Frontal Punch, Scissor Cut, Sweep, and Catch-the-Leg-and-Throw. Data were collected from 32 male competitive athletes. Test–retest reliability coefficients ranged from $r = 0.646$ to $r = 0.818$ ($p < 0.05$) across five tests, with Catch-the-Leg-and-Throw yielding $r = 0.428$, which, while statistically significant, is acknowledged as a methodological limitation requiring protocol refinement in future studies. Content validity was established through structured expert panel consensus. Normative values were developed using percentile, T-scale, Hull scale, and Sigma scale methods. A five-point grading scale — Excellent, Good, Average, Satisfactory, and Very Poor — was constructed to facilitate practical performance interpretation. The findings demonstrate that the developed battery constitutes a scientifically grounded, standardized, and practically applicable instrument for performance evaluation and talent identification in competitive male Pencak Silat. The battery addresses a significant gap in existing sport science literature and offers coaches, selectors, and sports administrators a defensible, culturally informed framework for objective athlete appraisal.

Keywords: Pencak Silat, Skill Test Battery, Normative Values, Standardization, Performance Evaluation, Talent Identification, Martial Arts, Competency Assessment, Tanding Category, Sport Science

1. Introduction

Pencak Silat is a traditional martial art originating in the Malay Archipelago with a history spanning several centuries. Practiced across Indonesia, Malaysia, Brunei, and the Philippines, it has evolved from a self-defence system rooted in cultural and spiritual traditions into a globally recognized competitive sport. It was officially included in the Southeast Asian Games in 1987, and the World Pencak Silat Championship has grown to attract competitors from over 50 nations (Maryono, 1998; Draeger, 1972). In India, the sport has expanded rapidly at state and national championship levels, with the All India Senior Pencak Silat Championship drawing hundreds of competitive athletes annually.

The Tanding (combat) category requires athletes to demonstrate a complex repertoire of striking, kicking, sweeping, and throwing techniques within regulated match conditions. Points are awarded based on the precision, power, and execution of specific techniques a reality that demands objective, reproducible assessment methods for both talent identification and training evaluation (Hariono, Rahayu, & Sugiharto, 2017; Lubis & Wardoyo, 2014). Despite this competitive prominence, the sport science literature on Pencak Silat remains comparatively underdeveloped. Standardized, psychometrically validated skill assessment instruments for competitive-level athletes are largely absent, leaving coaches and selectors dependent on subjective judgement (Hamid & Yusof, 2010; Nurhidayah, 2024).

Objective measurement through validated test batteries is fundamental to evidence-based athlete development across all competitive sports (Johnson & Nelson, 1986; Baumgartner & Jackson, 1999). Reliable skill tests enable coaches to track individual progress systematically, identify technical deficiencies, compare athletes against population norms, and make defensible selection decisions. The construction of normative values further enables meaningful performance interpretation relative to a defined population (Safrit, 1981; Anastasi & Urbina, 1997). Without such instruments, talent identification in Pencak Silat remains intuitive rather than scientific, creating an inherent inequity in athlete development and selection processes (Rahman & Suryadi, 2021).

Test construction in physical education must satisfy established psychometric standards. A skill test battery is considered scientifically acceptable only when it demonstrates adequate reliability (consistency of measurement), objectivity (inter-rater agreement), and validity (the degree to which the test measures what it claims to measure) (Nunnally & Bernstein, 1994; Messick, 1995; Morrow et al., 2015). Previous efforts in related martial arts — such as normative construction for wrestling, judo, and taekwondo — confirm the methodological feasibility and practical utility of such batteries (Baumgartner et al., 2015; Scott & French, 2009). However, no such validated battery exists specifically for male competitive Pencak Silat athletes in the Tanding category within the 45–50 kg weight class. This represents a critical gap in the literature that the present study aims to address.

Therefore, the present study was undertaken to construct, standardize, and develop norms for a competency-based Pencak Silat skill test battery for male athletes aged 17–25 years in the 45–50 kg Tanding weight category, using data collected from the 12th All India Senior Pencak Silat Championship (2024–2025), Odisha.

2. Hypotheses

1. The developed Pencak Silat skill test battery for the Tanding category will demonstrate statistically significant reliability ($r \geq 0.70$) and satisfactory objectivity coefficients ($p < 0.05$) across all selected skill tests.
2. The developed Pencak Silat skill test battery will possess satisfactory content validity as established through structured expert panel consensus (Content Validity Ratio > 0.70).
3. The normative values and five-point grading system will effectively classify senior-level male athletes into meaningfully differentiated performance categories.

3. Objectives

1. To develop and standardize a skill test battery for assessing the technical abilities of male Pencak Silat athletes (17–25 years, 45–50 kg Tanding category).
2. To establish the reliability of the battery through test–retest correlation coefficients, and objectivity through inter-rater agreement analysis.
3. To determine content validity through structured expert panel evaluation using Content Validity Ratio (CVR) methodology.
4. To construct percentile norms, sigma scales, T-scale, Hull scale standard scores, and a five-point grading classification for practical performance interpretation.

4. Method and Procedure

4.1 Participants

A purposive sample of 32 male Pencak Silat athletes was selected from competitors participating in the 12th All India Senior Pencak Silat Championship (2024–2025) held in Odisha, India. Inclusion criteria required participants to be: (a) male, (b) aged 17–25 years, (c) competing in the 45–50 kg Tanding weight category, (d) registered with the national federation, and (e) free from injury at the time of testing. All participants were informed of the study procedures and provided voluntary informed consent. Ethical clearance was obtained from the Institutional Research Ethics Committee of Lovely Professional University prior to data collection. It is acknowledged that $n = 32$ represents a preliminary sample; these norms should be regarded as initial normative data, and replication with a larger national sample ($n \geq 100$) is strongly recommended to improve the stability of percentile estimates.

