



Information Article

Development and Psychometric Standardization of a Sport-Specific Flexibility Assessment Battery for Junior Artistic Gymnasts in the Kurdistan Region of Iraq

Hiyam Sadiq Ahmed Hassan

Salahaddin University, College of Physical Education and Sports Sciences, Iraq

ARTICLE INFO ABSTRACT

Keywords:
Sport-Specific
Flexibility,
Gymnastics

The significance of this study lies in constructing sport-specific flexibility test batteries for junior gymnasts to establish standardized evaluative benchmarks for assessing flexibility profiles. The research problem emerges from the scarcity and lack of standardized assessments specifically tailored to measure sport-specific flexibility in junior gymnasts, which this study addresses through systematic test development.

The investigation aimed to construct sport-specific flexibility tests for junior gymnasts in the Kurdistan Region of Iraq and derive normative standard scores and proficiency levels for these assessments.

The researcher employed a descriptive survey design suited to the empirical nature of the study. A purposive sample of ($n = 60$) gymnasts was selected, constituting 80% of the total target population.

The findings demonstrated high validity and reliability for the constructed three-test battery, comprising:

1. The Star Flexibility Test (Hip Joint Flexibility)
2. The Back Bridge Test (Spinal Column Flexibility)
3. The Prone Backward Arm Extension Test (Shoulder Joint Flexibility)

The researcher recommends the formal adoption and standardized implementation of this sport-specific flexibility battery to evaluate junior gymnasts during athletic development and talent monitoring.

Corresponding Author

E-mail address:

DOI: <https://doi.org/10.26400/Sep/69/11>



This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/)



1. Introduction

Sports in most countries worldwide have witnessed noticeable development and progress across all levels. This advancement is the result of developments in the use of measurement and evaluation methods, which have elevated the performance levels of athletes and teams, enabling them to achieve optimal athletic accomplishments across various sports and competitions. This has been facilitated by employing modern methods and approaches that assist in the comprehensive physical preparation of athletes and teams, as prioritizing physical conditioning is regarded as the most critical means of enhancing athletic performance and outcomes.

The scientific literature indicates a clear gap in the methods employed to measure flexibility in artistic performance sports, particularly gymnastics among youth. Current practical testing applications focus primarily on general flexibility tests that do not adequately reflect the specialized ranges of motion required in gymnastics, such as hip flexibility, spinal extension, and shoulder joint mobility essential for executing precise technical movements.

Given the importance of tests and measurements across all sports disciplines as fundamental assessment tools that have garnered increased attention and application, "the data obtained through objective measurements are accurate and decisive, providing the capability to determine many fundamental measurement purposes in diagnosis, classification, prediction, and scientific research that serve practitioners in the sports field" (Al-Fartousi, 2015, p. 46).

Flexibility is considered one of the most critical physical determinants in gymnastics, directly contributing to movement quality, technical skill precision, and injury risk reduction during complex elements such as inversions, balance holds, and floor maneuvers. In the scientific training of youth gymnasts, "measuring and developing flexibility represents an essential step toward enhancing overall performance and bridging the gap between the current level of beginners and the standard required for official competitions" (Masaad et al., 2022, p. 10).

Despite the vital importance of flexibility, the tests conventionally used for measurement are often general and non-specialized. They do not accurately reflect the specific kinematic demands of gymnastics, nor do they necessarily align with the physiological characteristics of younger age categories (such as youth and novice athletes). "Therefore, constructing specific flexibility tests tailored to youth gymnastics aids in assessing the genuine capabilities of gymnasts and guides coaches toward developing more effective training programs" (Sabeeh & Mustafa, 2022, p. 14). Based on the foregoing, the significance of the present study lies in designing (constructing) specific flexibility tests for youth gymnasts in the Kurdistan Region of Iraq, serving as an objective reference framework for evaluating flexibility among youth gymnasts.

Research Problem

The fundamental problem stems from the fact that most tests utilized in youth gymnastics training are not methodologically designed to account for the specific characteristics of the sport, technical skill requirements, and age-related and physiological variations of youth categories. This leads to difficulties in accurately determining gymnasts' performance levels and benchmarking their progression during training, thereby hindering the efficacy of training programs and potentially impairing readiness for future competitions.





