Available online at : http://journal.unj.ac.id/unj/index.php/gjik
Gladi : Jurnal Ilmu Keolahragaan, 17 (04) 2026, 530-542
Permalink/DOI: https://doi.org/10.21009/GJIK.174.01

Sex-Specific Normative Reference Values for Physical Fitness and Basketball Skills in University Physical Education Students

Ali Al-Zakholi[1]

[1]Department of Physical Education and Sport Sciences, College of Education, University of Zakho, Kurdistan Region, Iraq

Correspondiing author. Email **:** Ali.al-zakholi@uoz.edu.krd

(Submission Track: Received: 23-09-2026, Final Revision: 12-10-2026, Available Online: 09-11-2026)

Abstract. This study aimed to develop sex-specific percentile norms and a classification scheme for a comprehensive battery of 33 physical-fitness tests and 12 offensive and defensive basketball-skill tests in university physical-education students, for whom such reference values are lacking, particularly in Kurdish and Iraqi settings. In a cross-sectional design, 179 students (111 men, 68 women) from the University of Zakho completed the batteries under standardized conditions; test–retest reliability established in an independent pilot sample was high to very high (r = 0.84–0.97). Empirical percentile values (P5, P10, P25, P50, P75, P90, and P95) and a five-level classification (very low to very high) anchored at the 10th, 25th, 75th, and 90th percentiles, with the scoring direction of every test made explicit, were computed separately for men and women. Men and women differed significantly on all 45 tests (Welch's *t*-tests, Holm-adjusted $p < 0.001$; |Hedges' g| = 0.95–6.91), with women superior only in flexibility, which justified the construction of sex-specific norms. Percentile reference tables are reported for every test; for example, the median male 20-m shuttle-run estimate of maximal oxygen uptake was 44.6 mL·kg^{-1}·min^{-1} (P5–P95: 40.3–49.3). The study provides the first sex-specific normative reference values for these batteries in Kurdish university physical-education students, offering a practical tool for screening, classification, and longitudinal monitoring. The single-institution sample may limit generalizability, and the extreme percentiles, particularly for women, should be regarded as provisional; the norms should be updated periodically and cross-validated in larger samples.

Keywords: athletic performance; basketball; norms; physical fitness; reference values; students

Copyright © 2026 Ali Al-Zakholi
This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License

INTRODUCTION

The interpretation of any fitness or skill test depends on a frame of reference. A raw score: a sprint time, a number of repetitions, a jump distance acquires meaning only when it is compared against the performance of a relevant peer group. Normative reference values, and especially percentile norms, provide this frame: they locate an individual's score within the distribution of a defined population and thereby support screening, selection, grouping, goal-setting, and the monitoring of change over time (Ramos-Sepúlveda et al., 2016). For norms to be valid, however, they must be derived from a population that resembles the individuals to whom they will be applied, and they must be kept current, because performance distributions shift across time, culture, and context (Guo et al., 2025).

Physical fitness and technical skill are both central to basketball, an intermittent, high-intensity sport that demands speed, agility, power, and rapid perceptual–motor responses alongside shooting, passing, dribbling, and defensive footwork (Mancha-Triguero et al., 2019; Stojanović et al., 2018). Fitness and skill testing is consequently routine in basketball and in physical-education programmes, where it informs selection and evaluation (Vaeyens et al., 2008; Ziv & Lidor, 2009). A wide range of field tests has been validated for basketball and for general fitness (Jeličić et al., 2020; Scanlan et al., 2021; Spiteri et al., 2014), and these tests are widely administered to students. Yet their usefulness in practice is constrained by a persistent gap: appropriate normative reference values against which student scores can be interpreted are frequently unavailable.

Two considerations make this gap acute for the present population. First, physical performance differs substantially between the sexes across most fitness and skill qualities, so a single set of norms applied to both men and women would misclassify large numbers of individuals; sex-specific norms are therefore required. Second, norms developed in North American, European, or East Asian cohorts may not transfer to Kurdish and Iraqi university physical-education students, whose training background, anthropometry, and lifestyle differ. To our knowledge, no sex-specific normative reference values exist for a comprehensive fitness-and-skill battery in this population.

The purpose of this study was therefore to develop sex-specific percentile norms and an accompanying five-level classification scheme for 33 physical-fitness tests and 12 basketball-skill tests in a sample of university physical-education students, and to confirm, through the magnitude of sex differences, the necessity of reporting the norms separately for men and women.

METHOD

A cross-sectional, descriptive design was used to construct normative reference values. Participants were 179 undergraduate students (111 men, 68 women) of the Department of Physical Education and Sport Sciences, College of Education, University of Zakho, distributed across the four academic years (Year 1, n = 43; Year 2, n = 46; Year 3, n = 49; Year 4, n = 41). A further 32 students who took part in pilot and reliability testing were not included in the normative sample. All participants were free of injury or illness, provided written informed consent, and were tested under standardized conditions; the study followed the Declaration of Helsinki and was approved by the institutional research ethics committee.

Two field-test batteries were administered. The basketball-skill battery comprised 12 tests in four domains :shooting, passing, dribbling, and defensive footwork drawn from established protocols. The physical-fitness battery comprised 33 tests, three for each of eleven components: cardiorespiratory endurance, muscular strength, muscular endurance, flexibility, body composition, speed, agility, power, balance, coordination, and reaction time, selected from recognized field batteries (American College of Sports Medicine, 2021; Mancha-Triguero et al., 2019). Testing was conducted by specialist lecturers after a standardized warm-up; skill tests were administered on a single day through four in-hall stations, and the fitness tests were distributed across three consecutive days (11 per day) using a fatigue-graded sequence in which low-demand measures preceded maximal-effort measures.