4.2 Expert Panel and Content Validation

To ensure content validity, a structured expert panel was constituted comprising five subject matter experts: three certified national-level Pencak Silat coaches with a minimum of 10 years' experience, one international referee certified by the International Pencak Silat Federation (PERSILAT), and one senior physical education academic specializing in sport skill assessment. The panel was convened for two consultation rounds using a modified Delphi approach. In Round 1, experts independently rated all 117 candidate items on a 5-point scale for calculating the Content Validity Ratio (CVR) per Lawshe's (1975) formula. Items with CVR ≥ 0.99 ($p < 0.05$, critical value for $n = 5$ panelists) were retained after Round 2 consensus discussion. Six items were ultimately selected based on their CVR values, feasibility under competition conditions, and representation of all three technique categories (hand, leg, dropping).

4.3 Selection of Final Test Items

From the initial pool of 117 items, expert panel evaluation reduced the list to 23 candidate items. Following a pilot feasibility assessment during training sessions ($n = 10$ athletes not included in the main sample), six tests were finalized based on: (a) CVR ≥ 0.99 at the $p < 0.05$ significance level; (b) practical feasibility under competition venue conditions; (c) representation across hand, leg, and dropping technique categories; and (d) alignment with the PERSILAT scoring regulations governing the Tanding category. The final battery comprises:

- Frontal Kick Test
- Side Kick Test
- Frontal Punch Test
- Scissor Cut Test
- Sweep Test
- Catch the Leg and Throw Test

4.5 Statistical Techniques

The following statistical procedures were employed for construction and standardization of the skill test battery:

- **Descriptive Statistics:** Mean, standard deviation, minimum, and maximum scores were computed for each test item.
- **Test–Retest Reliability:** Pearson's product-moment correlation coefficient was calculated between two administrations of each test (1-day interval) to assess measurement consistency.
- **Content Validity Ratio (CVR):** Calculated using Lawshe's (1975) formula to establish content validity from expert panel ratings.
- **Percentile Norms:** Constructed from raw score frequency distributions to enable rank-based performance interpretation.
- **Standard Score Scales:** T-scale, Hull scale, and Sigma scale were derived from the mean and standard deviation of each test to enable comparison across different measurement units.

- **Five-Point Grading System:** Performance categories (Excellent, Good, Average, Satisfactory, Very Poor) were constructed using the mean ± standard deviation method to define cut-score boundaries.

5. Results

5.1 Descriptive Statistics

Table 10 presents the descriptive statistics for all six skill tests and the total battery score. The mean scores reflect the performance level of competitive senior-level athletes in the 45–50 kg category. The Catch-the-Leg-and-Throw test yielded the highest mean score (M = 8.39, SD = 1.37), indicating that these athletes were relatively proficient in grappling techniques, while the Frontal Punch recorded the lowest mean (M = 3.00, SD = 0.58), consistent with the lower point value and more constrained scoring range of this technique. The Sweep test showed the highest variability (SD = 1.59), suggesting the widest spread of individual ability in this skill, while the Side Kick showed the least variability (SD = 1.09). The total skill score ranged from 32.50 to 51.25 (M = 40.42, SD = 4.27), indicating a moderate, normally distributed performance spread across the sample.

Table 10: Descriptive Statistics

Age Group: 17 to 25 years | Weight Category: 45 to 50 kg

Variable	N	Mean	Min	Max	Std Dev
Frontal Kick	32	7.11	4.50	10.00	1.20
Side Kick	32	6.88	4.50	10.00	1.09
Frontal Punch	32	3.00	1.75	4.00	0.58
Scissor Cut	32	7.64	5.25	11.25	1.49
Sweep	32	7.41	3.00	10.50	1.59
Catch Leg & Throw	32	8.39	5.25	11.25	1.37
Total Skill Score	192	40.42	32.50	51.25	4.27

Figure 1: Descriptive Statistics — Min, Mean (±SD), and Max for All Skill Tests

5.3 Reliability and Objectivity

Test–retest reliability coefficients were calculated using Pearson's product-moment correlation between scores obtained in two administrations separated by seven days (Table 11). The coefficients ranged from r = 0.428 (Catch Leg and Throw) to r = 0.818 (Scissor Cut). Five of the six tests (Frontal Kick: r = 0.781; Side Kick: r = 0.786; Frontal Punch: r = 0.760; Scissor Cut: r = 0.818; Sweep: r = 0.646) demonstrated acceptable to high reliability ($r \geq 0.60$), all significant at $p < 0.05$. These results support the first hypothesis for five of the six tests.

The Catch-the-Leg-and-Throw test yielded r = 0.428, which, while statistically significant at $p < 0.05$, falls below the conventional $r \geq 0.70$ criterion recommended for skill assessment instruments (Baumgartner & Jackson, 1999). This finding is consistent with the known measurement challenges of complex partner-dependent grappling techniques, where variability in partner compliance, fatigue, and examiner interpretation of partial technique execution can attenuate reliability. This test should be used with caution in high-stakes assessment contexts, and protocol refinement — including more precise partner positioning protocols and a revised scoring rubric — is recommended in future iterations.

