

Age- And Sex-Specific Physical Fitness Standards Across the Lifespan in Uzbekistan: A Comparative Analysis of Normative Design and International Reference Frameworks

Salimov Utkirjon Shaydullaevich

Doctor of Philosophy in Pedagogical Sciences (PhD), Associate Professor, Department of Physical Culture, Termez University of Economics and Service, Termez, Uzbekistan

Received: 20 August 2026; **Accepted:** 16 September 2026; **Published:** 10 October 2026

Abstract: Background: Age- and sex-specific physical fitness standards are widely used to evaluate population fitness, guide physical education and sport programmes, and support public-health surveillance. However, national systems differ substantially in test selection, age stratification, sex differentiation, normative design, and interpretation of performance thresholds. Uzbekistan has established a nationwide physical fitness testing system covering childhood, adolescence, adulthood, and older age, yet its age- and sex-specific standards have not been comprehensively examined in relation to major international reference frameworks.

Objective: This review aimed to analyse the age- and sex-specific normative structure of Uzbekistan's national physical fitness standards across the lifespan and to compare their design and interpretive characteristics with established international fitness assessment and reference frameworks.

Methods: A comparative review of the Uzbek national normative testing system was undertaken alongside international frameworks and evidence relevant to youth, adults, and older adults, including Eurofit, ALPHA-Fitness, FitBack, YFIT, FITNESSGRAM/NETFIT, adult field-based fitness evidence, and functional-fitness approaches for older populations. The comparison focused on age-group structure, sex-specific differentiation, test selection, fitness domains, normative versus criterion-referenced interpretation, threshold design, health relevance, and applicability to population surveillance.

Results: Uzbekistan's system provides a broad lifespan architecture comprising 11 age categories and age- and sex-specific Level I–III performance standards. Sex differentiation is evident from childhood and becomes increasingly pronounced in adulthood through both different thresholds and, in some cases, different test selection. Compared with international frameworks, the Uzbek system demonstrates greater breadth in combining health-related fitness, performance-oriented fitness, motor competence, sport-specific skills, and functional or adapted tasks. International systems, by contrast, more commonly employ smaller standardized test cores and more explicit distinctions between normative reference values, health-related criterion standards, and functional thresholds. The reviewed national documentation does not establish that Uzbekistan's Level I–III thresholds were derived from health outcomes or validated as health-risk cut-points.

Conclusions: Uzbekistan possesses a comprehensive age- and sex-differentiated physical fitness standards system with substantial potential for national monitoring. Future development should prioritize age-specific validation, transparent derivation of reference standards, clearer separation of performance norms from health-related criteria, sex-specific psychometric evaluation, and improved comparability with international population-fitness surveillance frameworks.

Keywords: Physical fitness; age-specific standards; sex-specific standards; normative reference values; fitness surveillance; criterion-referenced standards; lifespan assessment; Uzbekistan.

Introduction: Physical fitness is a multidimensional attribute that reflects the capacity to perform physical activity and encompasses components such as cardiorespiratory endurance, muscular strength and endurance, flexibility, speed, power, agility, and other motor-performance characteristics. Although physical activity and physical fitness are closely related, they are conceptually distinct: physical activity describes bodily movement and associated energy expenditure, whereas physical fitness represents a set of attributes that can be measured through standardized tests (Caspersen et al., 1985). This distinction is important for population monitoring because recommendations for physical activity, including the World Health Organization guidelines, do not themselves constitute physical fitness assessment systems (Bull et al., 2020). Consequently, national fitness-testing programmes require their own methodological foundations, age-appropriate protocols, reference values, and interpretive frameworks.

The public-health importance of physical fitness is well established. In children and adolescents, higher levels of cardiorespiratory and muscular fitness are associated with more favourable cardiometabolic, musculoskeletal, and other health indicators, and several components of youth fitness have demonstrated predictive relevance for later health outcomes (Ortega et al., 2008; Ruiz et al., 2009; Smith et al., 2014). In adults, cardiorespiratory fitness is strongly associated with morbidity, cardiovascular events, and all-cause mortality (Kodama et al., 2009; Lang et al., 2024). These findings have strengthened interest in the use of field-based fitness assessment not only for sport and physical education, but also for epidemiological surveillance and population health monitoring.

However, the value of a national fitness-testing system depends on more than the inclusion of commonly used exercises or performance tasks. The scientific meaning of test results is determined by how tests are selected, standardized, stratified, and interpreted. In particular, age and sex are central to normative design because physical growth, maturation, body composition, neuromuscular development, cardiorespiratory capacity, and functional performance change substantially across the lifespan. Sex-related differences also emerge and evolve across developmental stages and adulthood. Therefore, a single undifferentiated performance standard is unlikely to provide meaningful interpretation across children, adolescents, adults, and older adults. National systems commonly respond to this challenge by establishing age- and sex-specific test batteries, reference values, or performance thresholds.

At the same time, age- and sex-specific standards can be constructed according to different interpretive principles. Norm-referenced standards describe an individual's position relative to a reference population, commonly using percentiles or standardized distributions. Criterion-referenced standards classify performance against an external criterion, such as a health outcome, functional capacity, or clinically meaningful threshold. A third approach consists of administrative or achievement standards, in which performance categories are established for educational, sporting, military, or national programme purposes without necessarily being linked to a specific health outcome. These approaches are not interchangeable. A percentile does not automatically indicate health risk, and an official performance level should not be interpreted as a health-related cut-point unless its criterion validity has been established against an appropriate external outcome (Zhu et al., 2011; Welk et al., 2011).