Measurement standardization. All 45 tests were administered according to the procedures described in their source protocols and recognized field batteries, by specialist lecturers of the department, and under the same standardized conditions for every participant. Each testing session began with a standardized warm-up, and the order of testing was fixed: the 12 skill tests were completed on one day through four in-hall stations, and the 33 fitness tests were distributed over three consecutive days (11 per day) in a fatigue-graded sequence in which low-demand measures preceded maximal-effort measures, so that fatigue from one test did not affect the next. Men and women were tested with the same protocols, equipment, and sequence.

Reliability. Before the main study, the test–retest reliability of all 45 tests was established in the independent pilot sample (n = 32) using the Pearson correlation between two administrations one week apart; reliability was high to very high for every test (r = 0.84–0.97). Pearson's r describes the consistency of rank order between administrations but not absolute agreement, and it is insensitive to systematic change between trials (Atkinson & Nevill, 1998; Weir, 2005). The individual pilot scores were not retained in a form that allows re-analysis, so

intraclass correlation coefficients could not be computed retrospectively. As an indication of absolute measurement error, the standard error of measurement implied by the reported coefficients, SEM = SD × √(1 − r) (Weir, 2005), lies between approximately 0.17 SD (r = 0.97) and 0.40 SD (r = 0.84) of the between-subject standard deviation; for example, for the male 30-m sprint (SD = 0.20 s) this corresponds to an SEM of about 0.03–0.08 s and a minimal detectable change (MDC_{95} = 1.96 × √2 × SEM) of about 0.10–0.22 s. These values are approximate, because they assume that the variability of the pilot sample was similar to that of the normative sample. Normative values were computed separately for men and women. For each test, the mean, standard deviation, and the 5th, 10th, 25th, 50th, 75th, 90th, and 95th percentiles were calculated. Percentiles were estimated empirically, without smoothing or distributional assumptions, by linear interpolation between adjacent order statistics (Hyndman & Fan, 1996, definition 7, the default method of NumPy).

Data screening and distribution. Every participant contributed a score to every test, and no observations were excluded or replaced (n = 111 men and 68 women for all tests; Tables 1–4). Extreme scores were retained, because valid extreme performances belong to the distribution that norms are intended to describe, and because empirical percentiles are themselves robust to a small number of outlying values. Distribution shape was examined by comparing each empirical percentile with the value expected under a normal distribution (mean + z × SD). Agreement was close: across tests, the median of the largest discrepancy between empirical and normal-theory percentiles (P5–P95) was 0.18 SD in men and 0.21 SD in women, and the largest discrepancies (≈0.55 SD) occurred in tests with few discrete score values (e.g., Flamingo Balance and Free-Throw Accuracy). Because empirical percentiles do not depend on normality, the classification scheme was anchored on them rather than on the mean and standard deviation.

Sample size and precision of the norms. The normative sample was drawn from all four academic years of the department (Year 1–4, n = 41–49 each), so it covers the full undergraduate range of the population to which the norms will be applied. The precision of the norms was evaluated with the large-sample standard error of a percentile, $SE(P_p) \approx SD \times \sqrt{[p(1-p)/n]} / \varphi(z_p)$, where φ is the standard normal density. For the median, this gives approximately 0.12 SD for men (n = 111) and 0.15 SD for women (n = 68); for P10 and P90, 0.16 SD and 0.21 SD; and for P5 and P95, 0.20 SD and 0.26 SD, respectively. The central percentiles and the P10/P90 classification boundaries are therefore estimated with acceptable precision,

whereas P5 and P95 — particularly for women — should be treated as provisional. For this reason, P10 and P90 rather than P5 and P95 were used as the outer classification boundaries.

Classification scheme. A five-level scheme (very low, low, average, high, very high) was defined directly from the sex-specific empirical percentiles, with boundaries at P10, P25, P75, and P90 (Table 5). By construction, these boundaries place 10%, 15%, 50%, 15%, and 10% of the reference sample in the five bands and correspond, under a normal distribution, to $z = \pm 0.67$ and ± 1.28, so that the percentile and z-score definitions of each category are mutually consistent.

Scoring direction. For 21 of the 45 tests (all timed tests, step-test recovery heart rate, skinfold thickness, body-fat percentage, and the number of balance losses in the Flamingo test), a lower raw score indicates better performance; these tests are marked ↓ in Tables 1–4. The percentile columns of Tables 1–4 always list raw-score percentiles in ascending order, so for ↓ tests a low percentile value denotes a good performance. For classification, the percentile rank of ↓ tests was therefore reversed (performance percentile = 100 − raw-score percentile), and z-scores were computed as (mean − score)/SD, so that for every test a higher performance percentile and a higher z-score indicate better performance (Table 5). Body mass index is reported descriptively only and is not classified, because its relationship with health and performance is not monotonic.

Sex differences. To confirm the need for sex-specific norms, sex differences on every test were examined with Welch's independent-samples t-tests, which do not assume equal variances. For each test we report t, the Welch–Satterthwaite degrees of freedom, and the effect size Hedges' g (standardized mean difference based on the pooled SD, with small-sample correction) with its 95% confidence interval (Lakens, 2013), interpreted with the conventional thresholds of 0.2, 0.5, and 0.8 (Cohen, 1988). To control the family-wise error rate across the 45 comparisons, p values were adjusted with the Holm (1979) step-down procedure. Analyses were performed in Python (pandas, NumPy, SciPy), with statistical significance set at adjusted $p < 0.05$.

RESULT AND DISCUSSION

Result

Men and women differed significantly on all 45 tests, and every difference remained significant after Holm adjustment for multiple comparisons (all adjusted $p < 0.001$; Table 6). Effect sizes were large for every test ($|g| = 0.95$–6.91): the smallest differences were found in

the flexibility and shooting tests (|*g*| ≈ 0.95–1.22) and the largest in body composition, cardiorespiratory endurance, and strength (|*g*| ≈ 3.0–6.9). On the standardized metric, men outperformed women on every fitness and skill test except the three flexibility tests, on which women were superior (body mass index, which is not a performance measure, was higher in men). Because these differences were pervasive and, for most tests, substantial, separate normative reference values for men and women are clearly warranted; applying a single set of norms to both sexes would systematically misclassify individuals. Accordingly, all norms below are reported separately by sex.

