Tun J Sport Sci Med 2023; 1(1):32-47
10.61838/kman.tjssm.1.1.5
Tunisian Journal of Sports Science and Medicine
Review Article

Guidelines for Applying Psychometrics in Sports Science: Transitioning from Traditional Methods to the AI Era

Noomen Guelmami[1, 2, 4*], Mohamed Ben Aissa[1, 3, 4], Achraf Ammar[5, 6], Ismail Dergaa[1, 3, 5, 7], Khaled Trablesi[5], Haitham Jahrami[8, 9]

[1] Research Unit, Sportive Performance and Physical Rehabilitation, High Institute of Sports and Physical Education of Kef, University of Jendouba, 7100, Kef, Tunisia
[2] Postgraduate School of Public Health, Department of Health Sciences (DISSAL), University of Genoa, Genoa, Italy
[3] Research Unit Physical Activity, Sport, and Health, UR18JS01, National Observatory of Sport, Tunis 1003, Tunisia
[4] High Institute of Sports and Physical Education of Ksar Said, University of Manouba 2010, Tunis, Tunisia
[5] High Institute of Sport and Physical Education of Sfax, University of Sfax, Sfax 3000
[6] Johannes Gutenberg-University Mainz; Germany; Institute for Sport Science; Albert Schweitzer Straße 22; 55122 Mainz; Germany
[7] Primary Health Care Corporation (PHCC), Doha, Qatar
[8] College of Medicine and Medical Sciences, Arabian Gulf University, Manama, Kingdom of Bahrain
[9] Government Hospitals, Manama, Kingdom of Bahrain

*** Corresponding author email address**: mbenaissa.hs@gmail.com

Received: 2023-12-03 **Reviewed:** 2023-12-24 **Revised:** 2023-12-25 **Accepted:** 2023-12-26 **Published:** 2023-12-29

Abstract

Background: The transition from conventional methods to the AI era in sports science emphasizes a critical need for comprehensive guidance in effectively applying psychometrics. This imperative is highlighted by dynamic transformations, where the integration of structured guidelines becomes indispensable for adeptly navigating the challenges and seizing the opportunities presented by advancing technologies.

Aim: The aim of this study was to methodically outline and enhance the application of psychometrics in sports science, focusing on the transition from traditional methods to the artificial intelligence era. It sought to provide a clear, objective-driven framework for effectively utilizing psychometrics, particularly emphasizing integration with artificial intelligence technologies.

Methods: The research employed a dual-pronged approach, developing theoretical frameworks and practical considerations. It outlined key principles for selecting and implementing psychometric tools, prioritizing reliability and validity. Additionally, the study closely examined ethical considerations linked to AI-driven psychometrics in sports, focusing on areas like privacy, potential bias, and the importance of maintaining transparency in these practices.

Results: The study provided guidelines that bridged traditional psychometrics with emerging AI technologies in sports science. The results offered a roadmap for researchers, coaches, and practitioners, facilitating the transition to more robust assessment methodologies. Emphasis on data analysis and ethical considerations ensured responsible and effective integration of psychometrics and AI for athlete evaluation and development.

Conclusion: This research offers a comprehensive framework for navigating the intersection of psychometrics and AI in sports science. Addressing both theoretical and practical aspects, the guidelines empower stakeholders to transition responsibly to advanced assessment methodologies in the evolving technological landscape.

Keywords: *Artificial Intelligence, Machine Learning, Assessment, Validity, Reliability, Physical Activity, Data Analysis*

How to cite this article:

Guelmami N, Ben Aissa M, Ammar A, Dergaa, I, Trablesi K, Jahrami H. Guidelines for Applying Psychometrics in Sports Science: Transitioning from Traditional Methods to the AI Era. *Tun J Sport Sci Med*. 2023;1(1):32-47.

1. Introduction

Psychometrics, the scientific discipline dedicated to measuring psychological attributes, stands at the intersection of psychology and statistics (1). This scientific field began with the advent of intelligence testing in the early 20th century (2). Since then, it has expanded to include a wide array of assessments of different psychological constructs. It encompasses the rigorous development, validation, and application of assessments to evaluate various psychological constructs, including intelligence, personality traits, and mental health conditions (3).

Traditionally, the field has been firmly rooted in creating and validating assessments such as questionnaires and standardized tests, with a strong emphasis on ensuring the reliability, validity, and fairness of these measures (2). The process of developing a psychometric test is meticulous, beginning with a clear definition of the construct to be measured, followed by item generation and selection, where questions are carefully crafted to reflect various aspects of the construct accurately (4). This phase is often accompanied by pilot testing and refinement to ensure that the questions are understandable, relevant, and unbiased (5).

Developing comprehensive guidelines on the application of psychometrics in sports sciences involves a detailed exploration of the methods, ethics, and practical applications of psychological measurements in assessing and enhancing athletic performance and well-being. This extensive guide aims to provide sports psychologists, coaches, and sports organizations with a thorough understanding of how to effectively utilize psychometrics in the sports context.

2. Understanding Psychometrics in Sports Context

Psychometrics in sports involves the scientific measurement and analysis of mental capacities, personality traits, motivation, stress, anxiety levels, and other psychological aspects that can be correlated with the effects of physical activity (6). Identifying psychological strengths and weaknesses, formulating mental training programs, and aiding in the overall development of athletes are crucial. Mental aspects have a profound impact on physical performance, injury recovery, team dynamics, resilience, and the overall well-being of athletes. Understanding an athlete's mental state can help tailor training programs, enhance performance under pressure, and contribute to a healthier sporting environment (7).

3. Choosing Appropriate Psychometric Instruments

Instruments that have been rigorously tested and validated through scientific research were chosen. This means that the tool has been proven to accurately and consistently measure what it is intended to measure. For instance, a tool designed to assess mental health should have a strong ability to accurately identify characteristics associated with mental health in athletes (7).

4. Validity of the Questionnaires: An In-depth Analysis

Survey are fundamental tools for collecting data across various fields, from the social sciences to health research. The validity of a questionnaire significantly influences the reliability and applicability of its findings. Validity refers to the extent to which a questionnaire measures what it intends to measure. This comprehensive analysis delves into the different aspects of questionnaire validity, exploring its importance, types, challenges, and methods to enhance it.

Preliminary development and cross-cultural validation represent different stages in the process of developing and validating a measurement instrument, often in the context of psychological or educational assessments (8).

Preliminary or initial development is concerned with creating and refining a measurement instrument in its early stages, ensuring that it captures the intended construct effectively. Cross-cultural validation, on the other hand, is a later stage in which the instrument's performance and generalizability are assessed across diverse cultural contexts. Table 1 presents an explanation of each stage (8, 9).

Table 1. Differences between Preliminary (Initial Development) and Cross-Cultural Validation.

Aspect	Preliminary/Initial Development	Cross-cultural Validation
Purpose	Creation and refinement of a new measurement instrument	Assessment of the instrument’s performance across cultures
Methodology	Defining the construct, generating items, refining through expert reviews and pilot testing.	Adapting the instrument for different cultural contexts, considering language and cultural norms, and administering it to diverse samples
Objective	Ensure conceptual clarity, relevance, and comprehensive coverage of the intended construct.	Confirm the validity and reliability of the instrument across various cultural settings.
Focus	Early stages of instrument development.	Later stages, assessing cross-cultural applicability.
Key Processes	Conceptualization, item generation, expert reviews, cognitive interviews, and iterative pilot testing.	Translation, cultural adaptation, and systematic administration to diverse cultural groups.
Outcome	A refined instrument with enhanced conceptual validity and relevance of items	Validation of the instrument's psychometric properties across different cultures

5. Types of validity

5.1. Defining Content Validity

Content validity is a crucial aspect of questionnaire design, reflecting the extent to which the instrument represents all facets of the concept being measured. This approach is a fundamental step in ensuring that the questionnaire accurately captures the comprehensive scope of the subject matter (10).