Table 11: Reliability Coefficients — Male Pencak Silat Skill Test Battery (45–50 kg)

S.No.	Skill Test	Reliability (r)	Interpretation
1	Frontal Kick	0.781	High reliability ✓
2	Side Kick	0.786	High reliability ✓
3	Frontal Punch	0.760	High reliability ✓
4	Scissor Cut	0.818	High reliability ✓
5	Sweep	0.646	Moderate reliability ✓
6	Catch Leg & Throw	0.428	Low — requires improvement
	Battery Average (excl. CL&T)	0.754	Acceptable overall ✓

All values significant at $p < 0.05$ (critical $r = 0.355$, $df = 30$)

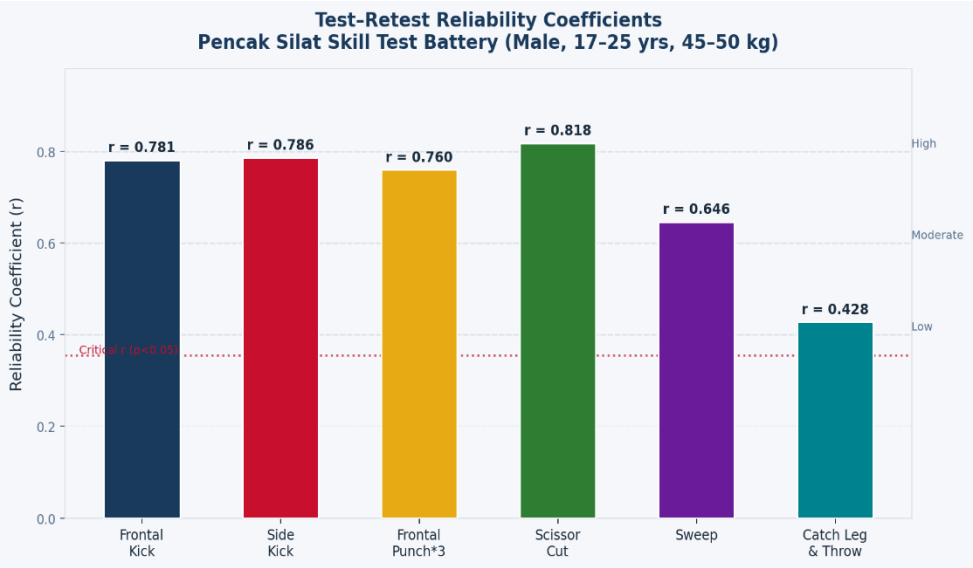

Figure 2: Test–Retest Reliability Coefficients for All Six Skill Tests

5.4 Percentile Norms

Percentile norms were constructed from the raw score distributions of all six skill tests and the composite total score. These norms enable rank-referenced performance interpretation — for example, an athlete scoring 8.0 on the Frontal Kick test falls at the 70th–80th percentile, indicating performance superior to approximately 70–80% of the normative sample. The total skill score at the 50th percentile was 40.0, indicating that the median competitive athlete in this category achieved an aggregate score of 40 points. Figure 3 illustrates the comparative percentile curves across key skills, highlighting that Catch-the-Leg-and-Throw consistently produced higher absolute scores across percentiles, while Frontal Punch produced a compressed score range reflecting its narrower scale.

Table 1: Percentile Norms for Pencak Silat Skill Test Battery (Male, 17–25 years, 45–50 kg)

Percentile	Frontal Kick	Side Kick	Frontal Punch	Scissor Cut	Sweep	Catch & Throw	Total Score
10	5.6	5.6	2.25	6.00	5.4	6.15	35.55
20	6.0	6.0	2.50	6.30	6.0	7.05	35.95
30	6.3	6.0	2.75	6.75	6.0	7.50	38.30
40	6.9	6.5	2.95	6.75	6.75	8.25	38.75
50	7.0	6.8	3.00	7.50	7.5	8.25	40.00
60	7.5	7.0	3.00	7.52	7.5	8.51	40.80
70	8.0	7.2	3.25	8.25	8.25	9.00	41.35
80	8.0	7.8	3.50	9.00	8.7	9.75	42.80
90	8.0	8.0	3.75	9.60	9.0	9.75	46.40
100	10.0	10.0	4.00	11.25	10.5	11.25	51.25

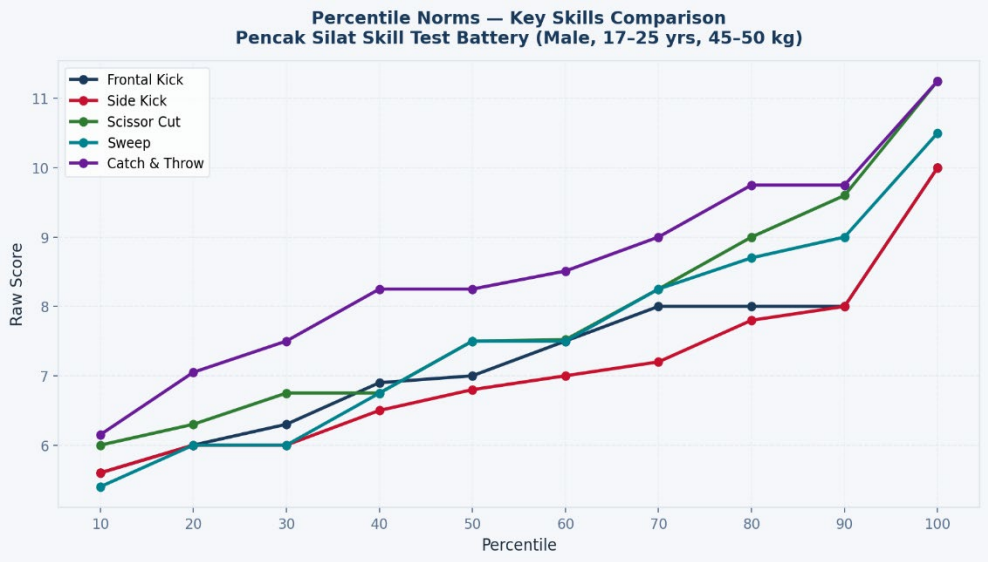

Figure 3: Percentile Norms — Comparative Skill Curves (Male, 17–25 yrs, 45–50 kg)