The sex-specific percentile norms for the 33 physical-fitness tests are presented in Table 1 (men) and Table 2 (women), and those for the 12 basketball-skill tests in Table 3 (men) and Table 4 (women). Each table reports, for every test, the sample size, mean, standard deviation, and the seven percentile values, enabling a practitioner to locate any student's raw score within the appropriate sex distribution. For instance, a male student whose 20-m shuttle-run estimate of maximal oxygen uptake is 44.6 mL·kg^{-1}·min^{-1} sits at approximately the 50th percentile of his peers, whereas a score of 49 or above would place him near the 95th percentile. For tests marked ↓ the reading is reversed: a male 30-m sprint time of 3.88 s lies near the 5th raw-score percentile and therefore corresponds to approximately the 95th performance percentil.

Table 1. *Percentile norms for the 33 physical-fitness tests — men (n = 111)*

Test (unit)	n	Mean	SD	P5	P10	P25	P50	P75	P90	P95
20-m Shuttle Run — VO2max (mL/kg/min)	111	44.7	2.83	40.2	41.3	42.6	44.6	47.0	48.4	49.2
Cooper 12-min Run (m)	111	2542.0	121.6	2357.5	2404.0	2466.0	2528.0	2637.5	2693.0	2736.5
YMCA Step Test — recovery HR (bpm) ↓	111	102.6	5.68	93.5	95.0	99.0	103.0	106.0	110.0	111.5
1-Repetition Maximum (kg)	111	75.2	8.81	61.4	63.7	68.2	74.1	82.2	88.0	89.2
Handgrip Strength (kg)	111	51.4	6.17	42.3	43.3	47.0	51.1	55.5	59.9	61.2
Isometric Mid-Thigh Pull (N)	111	2428.6	259.6	1988.0	2058.0	2247.0	2451.0	2607.5	2780.0	2808.5
Push-Up (reps)	111	29.3	4.17	23.0	24.0	26.5	29.0	32.0	34.0	36.5
Curl-Up (reps)	111	34.8	5.20	26.0	29.0	31.0	35.0	38.5	41.0	42.5
Prone Plank Hold (s)	111	74.9	9.59	60.0	63.0	68.5	74.0	82.0	88.0	90.5
Sit-and-Reach (cm)	111	27.9	2.69	23.1	24.7	26.4	27.9	29.6	30.7	32.1
Back-Scratch (cm)	111	2.78	1.22	0.95	1.40	2.05	2.80	3.50	4.20	4.45
Goniometric ROM (deg)	111	124.3	3.42	119.1	119.8	122.0	124.3	126.8	128.7	130.1
Skinfold Thickness (mm) ↓	111	63.1	5.43	54.4	56.2	59.6	62.9	66.7	70.2	71.3
Body Mass Index (kg/m2) §	111	23.1	0.80	21.8	22.1	22.6	23.0	23.6	24.2	24.3
Body Fat % (BIA) ↓	111	14.6	1.33	12.6	12.8	13.8	14.6	15.7	16.2	16.8
30-m Sprint (s) ↓	111	4.19	0.20	3.88	3.90	4.05	4.17	4.34	4.49	4.52
40-yard Dash (s) ↓	111	4.85	0.25	4.46	4.53	4.65	4.86	5.08	5.18	5.20
Flying 10-m Sprint (s) ↓	111	1.15	0.06	1.06	1.08	1.11	1.15	1.18	1.22	1.25
Illinois Agility Test (s) ↓	111	15.8	0.79	14.7	14.8	15.3	15.7	16.5	17.1	17.2
Agility T-Test (s) ↓	111	9.98	0.52	9.12	9.31	9.59	9.98	10.4	10.6	10.8
505 Change-of-Direction (s) ↓	111	2.40	0.13	2.20	2.23	2.30	2.40	2.51	2.57	2.59
Countermovement Jump (cm)	111	48.2	5.37	40.4	40.9	43.6	48.1	52.3	56.0	56.9
Standing Long Jump (cm)	111	247.4	25.3	211.5	215.0	223.5	248.0	263.5	284.0	287.5
Seated Medicine-Ball Throw (m)	111	7.22	0.76	6.05	6.23	6.62	7.24	7.70	8.24	8.52
Flamingo Balance (falls/60 s) ↓	111	6.30	0.96	5.00	5.00	6.00	6.00	7.00	7.00	8.00
Stork Stand (s)	111	32.0	5.28	24.4	25.5	28.3	31.7	35.9	39.1	40.0
Y-Balance composite (%)	111	91.2	3.61	85.4	86.5	89.3	91.3	93.5	95.5	97.2

Alternate Wall Toss (catches/30 s)	111	26.4	3.29	20.0	22.0	24.0	27.0	28.0	30.0	31.0
Plate Tapping (s) ↓	111	10.9	0.66	9.70	9.90	10.5	11.0	11.3	11.6	11.8
Soda-Pop Coordination (s) ↓	111	12.1	0.70	10.8	11.1	11.6	12.1	12.6	12.9	13.1
Ruler-Drop Reaction (ms) ↓	111	174.8	10.6	159.0	162.0	167.0	174.0	181.5	190.0	193.0
Nelson Foot Reaction (ms) ↓	111	197.0	12.2	175.0	180.0	189.0	197.0	205.0	211.0	215.5
Light/Computer Reaction (ms) ↓	111	216.2	13.2	196.0	199.0	206.0	216.0	226.0	232.0	236.0

Note. SD = standard deviation; Pxx = xxth percentile. ↓ = lower score indicates better performance; percentiles are listed in ascending order of the raw score, so for ↓ tests low percentile values represent good performance (see Table 5). § Reported descriptively; not classified. VO2max is expressed in $mL \cdot kg^{-1} \cdot min^{-1}$.