At its core, content validity concerns the relevance and representativeness of questionnaire items. It gauges whether the questions in the questionnaire cover the entire range of elements that constitute the concept or construct being measured. For instance, in a questionnaire assessing knowledge of healthy eating, content validity would require questions that address all key nutritional components, such as vitamins, minerals, macronutrients, and dietary guidelines (4).

Literature Review: Conducting a thorough review of the literature, research studies, and theoretical frameworks related to the concept ensures that the questionnaire encompasses all necessary content areas.

Expert Consultation: Involving subject matter experts in the questionnaire development process can provide valuable insights into the essential components of the concept. Experts can review and provide feedback on the questionnaire items, ensuring that they are comprehensive and relevant.

Item Development and Review: Developing a comprehensive list of items based on the literature review and expert input, followed by a meticulous review process to refine and validate each item.

Pilot Testing: Administering the questionnaire to a representative sample of the target population can reveal whether the items are understood as intended and whether any important aspects of the concept are missing.

5.2. Content validity indices

Content validity refers to the extent to which a test or measurement instrument adequately covers the full range of the construct it is intended to measure. It is a critical aspect of test development and validation, ensuring that the instrument truly reflects the theoretical components of the concept being measured. This concept is closely linked to the use of the content validity ratio (CVR), a specific statistical tool. The CVR helps quantify this aspect of content validity by measuring the degree to which individual test items are viewed as essential by subject matter experts (11).

The content validity ratio (CVR) is a statistical measure used to quantify the extent to which a test or survey item is deemed essential by subject matter experts (SMEs). It is a key component in establishing content validity and refers to how accurately an assessment or measurement tool represents the concept it is intended to measure. The CVR plays a crucial role in the development and evaluation of psychological tests, educational assessments, and research questionnaires.

The CVR is calculated based on the judgments of a panel of experts. A group of SMEs is selected. This group should have adequate knowledge or expertise in the relevant domain. Each expert is asked to rate each item in terms of its essentiality, often using a scale such as "essential," "useful but not essential," or "not essential". (12)

CVR for Each Item: The CVR for each item was calculated using the following formula:

$$\mathbf{CVR = (n_e - N/2)/(N/2)}$$

$\mathbf{n_e}$ is the number of experts indicating the item as "essential”

N is the total number of experts.

The CVR value ranges between -1 and +1. A positive CVR indicates that more than half of the experts consider the item essential. A negative value suggests less agreement on the item's essentiality.

The acceptable level of the CVR depends on the number of experts. Higher thresholds are generally required when fewer experts are involved. Lawshe developed a table to determine the minimum CVR value for an item to be considered valid based on the number of experts. Items with a CVR value above the threshold are typically retained, while those below the threshold may be revised or removed from the assessment (13).

Content validity index (CVI): Utilizing statistical measures such as the CVI, where experts rate the relevance of each item, can provide a quantitative assessment of content validity. The CVI was calculated based on the proportion of experts who agreed on the relevance of each item. Establishing content validity is not without challenges. A balance between comprehensiveness and practicality is required because too many items can lead to respondent fatigue, while too few may miss critical aspects of the concept. Additionally, the subjective nature of expert judgments can introduce variability in the assessment of content validity (14).

5.2.1. Item Content Validity Index (I-CVI)

The item content validity index (I-CVI) is another component of content validity assessment in psychometrics, specifically focusing on the individual items within a scale or instrument. The scale measures the agreement among experts on the content validity of each item separately.

The formula for I-CVI is often expressed as follows:

$$\boldsymbol{I-CVI} = \frac{number\ of\ experts\ who\ rated\ the\ item\ as\ content\ valid\ (rated\ 3\ or\ 4)}{total\ number\ of\ items\ in\ the\ scale}$$

In this formula, the "number of experts who rated the item as content valid" refers to the number of experts who agree that a specific item is relevant and representative of the construct being measured. The "total number of experts" is the overall number of experts who provided ratings for that item (15).

Typically, a rating of 3 or 4 is considered indicative of content validity for a particular item. The I-CVI is calculated for each item individually, providing insight into the level of agreement among experts on the relevance of each item (14).

Researchers may use the I-CVI in conjunction with the Scale Content Validity Index (S-CVI), such as the S-CVI/Ave and S-CVI/UA, to comprehensively assess the content validity of a scale or instrument. Specific guidelines and criteria for assessing content validity may vary; therefore, it is important to refer to the relevant literature or guidelines associated with a specific study or field (16).

5.2.2. Scale Content Validity (S-CVI)

The Scale Content Validity Index (S-CVI) is a measure used in psychometrics to assess the content validity of a scale or instrument, focusing on the relevance and representativeness of individual items. There are two common methods: the S-CVI/Ave (scale content validity index average) and the S-CVI/UA (scale content validity index universal agreement) (15, 16).

For S-CVI/Ave, the formula is:

$$\boldsymbol{S-CVI_{/ave}} = \frac{number\ of\ items\ rated\ as\ content\ valid\ (rated\ 3\ or\ 4)}{total\ number\ of\ items\ in\ the\ scale}$$

This method calculates the average content validity index across all items based on expert ratings. A rating of 3 or 4 often indicates that an item is considered content valid.

On the other hand, the S-CVI$_{UA}$ assesses the proportion of items with universal agreement among experts (15, 16). The formula is:

$$\boldsymbol{S-CVI_{/UA}} = \frac{number\ of\ items\ with\ universal\ agreement}{total\ number\ of\ items\ in\ the\ scale}$$

Universal agreement implies that all experts agree that a particular item is relevant and representative of the construct being measured. The goal is to achieve a high S-CVI/UA, indicating strong consensus among experts regarding the content validity of the items (16).

5.3. Construct validity

Construct validity refers to the degree to which a test measures what it claims, or purports, to be measuring. In the context of questionnaire design, whether the questionnaire accurately reflects the specific concept or theory it is based on should be evaluated (17). This validity plays a pivotal role in the development of a questionnaire, assessing whether it accurately measures the theoretical constructs it purports to. This form of validity is essential for ensuring that the questionnaire not only measures random or unrelated concepts but is also truly aligned with the specific theoretical constructs of interest. This approach is crucial because it underpins the questionnaire's credibility and relevance, ensuring that the inferences drawn from the questionnaire's results are legitimate and applicable to the intended theoretical framework (17).

5.3.1. Processes for Ensuring Construct Validity

Theoretical Framework Establishment: Initially, it was vital to define the theoretical constructs the questionnaire was intended to measure. This approach involves a clear understanding of the concept and its components, often grounded in a comprehensive review of relevant literature.

Item Alignment with Theoretical Constructs: Each item in the questionnaire must be carefully designed to align with the established theoretical framework. This ensures that the items collectively measure the construct as defined.

Statistical correlation and analysis: Employing statistical methods such as factor analysis, correlation coefficient analysis, and regression analysis helps in quantitatively assessing construct validity. These methods determine how well the questionnaire items correlate with each other and with the theoretical construct.

5.3.2. Construct Validity with Exploratory and Confirmatory Factor Analysis

Construct validity in questionnaire design is integral to verifying that the instrument accurately measures the intended theoretical constructs (18). A key part of establishing this validity is the use of exploratory factor analysis (EFA) and confirmatory factor analysis (CFA), two statistical methods that help in understanding and confirming the underlying structure of the questionnaire (19).