International physical fitness frameworks illustrate substantial variation in these design principles. Eurofit represents a broad multidimensional approach to youth fitness assessment and has generated large-scale international normative data across multiple components of physical performance (Tomkinson et al., 2018). ALPHA-Fitness was developed around health-related fitness and placed greater emphasis on the validity, reliability, feasibility, and health relevance of selected field tests (Ruiz et al., 2011). More recent systems such as FitBack have integrated standardized assessment, reference values, feedback, and population-surveillance functions, while the YFIT initiative has sought to define a concise, internationally applicable youth fitness-testing battery through expert consensus and evidence synthesis. FITNESSGRAM and NETFIT provide further examples of systems in which health-related or criterion-referenced interpretation is integrated with school-based assessment and digital data management. In older adults, by contrast, functional-fitness frameworks place greater emphasis on mobility, strength, balance, and the physical capacity required for independent living (Rikli and Jones, 1999, 2013).

These international models demonstrate that the optimal design of a fitness-testing system depends strongly on its intended purpose. A battery designed for school physical education may appropriately include motor competence and sport-specific tasks, whereas a population-health surveillance system may favour a smaller number of highly standardized tests with well-established validity, reliability, feasibility, and health relevance. Similarly, testing older adults requires different priorities from testing adolescents,

because maintenance of functional independence may be more important than sprint speed or sport-specific performance. The methodological challenge is therefore not simply to determine which tests are used, but to evaluate whether test selection, age stratification, sex differentiation, and threshold interpretation are aligned with the intended surveillance or performance objective.

Uzbekistan has developed a national physical fitness testing system that covers the population across the lifespan and applies age- and sex-specific standards within a formal national regulatory framework. The system includes 11 age categories extending from early school age to older adulthood and specifies graded Level I, II, and III performance standards. It also contains a broad range of endurance, strength, flexibility, speed, power, motor-skill, sport-specific, and adapted functional tasks. This breadth provides a substantial national structure for assessing physical preparedness across different population groups. At the same time, such a heterogeneous architecture raises important methodological questions concerning the basis on which age and sex categories were defined, the extent to which standards represent population norms or administrative achievement levels, and whether individual thresholds can be interpreted in relation to health or functional outcomes.

A further issue concerns comparability with international reference systems. Many contemporary fitness-surveillance frameworks increasingly emphasize smaller standardized core batteries, explicit test-selection criteria, psychometric validation, and clear separation between normative and criterion-referenced interpretation. By contrast, broader national programmes may combine health-related fitness with performance-oriented, motor-competence, and sport-specific objectives. Such differences do not necessarily indicate methodological weakness; rather, they reflect different policy and educational purposes. Nevertheless, they can limit cross-national comparability if the interpretive meaning of national thresholds is not clearly defined.

Despite the existence of a comprehensive national testing framework in Uzbekistan, its age- and sex-specific normative design has received limited attention in the international comparative literature reviewed for this study. A systematic examination of the architecture is therefore warranted, particularly in relation to the principles used by major international fitness assessment systems. Such an analysis can help distinguish structural strengths from evidence gaps and can clarify which elements are most suitable for performance evaluation, public-health surveillance, or

future criterion-based interpretation.

Accordingly, the aim of this review is to analyse the age- and sex-specific structure of Uzbekistan's national physical fitness standards across the lifespan and to compare their normative design with established international reference frameworks. Particular attention is given to age stratification, sex-specific differentiation, test selection, fitness domains, normative and criterion-referenced interpretation, health relevance, functional appropriateness across life stages, and suitability for population surveillance. The review also seeks to identify methodological strengths, areas requiring further validation, and priorities for improving the scientific interpretability and international comparability of Uzbekistan's national physical fitness standards.

The aim of this review was to analyse the age- and sex-specific structure of Uzbekistan's national physical fitness standards across the lifespan and to compare their normative design with established international reference frameworks in order to identify methodological strengths, interpretive limitations, evidence gaps, and priorities for future validation and population fitness surveillance.

LITERATURE REVIEW

Age- and sex-specific physical fitness standards are fundamental to the interpretation of human physical performance because fitness characteristics change substantially across growth, maturation, adulthood, and ageing. The methodological challenge is not merely to establish average performance values, but to determine how reference standards should be stratified by age and sex, how population distributions should be modelled, and whether thresholds are intended to represent relative performance or clinically meaningful levels of function. The available international literature demonstrates three major approaches to this problem: large-scale percentile-based normative reference systems, life-course modelling of fitness trajectories, and criterion-referenced standards linked to health or functional outcomes. Together, these approaches provide an important comparative foundation for evaluating the normative design of national physical fitness systems, including that of Uzbekistan.

Age- and sex-specific normative standards in children and youth. One of the strongest international examples of age- and sex-specific normative development in young populations is the study by Tomkinson et al. (2017). Using data from 1,142,026 children and adolescents representing 50 countries, the authors developed international normative values for the 20 m shuttle run. The scale of this dataset is particularly

important because it enabled the construction of reference distributions across a wide range of ages and both sexes, rather than relying on small national or regional samples.