Table 2. *Percentile norms for the 33 physical-fitness tests — women (n = 68)*

Test (unit)	n	Mean	SD	P5	P10	P25	P50	P75	P90	P95
20-m Shuttle Run — VO2max (mL/kg/min)	68	35.8	2.18	32.6	32.9	33.9	35.9	37.5	38.5	39.2
Cooper 12-min Run (m)	68	2053.3	87.6	1923.9	1951.4	1993.5	2042.5	2112.5	2165.5	2194.8
YMCA Step Test — recovery HR (bpm) ↓	68	119.7	5.67	111.0	112.7	115.8	120.0	124.0	126.3	128.0
1-Repetition Maximum (kg)	68	44.6	4.90	37.8	38.2	41.5	44.5	46.9	51.8	52.7
Handgrip Strength (kg)	68	33.3	3.54	27.9	28.7	30.7	33.1	36.1	37.8	38.9
Isometric Mid-Thigh Pull (N)	68	1573.5	146.2	1335.2	1370.4	1479.8	1570.5	1681.0	1768.7	1806.2
Push-Up (reps)	68	16.6	2.60	13.0	13.0	15.0	17.0	18.0	20.0	20.0
Curl-Up (reps)	68	23.6	3.08	19.0	20.0	21.0	23.0	26.0	27.0	28.0
Prone Plank Hold (s)	68	49.0	6.13	41.4	42.0	44.0	48.5	54.0	56.6	60.3
Sit-and-Reach (cm)	68	30.4	2.53	26.4	27.9	29.1	30.4	31.4	33.1	34.6
Back-Scratch (cm)	68	4.15	1.14	2.47	2.70	3.48	4.25	4.93	5.53	5.86
Goniometric ROM (deg)	68	127.6	3.37	123.0	123.8	125.0	127.7	130.0	132.4	133.2
Skinfold Thickness (mm) ↓	68	90.2	5.90	81.8	83.7	86.9	90.0	94.0	96.8	100.0
Body Mass Index (kg/m2) §	68	22.2	0.71	21.2	21.3	21.6	22.3	22.5	23.1	23.4
Body Fat % (BIA) ↓	68	23.6	1.24	21.6	22.0	22.7	23.6	24.6	25.0	25.5
30-m Sprint (s) ↓	68	4.76	0.23	4.42	4.43	4.61	4.76	4.86	5.14	5.18
40-yard Dash (s) ↓	68	5.52	0.24	5.14	5.19	5.40	5.52	5.64	5.89	5.93
Flying 10-m Sprint (s) ↓	68	1.31	0.06	1.21	1.22	1.27	1.32	1.35	1.39	1.41
Illinois Agility Test (s) ↓	68	18.0	0.85	16.7	17.0	17.5	18.0	18.4	19.2	19.6
Agility T-Test (s) ↓	68	11.4	0.53	10.6	10.6	11.1	11.4	11.7	12.1	12.3
505 Change-of-Direction (s) ↓	68	2.72	0.15	2.49	2.53	2.61	2.73	2.79	2.94	2.97
Countermovement Jump (cm)	68	32.9	3.50	27.1	28.8	31.1	33.0	35.1	37.9	38.5
Standing Long Jump (cm)	68	186.8	17.4	159.4	163.7	178.8	187.0	198.2	210.3	214.0
Seated Medicine-Ball Throw (m)	68	4.91	0.50	4.04	4.29	4.65	4.92	5.18	5.50	5.77
Flamingo Balance (falls/60 s) ↓	68	7.35	0.89	6.00	6.00	7.00	7.00	8.00	8.00	9.00
Stork Stand (s)	68	23.7	4.19	17.7	18.2	21.0	23.5	27.3	28.7	30.0
Y-Balance composite (%)	68	85.1	2.95	80.6	81.4	83.0	85.1	87.1	89.0	90.0
Alternate Wall Toss (catches/30 s)	68	18.8	2.54	15.0	15.0	17.0	19.0	20.0	22.0	23.0
Plate Tapping (s) ↓	68	12.4	0.67	11.2	11.5	11.9	12.3	12.9	13.2	13.4
Soda-Pop Coordination (s) ↓	68	13.6	0.72	12.4	12.6	13.1	13.6	14.2	14.5	14.9
Ruler-Drop Reaction (ms) ↓	68	199.1	11.1	181.7	186.0	192.0	200.5	205.2	212.0	215.7
Nelson Foot Reaction (ms) ↓	68	224.9	12.4	206.0	207.7	216.8	223.0	234.2	241.0	245.0
Light/Computer Reaction (ms) ↓	68	248.0	12.9	222.3	231.0	240.8	249.0	258.0	263.0	264.0

Note. SD = standard deviation; Pxx = xxth percentile. ↓ = lower score indicates better performance; percentiles are listed in ascending order of the raw score, so for ↓ tests low percentile values represent good performance (see Table 5). § Reported descriptively; not classified.

The basketball-skill norms show the same broad pattern of male advantage, though the sex gap in the skill domains was, in relative terms, somewhat smaller than in the most athletic fitness qualities. This is consistent with the observation that skill performance combines technical execution with physical capacity, so that technique partially attenuates the physically driven sex difference. The skill percentile tables nonetheless

differ enough between the sexes to require separate norms. To translate a percentile into an interpretable verbal category, the accompanying five-level classification scheme (Table 5 and Figure 1) assigns each score to one of five bands — very low, low, average, high, or very high — defined by the P10, P25, P75, and P90 values of the relevant sex distribution. By construction, the central band contains the middle 50% of students and each extreme band 10%, so the extreme bands flag the small proportions of markedly low or high performers who may warrant remedial attention or advanced development, respectively. For ↓ tests the bands are applied in reverse: for example, a male 30-m sprint time below P10 (3.90 s) is classified as very high, and a time above P90 (4.49 s) as very low.