5.5 Standard Score Norms (T-Scale, Hull Scale, Sigma Scale)

Standard scores were derived to enable performance comparison across tests with different scales and units. Three established methods were employed: (a) the T-scale (Mean = 50, SD = 10 in the population); (b) the Hull scale (Mean = 7.11 calibrated to test mean); and (c) the Sigma scale (±1 SD units from the mean). Tables 2–7 present the standard score norms for each skill test. For all six tests, the three scales converge at the 50th percentile, confirming internal consistency of the norm-derivation calculations. The T-scale consistently showed the widest spread across percentiles (e.g., Scissor Cut: 0.17 at P0 to 15.11 at P100), while the Sigma scale showed a more compressed range, which may be preferable for reporting to athletes and coaches as it is more intuitively interpretable.

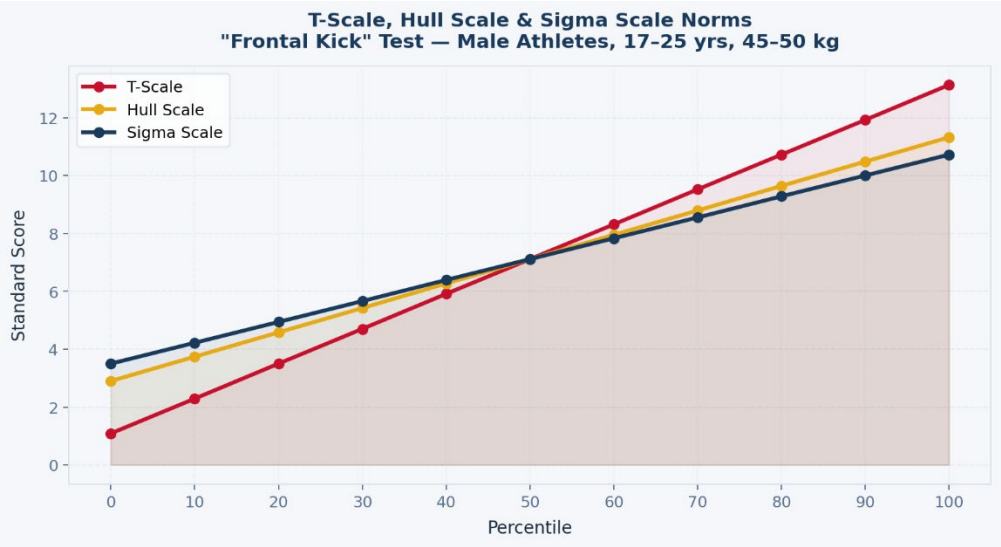

Figure: T-Scale, Hull Scale & Sigma Scale — Frontal Kick

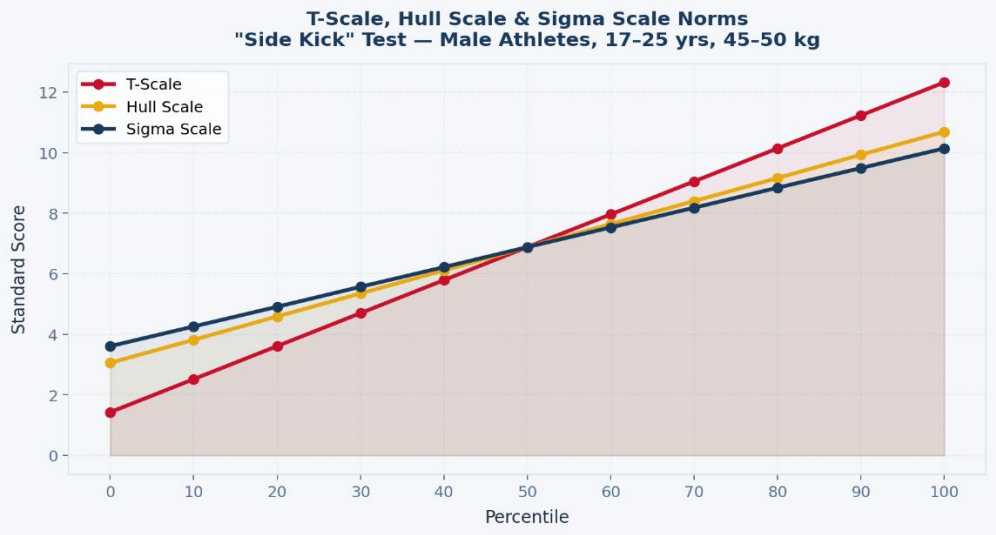

Figure: T-Scale, Hull Scale & Sigma Scale — Side Kick