The methodological importance of this work extends beyond the 20 m shuttle run itself. Tomkinson et al. demonstrated how physical fitness performance can be interpreted using age- and sex-specific percentile values. Such an approach recognizes that a raw performance value has limited meaning unless it is interpreted in relation to individuals of comparable developmental stage and sex. A shuttle-run result that represents high performance at one age may represent only average performance at another age, while sex-related differences become increasingly relevant during adolescence.

This normative model is particularly useful for international benchmarking because it provides a common comparative reference rather than relying exclusively on country-specific cut-points. However, its interpretive function remains primarily norm-referenced. Percentiles indicate where an individual stand relative to a reference distribution; they do not necessarily indicate whether that level of fitness corresponds to increased or reduced health risk. This distinction is essential when comparing such international reference systems with national achievement standards.

For Uzbekistan, the methodological relevance of the Tomkinson et al. approach lies in its explicit separation of age and sex and in the use of population distributions to construct interpretable reference values. If national performance standards are intended to describe relative population fitness, a percentile-based framework provides a transparent model for determining whether a given threshold corresponds, for example, to the 25th, 50th, 75th, or 90th percentile of national performance. Such an approach could also facilitate future comparisons between Uzbek populations and international youth reference distributions.

Physical fitness standards across the life course. The importance of age-specific interpretation is not limited to children and adolescents. Dodds et al. (2014) provide a valuable life-course model through their analysis of grip strength across 12 British studies. Rather than examining a narrow age range, the study considered normative patterns across the life course, allowing the authors to describe how muscular strength changes from earlier adulthood into later life.

This life-course perspective is methodologically important because physical fitness does not change in a linear manner. Many physical attributes develop

during childhood and adolescence, reach their highest levels during early or middle adulthood, and subsequently decline with ageing. Therefore, normative standards designed for national use should reflect the shape of these age-related trajectories rather than simply applying uniform intervals or assuming constant rates of change.

Dodds et al. also demonstrate the importance of sex-specific analysis. Grip strength differs substantially between males and females, and the magnitude and trajectory of these differences vary across age. Consequently, combined-sex normative values would obscure meaningful biological and functional variation. This finding supports the broader principle that sex-specific normative design is not merely an administrative convention but a methodological requirement when the underlying physical characteristic shows systematic sex-related variation.

For a lifespan national fitness system such as Uzbekistan's, this evidence suggests that age categories should ideally correspond to meaningful stages in the development and decline of physical capacities. It also raises the question of whether all fitness components should use identical age-group boundaries. Different components may follow different age-related trajectories, and the most appropriate grouping for muscular strength may not necessarily be identical to that for endurance, speed, flexibility, or functional mobility.

Normative functional fitness in older adults. The interpretation of physical fitness becomes particularly important in older populations, where the purpose of assessment increasingly shifts from athletic or performance-oriented evaluation toward functional capacity and independence. Rikli and Jones (1999) made a major contribution to this field by developing normative functional fitness scores from 7,183 community-dwelling adults aged 60–94 years.

A notable methodological strength of this work is the use of sex-specific and five-year age-group norms. Such relatively narrow age intervals are especially relevant in older populations because functional capacity may change substantially over a relatively short period. Grouping all individuals aged 60 years and older into a single category could therefore conceal important differences between, for example, adults aged 60–64 and those aged 85–89.

The Rikli and Jones normative model also illustrates how the content of physical fitness assessment should evolve with age. In older adults, tests related to functional strength, mobility, flexibility, and endurance become especially important because these characteristics are closely connected to the ability to

perform everyday activities. This differs conceptually from younger-age test systems in which sprint speed, explosive power, sport skill, or competitive performance may receive greater emphasis.

The implication for lifespan national standards is that age adaptation should involve more than changing numerical thresholds. The structure of the battery itself may need to change. A scientifically coherent lifespan system should therefore consider whether the selected test at each age remains appropriate to the functional priorities of that life stage.

From percentile norms to criterion-referenced standards. An important development in the literature is the transition from purely normative interpretation toward criterion-referenced standards. Rikli and Jones (2013) addressed this issue by developing and validating clinically relevant fitness standards associated with maintaining physical independence in later life.

This distinction is fundamental. Norm-referenced standards answer the question:

How does this individual perform relative to others of the same age and sex?

Criterion-referenced standards address a different question:

Is this level of fitness sufficient to achieve or maintain a meaningful health or functional outcome?

The Rikli and Jones approach is therefore particularly relevant to the interpretation of national physical fitness standards. A performance threshold can only be described as a health- or function-related criterion if it has been linked empirically to an external outcome. Administrative categories, medals, grades, or achievement levels do not become criterion-referenced health standards merely because they are age- and sex-specific.

For Uzbekistan, this conceptual distinction is especially important when interpreting graded national standards. If national levels were established primarily as achievement classifications, they should be described as performance standards unless evidence demonstrates that specific thresholds predict health status, disease risk, functional limitation, or another relevant external criterion. International evidence therefore suggests that future validation should distinguish between the reliability and validity of the test itself and the validity of the threshold used to interpret the test.