Table 3. *Percentile norms for the 12 basketball-skill tests — men (n = 111)*

Test (unit)	n	Mean	SD	P5	P10	P25	P50	P75	P90	P95
AAHPERD Speed Spot Shooting (pts)	111	31.1	4.88	23.0	25.0	27.5	31.0	34.5	38.0	38.0
Free-Throw Accuracy (/10)	111	7.01	1.75	4.00	4.00	6.00	7.00	8.00	9.00	9.00
Dynamic Perimeter Shooting (pts)	111	10.9	2.55	7.00	7.00	9.00	11.0	12.0	15.0	15.0
AAHPERD Wall Chest Pass (pts)	111	38.8	6.09	29.5	31.0	33.5	39.0	43.5	47.0	48.5
Johnson Baseball-Pass Accuracy (pts)	111	20.2	3.85	14.0	15.0	17.0	20.0	23.0	26.0	26.0
Wall Speed Pass (count/30 s)	111	30.1	4.83	22.0	23.0	26.5	31.0	33.5	36.0	38.5
AAHPERD Control Dribble (s) ↓	111	15.8	1.35	13.5	13.9	14.8	15.9	16.9	17.6	17.9
I-Dribble Agility (s) ↓	111	17.9	1.62	15.6	16.1	16.6	17.7	19.1	20.2	20.6
Straight-Line Speed Dribble (s) ↓	111	4.33	0.35	3.80	3.87	4.08	4.29	4.52	4.84	4.96
AAHPERD Defensive Movement (s) ↓	111	23.4	1.62	20.9	21.4	22.1	23.3	24.4	25.4	25.9
Lane Agility Drill (s) ↓	111	13.7	1.16	11.8	12.2	13.0	13.6	14.4	15.4	15.8
Reactive Defensive Shuttle (s) ↓	111	2.95	0.27	2.50	2.58	2.76	2.95	3.15	3.35	3.39

Note. SD = standard deviation; Pxx = xxth percentile. ↓ = lower score indicates better performance; percentiles are listed in ascending order of the raw score, so for ↓ tests low percentile values represent good performance (see Table 5).

Table 4. *Percentile norms for the 12 basketball-skill tests — women (n = 68)*

Test (unit)	n	Mean	SD	P5	P10	P25	P50	P75	P90	P95
AAHPERD Speed Spot Shooting (pts)	68	25.2	4.72	17.4	19.0	22.0	25.5	29.0	31.0	32.0
Free-Throw Accuracy (/10)	68	5.34	1.76	2.35	3.00	4.00	5.00	7.00	7.00	8.00
Dynamic Perimeter Shooting (pts)	68	8.53	1.88	5.35	6.00	7.00	9.00	10.0	11.0	11.7
AAHPERD Wall Chest Pass (pts)	68	30.7	5.08	22.4	24.0	26.8	31.5	34.0	38.0	38.0
Johnson Baseball-Pass Accuracy (pts)	68	16.4	3.10	11.0	12.0	14.0	17.0	19.0	21.0	21.0
Wall Speed Pass (count/30 s)	68	25.6	4.24	19.4	20.0	22.0	26.0	29.0	31.0	31.6
AAHPERD Control Dribble (s) ↓	68	17.8	1.27	15.7	16.1	17.1	17.8	18.6	19.4	19.9
I-Dribble Agility (s) ↓	68	20.1	1.58	17.5	18.0	19.3	19.8	20.7	22.2	23.1
Straight-Line Speed Dribble (s) ↓	68	4.74	0.35	4.22	4.29	4.50	4.77	4.93	5.14	5.35
AAHPERD Defensive Movement (s) ↓	68	26.2	1.93	23.4	23.8	25.0	26.0	27.0	28.9	29.5
Lane Agility Drill (s) ↓	68	15.5	1.13	13.9	14.1	14.7	15.5	16.0	17.5	17.7
Reactive Defensive Shuttle (s) ↓	68	3.34	0.26	2.94	3.00	3.18	3.32	3.50	3.66	3.74

Note. SD = standard deviation; Pxx = xxth percentile. ↓ = lower score indicates better performance; percentiles are listed in ascending order of the raw score, so for ↓ tests low percentile values represent good performance (see Table 5).

Table 5. *Five-level normative classification scheme*

Category	Performance-percentile band (expected %)	Equivalent *z* band†	Raw score, higher-is-better tests	Raw score, lower-is-better tests (↓)
Very low	< P10 (10%)	$z < -1.28$	< P10	> P90
Low	P10 to < P25 (15%)	−1.28 to < −0.67	P10 to < P25	> P75 to P90
Average	P25 to P75 (50%)	−0.67 to +0.67	P25 to P75	P25 to P75
High	> P75 to P90 (15%)	> +0.67 to +1.28	> P75 to P90	P10 to < P25
Very high	> P90 (10%)	$z > +1.28$	> P90	< P10

Note. Bands are defined from the empirical, sex-specific percentiles reported in Tables 1–4 (columns P10, P25, P75, P90). For lower-is-better tests (↓), the raw-score bands are reversed so that "very high" always denotes the best performance. † *z*-score equivalents assume an approximately normal distribution; for ↓ tests z = (mean − score)/SD, otherwise z = (score − mean)/SD. Body mass index is not classified.