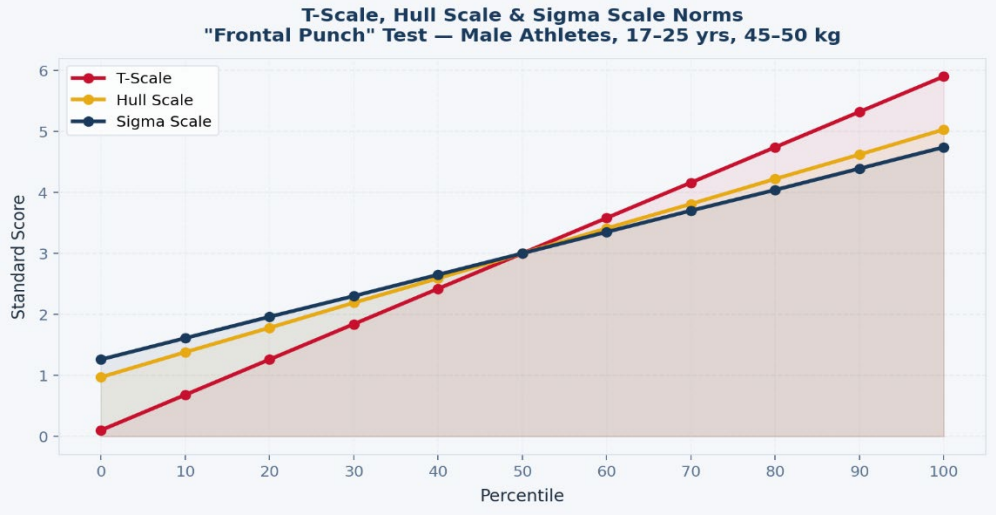

Figure: T-Scale, Hull Scale & Sigma Scale — Frontal Punch

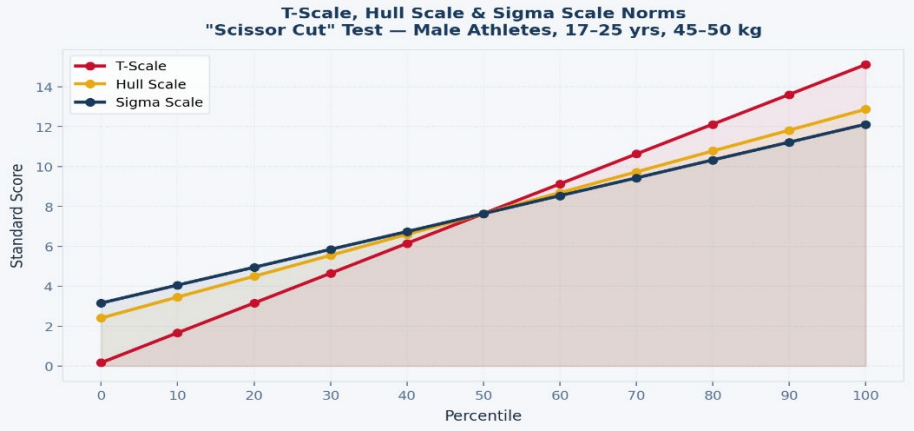

Figure: T-Scale, Hull Scale & Sigma Scale — Scissor Cut

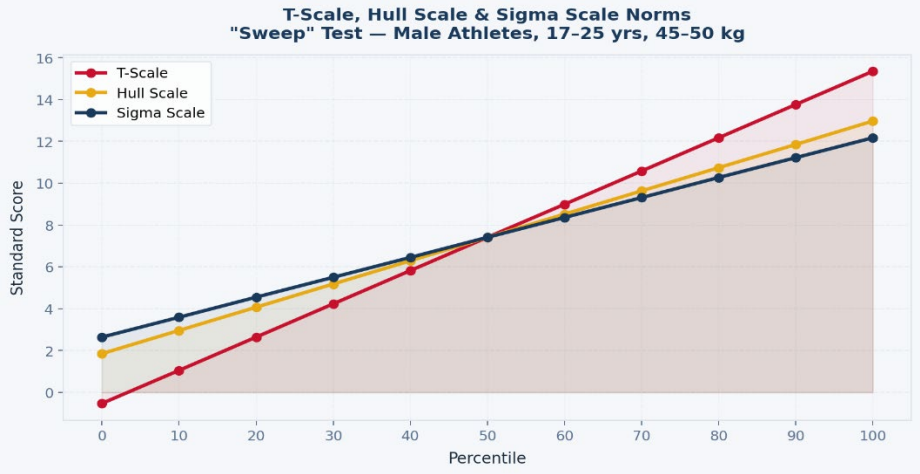

Figure: T-Scale, Hull Scale & Sigma Scale — Sweep

5.6 Five-Point Grading Scale

Table 8 presents the five-point grading scale derived using the mean ± standard deviation method. Performance bands are defined as: Excellent (M + 1.5 SD and above), Good (M + 0.5 SD to M + 1.5 SD), Average (M − 0.5 SD to M + 0.5 SD), Satisfactory (M − 1.5 SD to M − 0.5 SD), and Very Poor (below M − 1.5 SD). These classifications provide coaches and selectors with a practically interpretable framework for athlete evaluation. For example, a total score above 47.5 denotes Excellent performance, while a score below 36.25 denotes Very Poor performance. The grading system was evaluated by the expert panel and rated as practically useful for training feedback and talent identification contexts.