The researcher identified the problem in the scarcity and lack of specific flexibility tests designed for youth gymnasts. Consequently, the present study attempts to address this gap by constructing specific flexibility tests for youth gymnasts.

Research Objectives

1. Constructing specific flexibility tests for youth gymnasts in the Kurdistan Region of Iraq.
2. Establishing standard scores and normative levels for the specific flexibility tests for youth gymnasts.

Research Scope

- **Human Scope:** Youth gymnasts in the Kurdistan Region of Iraq for the 2025/2026 sports season, numbering 60 gymnasts.
- **Temporal Scope:** The period from November 25, 2025, to February 15, 2026.
- **Spatial Scope:** Specialized gymnasiums of sports clubs and youth centers across the governorates of the Kurdistan Region (Erbil, Duhok, Sulaymaniyah).

Definition of Terms

- **Specific Flexibility:** "The ability of an athlete to perform movements through a wide range of motion that aligns with the kinematic demands of a specific sport skill or event" (Al-Fateh, 2016, p. 54).
- **Operational Definition:** The ability of gymnasts to execute technical movements and skills across an extensive range of motion in the joints and muscles recruited during performance, measured via specific flexibility tests linked to gymnastics skills, including spinal, hip, and shoulder flexibility.

2. Research Methodology and Procedures

2.1 Research Method

The researcher employed the descriptive method using a survey approach due to its appropriateness for the nature of the current investigation.

2.2 Research Population and Sample

The research population was defined as youth gymnasts aged 14–15 years in the Kurdistan Region for the 2025/2026 season, totaling 75 gymnasts representing first-division clubs and youth centers across the governorates of Erbil, Duhok, and Sulaymaniyah. The research sample was selected purposively to achieve the research objectives. The sample comprised 60 gymnasts after excluding the pilot and reliability testing sample (15 gymnasts). Thus, the core research sample represented 80% of the total population. Tables 1 and 2 present the distribution of the research population and sample.

Table 1. Distribution of the Research Population by Governorates and Clubs

No.	Governorate	Clubs	Number
1	Erbil	Erbil	10
		Sirwan	10
		Akad Ankawa	12
2	Duhok	Duhok	12
		Semel	13
3	Sulaymaniyah	Sulaymaniyah	10
		Newroz	8

Table 2. Distribution of the Research Sample

No.	Sample	Number	Percentage (%)
1	Construction and Normative Sample	60	80%
2	Pilot and Reliability Sample	15	20%
3	Total Population	75	100%





Prior to constructing and administering the tests to the primary research sample, the researcher verified the homogeneity of the sample across the variables of age, height, and body mass (with club gymnasts' ages ranging between 16–17 years), as presented in Table 3.

Table 3. *Descriptive Characteristics and Homogeneity of the Research Sample in Age, Height, and Body Mass*

No.	Variables	Unit of Measurement	Mean ($\bar{X}$)	$\pm$ SD	Coefficient of Variation (CV%)
1	Age	Month	196.24	3.14	1.60%
2	Height	m	168.41	8.16	4.84%
3	Body Mass	kg	66.54	4.47	6.71%

Table 3 shows the homogeneity of the sample across age, height, and body mass, as indicated by the coefficients of variation (1.60%, 4.84%, and 6.71%, respectively), all of which were below 30%. As noted by Al-Tikriti and Al-Obeidi (1999), "a coefficient of variation below 30% indicates homogeneity of the sample and the absence of dispersion in the studied variable" (Al-Tikriti & Al-Obeidi, 1999, p. 212).

2.3 Data Collection Instruments

The researcher used the following instruments to collect data:

- Questionnaire.
- Specific flexibility tests designed and constructed by the researcher.

2.3.1 Questionnaire

A questionnaire was distributed to expert judges and specialists requesting their evaluation regarding the validity and suitability of the designed specific flexibility tests prior to administering them to the main research sample, as detailed in the face validity section.

2.3.2 Construction of the Specific Flexibility Tests

The researcher constructed and designed three specific flexibility tests for youth gymnasts:

1. Star Flexibility Test (Hip Flexibility).
2. Back Bridge Test (Spinal Flexibility).
3. Prone Rear Arm Raise Test (Shoulder Flexibility).