Cross-cultural differences in normative fitness values. National standards cannot always be transferred directly from one country to another. Marques et al. (2014), using data from 4,712 Portuguese adults aged

65–103 years, developed normative functional fitness standards and examined cross-cultural comparisons. Their percentile-based analysis demonstrates that functional performance varies not only by age and sex but also across populations.

Cross-cultural variation may reflect differences in physical activity patterns, body composition, socioeconomic conditions, health status, environmental context, lifestyle, and testing practices. Consequently, international reference values provide useful benchmarks but should not automatically replace national reference data.

This issue is particularly important for Uzbekistan. International norms can help identify broad methodological expectations and provide comparative context, but national standards should ideally be supported by representative national data. A threshold established in a British, Portuguese, North American, or multinational sample may not have the same distributional meaning in the Uzbek population.

Therefore, the most defensible role of international reference systems is comparative rather than substitutive. They can inform test selection, statistical modelling, age-group design, percentile construction, and validation strategies, while national normative values should ultimately reflect the characteristics of the population to which they are applied.

Contemporary evidence on benchmark values in older adults. The systematic review by Cossio-Bolaños et al. (2024) provides a broader synthesis of functional fitness benchmark values for older adults. By integrating findings across multiple studies, this review highlights both the usefulness and the heterogeneity of existing reference standards.

The international literature reviewed by Cossio-Bolaños et al. confirms that benchmark development in older adults commonly relies on age- and sex-specific values. At the same time, variation across studies in test protocols, age categories, sample characteristics, statistical approaches, and reference classifications limits direct comparability. This methodological heterogeneity is an important lesson for national fitness systems: the value of a reference standard depends on rigorous standardization of the test protocol used to generate and apply it.

The review also reinforces the need to distinguish between descriptive benchmark values and clinically or functionally meaningful thresholds. Percentiles are valuable for identifying relative position within a population, whereas criterion-referenced cut-points require independent validation against outcomes such as mobility, functional independence, disability, or other clinically meaningful endpoints.

Implications for normative design across the lifespan.

Taken together, the reviewed studies demonstrate several principles that are directly relevant to the design and evaluation of national physical fitness standards.

First, age-specific reference values are indispensable. The studies of Tomkinson et al., Dodds et al., Rikli and Jones, and Marques et al. all show substantial age-related variation in physical performance. However, the appropriate width of age categories may differ across the lifespan. Broad categories may be sufficient during periods of relative stability, whereas narrower categories may be preferable during rapid growth or advanced ageing.

Second, sex-specific standards are methodologically justified when systematic differences exist in the underlying fitness characteristic. The international evidence consistently models males and females separately rather than assuming that a single threshold can be applied to both groups.

Third, normative and criterion-referenced standards must remain conceptually separate. Tomkinson et al., Dodds et al., Rikli and Jones (1999), and Marques et al. primarily provide population-relative normative information. Rikli and Jones (2013), by contrast, demonstrate how thresholds can be linked to an external functional criterion. These two approaches answer different scientific questions and should not be conflated.

Fourth, test selection should change with life stage. Youth assessment may appropriately emphasize cardiorespiratory fitness and performance-oriented measures, while older-adult assessment should increasingly prioritize functional strength, mobility, flexibility, endurance, and independence-related capacities.

Fifth, international comparability requires protocol standardization. A national test that resembles an international test cannot automatically be assumed to have identical normative, validity, or criterion properties. Differences in protocol, distance, timing method, equipment, scoring rules, or participant preparation may materially affect the interpretation of results.

Relevance to the Uzbekistan national system. Against this international evidence base, Uzbekistan's age- and sex-specific physical fitness standards can be evaluated along three distinct dimensions: normative structure, threshold interpretation, and lifespan appropriateness.

From a normative perspective, the existence of age- and sex-specific standards is consistent with international practice. The central question, however,

is how the national thresholds were derived. International best practice demonstrates the value of large representative samples, sex-specific modelling, age-specific distributions, and clearly reported statistical procedures.

From an interpretive perspective, the reviewed literature shows that national performance levels should not be assumed to represent health-related criteria unless they have been validated against relevant external outcomes. Percentile position, achievement level, and health-related cut-point are analytically distinct categories.

Finally, from a lifespan perspective, international evidence suggests that the content and meaning of physical fitness assessment should evolve with age. A national system extending from childhood into older adulthood therefore requires not only different numerical standards but also age-appropriate measurement priorities and validation strategies.

Summary. The international literature provides a strong methodological basis for evaluating age- and sex-specific physical fitness standards. Tomkinson et al. (2017) illustrate large-scale international percentile modelling in youth; Dodds et al. (2014) demonstrate life-course normative modelling of muscular strength; Rikli and Jones (1999) establish detailed age- and sex-specific functional norms for older adults; Rikli and Jones (2013) show how criterion-referenced thresholds can be linked to physical independence; Marques et al. (2014) demonstrate the importance of national and cross-cultural normative differences; and Cossio-Bolaños et al. (2024) synthesize contemporary international evidence on functional-fitness benchmarks in older adults.