Exploratory Factor Analysis (EFA)

EFA is often the first step in assessing construct validity. It is used to explore the underlying factor structure of a questionnaire without preconceived notions of what this structure should be. EFA involves identifying groups of items that are correlated with each other but less so with items in other groups. These groups represent different factors or constructs within the questionnaire. The main goal is to uncover the number of latent factors and the pattern of relationships among items, providing insights into whether the items collectively measure the intended constructs. EFA is particularly useful in the initial stages of questionnaire development when the factor structure is unknown or when theoretical assumptions have yet to be fully developed (20).

EFA is used to uncover the underlying structure of a relatively large set of variables. Commonly, the underlying relationships between measured variables are identified, and the data are reduced to a smaller set of summary variables (20). EFA is particularly useful when the researcher does not have an a priori hypothesis about the relationships among variables (21). There are several techniques for exploratory factor analysis, each with its own unique approach and application:

Parallel analysis

Parallel analysis is a statistical technique employed in factor analysis to ascertain the appropriate number of factors to retain in a dataset (22). This method involves a comparison between the eigenvalues derived from actual data and those obtained from randomly generated data, often through Monte Carlo simulations. Eigenvalues, which indicate the variance explained by each factor in factor analysis, are examined to identify factors that exceed the values expected by chance in random datasets. The primary aim is to avoid the pitfalls of either over-extracting or under-extracting factors, ensuring that the retained factors are likely to represent the genuine underlying patterns present in the observed data (23).

Principal component analysis (PCA)

Description: PCA is a technique that converts a set of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. This method is used primarily for data reduction. It is used to summarize the data, reduce its dimensionality, and highlight patterns (24).

Common Factor Analysis (CFA)

Common factor analysis is utilized to identify latent constructs underlying a set of observed variables. The process involves extracting factors from the data using principal component analysis, followed by factor rotation (orthogonal or oblique) for clearer interpretability. It assesses factor loadings and communalities to understand the relationships between observed variables and underlying factors. Determining the appropriate number of factors to retain is a crucial step, guided by criteria such as Kaiser’s criterion or the scree test (25).

Principal Axis Factoring (PAF)

The PAF attempts to explain the variance–covariance structure of a set of variables through a few underlying but unobservable random quantities known as factors. This method is commonly used in cases where the goal is to identify the underlying latent constructs (26).

Maximum Likelihood Factor Analysis (MLFA)

MLFA is a method of factor analysis in which the factors are estimated by maximizing the likelihood of the observed correlations among variables, assuming multivariate normality. It is used when the sample size is large, and the assumption of multivariate normality can be reasonably met (27).

Image Factoring

This method involves calculating the 'image' of a variable, which is the part predicted by all the other variables, and then factoring these images. Image factoring is less commonly used but can be effective in specific contexts where partial correlations are of interest (28).

Alpha Factoring

Alpha factoring is based on maximizing the reliability (alpha) of the factors. This technique focuses on maximizing the internal consistency of the identified factors. This method is useful where reliability is a key concern (29).

Unweighted least squares (ULS)

ULS factor analysis is a method in which factors are derived based on minimizing the sum of squared discrepancies between the observed and estimated correlation matrices. The model was used when the normality assumption was not met, and the goal was to minimize overall discrepancies in correlations (30).

Generalized least squares (GLS)

In GLS, factors are derived by minimizing the weighted sum of squared discrepancies between the observed and estimated correlation matrices. This method is more efficient than the ULS under multivariate normality and is used when specific weighting of discrepancies is desired (31).

Rotation methods: Quartimax, Varimax, and Oblimin rotation

These are rotation methods applied after the extraction of factors. Quartimax and varimax rotation are orthogonal rotations (factors are uncorrelated), while oblimin rotation allows for correlated factors (32). These rotations are used to make the factor structure more interpretable. Varimax rotation is the most commonly used method for determining orthogonal rotation. Each type of EFA has its strengths and is suitable for different research scenarios. The choice of method depends on the specific objectives of the study, the nature of the data, and the underlying assumptions that can be reasonably met (32).

Confirmatory Factor Analysis

CFA is used after EFA or when there is a preexisting theory regarding the factor structure of a questionnaire (19). Unlike EFA, CFA tests a specific hypothesis about the factor structure, usually based on the results of EFA or theoretical considerations. This process involves specifying a model and then assessing how well the data fit this model. CFA is used to confirm or refute the factor structure suggested by EFA or theory (33). This approach provides a means to test the construct validity of a questionnaire by verifying whether the hypothesized factor structure is consistent with the data. This method is crucial for fine-tuning the questionnaire and ensuring that it accurately reflects the theoretical constructs. It is particularly valuable to finalize the questionnaire for use in larger-scale studies (33, 34).

Confirmatory factor analysis (CFA) is a statistical technique used in the field of structural equation modeling (SEM) to test whether a set of observed variables measures the number of constructs (factors) that are expected theoretically. It differs from exploratory factor analysis (EFA) in that CFA is hypothesis-driven, testing a prespecified structure, whereas EFA is more exploratory in nature and is used to identify potential structures. In CFA, a researcher starts with a hypothesis about how many factors are present and which observed variables are related to which factors (33). This hypothesized model was then tested statistically. CFA deals with latent variables (unobserved variables) that are inferred from observed variables. These latent variables represent constructs such as intelligence, satisfaction, or motivation. The model includes factor loadings, which indicate the strength and direction of the relationship between the observed variables and their underlying latent factors. CFA models also incorporate error terms for each observed variable, acknowledging measurement error or variance in the observed variables not explained by the latent factors (34).

Model Types in Confirmatory Factor Analysis (CFA)

In confirmatory factor analysis (CFA), different model types are used to represent the complexity and structure of the data (33). Understanding these model types is crucial in sports science research for appropriately analyzing and interpreting various constructs, such as psychological traits, team dynamics, or physical performance indicators. The main types are as follows:

Each of these models serves a specific purpose and is chosen based on the research question and the nature of the data used in sports science research. The complexity of the model increases from first order to third order, with each providing a more detailed representation of the underlying constructs and their interrelationships.

Table 2. CFA models with descriptions and examples

Type of CFA Model	Description	Example in Sports Science and Physical Education
First-Order CFA Model	Basic form where observed variables are directly related to their underlying latent factors.	Arabic Mood Scale (ARAMS) (35). teacher of physical education job satisfaction inventory (TPEJSI) (36).
Second-Order CFA Model	Involves first-order factors influenced by a higher-order (second-order) factor.	Physical Education Grit Scale (PE-GRIT) (37). Arabic Coach-Athlete Relationship Questionnaire (ACART-Q) (38).
Third-Order CFA Model	Hierarchical model with a third-order factor influencing several second-order factors, which then influence first-order factors.	Arabic Ottawa mental skills assessment tool (AOMSAT-3) (39).
Multidimensional CFA Model	Used when a construct is best represented by multiple, correlated factors.	Arabic Questionnaire on Teacher Interaction (AQTI) (40). Arabic Game Experience Questionnaire (A-GEQ) (41).
Multitrait-Multimethod (MTMM) CFA Model	Analyzes data where each trait is measured by multiple methods, assessing convergent and discriminant validity.	Physical Inactivity Perceived Experience Scale (PIPES) (42).
Longitudinal CFA Model	Used for data collected at multiple time points to assess stability and change over time.	Beck Scale for Suicide Ideation (BSS) (43).