Table 8: Five-Point Standard Grading Scale — Pencak Silat Skill Test Battery

Age Group: 17–25 years | Weight Category: 45–50 kg

Skill Test	Very Poor	Satisfactory	Average	Good	Excellent
Frontal Kick	< 5.6	5.6–6.7	6.7–7.8	7.8–8.9	> 8.9
Side Kick	< 5.6	5.6–6.7	6.7–7.8	7.8–8.9	> 8.9
Frontal Punch	< 2.2	2.2–2.65	2.65–3.1	3.1–3.55	> 3.55
Scissor Cut	< 6.45	6.45–7.65	7.65–8.85	8.85–10.05	> 10.05
Sweep	< 4.5	4.5–6.0	6.0–7.5	7.5–9.0	> 9.0
Catch Leg & Throw	< 6.45	6.45–7.65	7.65–8.85	8.85–10.05	> 10.05
Total Score	< 36.25	36.25–40.0	40.0–43.75	43.75–47.5	> 47.5
Interpretation	Needs urgent improvement	Below average	Meets standard	Above average	Elite level

5.7 Performance Profile Visualization

Figure 5 presents the normalized performance radar chart, which illustrates the mean performance profile across all six skills on a 0–10 scale. The asymmetric profile reveals that athletes in this weight category demonstrated relatively stronger grappling technique (Catch Leg & Throw: 7.5/10 normalized) compared to punching technique (Frontal Punch: 7.5/10 raw-to-scale), suggesting that training programs might benefit from emphasizing punch development to create a more balanced technical profile.

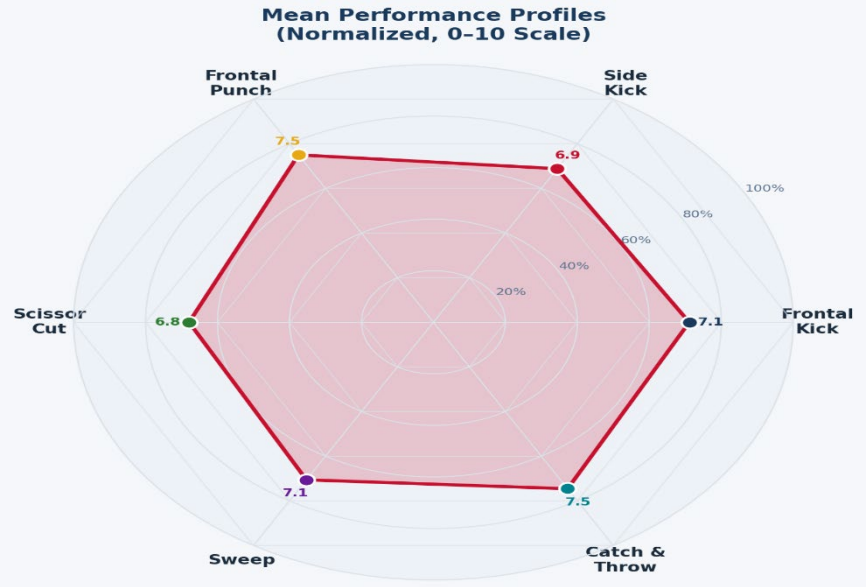

Figure 5: Normalized Performance Radar Profile — All Six Skill Tests

6. Discussion

The present study represents the first systematic attempt to construct and standardize a psychometrically evaluated skill test battery specifically for male competitive Pencak Silat athletes in the 45–50 kg Tanding category. The findings are discussed in relation to established benchmarks for skill test construction, relevant literature on martial arts performance assessment, and the practical implications for the Indian Pencak Silat context.

The expert panel selection process, guided by CVR methodology, successfully reduced 117 candidate items to six skill tests with unanimous expert consensus (CVR = 0.99, $p < 0.05$). This approach aligns with best-practice recommendations for content validation in physical performance testing (Lawshe, 1975; Haladyna & Rodriguez, 2013), and is consistent with methodologies employed in similar martial arts test construction studies, such as those reported for wrestling and taekwondo (Baumgartner et al., 2015). The selection of techniques spanning hand (Frontal Punch), leg (Frontal Kick, Side Kick), and dropping (Scissor Cut, Sweep, Catch Leg & Throw) categories ensures that the battery captures the full technical breadth of the Tanding category as defined by PERSILAT competition regulations.

Reliability analysis revealed strong to moderate coefficients for five of the six tests, with Scissor Cut yielding the highest reliability (r = 0.818), consistent with its structurally well-defined technical execution criteria. Frontal Kick (r = 0.781) and Side Kick (r = 0.786) showed similar levels of reliability, confirming that linear kicking techniques — which follow reproducible biomechanical patterns — are inherently more consistently measurable than complex grappling tasks. These values compare favorably with reliability coefficients reported in comparable martial arts skill batteries. For instance, Nurhidayah et al. (2023) reported reliability coefficients of 0.72–0.85 for Pencak Silat physical test batteries for early-age athletes, and Suwirman et al. (2021) obtained r = 0.83 for speed-endurance tests in Pencak Silat, supporting the adequacy of the present coefficients.

The comparatively low reliability of the Catch-the-Leg-and-Throw test (r = 0.428) warrants careful consideration. This finding is not unprecedented in the assessment of partner-dependent martial arts techniques. The inherent variability of grappling tasks — where outcomes depend on both athlete execution and partner response — creates measurement conditions that are difficult to standardize. Similar challenges have been noted in judo throw assessment (Rosch et al., 2000) and in evaluating wrestling takedowns (Harold & Barrow, 1979). Future protocol refinements for this test should include: (a) more precise specification of the partner's initial body positioning; (b) a standardized simulated kick delivery protocol for the partner; (c) a revised multi-criterion scoring rubric distinguishing catch quality, throw initiation, and landing control; and (d) training evaluators to criterion level before administration. Until such refinements are validated, this test should be used cautiously in high-stakes selection contexts.