Collectively, these studies suggest that the scientific evaluation of Uzbekistan's national standards should focus not only on whether standards differ by age and sex, but also on how those standards were derived, what type of reference they represent, whether the underlying protocols have been standardized and validated, and whether their interpretation is appropriate to the functional priorities of each stage of the lifespan.

RESULTS AND DISCUSSION

1. Lifespan structure of Uzbekistan's age-specific physical fitness standards. The comparative analysis shows that Uzbekistan has established a broad age-differentiated physical fitness testing system extending from early school age to older adulthood. The national framework is organized into 11 age categories: 7–8, 9–10, 11–12, 13–15, 16–17, 18–29, 30–39, 40–49, 50–59, 60–69, and 70 years and older. This structure provides continuity across the lifespan and differs from many

international reference systems that focus on a narrower developmental stage, most commonly children and adolescents or older adults.

The principal strength of this approach is structural continuity. Rather than treating school-age fitness, adult fitness, and older-adult functional capacity as entirely separate policy areas, the Uzbek system places them within a common national architecture. From a surveillance perspective, such continuity could provide a useful basis for longitudinal population monitoring if the protocols, interpretation rules, and reference standards are sufficiently standardized.

However, the existence of age categories alone does not establish that the category boundaries are physiologically optimal. International evidence suggests that age-related changes in physical fitness are component-specific. Cardiorespiratory endurance, muscular strength, speed, flexibility, and functional capacity do not follow identical trajectories across the life course. Dodds et al. (2014), for example, demonstrated that grip strength develops, peaks, and declines according to a characteristic age-related pattern, with clear differences between men and women. Similarly, the work of Rikli and Jones (1999) indicates that relatively narrow five-year age bands are particularly informative in later life, when physical function may decline substantially over comparatively short periods.

Accordingly, Uzbekistan's use of predefined age categories represents a strong administrative framework, but future empirical validation should determine whether the same age-group boundaries are equally appropriate for all fitness components.

2. Sex-specific differentiation in the national standards. A prominent characteristic of the Uzbek system is the extensive differentiation of standards by sex. This differentiation begins in childhood and becomes increasingly evident through adolescence and adulthood.

In the healthy population section, the great majority of youth test rows contain different Level III, II, and I values for boys and girls. At ages 7–8, sex-specific values differ in 18 of 19 test rows; at ages 9–10, in 18 of 20; at ages 11–12, in 21 of 22; at ages 13–15, in all 22; and at ages 16–17, in all 25 relevant rows. Thus, sex differentiation is not limited to late adolescence but is embedded throughout most of the childhood and adolescent normative architecture.

The differentiation becomes more pronounced in adulthood because the system sometimes changes not only the performance threshold but also the test itself. For example, among adults aged 18–29 years, men are assessed using a 3000 m endurance run and strength

tasks such as pull-ups or kettlebell exercises, whereas women may be assessed using a 2000 m run and alternative upper-body strength tasks.

From a normative perspective, this design is broadly consistent with international practice. Tomkinson et al. (2017), in developing international 20 m shuttle-run norms from more than 1.1 million children and adolescents in 50 countries, modelled performance separately by age and sex. Dodds et al. (2014) likewise demonstrated pronounced sex-specific differences in grip strength across the life course. Rikli and Jones (1999) and Marques et al. (2014) also used sex-specific reference values when evaluating functional fitness in older adults.

These international studies support the methodological principle underlying the Uzbek system: sex-specific interpretation is necessary where meaningful differences exist in the distribution and trajectory of physical performance. Nevertheless, sex differentiation should ideally be based on empirical population distributions rather than solely on predetermined administrative expectations. The key methodological question is therefore not whether male and female standards should differ, but how the magnitude of those differences was established.

3. Comparison with percentile-based international normative models. One of the clearest differences between the Uzbek system and major international reference frameworks concerns the method of interpreting performance.

The international normative model developed by Tomkinson et al. (2017) provides percentile values for 20 m shuttle-run performance across age and sex groups. Such percentile-based systems have an important advantage: they show where an individual lies within a reference distribution. A child performing at the 75th percentile is understood to perform better than approximately three quarters of the reference population of the same age and sex.

A comparable statistical logic is evident in Dodds et al. (2014), who generated age- and sex-specific normative grip-strength values across the life course, and in Marques et al. (2014), who developed percentile distributions for functional fitness among 4,712 Portuguese adults aged 65–103 years.

The Uzbek national system uses a different interpretive architecture. Instead of continuous percentile distributions, performance is classified into three achievement categories: Level III, Level II, and Level I. This approach is administratively straightforward and may be useful for certification, motivation, physical education, or national achievement programmes. However, it provides less information about the

position of an individual within the population distribution.

For example, a Level II standard does not inherently indicate whether the threshold corresponds to the 40th, 60th, or 80th percentile of the national population. Without an underlying national reference distribution, the statistical meaning of each level remains uncertain.

This distinction is important because categorical achievement standards and percentile-based norms serve different functions. Achievement standards identify whether a predefined performance level has been reached, whereas percentile norms describe relative standing within a population.

4. Age-specific interpretation from childhood to adulthood. The Uzbek standards show a clear shift in test content across age groups. In younger age categories, the system contains a relatively high proportion of sprinting, jumping, throwing, coordination, and motor-skill tasks. As age increases, endurance and strength tasks become more prominent, while some highly performance-oriented tasks are reduced or modified.