Model evaluation

Goodness of Fit: The key part of CFA is assessing how well the hypothesized model fits the actual data (44). This is done using various fit indices:

Table 3. Fit indices of CFA

Fit Index	Description
Chi-Square Statistic (χ^2)	Assesses the discrepancy between observed and expected covariance matrices. A nonsignificant value ($p > 0.05$) indicates good model fit but is sensitive to sample size.
Comparative Fit Index (CFI)	Compares the hypothesized model with a null model. Values close to 1 (≥ 0.95) indicate a good fit.
Tucker–Lewis Index (TLI)	Also known as the Non-Normed Fit Index (NNFI), it penalizes model complexity. Values close to 1 (≥ 0.95) suggest a good model fit.
Root Mean Square Error of Approximation (RMSEA)	Assesses model fit per degree of freedom, accounting for model complexity. Values ≤ 0.06 indicate a good fit; values up to 0.08 are acceptable.
Standardized Root Mean Square Residual (SRMR)	The standardized difference between observed and predicted correlations. Values less than 0.08 typically indicate a good fit.
Incremental Fit Index (IFI)	Measures the relative improvement in fit of the hypothesized model compared to a null model. Values close to 1 (≥ 0.95) indicate a good fit.
Bentler-Bonett Normed Fit Index (NFI)	Compares the chi-square value of the model with that of the null model. Values above 0.90 indicate a good fit.
Relative Fit Index (RFI)	Similar to NFI but adjusts for degrees of freedom. Values above 0.90 suggest a good model fit.
Adjusted Goodness-of-Fit Index (AGFI)	Adjusts the Goodness-of-Fit Index for the number of degrees of freedom. Values above 0.85 indicate a good fit.
Parsimony Normed Fit Index (PNFI)	Considers the parsimony of the model, rewarding simpler models. Higher values indicate better parsimonious fit.
McDonald's Noncentrality Index (NCI)	Based on noncentral chi-square distribution and sensitive to sample size. Higher values indicate a better fit.
Akaike Information Criterion (AIC)	Used for model comparison; lower values indicate a better fit.
Bayesian Information Criterion (BIC)	Similar to AIC, used for model comparison with lower values indicating a better fit.

Model Modification: Based on the fit indices, researchers may modify the model (e.g., by adding or removing paths) to improve fit. However, such modifications should be theory-driven and not just for the sake of achieving a better fit.

5.3.3. Combining EFA and CFA in Establishing Construct Validity

The combination of EFA and CFA offers a comprehensive approach to establishing construct validity.

EFA can be used initially to explore the possible factor structure, followed by CFA to confirm this structure (17).

This two-step approach allows for both the discovery of underlying constructs and the rigorous testing of a hypothesized model, ensuring a robust assessment of the questionnaire's validity (17).

5.3.4. Convergent and discriminant validity

Convergent validity was established when the questionnaire showed a high correlation with other measures of the same construct. Discriminant validity, on the other hand, is established when the questionnaire does not correlate with unrelated constructs, thereby confirming that the questionnaire measures the intended construct rather than something else (45).

5.3.5. Criterion-related validity

While closely related to construct validity, criterion-related validity (both concurrent and predictive) involves correlating the questionnaire with external measures. If these correlations align with theoretical expectations, this further substantiates the questionnaire’s construct validity.

5.3.6. Factorial Invariance in Sports Science Research

Introduction to Factorial Invariance Factorial invariance is a statistical concept crucial in validating that a measurement tool or test has the same meaning and structure across different groups or conditions. In sports science, instruments such as performance tests, psychological scales, or health questionnaires are interpreted similarly across various groups, such as athletes from different sports, genders, age groups, or cultural backgrounds (46).

Factorial invariance confirmed that a tool measure was constructed consistently across different athlete populations. For instance, a mental toughness scale must be invariant to be equally applicable to both male and female athletes. This approach allows for valid comparisons between groups. Without invariance, differences in scores might reflect measurement biases rather than true differences in the construct.

Factorial invariance is a cornerstone in the psychometric evaluation of measurements. This ensures that the constructs being measured hold the same meaning across different groups, which is essential for the validity of cross-group comparisons and interpretations (46). Understanding and testing factorial invariance helps researchers and practitioners in sports science make accurate and meaningful comparisons across diverse athlete populations.

Testing Factorial Invariance

Configural Invariance: The basic structure (the number and pattern of factors) of the questionnaire or test was the same across groups. This is the most basic level of invariance (47).

Metric (or Weak) Invariance: Factor loadings are the same across groups. This implies that the construct is conceptualized in the same way across different groups (48).

Scalar (or Strong) Invariance: Both factor loadings and intercepts are the same across groups. This level is necessary for comparing means across groups (48).

Strict Invariance: Involves the equality of factor loadings, intercepts, and residuals. This is the strongest form of invariance, indicating that even measurement errors are equal across groups (48).

Factorial invariance is typically tested using multigroup confirmatory factor analysis (CFA). This involves comparing the fit of a measurement model across different groups and testing progressively more constrained models (configural, metric, scalar, strict).

Fit indices such as the comparative fit index (CFI), Tucker–Lewis index (TLI), and root mean square error of approximation (RMSEA) are used to assess model fit at each level (49). Complex constructs such as psychological traits or physical abilities can have varying interpretations across cultures or genders, making invariance testing challenging. Large and diverse samples are often needed to accurately test invariance, which can be a limitation in some sports science studies.

5.4. Estimating Reliability in Sports Science

Reliability in sports science research is critical for ensuring that measurements, such as athletic performance tests or psychological assessments, are consistently accurate. This parameter refers to the degree to which these measurements provide stable and error-free results over time and across various conditions. In this context, reliability is crucial for both the precision of the data and for drawing valid conclusions (50).

Test-retest reliability is a fundamental method in sports science for assessing the consistency of measurements over time. This form of reliability is critical, especially when evaluating psychological traits or physical performance that can vary with time (50).

In sports science, test-retest reliability involves administering the same psychometric or physical performance test to the same group of athletes at two different points in time. The consistency of the results is subsequently evaluated.

5.4.1. *Coefficient calculation and interpretation*

Reliability coefficients, ranging from 0 (no reliability) to 1 (perfect reliability), quantify the degree of consistency. A high reliability coefficient, closer to 1, indicates that the measurement is stable over time. For example, in a study measuring the reaction times of sprinters, a reliability coefficient of 0.92 suggested a high level of consistency in the measurement across the two testing sessions.

5.4.2. *Methods to Confirm Test-Retest Reliability*

Statistical analysis: The most common method for confirming test-retest reliability is through correlation coefficients, such as Pearson’s or Spearman’s correlation coefficients, depending on the data type. These statistical tests assess the relationship between the scores from the first and second tests.

Appropriate Time Interval: The interval between the two tests is crucial. The duration should be long enough to prevent recall or learning effects but short enough to ensure that no significant change in the athlete's condition or psychological state has occurred.

Consistency in Test Conditions: To ensure reliability, the testing conditions (e.g., environment, equipment, time of day) must be as consistent as possible across both sessions.

Sample Size Consideration: A sufficiently large sample size can provide more accurate reliability estimates, as it reduces the impact of outliers or individual variances on the overall results.

Test-retest reliability is an essential measure in sports science research, particularly when assessing variables that may fluctuate over time, such as psychological resilience or physical endurance. A high reliability coefficient is indicative of the measurement's stability, enhancing the credibility of the research findings. Employing robust statistical methods, maintaining consistent testing conditions, and choosing an appropriate time interval are all crucial for accurately determining test-retest reliability in sports science studies.