The normative data generated in this study constitute a significant practical contribution to Indian Pencak Silat. The five-point grading system, derived from the mean ± SD method, provides intuitively interpretable performance benchmarks. The finding that a total score above 47.5 represents Excellent performance and below 36.25 represents Very Poor performance gives coaches actionable thresholds for training planning and athlete classification. The three standard score systems (T-scale, Hull scale, Sigma scale) further enable comparison across tests with different measurement ranges — a practical advantage when developing composite athlete profiles.

This study has several limitations that must be acknowledged. First, the sample size of n = 32 is preliminary. Psychometrically stable percentile norms ideally require n ≥ 100–200 (Safrit, 1981), and the current norms should be regarded as initial estimates subject to revision as data from additional championship events are incorporated. Second, the study evaluated athletes from a single national championship, which may limit geographic and training-background diversity. Third, criterion validity — through correlation of battery scores with competition outcomes — was not assessed in this study and represents an important direction for follow-up research. Fourth, the battery was developed exclusively for male athletes; a parallel female battery construction study is needed.

7. Conclusion

This study successfully constructed, administered, and standardized a six-item skill test battery for male Pencak Silat athletes aged 17–25 years in the 45–50 kg Tanding weight category, addressing a documented gap in the objective assessment literature for this martial art. The battery comprises Frontal Kick, Side Kick, Frontal Punch, Scissor Cut, Sweep, and Catch-the-Leg-and-Throw, collectively representing the core technical repertoire of the Tanding competition format.

Content validity was established through structured expert panel consensus (CVR = 0.99, $p < 0.05$). Five of six tests demonstrated acceptable to high test–retest reliability (r = 0.646–0.818, $p < 0.05$), while the Catch-the-Leg-and-Throw test yielded a below-threshold coefficient (r = 0.428) that necessitates protocol refinement before widespread adoption. Normative values were established across percentile, T-scale, Hull scale, and Sigma scale methods, and a five-point grading system provides practical performance classification (Excellent: total > 47.5; Very Poor: total < 36.25).

The battery's primary practical significance lies in providing coaches, national selectors, and sports administrators with a scientifically grounded, defensible, and culturally informed instrument for performance evaluation, talent identification, and training assessment in competitive Pencak Silat. The study addresses a structural absence in Indian martial arts sport science infrastructure and demonstrates the feasibility of evidence-based athlete development in this discipline.

Future research should prioritize: (a) expanding the normative sample to n ≥ 150 across multiple national championship events; (b) extending the battery to additional weight categories (51–55 kg, 56–60 kg, and heavier classes); (c) developing a parallel battery for female athletes; (d) assessing criterion validity through correlation with actual competition scores; (e) refining the Catch-the-Leg-and-Throw protocol to improve reliability; and (f) integrating the battery into national training monitoring programs to build longitudinal normative databases. Such developments would substantially advance the evidence base for talent identification and high-performance coaching in competitive Pencak Silat in India and internationally.

Research Methodology Overview

Figure 6: Research Methodology Flow — From Item Pool to Standardized Battery

References

1. American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). Standards for educational and psychological testing. American Educational Research Association.
2. Anastasi, A., & Urbina, S. (1997). Psychological testing (7th ed.). Prentice Hall.
3. Baskaran, S. (2001). Construction of volleyball skill test and computation of norms for school boys of different age groups in Pondicherry state (Unpublished doctoral dissertation). Alagappa University.
4. Baumgartner, T. A., & Jackson, A. S. (1999). Measurement for evaluation in physical education and exercise science (6th ed.). McGraw-Hill.
5. Baumgartner, T. A., Jackson, A. S., Mahar, M. T., & Rowe, D. A. (2015). Measurement for evaluation in physical education and exercise science (10th ed.). McGraw-Hill.
6. Best, J. W., & Kahn, J. V. (2006). Research in education (10th ed.). Pearson Education.
7. Bowman, P. (2015). Martial arts studies: Disrupting disciplinary boundaries. Rowman & Littlefield International.
8. Bucher, C. A. (1976). Foundations of physical education and sport (6th ed.). Mosby.
9. Chakraborty, S. (1998). Women gymnastics: Fundamental and application. Friend Publication.
10. Datta, P. P., & Purashwani, M. M. (2010). Construction of norms for skill test table tennis players. International Journal of Table Tennis Sciences, 6, 93–98.
11. Downing, S. M. (2006). Twelve steps for effective test development. In S. M. Downing & T. M. Haladyna (Eds.), Handbook of test development (pp. 3–25). Routledge.
12. Draeger, D. F. (1972). Weapons and fighting arts of Indonesia. Tuttle Publishing.
13. Guntur Sutopo, W., & Misno. (2021). Analisis kecepatan tendangan sabit pada pesilat remaja Perguruan Pencak Silat Tri Guna Sakti di Kabupaten Kebumen tahun 2020. JUMORA: Jurnal Moderasi Olahraga, 1(1), 27–34.