This progression reflects a reasonable lifespan logic. Childhood and adolescence are periods of motor development, neuromuscular maturation, and acquisition of fundamental movement skills. Consequently, speed, power, coordination, and object-control tasks may be relevant within a broader physical education framework.

However, international normative systems increasingly distinguish between health-related fitness and broader motor performance. Tomkinson et al. (2017), for instance, focused on cardiorespiratory fitness using the 20 m shuttle run, whereas the Uzbek system incorporates a much wider range of tasks. This makes the national system more comprehensive from a physical-performance perspective but also creates interpretive complexity.

The broader the test battery becomes, the more important it is to clarify what each test is intended to represent. A sprint test may primarily reflect speed, a standing long jump explosive power, and a ball-accuracy task motor competence or sport-related skill. These outcomes should not automatically be interpreted as equivalent markers of health-related fitness.

Thus, one of the central methodological issues is whether the national system is intended primarily as a physical preparedness programme, a health surveillance system, a motor-development assessment system, or a combination of these purposes. The

present analysis suggests that it currently combines several functions.

5. Adults: the need for life-course normative modelling. The transition from adolescence to adulthood is particularly important in normative design. In the Uzbek framework, the 18–29-year category is relatively broad, followed by 30–39, 40–49, and subsequent ten-year categories.

The life-course data presented by Dodds et al. (2014) demonstrate that physical capacity changes systematically across adulthood. Grip strength, for example, follows a trajectory of development, peak performance, and subsequent decline. This evidence suggests that adult standards should ideally be based on empirically observed age-related trajectories rather than assumed uniform differences between decades.

For Uzbekistan, this has two implications. First, broad adult age groups may be administratively efficient but may conceal meaningful differences within the category. A 19-year-old and a 29-year-old, for example, may not necessarily belong to the same normative distribution for every fitness component. Second, different physical fitness components may require different modelling strategies.

Future national reference studies could therefore examine continuous age functions in addition to categorical age groups. This would permit the development of smoothed age-specific centiles and provide a stronger empirical basis for deciding where category boundaries should be placed.

6. Older adults: transition from performance to functional fitness. The international literature provides particularly strong evidence that the interpretation of physical fitness should change in older age.

Rikli and Jones (1999), using a sample of 7,183 community-dwelling adults aged 60–94 years, developed age- and sex-specific normative standards for functional fitness. Their approach emphasized capacities relevant to everyday functioning rather than athletic performance alone. Marques et al. (2014) reached similar conclusions in Portuguese adults aged 65–103 years, demonstrating the value of age- and sex-specific percentile norms for functional fitness.

The systematic review by Cossio-Bolaños et al. (2024) further confirms that functional fitness benchmarks in older populations are strongly dependent on age, sex, test protocol, and population characteristics.

The Uzbek system does demonstrate an age-related transition toward walking, modified strength, flexibility, swimming, and other less performance-intensive tasks. This is a positive structural feature. However, the international literature suggests that the

key issue in older populations is not merely whether a test becomes easier with age, but whether the test measures a function that is relevant to mobility, independence, and daily living.

This distinction is especially important when considering the 60–69 and 70+ categories. At these ages, a national fitness system may derive greater public-health value from identifying functional limitation than from ranking athletic performance.

7. Norm-referenced standards versus criterion-referenced cut-points. One of the most important findings of the comparative analysis is the need to distinguish the three principal forms of performance interpretation: norm-referenced values, criterion-referenced thresholds, and administrative achievement standards.

Tomkinson et al. (2017), Dodds et al. (2014), Rikli and Jones (1999), and Marques et al. (2014) primarily demonstrate norm-referenced approaches. Their standards describe how individuals perform relative to age- and sex-specific populations.

Rikli and Jones (2013), by contrast, developed criterion-referenced standards linked to the maintenance of physical independence in later life. This represents a fundamentally different methodological approach. A cut-point is meaningful because it is related to an external functional outcome, not simply because it separates higher and lower performers.

The Uzbek Level I–III standards currently align most closely with an achievement-standard model. The national documentation specifies performance thresholds for different age and sex groups, but the reviewed evidence does not demonstrate that these thresholds were derived from health outcomes, disability risk, functional limitation, or other external criteria.

Therefore, Level I–III values should not be described as health-risk cut-points unless such validation is subsequently demonstrated.

This distinction has practical consequences. A person who fails to achieve a particular national level cannot automatically be classified as unhealthy, just as a person who achieves Level I cannot automatically be classified as being at low health risk. The system currently provides information about performance relative to predetermined national standards, not necessarily health status.

8. International comparability and transferability of reference values. The reviewed literature also highlights the limitations of directly transferring international norms to Uzbekistan.

Tomkinson et al. (2017) provide a powerful

international reference model, but the sample represents pooled data from 50 countries. Dodds et al. (2014) analysed British populations, while Marques et al. (2014) focused on Portuguese older adults. These datasets are valuable methodological comparators but cannot automatically be treated as national normative standards for Uzbekistan.

Population-specific differences may arise from physical activity patterns, body size, maturation, nutrition, socioeconomic conditions, environmental context, cultural practices, health status, and methodological differences in testing.

Accordingly, international norms are most appropriately used as comparative benchmarks, not substitutes for nationally derived reference values.