5.4.3. *Interrater reliability*

Interrater reliability refers to the degree of agreement or consistency between different raters or judges when they evaluate, score, or assess the same phenomenon. It is a crucial tool in research and clinical settings where subjective judgments or assessments are made, as it ensures that the evaluations are not dependent on a single individual and can be replicated. This approach ensures that different individuals assessing the same phenomenon arrive at similar conclusions. This approach enhances the generalizability of the results, as it indicates that the findings are not unique to a specific rater (51).

5.4.4. *Internal consistency: Ensuring uniformity within measurements*

Understanding Internal Consistency Internal consistency is a critical aspect of reliability in sports science research, particularly when dealing with psychometric assessments or physical performance tests. It refers to the extent to which different parts of a measurement tool (such as a questionnaire or a performance test) consistently measure the same construct (29).

Significance in Sports Science In the context of sports science, internal consistency ensures that various items or components of a test reliably assess the same underlying factor. For instance, in a questionnaire evaluating an athlete's mental health, all items should collectively measure aspects of mental health resilience.

5.4.5. *Coefficient calculation and interpretation*

The most common measure of internal consistency is the Cronbach's alpha coefficient, with values ranging from 0 (no internal consistency) to 1 (perfect internal consistency). A high Cronbach's alpha (typically above 0.7) indicates that the items within the test reliably measure the same underlying construct. For example, a Cronbach's alpha of 0.85 on a team cohesion scale suggests a high level of internal consistency among the scale's items (52).

In sports science research, along with Cronbach's alpha, McDonald’s omega and Guttman’s lambda are also critical in assessing the internal consistency of measurement tools. These methods provide a more nuanced understanding of reliability, which is particularly useful when dealing with complex and multidimensional constructs.

Table 4. Descriptions and typical thresholds of reliability coefficients

Reliability Coefficient	Description	Typical Thresholds
Cronbach's Alpha (α)	Measures internal consistency of a test or scale. Higher values indicate better reliability.	Acceptable: >0.7, Good: >0.8, Excellent: >0.9
Guttman's Lambda-6	A measure of internal consistency, similar to Cronbach's Alpha, but often considered more robust.	Similar to Cronbach's Alpha
Guttman's Lambda-2	Another variant of Guttman's lambda for internal consistency. Less commonly used than Lambda-6.	Similar to Cronbach's Alpha
McDonald's Omega (ω)	Estimates the test score's reliability based on a model of the factor structure of the test items.	Acceptable: >0.7, Good: >0.8, Excellent: >0.9
Cohen's Kappa (κ)	Assesses the agreement between two raters, adjusting for agreement occurring by chance.	Slight: 0.01-0.20, Fair: 0.21-0.40, Moderate: 0.41-0.60, Substantial: 0.61-0.80, Almost Perfect: 0.81-1.0

5.5. *Relevance to Sport*

The selected tool must be relevant to the particular requirements of the sport. For example, team sports such as soccer or basketball might require tools that measure group dynamics, communication skills, and cooperative behavior. In contrast, individual sports such as tennis or gymnastics might demand tools focused on self-motivation, stress management, and concentration.

Adaptability to High-Pressure Situations: Sports often involves high-stress, high-stakes environments. Therefore, psychometric instruments should be capable of assessing how an athlete handles pressure, makes decisions under stress, and cope with the mental challenges of competitive sports.

Table 5. Recommended software and packages

Software Tool/Package	EFA	CFA	Reliability Analysis	Sensitivity Analysis	IRT Analysis	AI Capabilities
R (with psych package)	Extensive (Various rotation methods)	Yes (With lavaan)	Yes (Various methods)	Yes (With additional packages)	Yes (With mirt, ltm)	Advanced (With various machine learning packages)
JASP	Yes	Limited (Basic support)	Yes	Limited	No	Limited
PSPP	Yes	No	Yes	No	No	No
Scikit-learn (Python)	Yes (Basic)	No	No	Yes (General libraries)	No	Advanced (Machine learning and data analysis)
Jamovi	Yes	Limited (R integration)	Yes	No	Limited (R integration)	Limited (Through R integration)
AMOS (IBM SPSS)	No	Extensive	Yes	No	No	No
Mplus	Yes	Extensive	Yes	Yes	Yes	No
TensorFlow/Keras (Python)	No	No	No	Advanced (General libraries)	No	Advanced (Deep learning)
PyTorch (Python)	No	No	No	Advanced (General libraries)	No	Advanced (Deep learning)
SAS/STAT	Yes	Extensive	Yes	Yes	Yes	Limited (Machine learning capabilities)

5.6. *Cross-cultural validation*

In sports science, cross-cultural validation is vital for ensuring that tests, surveys, or measurement tools are equally valid and reliable across various cultural contexts. This practice is crucial in research that spans different countries or cultural groups, particularly in a field where athletes, teams, or sports phenomena are compared globally. The goal is to ensure that these instruments measure the intended constructs similarly across cultures without being influenced by cultural biases (9).

The process typically begins with the translation and back-translation of the instrument, such as questionnaires on team dynamics or mental toughness, into the language of the target culture. This step is essential for maintaining the consistency of the content across different languages. Next, cultural adaptation is undertaken, where items are adjusted to ensure that they are culturally relevant and appropriate. This might involve modifying specific terms or references that do not translate well culturally, such as idiomatic expressions or culturally specific sports terminology (53).

Following adaptation, the instrument undergoes pilot testing with a sample from the target culture, which could include athletes, coaches, or sports teams. This step helps identify any issues or misunderstandings specific to that cultural context (54). Subsequently, psychometric testing was conducted to assess the reliability and validity of the instrument within the new cultural setting. This approach ensures that key constructs are accurately measured across cultures.

Finally, a comparative analysis is performed, comparing the results across different cultural groups. This approach is crucial for assessing the cross-cultural applicability of the instrument and identifying any significant cultural influences on sports-related behaviors or attitudes. Through these steps, cross-cultural validation in sports science contributes to making research findings universally applicable and inclusive, enhancing understanding and communication among diverse sports communities.

5.7. Ethical considerations

Informed Consent: Athletes should be fully informed about the nature, purpose, and potential use of the psychometric assessments, and their consent should be obtained, especially in the case of minors.

Confidentiality: Athletes' psychological data must be treated with utmost confidentiality. Only authorized personnel should have access to this information, and it should be used solely for the intended purpose of enhancing athlete performance and well-being.

Nondiscrimination: Psychometric assessments should be used as tools for positive development and not as a basis for discrimination or exclusion from sports activities.

5.8. Administering psychometric assessments

Professional Administration: Only qualified professionals, such as sports psychologists, should administer these assessments. They should have a deep understanding of both the psychological aspects and the specific sports dynamics.

Environment: Assessments should be conducted in a comfortable and nonthreatening environment to ensure that athletes can respond authentically and without bias. Timing: The timing of the assessment should be strategically planned to avoid interference with the athlete's training or competition schedule and to ensure the accuracy of the data.

5.9. Artificial Intelligence (AI) in Psychometrics: An Emerging Paradigm

The roles of AI in psychometrics can be categorized into several key areas:

The role of AI in psychometrics is transformative, offering enhanced precision, efficiency, and scope in the measurement of psychological constructs. Its integration is pushing the boundaries of traditional psychometric practices, opening up new possibilities for research, assessment, and intervention in the field of psychology (55).

Psychometrics contributes significantly to AI through the development of reliable and valid assessments. These assessments are used to measure cognitive abilities, personality traits, emotional states, and other psychological constructs. It offers rigorous methodologies for validating the data collected, especially in contexts where AI systems are designed to make predictions about human behaviors or traits (55).