The specifications of the constructed tests are described below:

Test 1: Star Flexibility Test (Hip Flexibility)

- **Objective:** To measure hip and pelvic joint flexibility.
- **Equipment:** A five-pointed star marked on the floor, measuring tape.
- **Administration:** The gymnast stands at the center point of the five-pointed star and moves the entire lower limb from the hip joint along the five radial directions of the star, attempting to reach the maximum possible distance with the foot in all directions.
- **Instructions:** A five-pointed star is drawn on the ground with radial lines extending 1.0 m from the center in all directions. The gymnast stands at the center, extending the leg to reach the furthest possible point along each line.
- **Scoring:** Five reach distances are recorded for each leg. The arithmetic mean is calculated for each leg separately, followed by the grand mean of both legs, representing hip joint flexibility. Higher scores indicate greater flexibility. Distances are recorded in meters and sub-units.

Test 2: Back Bridge Test (Spinal Flexibility)

- **Objective:** To measure spinal flexibility.





- **Equipment:** Measuring tape.
- **Administration:** The gymnast stands upright and performs a backward bridge by extending the back and head backward to the maximum possible extent without falling.
- **Instructions:** A measuring tape is positioned relative to the posterior region of the head to measure the vertical distance from the head to the floor during backward extension.
- **Scoring:** The distance from the back of the head to the floor is recorded. Smaller distances reflect greater spinal flexibility. Measurements are recorded in meters and sub-units.

Test 3: Prone Rear Arm Raise Test (Shoulder Flexibility)

- **Objective:** To measure shoulder joint flexibility.
- **Equipment:** Measuring tape, level floor surface.
- **Administration:** The gymnast lies prone on a flat surface with arms extended forward. Upon preparation, the gymnast raises both arms upward to the maximum achievable height while maintaining a fully extended body without moving the trunk or lifting the head.
- **Instructions:** The measuring tape is attached to one hand of the participant. The vertical displacement is measured when raising the arms upward, ensuring no compensatory movement occurs in the head or torso, movement is isolated to the glenohumeral joint, and the elbows remain fully extended.
- **Scoring:** The vertical distance from the highest point of the gymnast's hands to the floor is recorded, representing shoulder flexibility. Greater distances indicate superior shoulder flexibility. Measurements are recorded in meters and sub-units.

2.3.2.1 Pilot Study of the Specific Flexibility Tests

To assess the feasibility and practical implementation of the designed physical tests, a pilot study was conducted on a subsample of 15 gymnasts from Sulav Sports Club on January 15, 2026, at 4:00 PM. The pilot study allowed the researcher to verify the procedural feasibility of test administration and confirm the absence of execution obstacles.

2.3.2.2 Scientific Rigor / Psychometric Properties of the Tests

1. Test Validity

- **Face Validity:** Face validity is defined as "a facet of content validity established by presenting the instrument to a panel of experts to inspect the test superficially and conclude whether the test genuinely measures what it was designed to measure" (Al-Abbasi, 2018, p. 286).

This form of validity depends on the test content and its relevance to the measured phenomenon. Bloom et al. (1983) noted that "a 75% consensus among expert arbiters can be relied upon in this type of validity" (Bloom et al., 1983, p. 126).



The constructed specific flexibility tests were presented to a panel of measurement and evaluation experts* to ascertain their face validity and suitability. An agreement threshold of 75% was adopted. Table 4 presents the expert approval percentages.

Table 4. Expert Approval Percentages for the Designed Tests

No.	Tests	Approving	Rejecting	Total Panel	Percentage (%)
1	Star Flexibility Test (Hip Flexibility)	9	1	10	90%
2	Back Bridge Test (Spinal Flexibility)	10	0	10	100%
3	Prone Rear Arm Raise Test (Shoulder Flexibility)	8	2	10	80%

Table 4 shows that all designed physical tests achieved expert approval rates ranging from 80% to 100%, exceeding the 75% threshold, thereby confirming their validity.

- **Discriminant Validity:** "Also termed construct validity, it refers to the degree to which a test measures a hypothetical construct or specific trait, representing a psychological or physical characteristic inferred through associated behaviors" (Melhem, 2010, p. 319).