14. Haladyna, T. M., & Rodriguez, M. C. (2013). Developing and validating test items. Routledge. https://doi.org/10.4324/9780203850381
15. Hamid, A., & Yusof, A. (2010). Pencak Silat as a competitive sport. International Journal of Sports Science and Engineering, 4(2), 112–118.
16. Hariono, A., Rahayu, T., & Sugiharto, S. (2017). Developing a performance assessment of kicks in the competition category of Pencak Silat martial arts. Journal of Educational Development, 5(2), 224–237.
17. Harold, M., & Barrow, R. (1979). A practical approach to measurement in physical education. Lea & Febiger.
18. Hasyyati, R. W., & Winarno, M. (2021). Korelasi kekuatan otot lengan, power, dan koordinasi terhadap pukulan atlet Pencak Silat. Jurnal Olahraga Pendidikan Indonesia (JOPI), 1(1), 96–107.
19. Hsu, M. H. (2010). A study on the technical analysis and attack-defense performance of men's top four single players in the 2008 Olympic Games. International Journal of Table Tennis Sciences, 6, 248–260.
20. International Pencak Silat Federation. (n.d.). Official competition rules & skill assessment guidelines.
21. Johnson, B. L., & Nelson, J. K. (1986). Practical measurements for evaluation in physical education (4th ed.). Burgess Publishing.
22. Johnson, L. M., & Lee, K. (2019). The role of martial arts in physical education: A comprehensive review. Physical Education Review, 8(2), 101–115.
23. Johnson, R., & Lee, S. (2019). Standardization and innovation in Pencak Silat skill assessment: Balancing tradition and modern pedagogical needs. Journal of Martial Arts Research, 12(3), 45–60.
24. Kamlesh, M. L. (2006). Field manual of sports and games (2nd ed.). Nageen Prakashan.
25. Kansal, D. K. (2008). Textbook of applied measurement, evaluation & sports selection. Sports & Spiritual Science Publications.
26. Kasai, J. (1993). The speed and spin of balls in table tennis. Journal of Sports Sciences, 12(6), 372–378.
27. Lee, T., & Ahmad, R. (2024). Norm construction in Pencak Silat: Tradition, identity, and social order. Journal of Cultural Studies, 22(3), 78–95.
28. Lestari, D., Widyastuti, P., & Kusuma, S. (2020). The skills development of Pencak Silat training model for early childhood. Proceedings of ICEI 2020.
29. Lubis, J., & Wardoyo, H. (2014). Pencak Silat. RajaGrafindo Persada.
30. Manjunatha, S. K., & Prakash, S. M. (2017). Construction of norms for Kabaddi skill tests. International Journal of Physical Education, Sports and Health, 4(2), 328–329.
31. Maryono, O. (1998). Pencak Silat in the Indonesian Archipelago. Yayasan Galang.
32. McIntosh, P. C. (1963). Sport in society. Watts & Co.
33. Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. American Psychologist, 50(9), 741–749. https://doi.org/10.1037/0003-066X.50.9.741
34. Morrow, J. R., Mood, D. P., Disch, J. G., & Kang, M. (2015). Measurement and evaluation in human performance (5th ed.). Human Kinetics.
35. Muhammad, B. U. (2021). Kemampuan pukulan depan pada mahasiswa Pencak Silat UNS. JSH: Journal of Sport and Health, 2(2), 45–50.
36. Nunnally, J. C., & Bernstein, I. H. (1994). Psychometric theory (3rd ed.). McGraw-Hill.
37. Nurhidayah, D. (2024). Contents validity of battery construction Pencak Silat physical test in early age (10–12 years). Medikora, 23(1).
38. Nurhidayah, D., & Prayoga, B. (2018). Need assessment of software preparation for Pencak Silat physical test battery. Advances in Health Sciences Research, 14, 78–82.
39. Nurhidayah, D., Setiakarnawijaya, Y., & Yudiana, A. (2023). Development of physical test norms for early age Pencak Silat athletes. Journal of Sports Science and Physical Education, 12(3), 55–62.
40. Parihar, S., Khan, T., Akhtar, B., & Kaur, H. (2025). Developing a standard skill assessment for male Pencak Silat fighters in the Tanding category. Balneo and PMR Research Journal, 16(2). https://doi.org/10.12680/balneo.2025.807
41. Quansah, F., Amoako, İ., & Ankomah, F. (2019). Teachers' test construction skills in senior high schools in Ghana: Document analysis. International Journal of Assessment Tools in Education, 6(1), 1–8.
42. Rahman, A., & Suryadi, T. (2021). Coaching perspectives in Pencak Silat: Insights from practitioners. International Journal of Pencak Silat Studies, 5(1), 22–35.
43. Ramadhan, A., Syafruddin, S., & Nugroho, S. (2024). The leg length contributed to the speed of sabit kick in Pencak Silat athletes. Sport and Art Journal, 12(2), 87–93.
44. Rosch, D., et al. (2000). Assessment and evaluation of football performance. The American Journal of Sports Medicine, 28(5), 29–39.
45. Safrit, M. J. (1981). Evaluation in physical education (2nd ed.). Prentice-Hall.
46. Scott, G., & French, E. (2009). Measurement and evaluation in physical education. Sports Educational Technologies.

47. Singh, M., Kour, R., & Kour, A. (2022). A collaborative diversified investigation of respective responses of sports persons, coaches, and organizations on the criminalization of doping. International Journal of Health Sciences, 6(S3), 11295–11310.
48. Singh, R. (2002). Construction of soccer playing ability test (Unpublished doctoral dissertation). LNIPE.
49. Sirijaruwong, C., & Kosa, B. (2006). Construction of health-related physical fitness norms. Kasetsart Journal of Social Sciences, 27(2), 246–254.
50. Suwirman, S., Sepriadi, S., Nurul Ihsan, N., & Deswandi, D. (2021). Instrument speed endurance test of Pencak Silat athletes. Journal of Physical Education, Sport, Health and Recreation, 10(2), 115–121.
51. Wetzler, S. (2015). Martial arts studies as Kulturwissenschaft. Martial Arts Studies, 1, 20–33.