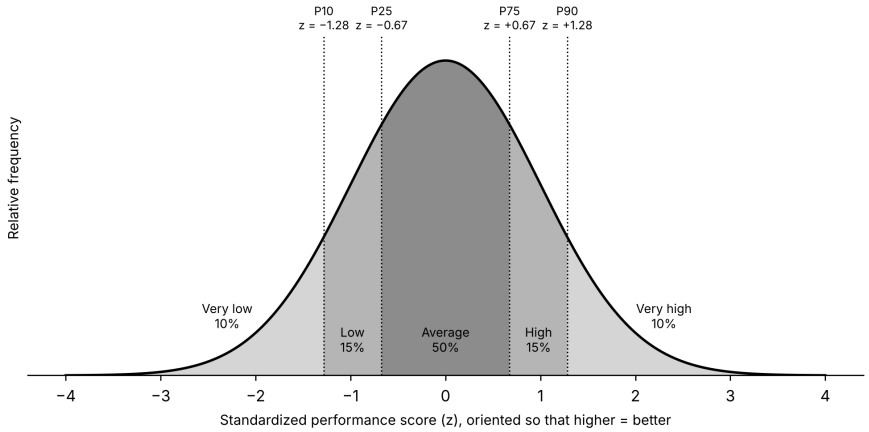


Figure 1. *Five-level normative classification model*

*Category boundaries at the 10th, 25th, 75th, and 90th percentiles (*z *= ±0.67 and ±1.28 under a normal distribution) partition the distribution into very low (10%), low (15%), average (50%), high (15%), and very high (10%) bands. For lower-is-better tests,* z *is oriented so that higher values indicate better performance.*

Table 6. *Sex differences on the 45 tests (Welch's t-tests, Holm-adjusted)*

Test (unit)	Men, M (SD)	Women, M (SD)	*t*	*df*	p_{Holm}	Hedges' *g* [95% CI]
Physical fitness						
20-m Shuttle Run — VO2max (mL/kg/min)	44.7 (2.83)	35.8 (2.18)	23.61	167.8	< 0.001	3.40 [2.94, 3.87]
Cooper 12-min Run (m)	2542.0 (121.6)	2053.3 (87.6)	31.15	172.3	< 0.001	4.42 [3.88, 4.97]
YMCA Step Test — recovery HR (bpm) ↓	102.6 (5.68)	119.7 (5.67)	−19.57	142.0	< 0.001	−3.00 [−3.43, −2.57]
1-Repetition Maximum (kg)	75.2 (8.81)	44.6 (4.90)	29.83	175.6	< 0.001	4.02 [3.51, 4.54]
Handgrip Strength (kg)	51.4 (6.17)	33.3 (3.54)	24.93	176.4	< 0.001	3.38 [2.92, 3.84]
Isometric Mid-Thigh Pull (N)	2428.6 (259.6)	1573.5 (146.2)	28.17	176.0	< 0.001	3.81 [3.31, 4.31]
Push-Up (reps)	29.3 (4.17)	16.6 (2.60)	25.10	176.9	< 0.001	3.46 [2.99, 3.93]
Curl-Up (reps)	34.8 (5.20)	23.6 (3.08)	18.09	176.8	< 0.001	2.47 [2.07, 2.87]
Prone Plank Hold (s)	74.9 (9.59)	49.0 (6.13)	22.04	176.7	< 0.001	3.05 [2.62, 3.49]
Sit-and-Reach (cm)	27.9 (2.69)	30.4 (2.53)	−6.26	148.5	< 0.001	−0.95 [−1.26, −0.63]
Back-Scratch (cm)	2.78 (1.22)	4.15 (1.14)	−7.60	149.2	< 0.001	−1.15 [−1.47, −0.82]
Goniometric ROM (deg)	124.3 (3.42)	127.6 (3.37)	−6.32	143.4	< 0.001	−0.97 [−1.28, −0.65]
Skinfold Thickness (mm) ↓	63.1 (5.43)	90.2 (5.90)	−30.73	132.8	< 0.001	−4.81 [−5.39, −4.23]
Body Mass Index (kg/m2) §	23.1 (0.80)	22.2 (0.71)	7.84	154.7	< 0.001	1.17 [0.84, 1.49]
Body Fat % (BIA) ↓	14.6 (1.33)	23.6 (1.24)	−45.84	149.5	< 0.001	−6.91 [−7.69, −6.13]
30-m Sprint (s) ↓	4.19 (0.20)	4.76 (0.23)	−16.89	126.9	< 0.001	−2.68 [−3.09, −2.27]
40-yard Dash (s) ↓	4.85 (0.25)	5.52 (0.24)	−17.84	146.3	< 0.001	−2.71 [−3.12, −2.30]
Flying 10-m Sprint (s) ↓	1.15 (0.06)	1.31 (0.06)	−17.32	141.8	< 0.001	−2.66 [−3.06, −2.25]