An independent-samples *t*-test between two equal extreme groups was used to evaluate the tests' discriminatory power between high- and low-performing gymnasts on the measured traits. Following administration to the construction sample (N=60), scores were ranked in descending order. The upper 50% (n=30) and lower 50% (n=30) extreme groups were identified. This procedure is established in physical education research when construction samples are under 100 participants (Al-Zobaie et al., 1981, p. 75). Means, standard deviations, and *t*-values between the upper and lower groups were computed (Table 5).

Table 5. Discriminant Power of the Designed Specific Flexibility Tests

No.	Test Name	Unit	Upper Group Mean (X ⁻)	Upper Group $\pm$ SD	Lower Group Mean (X ⁻)	Lower Group $\pm$ SD	t-value	Sig. (p)
1	Star Flexibility Test (Hip Flexibility)	cm	85.26	2.98	74.16	2.52	15.57	0.001
2	Back Bridge Test (Spinal Flexibility)	m	140	2.34	151.73	4.53	12.58	0.001
3	Prone Rear Arm Raise Test (Shoulder Flexibility)	cm	41.76	2.59	34.20	2.49	11.50	0.001

Note. Significant at $p < 0.05$.

*** Expert Panel Members:**

- Prof. Dr. Thailam Younis Allawi — University of Mosul / College of Physical Education and Sports Sciences
- Prof. Dr. Makki Mahmoud Hussein — University of Mosul / College of Physical Education and Sports Sciences
- Prof. Dr. Majeed Khuda Bakhsh — University of Sulaimani / College of Physical Education and Sports Sciences
- Prof. Dr. Jalal Kamal Mohammed — University of Duhok / College of Physical Education and Sports Sciences
- Prof. Dr. Saad Fadhel Abdulqader — Al-Hadba University / College of Physical Education and Sports Sciences
- Prof. Dr. Hardan Aziz Salman — Mustansiriyah University / College of Physical Education and Sports Sciences
- Assoc. Prof. Dr. Ahmed Muayad Hussein — University of Mosul / College of Basic Education
- Assoc. Prof. Dr. Jameel Ahmed Hussein — University of Duhok / College of Physical Education and Sports Sciences
- Assoc. Prof. Dr. Ahmed Hisham Ahmed — University of Mosul / College of Basic Education
- Assoc. Prof. Dr. Wiam Amer Ahmed — University of Kirkuk / College of Physical Education and Sports Sciences





Table 5 reveals that all candidate tests possess high discriminant power, with calculated t -values ranging from 11.50 to 15.57 ($p=0.001<0.05$), confirming their capacity to differentiate between high- and low-flexibility performance levels.

- **Intrinsic Validity:** Intrinsic validity serves as an index of empirical validity, defined as "the validity of experimental scores freed from measurement errors, calculated as the square root of the reliability coefficient" (Farahat, 2001, p. 123), as presented in Table 6.

2. Test Reliability "Reliability refers to the ability of an instrument or test to yield consistent or comparable results upon repeated administration to the same sample under identical conditions" (Al-Azzawi, 2008, p. 129).

The test-retest method was employed to evaluate stability across time (Mikhail, 2016, p. 208). Tests were initially administered to the reliability sample ($n=15$) on January 16, 2026, and readministered on January 22, 2026 (a 6-day interval). Pearson's product-moment correlation coefficient was calculated between both administrations (Table 6). A reliability coefficient of ≥ 0.71 (corresponding to an alienation coefficient of ≤ 0.70) was adopted as the criterion for acceptance (Al-Zubair, 2019, p. 39).

Table 6. Reliability, Alienation, and Intrinsic Validity Coefficients for the Designed Tests

No.	Test Name	Unit	Correlation Coefficient (r)	Coefficient of Alienation (k)	Intrinsic Validity	Significance (p)
1	Star Flexibility Test (Hip Flexibility)	cm	0.82	0.56	0.90	0.001
2	Back Bridge Test (Spinal Flexibility)	m	0.80	0.60	0.89	0.001
3	Prone Rear Arm Raise Test (Shoulder Flexibility)	cm	0.81	0.58	0.90	0.001

Note. Significant at $p<0.05$.

Table 6 confirms that all constructed tests demonstrate high stability and reliability coefficients, supporting their suitability for evaluating specific flexibility in youth gymnasts.