5.10. AI's Contribution to Psychometrics

Data analysis and pattern recognition: AI, particularly machine learning, excels at analyzing large datasets. In psychometrics, AI can identify complex patterns and relationships within assessment data that might not be evident through traditional statistical methods (55).

5.10.1. Enhancing Test Development and Item Analysis

Automated Item Generation: AI can assist in generating a large pool of test items, ensuring a diverse and comprehensive set of questions.

Item Selection and Validation: AI algorithms can analyze item responses to determine the most effective items, helping in refining tests to improve their reliability and validity.

5.10.2. Facilitating Adaptive Testing

Tailored Assessments: AI enables the development of adaptive testing, where the difficulty of questions is adjusted in real time based on the test-taker's previous responses. This leads to more efficient and personalized assessments.

Improved accuracy: By adapting to an individual's ability level, AI-driven tests can more accurately measure a range of abilities or traits.

5.10.3. Data analysis and pattern recognition

AI, particularly machine learning, can handle large and complex datasets, extracting meaningful patterns and insights. It can be used to predict future trends or outcomes based on psychometric data, aiding in various applications such as educational planning or mental health interventions. However, AI can uncover subtle and complex relationships in psychometric data that may be missed by traditional methods. It introduces new methodologies and approaches for psychometric research, allowing for more innovative studies (56, 57).

AI systems can provide instant scoring and interpretation of test results, enhancing the efficiency of psychological assessments. Moreover, AI-driven online assessments can reach a broader audience, making psychometric testing more accessible(55).

5.10.4. Ethical and Fair Assessment

Bias Detection: AI can be used to analyze test items and responses for potential biases, contributing to fairer assessments.

Customization for Diverse Populations: AI can help in designing tests that are culturally and linguistically appropriate for diverse populations.

5.10.5. Real-time feedback and intervention

Immediate Results: AI-enabled systems can provide immediate feedback to test-takers, which is beneficial in educational and clinical settings.

Dynamic Interventions: Based on the assessment results, AI can suggest personalized interventions or learning paths.

5.10.6. Integration with Other Technologies

In combination with VR/AR, AI can be integrated with virtual and augmented reality for immersive and interactive psychological assessments.

Wearable Technology: AI can analyze data from wearable devices for continuous psychological monitoring.

5.10.7. Challenges and Considerations

While AI offers numerous advantages in psychometrics, challenges such as ensuring privacy, ensuring data security, ensuring the ethical use of AI, and maintaining the interpretability of AI models are critical areas that need continuous attention and development.

5.11. Qualitative Psychometrics: Exploring Subjective athlete Experiences

The integration of qualitative psychometrics into sports science offers a unique lens through which to understand athletes' psychological experiences, behaviors, and perceptions. Unlike quantitative methods that provide numerical data, qualitative psychometrics delve into the rich, descriptive aspects of athletes' mental states, motivations, and experiences (58).

In the realm of sports science, the integration of qualitative psychometrics offers profound insights into the psychological aspects of athletes' performances. While quantitative data provide measurable indicators of performance, qualitative psychometrics delve deeper into athletes' mental states, motivations, and interpersonal dynamics, offering a more nuanced understanding crucial for holistic athlete development.

5.12. Application in Sports Science

Athlete Mental Health: The rigorous demands and high-pressure environments in sports make understanding athletes' mental health crucial. Qualitative methods allow for exploring personal narratives related to stress, anxiety, and coping mechanisms, providing insights beyond what standard psychological tests can offer.

Team Dynamics: Team sports thrive on cohesion and effective interpersonal dynamics. Qualitative psychometrics help unravel the subtleties of team interactions, leadership styles, and group dynamics, which are essential for fostering a supportive and collaborative team environment.

Motivation and Drive: Diverse factors drive athletes, from personal goals to external rewards. Through qualitative analysis, researchers and coaches can gain a deeper understanding of these motivating factors by tailoring approaches to enhance individual and team performances.

Coaching Techniques: Evaluating coaching methodologies from athletes' perspectives is vital in assessing their effectiveness. Qualitative feedback provides invaluable insights into how different coaching styles impact athletes, guiding the development of more effective training strategies.

5.12.1. Methodologies

In-depth interviews: Personal interviews with athletes serve as a window into their individual experiences, emotions, and perceptions, offering a depth of understanding that quantitative measures might miss.

Focus Groups: Group discussions can uncover collective experiences and insights that are particularly useful in understanding team dynamics and shared challenges or motivations.

Observational studies: Observing athletes in their natural training and competitive environments offers real-time insights into their behaviors, interactions, and coping strategies in various scenarios.

Narrative analysis: Analyzing personal stories and accounts from athletes can reveal underlying themes and patterns related to their psychological state, informing tailored support and intervention strategies.

5.12.2. Benefits challenges and considerations

Qualitative psychometrics provide a comprehensive view of an athlete's mental and emotional state, which is crucial for all-rounded athlete development. This insight can guide the development of customized psychological interventions and support strategies to cater to individual athlete needs. Moreover, it fosters better communication between coaches, psychologists, and athletes, leading to improved understanding and collaboration. However, ensuring objectivity in analyzing qualitative data is challenging due to the inherently subjective nature of the method. Researchers must employ rigorous methods to validate their interpretations.

Qualitative psychometrics play a pivotal role in sports science, offering in-depth insights that quantitative methods alone cannot provide. By carefully applying and interpreting qualitative data, sports professionals can significantly enhance athlete support systems, coaching methodologies, and overall team dynamics. This integrated approach is essential for advancing the field of sports science and fostering the development of well-rounded, mentally robust athletes. Combining qualitative insights with quantitative data is fundamental for a comprehensive understanding of athlete performance and well-being.

Nevertheless, confidentiality and ethical handling of sensitive information are paramount in qualitative research, especially when dealing with personal and potentially vulnerable disclosures.

5.12.3. Interpretation Results

Expert analysis: Interpretation of the results should be performed by professionals who can understand and analyze the data in the context of sports performance.

Holistic View: Results should be viewed as part of a larger picture that includes the athlete's physical health, performance data, personal background, and environmental factors.

Avoid Overreliance: Psychometric results should be used in conjunction with other assessment methods, including physical assessments, performance statistics, and direct observations.

5.12.4. Applications in Training and Development

Individualized Strategies: Based on psychometric insights, tailor training, motivation, and mental strength strategies to suit individual athletes. This can include personalized mental skills training, stress management techniques, and goal-setting exercises.

Team Dynamics: For team sports, results can be used to understand team dynamics and improve communication, leadership styles, and group cohesion. This approach can help in building a more effective and harmonious team environment.

Performance Enhancement: Incorporate findings into mental skills training, such as visualization, relaxation techniques, and cognitive-behavioral strategies, to help athletes enhance their performance, particularly under pressure.

5.12.5. Continuous Monitoring and Feedback

Regular Assessment: Conducting assessments regularly helps in tracking the psychological development of athletes and provides insights for timely interventions.

Feedback to Athletes: Provide athletes with constructive and comprehensible feedback from their assessments. This helps them understand their mental strengths and areas for improvement.

Adjusting training programs: Use ongoing psychometric data to adjust and refine training programs, ensuring that they align with the athletes' current mental states and needs.

5.12.6. Research and Development

Stay Updated: Keep abreast of the latest research and developments in the field of sports psychometrics to ensure the use of the most current and effective tools and methods.

Contribute to Research: Encourage and participate in research studies to further the understanding of psychological factors in sports and their impact on performance.

Integrating With Other Disciplines

The staff should collaborate with other professionals, such as coaches, medical staff, nutritionists, and physiotherapists, to ensure a comprehensive support system for athletes.