Test (unit)	Men, M (SD)	Women, M (SD)	*t*	*df*	p_{Holm}	Hedges' *g* [95% CI]
Illinois Agility Test (s) ↓	15.8 (0.79)	18.0 (0.85)	−17.26	133.8	< 0.001	−2.69 [−3.10, −2.28]
Agility T-Test (s) ↓	9.98 (0.52)	11.4 (0.53)	−17.52	139.7	< 0.001	−2.70 [−3.11, −2.29]
505 Change-of-Direction (s) ↓	2.40 (0.13)	2.72 (0.15)	−14.56	126.5	< 0.001	−2.31 [−2.70, −1.93]
Countermovement Jump (cm)	48.2 (5.37)	32.9 (3.50)	23.07	176.3	< 0.001	3.21 [2.76, 3.66]
Standing Long Jump (cm)	247.4 (25.3)	186.8 (17.4)	18.96	174.6	< 0.001	2.67 [2.26, 3.07]
Seated Medicine-Ball Throw (m)	7.22 (0.76)	4.91 (0.50)	24.51	176.1	< 0.001	3.42 [2.95, 3.88]
Flamingo Balance (falls/60 s) ↓	6.30 (0.96)	7.35 (0.89)	−7.43	150.1	< 0.001	−1.12 [−1.44, −0.80]
Stork Stand (s)	32.0 (5.28)	23.7 (4.19)	11.63	165.4	< 0.001	1.69 [1.34, 2.04]
Y-Balance composite (%)	91.2 (3.61)	85.1 (2.95)	12.31	162.8	< 0.001	1.80 [1.44, 2.15]
Alternate Wall Toss (catches/30 s)	26.4 (3.29)	18.8 (2.54)	17.33	167.6	< 0.001	2.50 [2.10, 2.90]
Plate Tapping (s) ↓	10.9 (0.66)	12.4 (0.67)	−14.62	140.2	< 0.001	−2.25 [−2.63, −1.87]
Soda-Pop Coordination (s) ↓	12.1 (0.70)	13.6 (0.72)	−13.67	138.7	< 0.001	−2.11 [−2.48, −1.74]
Ruler-Drop Reaction (ms) ↓	174.8 (10.6)	199.1 (11.1)	−14.46	136.8	< 0.001	−2.24 [−2.62, −1.86]
Nelson Foot Reaction (ms) ↓	197.0 (12.2)	224.9 (12.4)	−14.70	140.0	< 0.001	−2.26 [−2.65, −1.88]
Light/Computer Reaction (ms) ↓	216.2 (13.2)	248.0 (12.9)	−15.87	144.3	< 0.001	−2.42 [−2.81, −2.03]
Basketball skills						
AAHPERD Speed Spot Shooting (pts)	31.1 (4.88)	25.2 (4.72)	8.01	145.5	< 0.001	1.22 [0.89, 1.55]
Free-Throw Accuracy (/10)	7.01 (1.75)	5.34 (1.76)	6.17	141.2	< 0.001	0.95 [0.63, 1.27]
Dynamic Perimeter Shooting (pts)	10.9 (2.55)	8.53 (1.88)	7.13	170.9	< 0.001	1.02 [0.70, 1.34]
AAHPERD Wall Chest Pass (pts)	38.8 (6.09)	30.7 (5.08)	9.59	160.9	< 0.001	1.41 [1.07, 1.74]
Johnson Baseball-Pass Accuracy (pts)	20.2 (3.85)	16.4 (3.10)	7.25	164.2	< 0.001	1.06 [0.73, 1.38]
Wall Speed Pass (count/30 s)	30.1 (4.83)	25.6 (4.24)	6.53	155.9	< 0.001	0.97 [0.65, 1.29]
AAHPERD Control Dribble (s) ↓	15.8 (1.35)	17.8 (1.27)	−9.98	148.5	< 0.001	−1.51 [−1.85, −1.17]
I-Dribble Agility (s) ↓	17.9 (1.62)	20.1 (1.58)	−8.96	144.6	< 0.001	−1.36 [−1.70, −1.03]
Straight-Line Speed Dribble (s) ↓	4.33 (0.35)	4.74 (0.35)	−7.61	141.8	< 0.001	−1.17 [−1.49, −0.84]
AAHPERD Defensive Movement (s) ↓	23.4 (1.62)	26.2 (1.93)	−10.00	123.3	< 0.001	−1.60 [−1.94, −1.25]
Lane Agility Drill (s) ↓	13.7 (1.16)	15.5 (1.13)	−10.24	144.7	< 0.001	−1.56 [−1.90, −1.22]
Reactive Defensive Shuttle (s) ↓	2.95 (0.27)	3.34 (0.26)	−9.60	146.0	< 0.001	−1.46 [−1.80, −1.12]

Note. Men: $n = 111$; women: $n = 68$. *t* and *g* are computed as men minus women; *df* = Welch–Satterthwaite degrees of freedom; p_{Holm} = *p* adjusted for 45 comparisons with the Holm (1979) procedure (all unadjusted $p < 10^{-8}$). For ↓ tests (lower = better), a negative *t* and *g* indicate better performance by men. § Not a performance measure. $|g| \geq 0.8$ = large effect (Cohen, 1988).

Discussion

The present norms extend the small body of population-specific fitness reference values into a previously undocumented group. Prior normative studies have generally addressed schoolchildren or general university populations and have concentrated on health-related fitness (Guo et al., 2025; Ramos-Sepúlveda et al., 2016); the present study adds a comprehensive, sex-specific set of reference values that spans both health- and skill-related fitness and, uniquely, offensive and defensive basketball skills, in Kurdish university physical-education students. The pervasive sex differences observed here echo the wider literature on physical performance in basketball, in which male players consistently exhibit superior speed, power, and change-of-direction ability (Mancha-Triguero et al., 2019; Ziv & Lidor, 2009), and they provide the empirical justification for sex-specific norming.

The single exception to the male advantage , flexibility is consistent with repeated reports that women exhibit greater joint range of motion, and it reinforces the more

general point that fitness qualities do not move together uniformly across the sexes. In practical terms, the norms and classification scheme give teachers and coaches an immediate, interpretable tool: a student's raw score on any of the 45 tests can be converted to a sex-specific percentile and verbal category, supporting transparent feedback, criterion-referenced grouping, talent screening, and the tracking of progress across a programme (Vaeyens et al., 2008). Where testing time is limited, the same tables can be applied to any subset of tests without recalculation.

Several limitations should be noted. The norms derive from a single institution and a moderate sample, so they are best regarded as local reference values pending cross-validation in larger, multi-centre samples; percentile estimates at the extremes (P5 and P95), in particular, are less stable in samples of this size, especially for women, whose subsample was smaller (estimated standard error ≈ 0.26 SD for women and 0.20 SD for men; see Research Method). These extreme percentiles should therefore be interpreted as provisional, and the classification uses the more precise P10 and P90 values as its outer boundaries. Reliability was established in a modest pilot sample (n = 32) and is available only as Pearson correlations; future normative updates should report test-by-test intraclass correlation coefficients, SEM, and MDC computed from retained test–retest data. The cross-sectional design captures a single cohort at one time point, and because performance distributions drift, the norms should be revised periodically. Finally, the sample comprised physical-education students, whose athletic background may render the values inappropriate for the general student population. Future work should expand the sample across institutions and regions, examine age- and year-specific norms, and re-estimate the values at regular intervals

CONCLUSION

This study developed the first sex-specific normative reference values and a five-level classification scheme for a comprehensive battery of 33 physical-fitness tests and 12 offensive and defensive basketball-skill tests in Kurdish university physical-education students. Because men and women differed significantly on all 45 tests — men superior on every test except flexibility — the norms were, and should be, reported separately by sex. The resulting percentile tables and classification bands provide practitioners with a practical, interpretable instrument for screening, criterion-referenced evaluation, talent identification, and longitudinal monitoring in this population. Given the single-

institution, cross-sectional basis of the values, they should be treated as local reference standards, cross-validated in larger and more diverse samples, and updated periodically to preserve their accuracy.