2.4 Main Administration of the Specific Flexibility Tests

Following the completion of test design and psychometric verification, the construction sample ($N=60$) data obtained during the discriminant validation phase were utilized to derive standard scores and normative levels. This procedure follows established methodology (Al-Zubair, 2019; Yohana, 2023) to prevent repetitive testing on small populations. The primary testing period spanned January 25, 2026, to February 12, 2026, following these standardized protocols:

1. Administering tests following standardized warm-up supervised by coaches.
2. Providing sufficient rest intervals between successive tests to ensure physiological recovery.
3. Conducting testing across Erbil, Duhok, and Sulaymaniyah according to the schedule in Table 7.

Table 7. Schedule of Test Administration Across Clubs and Governorates

No.	Governorate	Clubs	Sample (n)	Date
1	Erbil	Erbil	9	January 25–26, 2026
		Sirwan	8	January 28–29, 2026
		Akad Ankawa	9	February 2–3, 2026
2	Duhok	Duhok	9	February 4–5, 2026
		Semel	8	February 6–7, 2026
3	Sulaymaniyah	Sulaymaniyah	9	February 8–9, 2026
		Newroz	8	February 11–12, 2026





2.5 Statistical Analysis

Data were analyzed using the Statistical Package for the Social Sciences (SPSS).

3. Results

Following data collection from the administration sample (N=60), statistical parameters were computed as shown in Table 8.

Table 8. Descriptive Statistical Parameters of the Specific Flexibility Tests

No.	Test Name	Mean (X ⁻)	± SD	Mode	Skewness	Distribution Shape
1	Star Flexibility Test (Hip Flexibility)	79.71	6.23	74	0.91	Moderate Positive
2	Back Bridge Test (Spinal Flexibility)	145.86	6.91	141	-0.70	Moderate Negative
3	Prone Rear Arm Raise Test (Shoulder Flexibility)	37.98	4.57	41	-0.66	Moderate Negative

Table 8 indicates that the skewness coefficients for all tests fell within the range of ± 1 , confirming a normal distribution suitable for establishing standard scores and generalizable to similar populations (Al-Tikriti & Al-Obeidi, 1999, p. 179).

Normative levels and standard scores (6σ) were subsequently derived, as shown in Tables 9 through 14.

Table 9. Normative Levels for the Star Flexibility Test (Hip Flexibility)

Normative Level	Raw Scores	Standard Scores	Frequency	Percentage (%)
Very Good	92.18 and above	82–100	1	1.66%
Good	85.95–92.17	64–81	6	10.00%
Moderate	79.72–85.94	48–63	22	36.66%
Acceptable	73.49–79.71	32–47	25	41.66%
Poor	67.26–73.48	15–31	4	6.66%
Very Poor	67.25 and below	0–14	2	3.33%
Total			60	100%

Table 10. Raw Scores and Six-Sigma (6σ) Standard Scores for the Star Flexibility Test (Hip Flexibility)

Raw Score	Standard Score	Raw Score	Standard Score	Raw Score	Standard Score
61	0	74	34	87	69
62	2	75	37	88	72
63	5	76	40	89	74
64	7	77	42	90	77
65	10	78	45	91	80
66	13	79	48	92	82
67	15	80	50	93	85
68	18	81	53	94	88
69	21	82	56	95	90
70	24	83	59	96	93
71	26	84	61	97	96
72	29	85	64	98	98
73	32	86	66	99	100

Table 11. Normative Levels for the Back Bridge Test (Spinal Flexibility)

Normative Level	Raw Scores	Standard Scores	Frequency	Percentage (%)
Very Good	132.03 and below	83–100	1	1.66%
Good	138.94–132.04	68–82	4	6.66%
Moderate	145.85–138.95	52–67	27	45.00%
Acceptable	152.76–145.86	35–51	22	36.66%
Poor	159.67–152.77	18–34	5	8.33%
Very Poor	159.68 and above	0–17	1	1.66%
Total			60	100%



Table 12. Raw Scores and Six-Sigma (6σ) Standard Scores for the Back Bridge Test (Spinal Flexibility)

Raw Score	Standard Score	Raw Score	Standard Score	Raw Score	Standard Score
166	1	152	35	138	68
165	3	151	37	137	71
164	6	150	40	136	73
163	8	149	42	135	76
162	11	148	44	134	78
161	13	147	47	133	81
160	15	146	49	132	83
159	18	145	52	131	85
158	20	144	54	130	88
157	23	143	56	129	90
156	25	142	59	128	93
155	27	141	61	127	95
154	30	140	64	126	97
153	32	139	66	125	100