Holistic athlete development: Emphasize the importance of integrating mental training with physical, technical, and tactical training, recognizing the interdependence of mental and physical aspects in sports performance.

6. Conclusion

This extensive guide underscores the significance of psychometrics in sports sciences, highlighting the need for an ethical, professional, and informed approach in its application. By incorporating psychological assessments and interventions into sports training and development programs, athletes can achieve not only improved performance but also a balanced and healthy sporting career. The field of sports psychometrics is dynamic and evolving, and its effective application requires continuous learning, adaptation, and integration with other aspects of sports sciences.

Ethical Approval and Consent to Participate

Not applicable.

Consent for Publication

Not applicable.

Competing Interests

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as potential conflicts of interest.

Funding

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Authors' Contributions

N.G, M.B.A and H.J: conception and design. A.A, I.D and K.T: analysis and interpretation of the data. N.G, M.B.A, A.A and H.J: drafting of the article. N.G, M.B.A, A.A, I.D, K.T and H.J: revising it critically for intellectual content. All authors gave their final approval to the version that will be published.

Declaration

ChatGPT was employed to meticulously review language imperfections, enhancing the manuscript's fluency and overall quality (59, 60).

References

1. Rust J, Golombok S. Modern psychometrics: The science of psychological assessment2014 2014. [DOI]
2. Kaufman AS. Contemporary intellectual assessment: Theories, tests, and issues2018 2018.
3. Raykov T, Marcoulides GA. Introduction to psychometric theory2011 2011. [DOI]
4. Truijens FL, Cornelis S, Desmet M, De Smet MM, Meganck R. Validity beyond measurement: Why psychometric validity is insufficient for valid psychotherapy research. Frontiers in psychology. 2019;10:532. [PMID: 30915008] [PMCID: PMC6423000] [DOI]
5. Malmqvist J, Hellberg K, Möllås G, Rose R, Shevlin M. Conducting the Pilot Study: A Neglected Part of the Research Process? Methodological Findings Supporting the Importance of Piloting in Qualitative Research Studies. International Journal of Qualitative Methods. 2019;18:160940691987834. [DOI]
6. Hashim HA. Application of Psychometrics in Sports and Exercise (Penerbit USM)2019 2019.
7. Collins D, Cruickshank A. Psychometrics in Sport: The Good, the Bad and the Ugly. In: Cripps B, Cripps B, editors. Psychometric Testing2017. p. 145-56. [PMID: 28802836] [DOI]
8. Braun V, Clarke V, Boulton E, Davey L, McEvoy C. The online survey as a qualitative research tool. International Journal of Social Research Methodology. 2021;24(6):641-54. [DOI]
9. van de Vijver FJR, Leung K. Methods and Data Analysis for Cross-Cultural Research2021 2021. [DOI]
10. Devi Prasad B, editor Qualitative content analysis: Why is it still a path less taken? Forum Qualitative Sozialforschung/Forum: Qualitative Social Research; 2019 20192019.
11. Hong QN, Pluye P, Fàbregues S, Bartlett G, Boardman F, Cargo M, et al. Improving the content validity of the mixed methods appraisal tool: a modified e-Delphi study. Journal of clinical epidemiology. 2019;111:49–59. [PMID: 30905698] [DOI]
12. Madadizadeh F, Bahariniya S. Tutorial on how to Calculating content validity of scales in medical research. Perioperative Care and Operating Room Management. 2023:100315. [DOI]

13. Wilson FR, Pan W, Schumsky DA. Recalculation of the Critical Values for Lawshe's Content Validity Ratio. Measurement and Evaluation in Counseling and Development. 2012;45(3):197-210. [DOI]
14. Rutherford-Hemming T. Determining content validity and reporting a content validity index for simulation scenarios. Nursing Education Perspectives. 2015;36(6):389–93. [PMID: 26753299] [DOI]
15. Yusoff MSB. ABC of content validation and content validity index calculation. Education in Medicine Journal. 2019;11(2):49–54. [DOI]
16. Polit DF, Beck CT. The content validity index: Are you sure you know what's being reported? critique and recommendations. Res Nurs Health. 2006;29(5):489-97. [PMID: 16977646] [DOI]
17. Flake JK, Davidson IJ, Wong O, Pek J. Construct validity and the validity of replication studies: A systematic review. American Psychologist. 2022;77(4):576. [PMID: 35482669] [DOI]
18. Neumann CS, Kosson DS, Salekin RT. Exploratory and confirmatory factor analysis of the psychopathy construct: Methodological and conceptual issues. The Psychopath2017. p. 79–104.
19. Orcan F. Exploratory and Confirmatory Factor Analysis: Which One to Use First? Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi. 2018;9(4):414-21. [DOI]
20. Schreiber JB. Issues and recommendations for exploratory factor analysis and principal component analysis. Research in Social and Administrative Pharmacy. 2021;17(5):1004–11. [PMID: 33162380] [DOI]
21. Watkins MW. Exploratory Factor Analysis: A Guide to Best Practice. Journal of Black Psychology. 2018;44(3):219-46. [DOI]
22. Lim S, Jahng S. Determining the number of factors using parallel analysis and its recent variants. Psychological methods. 2019;24(4):452. [PMID: 31180694] [DOI]
23. Ledesma RD, Valero-Mora P. Determining the number of factors to retain in EFA: An easy-to-use computer program for carrying out parallel analysis. Practical assessment, research, and evaluation. 2019;12(1):2.
24. Labrín C, Urdinez F. Principal component analysis. R for Political Data Science2020. p. 375–93. [DOI]
25. Beavers AS, Lounsbury JW, Richards JK, Huck SW, Skolits GJ, Esquivel SL. Practical considerations for using exploratory factor analysis in educational research. Practical Assessment, Research, and Evaluation. 2019;18(1):6.
26. DiStefano C, Zhu M, Mindrila D. Understanding and using factor scores: Considerations for the applied researcher. Practical assessment, research, and evaluation. 2019;14(1):20.
27. Damayanti S, Durachman Y, Khudzaeva E, editors. Factors Analysis of Interest in Using Pay Later in E-commerce Applications Using Principal Component Analysis and Maximum Likelihood Estimation Methods. 2022 10th International Conference on Cyber and IT Service Management (CITSM); 2022 20222022. [DOI]
28. Liu K, Huang K-L, Sfarra S, Yang J, Yi L, Yuan Y. Factor analysis thermography for defect detection of panel paintings. Quantitative InfraRed Thermography Journal. 2023;20(1):25-37. [DOI]
29. Kalkbrenner MT. Alpha, Omega, and Internal Consistency Reliability Estimates: Reviewing These Options and When to Use Them. Counseling Outcome Research and Evaluation. 2023;14(1):77-88. [DOI]
30. Forootani A, Iervolino R, Tipaldi M. Applying unweighted least-squares based techniques to stochastic dynamic programming: theory and application. IET Control Theory & Applications. 2019;13(15):2387-98. [DOI]
31. Channark S, Kumam P, Martinez-Moreno J, Chaipunya P, Jirakitpuwapat W. Generalized and optimal sequence of weights on a progressive-iterative approximation method with memory for least square fitting. Math Methods in App Sciences. 2022;45(17):11013-30. [DOI]
32. Osborne JW. What is rotating in exploratory factor analysis? Practical Assessment, Research, and Evaluation. 2019;20(1).
33. Marsh HW, Guo J, Dicke T, Parker PD, Craven RG. Confirmatory Factor Analysis (CFA), Exploratory Structural Equation Modeling (ESEM), and Set-ESEM: Optimal Balance Between Goodness of Fit and Parsimony. Multivariate Behavioral Research. 2020;55(1):102-19. [PMID: 31204844] [DOI]
34. Hox JJ. Confirmatory Factor Analysis. In: Barnes JC, Forde DR, Barnes JC, Forde DR, editors. The Encyclopedia of Research Methods in Criminology and Criminal Justice2021. p. 830-2. [DOI]
35. Sahli H, Sahli F, Saidane M, Rebhi M, Guelmami N, Trabelsi K, et al. Testing the psychometric properties of an Arabic version of the Brunel Mood Scale among physical education students. European Journal of Investigation in Health, Psychology and Education. 2023;13(8):1539–52. [PMID: 37623309] [PMCID: PMC10453414] [DOI]
36. Chalghaf N, Guelmami N, Re TS, Maldonado Briegas JJ, Garbarino S, Azaiez F, Bragazzi NL. Trans-cultural adaptation and validation of the "teacher job satisfaction scale" in Arabic language among sports and physical education teachers ("teacher of physical education job satisfaction inventory"—TPEJSI): insights for sports, educational, and occupational psychology. Frontiers in psychology. 2019;10:2234. [PMID: 31695634] [PMCID: PMC6817567] [DOI]
37. Guelmami N, Chalghaf N, Tannoubi A, Puce L, Azaiez F, Bragazzi NL. Initial development and psychometric evidence of physical education grit scale (PE-GRIT. Frontiers in Public Health. 2022;10:818749. [PMID: 35309217] [PMCID: PMC8927648] [DOI]
38. Jelleli H, Hindawi O, Kaddech N, Rebhi M, Ben Aissa M, Saidane M, et al. Reliability and validity of the Arabic version of coach-athlete relationship questionnaire: ACART-Q. Cogent Social Sciences. 2023;9(2):2282275. [DOI]
39. Noômen G, Mohamed J, Nasri H, Brahim A. Sensitivity, internal consistency and factorial structure of the Arabic version of OMSAT-3. Advances in Physical Education. 2015;5(01):18. [DOI]
40. Boukari S, Guelmami N, Chotrane SG, Bouzid S, Khemiri A, Muscella A, Khalifa R. Adaptation of the Questionnaire on Teacher Interaction in Tunisia: Teaching Strategies to Promote Sustainable Education in Schools. Sustainability. 2022;14(5):2489. [DOI]
41. Rebhi M, Ben Aissa M, Tannoubi A, Saidane M, Guelmami N, Puce L, et al. Reliability and Validity of the Arabic Version of the Game Experience Questionnaire: Pilot Questionnaire Study. JMIR Formative Research. 2023;7:e42584. [PMID: 36482747] [PMCID: PMC10131659] [DOI]
42. Guelmami N, Chalghaf N, Tannoubi A, Puce L, Azaiez F, Bragazzi NL. Initial Psychometric Evidence of Physical Inactivity Perceived Experience Scale (Pipes): COVID-19 Pandemic as a Pilot Study. Frontiers in Public Health. 2022;10:819052. [PMID: 35392464] [PMCID: PMC8980326] [DOI]