REFERENCES

American College of Sports Medicine. (2021). *ACSM's guidelines for exercise testing and prescription* (11th ed.). Wolters Kluwer.

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

Bae, J.-Y. (2022). Positional differences in physique, physical strength, and lower-extremity stability in Korean male elite high-school basketball athletes. *International Journal of Environmental Research and Public Health, 19*(6), 3416. https://doi.org/10.3390/ijerph19063416

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Guo, L., Jiang, L., Huang, H., & Said, M. A. (2025). A longitudinal study of four-year changes in physical fitness among university students before and after COVID-19: 2019–2022. *PLOS ONE, 20*(10), e0334088. https://doi.org/10.1371/journal.pone.0334088

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics, 6*(2), 65–70.

Hyndman, R. J., & Fan, Y. (1996). Sample quantiles in statistical packages. *The American Statistician, 50*(4), 361–365. https://doi.org/10.1080/00031305.1996.10473566

Jeličić, M., Ivančev, V., Čular, D., Čović, N., Stojanović, E., Scanlan, A. T., & Milanović, Z. (2020). The 30–15 intermittent fitness test: A reliable, valid, and useful tool to assess aerobic capacity in female basketball players. *Research Quarterly for Exercise and Sport, 91*(1), 83–91. https://doi.org/10.1080/02701367.2019.1648743

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology, 4*, 863. https://doi.org/10.3389/fpsyg.2013.00863

Mancha-Triguero, D., García-Rubio, J., Calleja-González, J., & Ibáñez, S. J. (2019). Physical fitness in basketball players: A systematic review. *The Journal of Sports Medicine and Physical Fitness, 59*(9), 1513–1525. https://doi.org/10.23736/S0022-4707.19.09180-1

Munshi, P., Khan, M. H., Arora, N. K., Nuhmani, S., Anwer, S., Li, H., & Alghadir, A. H. (2022). Effects of plyometric and whole-body vibration on physical performance in collegiate basketball players: A crossover randomized trial. *Scientific Reports, 12*, 5651. https://doi.org/10.1038/s41598-022-09142-8

Palma-Muñoz, I., Ramírez-Campillo, R., Azocar-Gallardo, J., Álvarez, C., Asadi, A., Moran, J., & Chaabene, H. (2021). Effects of progressed and nonprogressed volume-

based overload plyometric training on components of physical fitness and body composition variables in youth male basketball players. *Journal of Strength and Conditioning Research, 35*(6), 1642–1649. https://doi.org/10.1519/JSC.0000000000002950

Ramos-Sepúlveda, J. A., Ramírez-Vélez, R., Correa-Bautista, J. E., Izquierdo, M., & García-Hermoso, A. (2016). Physical fitness and anthropometric normative values among Colombian-Indian schoolchildren. *BMC Public Health, 16*, 962. https://doi.org/10.1186/s12889-016-3652-2

Scanlan, A. T., Madueno, M. C., Guy, J. H., Giamarelos, K., Spiteri, T., & Dalbo, V. J. (2021). Measuring decrement in change-of-direction speed across repeated sprints in basketball: Novel versus traditional approaches. *Journal of Strength and Conditioning Research, 35*(3), 841–845. https://doi.org/10.1519/JSC.0000000000002828

Spiteri, T., Nimphius, S., Hart, N. H., Specos, C., Sheppard, J. M., & Newton, R. U. (2014). Contribution of strength characteristics to change of direction and agility performance in female basketball athletes. *Journal of Strength and Conditioning Research, 28*(9), 2415–2423. https://doi.org/10.1519/JSC.0000000000000547

Stojanović, E., Stojiljković, N., Scanlan, A. T., Dalbo, V. J., Berkelmans, D. M., & Milanović, Z. (2018). The activity demands and physiological responses encountered during basketball match-play: A systematic review. *Sports Medicine, 48*(1), 111–135. https://doi.org/10.1007/s40279-017-0794-z

Vaeyens, R., Lenoir, M., Williams, A. M., & Philippaerts, R. M. (2008). Talent identification and development programmes in sport: Current models and future directions. *Sports Medicine, 38*(9), 703–714. https://doi.org/10.2165/00007256-200838090-00001

Weir, J. P. (2005). Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *Journal of Strength and Conditioning Research, 19*(1), 231–240. https://doi.org/10.1519/15184.1

Yilmaz, N. (2022). Investigation of the effect of isometric core strength training in addition to basic basketball training on explosive power in children aged 9–17. *Pedagogy of Physical Culture and Sports, 26*(1), 4–11. https://doi.org/10.15561/26649837.2022.0201

Zhang, M., Liang, X., Huang, W., Ding, S., Li, G., Zhang, W., Li, C., Zhou, Y., Sun, J., & Li, D. (2023). The effects of velocity-based versus percentage-based resistance training on athletic performances in sport-collegiate female basketball players. *Frontiers in Physiology, 13*, 992655. https://doi.org/10.3389/fphys.2022.992655

Ziv, G., & Lidor, R. (2009). Physical attributes, physiological characteristics, on-court performances and nutritional strategies of female and male basketball players. *Sports Medicine, 39*(7), 547–568. https://doi.org/10.2165/00007256-200939070-00003