Table 13. Normative Levels for the Prone Rear Arm Raise Test (Shoulder Flexibility)

Normative Level	Raw Scores	Standard Scores	Frequency	Percentage (%)
Very Good	47.13 and above	82–100	2	3.33%
Good	42.56–47.12	64–81	5	8.33%
Moderate	37.97–42.55	46–63	25	41.66%
Acceptable	33.42–37.98	31–45	22	36.66%
Poor	28.85–33.41	13–30	5	8.33%
Very Poor	28.84 and below	0–12	1	1.66%
Total			60	100%

Table 14. Raw Scores and Six-Sigma (6σ) Standard Scores for the Prone Rear Arm Raise Test (Shoulder Flexibility)

Raw Score	Standard Score	Raw Score	Standard Score	Raw Score	Standard Score
24	0	34	35	44	71
25	2	35	39	45	75
26	6	36	42	46	79
27	9	37	46	47	82
28	13	38	50	48	86
29	17	39	53	49	90
30	20	40	57	50	93
31	24	41	61	51	97
32	28	42	64	52	100
33	31	43	68	—	—

4. Conclusion and Recommendations

The three constructed specific flexibility tests—namely, the Star Flexibility Test (Hip Flexibility), the Back Bridge Test (Spinal Flexibility), and the Prone Rear Arm Raise Test (Shoulder Flexibility)—demonstrated high validity and reliability in assessing specific flexibility in youth gymnasts. Standard scores and normative levels were successfully established, with the majority of gymnasts scoring within the "Acceptable" and "Moderate" categories across all tests.

Based on these findings, the researcher recommends:

1. Adopting the specific flexibility tests constructed in the present study to evaluate youth gymnasts in the Kurdistan Region of Iraq.
2. Utilizing the derived standard scores and normative levels when assessing youth gymnasts on specific flexibility measures.





3. Urging gymnastics coaches to focus on developing specific flexibility during regular training sessions due to its fundamental role in enhancing athletic performance.

References

- Al-Abbasi, K. F. K. (2018). *Scientific research methods* (1st ed.). Dar Noon for Printing and Publishing.
- Al-Azzawi, R. Y. (2008). *Introduction to scientific research methodology* (1st ed.). Dar Dijlah for Printing and Publishing.
- Al-Fartousi, A. S., et al. (2015). *Measurement, testing, and evaluation in the sports field*. Al-Muhaimin Press.
- Al-Fateh, W. M. (2016). *Foundations of sports training for youth athletes*. Arab Foundation for Science and Culture.
- Al-Tikriti, W. Y., & Al-Obaidi, H. M. (1999). *Statistical applications and computer use in physical education research*. Directorate of Dar Al-Kutub for Printing and Publishing.
- Al-Zubair, A. M. H. (2019). *Constructing physical-technical tests in freestyle swimming for seniors* [Unpublished doctoral dissertation, College of Basic Education – University of Mosul].
- Bloom, B. S., et al. (1983). *Handbook on formative and summative evaluation of student learning* (M. A. Al-Mufti et al., Trans.). McGraw-Hill.
- Farahat, L. E. (2001). *Measurement and testing in physical education*. Book Publishing Center.
- Melhem, S. M. (2010). *Research methods in education and psychology* (6th ed.). Dar Al-Maseera for Publishing, Distribution, and Printing.
- Mikhael, E. N. (2016). *Construction and standardization of psychological and educational tests and scales* (1st ed.). Dar Al-I'sar Al-Ilmi for Publishing and Distribution.
- Mosaad, W. S. A., et al. (2022). Constructing a test battery for specific physical determinants in female artistic gymnasts aged 10 years. *Egyptian Journals*, Helwan University.
- Sabeeh, S. M. R., & Mustafa, A. A. A. (2022). Constructing a test battery for physical determinants (muscular power and flexibility) for female artistic gymnasts in light of under-10 compulsory routines. *Scientific Journal of Sports Science and Arts*, Helwan University.
- Youhanna, F. E. M. (2023). *Contribution percentages of selected technical skill tests to offensive tactical thinking among senior female volleyball players in Iraq* [Unpublished master's thesis, College of Basic Education – University of Mosul].