For Uzbekistan, future large-scale data collection could allow each national standard to be located within both a national and international distribution. Such dual interpretation would provide two complementary forms of information: relative position within Uzbekistan and relative position within an international reference framework.

9. Sex-specific standards: threshold differences versus test differences. An important methodological issue emerging from the Uzbek system is that sex differentiation occurs in two forms.

The first is threshold differentiation, in which males and females perform the same test but are assigned different standards. This approach is common internationally and is generally compatible with normative design.

The second is test differentiation, in which men and women perform different tests or different distances. This is more difficult to interpret scientifically because the resulting scores may represent related but non-equivalent constructs.

For example, a 3000 m run and a 2000 m run both measure endurance-related performance but are not directly equivalent. Similarly, pull-ups and low-bar pull exercises assess related dimensions of upper-body muscular performance but differ biomechanically and in relative loading.

This does not mean that sex-specific test selection is necessarily inappropriate. It may be justified by safety, feasibility, tradition, or population characteristics. However, it reduces direct comparability between male and female results.

International normative systems such as those described by Tomkinson et al. and Dodds et al. more commonly preserve the same measurement construct and protocol while modelling sex-specific reference distributions. From a measurement perspective, this

facilitates clearer comparison across sexes.

Therefore, future refinement of the Uzbek system may benefit from distinguishing circumstances in which sex-specific thresholds are sufficient from those in which genuinely different tests are required.

10. Threshold derivation and the need for transparent methodology. The strongest methodological difference between the international literature and the Uzbek national framework concerns transparency in the derivation of standards.

The studies reviewed here generally report their sample characteristics, age and sex stratification, statistical methods, distribution modelling, and validation procedures. Tomkinson et al. (2017) provide an example of large-scale international normative modelling; Dodds et al. (2014) demonstrate life-course modelling; Rikli and Jones (2013) explicitly link thresholds to an external functional criterion.

In contrast, the reviewed Uzbek normative documentation specifies the performance standards themselves but does not report the empirical procedure by which the Level I–III thresholds were derived. The document therefore allows the analyst to determine what the standards are, but not fully why those particular numerical thresholds were selected.

This represents an evidence gap rather than proof that no underlying empirical work exists. Additional technical reports, unpublished national datasets, expert-panel documentation, or institutional analyses may exist outside the reviewed source.

Nevertheless, from the standpoint of scientific transparency, future editions of the national standards would benefit from explicit documentation of sample size, sampling strategy, age and sex distribution, statistical modelling, percentile or criterion derivation, and validation procedures.

11. Internal consistency of age- and sex-specific standards. The documentary analysis also identified several internal inconsistencies that are relevant to normative interpretation.

In some disability-specific sections, an age-category title refers to 30–39 years while the underlying columns specify 40–44 and 45–49 years. The same structural mismatch appears across several population groups. In another section, the final category is phrased as “up to 70 years,” although the overall national sequence indicates a 70+ category.

A further example occurs among healthy boys aged 16–17 years in the voluntary 100 m test, where the reported values are Level III = 16.5 s, Level II = 16.3 s, and Level I = 16.5 s. This produces a non-monotonic sequence because Level I is not superior to Level II.

Such issues should not be silently corrected without an authoritative source or erratum. However, they demonstrate why national normative systems require formal quality-control procedures before publication and periodic revision thereafter.

In a standards-based system, even a small numerical inconsistency can alter classification, particularly near threshold values. Consequently, transparent version control and technical validation are essential components of normative governance.

12. Implications for future national reference development. The international evidence reviewed in this study suggests that the next stage in the development of Uzbekistan’s physical fitness standards should focus less on adding new tests and more on strengthening the empirical basis of the existing framework.

A scientifically robust national reference system could incorporate several complementary stages. First, representative population data should be collected using standardized protocols. Second, age- and sex-specific distributions should be modelled, preferably using methods capable of generating smoothed centile curves. Third, current Level I–III thresholds could be mapped onto these national distributions. This would reveal whether each level corresponds to a stable percentile range across age and sex.

Fourth, where health or functional interpretation is intended, candidate thresholds should be tested against independent outcomes. The methodology used by Rikli and Jones (2013) provides a useful conceptual model for such criterion validation in older adults.

Fifth, the appropriateness of sex-specific test substitution should be examined empirically. Where feasible, using the same construct and protocol with sex-specific norms may improve comparability.

Finally, standards should be periodically updated. Normative values can change across generations because of secular trends in body size, lifestyle, physical activity, nutrition, and health. Consequently, national standards should be viewed as dynamic reference systems rather than permanent fixed values.

13. Overall interpretation. The comparative evidence indicates that Uzbekistan possesses an unusually broad age- and sex-differentiated national fitness architecture with substantial potential for population-level assessment. Its major strength lies in its lifespan coverage and formal differentiation across demographic groups.

At the same time, the international literature demonstrates that the scientific quality of normative standards depends not only on the existence of age-

and sex-specific thresholds but on the transparency of their derivation and the clarity of their intended interpretation.

The principal methodological distinction is therefore between having age- and sex-specific standards and having empirically derived age- and sex-specific reference standards.