43. de Beurs DP, Fokkema M, de Groot MH, de Keijser J, Kerkhof AJFM. Longitudinal measurement invariance of the Beck Scale for Suicide Ideation. Psychiatry Research. 2015;225(3):368-73. [PMID: 25571773] [DOI]
44. Nattino G, Pennell ML, Lemeshow S. Assessing the goodness of fit of logistic regression models in large samples: A modification of the Hosmer-Lemeshow test. Biometrics. 2020;76(2):549-60. [DOI]
45. Cheung GW, Cooper-Thomas HD, Lau RS, Wang LC. Reporting reliability, convergent and discriminant validity with structural equation modeling: A review and best-practice recommendations. Asia Pac J Manag. 2023. [DOI]
46. Wu AD, Li Z, Zumbo BD. Decoding the meaning of factorial invariance and updating the practice of multi-group confirmatory factor analysis: A demonstration with TIMSS data. Practical Assessment, Research, and Evaluation. 2019;12(1):3.
47. Jorgensen TD, Kite BA, Chen P-Y, Short SD. Finally! A Valid Test of Configural Invariance Using Permutation in Multigroup CFA. In: Van Der Ark LA, Wiberg M, Culpepper SA, Douglas JA, Wang W-C, Van Der Ark LA, et al., editors. Quantitative Psychology. 196. Cham2017. p. 93-103. [DOI]
48. Cheung GW, Lau RS. A Direct Comparison Approach for Testing Measurement Invariance. Organizational Research Methods. 2012;15(2):167-98. [DOI]
49. Shi D, Lee T, Maydeu-Olivares A. Understanding the Model Size Effect on SEM Fit Indices. Educational and Psychological Measurement. 2019;79(2):310-34. [PMID: 30911195] [PMCID: PMC6425088] [DOI]
50. Noble S, Scheinost D, Constable RT. A decade of test-retest reliability of functional connectivity: A systematic review and meta-analysis. Neuroimage. 2019;203:116157. [PMID: 31494250] [PMCID: PMC6907736] [DOI]
51. Belur J, Tompson L, Thornton A, Simon M. Interrater Reliability in Systematic Review Methodology: Exploring Variation in Coder Decision-Making. Sociological Methods & Research. 2021;50(2):837-65. [DOI]
52. Viladrich C, Angulo-Brunet A, Doval E. A journey around alpha and omega to estimate internal consistency reliability. Anales de psicología. 2017;33(3):755–82. [DOI]
53. Schwartz SH, Melech G, Lehmann A, Burgess S, Harris M, Owens V. Extending the Cross-Cultural Validity of the Theory of Basic Human Values with a Different Method of Measurement. Journal of Cross-Cultural Psychology. 2001;32(5):519-42. [DOI]
54. Prince M. Measurement validity in cross-cultural comparative research. Epidemiology and Psychiatric Sciences. 2008;17(3):211–20. [PMID: 18924560] [PMCID: PMC10454710] [DOI]
55. Wang X, Jiang L, Hernandez-Orallo J, Sun L, Stillwell D, Luo F, Xie X. Evaluating General-Purpose AI with Psychometrics. 2023.
56. Liem CCS, Langer M, Demetriou A, Hiemstra AMF, Sukma Wicaksana A, Born MP, König CJ. Psychology Meets Machine Learning: Interdisciplinary Perspectives on Algorithmic Job Candidate Screening. In: Escalante HJ, Escalera S, Guyon I, Baró X, Güçlütürk Y, Güçlü U, et al., editors. Explainable and Interpretable Models in Computer Vision and Machine Learning. Cham2018. p. 197-253. [DOI]
57. Oliveira Chaves L, Gomes Domingos AL, Louzada Fernandes D, Ribeiro Cerqueira F, Siqueira-Batista R, Bressan J. Applicability of machine learning techniques in food intake assessment: A systematic review. Critical Reviews in Food Science and Nutrition. 2023;63(7):902–19. [PMID: 34323627] [DOI]
58. Wijsen LD, Borsboom D, Alexandrova A. Values in Psychometrics. Perspect Psychol Sci. 2022;17(3):788-804. [PMID: 34846952] [PMCID: PMC9069645] [DOI]
59. Dergaa I, Saad HB. Joining hands for better health: The inception of the New Asian Journal of Medicine. New Asian Journal of Medicine. 2023;1(1):1-3. [DOI]
60. Chtourou H, Guelmami N, Trabelsi K, Dergaa I. The Beginning of Our Journey: The Launch of the Tunisian Journal of Sports Science and Medicine. Tunisian Journal of Sports Science and Medicine. 2023;1(1):1-4. [DOI]