Uzbekistan clearly possesses the former. The next research priority is to establish the latter through representative population data, explicit statistical modelling, psychometric evaluation, and, where appropriate, validation against health or functional outcomes.

In this respect, the international frameworks reviewed here should not be regarded as templates to be copied directly. Rather, they provide methodological reference points for strengthening a national system that already has substantial structural breadth. A future Uzbek reference framework combining national percentile distributions, carefully defined achievement standards, and health- or function-related criteria where empirically justified could substantially improve both domestic interpretation and international comparability.

CONCLUSIONS

This review demonstrates that Uzbekistan has developed a broad age- and sex-specific physical fitness standards system covering the lifespan from early school age to older adulthood. The national framework comprises 11 age categories and applies differentiated Level I–III performance standards for males and females, with sex differences evident from childhood and becoming more pronounced in adulthood through both threshold differentiation and, in some cases, different test selection.

The comparative analysis shows that the Uzbek system has important structural strengths, particularly its lifespan coverage, demographic differentiation, and inclusion of a wide range of endurance, strength, flexibility, speed, power, motor-skill, sport-specific, and functional tasks. However, its normative design differs from major international reference frameworks. International systems more commonly rely on empirically derived percentile distributions, standardized test protocols, life-course modelling, or criterion-referenced thresholds linked to health or functional outcomes. In contrast, the reviewed Uzbek documentation specifies Level I–III standards but does not report the statistical or health-related procedures used to derive these thresholds.

Accordingly, the current Level I–III values should be interpreted primarily as national performance or

achievement standards, rather than as population percentiles or health-related criterion cut-points, unless additional validation evidence is available. This distinction is particularly important for preventing the inappropriate classification of individuals as healthy, unhealthy, functionally independent, or at risk solely on the basis of achievement-level performance.

The analysis also indicates that age-specific assessment should involve more than adjusting numerical thresholds. Different stages of the lifespan require different measurement priorities. In children and adolescents, fitness assessment may legitimately include speed, power, coordination, and motor competence alongside health-related fitness. In adults, standardized endurance and muscular-fitness measures become increasingly important, while in older adults the emphasis should progressively shift toward functional capacity, mobility, strength, flexibility, endurance, and maintenance of physical independence.

Several practical priorities emerge from these findings. Future development of Uzbekistan's national standards should focus on collecting representative national fitness data; constructing age- and sex-specific percentile distributions; examining whether existing age-category boundaries reflect actual life-course changes in performance; evaluating whether sex-specific test substitution is necessary or whether common protocols with sex-specific reference values would improve comparability; and documenting the reliability, validity, and feasibility of national test protocols. Where health or functional interpretation is intended, candidate thresholds should be validated against independent health, mobility, or functional outcomes rather than inferred from performance categories alone.

Particular attention should be given to older populations, for whom functional independence may provide a more meaningful criterion than performance ranking, and to periodic revision of normative values in response to secular changes in physical fitness, lifestyle, and population characteristics. Internal inconsistencies in age labels, units, and threshold sequences identified in the national documentation also indicate the need for formal technical quality control and version management when standards are updated.

Overall, Uzbekistan already possesses a substantial organizational foundation for nationwide physical fitness assessment. The principal next step is not simply to expand the number of tests, but to strengthen the empirical derivation, psychometric validation, interpretive transparency, and international

comparability of the existing standards. A future system that combines nationally representative percentile references, clearly defined achievement standards, and health- or function-related criterion thresholds where empirically justified would provide a stronger basis for physical education, population fitness monitoring, public-health surveillance, and evidence-informed national policy.

REFERENCES

1. Ministry of Tourism and Sports of the Republic of Uzbekistan. Jismoniy tayyorgarlik darajasi sport tests: types and criteria by age categories. Official normative document; 2021. Developed pursuant to Presidential Resolution No. PQ-5148 of 16 June 2021. Uzbek.
2. Tomkinson GR et al. (2017). International normative 20 m shuttle run values from 1,142,026 children and youth representing 50 countries. *British Journal of Sports Medicine*, 51(21), 1545–1554. DOI: 10.1136/bjsports-2016-095987.
3. Dodds RM et al. (2014). Grip Strength across the Life Course: Normative Data from Twelve British Studies. *PLOS ONE*, 9(12), e113637. DOI: 10.1371/journal.pone.0113637.
4. Rikli RE, Jones CJ. (1999). Functional Fitness Normative Scores for Community-Residing Older Adults, Ages 60–94. *Journal of Aging and Physical Activity*, 7(2), 162–181. DOI: 10.1123/japa.7.2.162.
5. Rikli RE, Jones CJ. (2013). Development and Validation of Criterion-Referenced Clinically Relevant Fitness Standards for Maintaining Physical Independence in Later Years. *The Gerontologist*, 53(2), 255–267. DOI: 10.1093/geront/gns071.
6. Marques EA et al. (2014). Normative functional fitness standards and trends of Portuguese older adults: cross-cultural comparisons. *Journal of Aging and Physical Activity*, 22(1), 126–137. DOI: 10.1123/japa.2012-0203.
7. Cossio-Bolaños M et al. (2024). Functional fitness benchmark values for older adults: a systematic review. *Frontiers in Public Health*, 12, 1335311. DOI: 10.3389/fpubh.2024.1335311.