

Article

Not peer-reviewed version

Adaptation and Validation of Psychological Assessment Questionnaires Using Confirmatory Factor Analysis: A Tutorial for Planning and Reporting Analysis

[Andrés Ramírez *](#) , [Luis Burgos-Benavides](#) , [Hugo Sinchi](#) , [Jessica Vanessa Quito-Calle](#) , [Francisco Herrero Díez](#) , [Francisco Javier Rodríguez-Díaz](#)

Posted Date: 17 February 2025

doi: 10.20944/preprints202502.1192.v1

Keywords: Adaptation; Psychometric Validation; Reliability; Factor Analysis; Psychometrics



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Article

Adaptation and Validation of Psychological Assessment Questionnaires Using Confirmatory Factor Analysis: A Tutorial for Planning and Reporting Analysis

Andrés Ramírez ^{1,*}, Luis Burgos-Benavides ², Hugo Sinchi ³, Jessica Vanessa Quito-Calle ¹, Javier Herrero Díez ² and Francisco Javier Rodríguez-Díaz ²

¹ Department of Clinical Psychology, Universidad Politécnica Salesiana, Cuenca 010107, Ecuador

² Department of Psychology, Universidad de Oviedo, 33003 Oviedo, Spain; burgosluis@uniovi.es (L.B.-B.); herrero@uniovi.es (J.H.D.); gallego@uniovi.es (F.J.R.-D.)

³ Department of Psychology, Pontificia Universidad Católica del Ecuador, Esmeraldas 080101, Ecuador; hfsinchi@pucese.edu.ec

* Correspondence: aramirez1@ups.edu.ec

Abstract: This article aims to review the linguistic adaptation and psychometric validation of psychological assessment instruments through a narrative review. Linguistic adaptation involves translating and culturally adjusting instruments to fit the target population, while psychometric validation assesses reliability using indicators such as Cronbach's Alpha ($\alpha \geq 0.70$) and McDonald's Omega ($\omega \geq 0.70$). Additional reliability measures include Guttman's Lambda ($\lambda \geq 0.70$) and Composite Reliability ($CR \geq 0.70$). Validity is evaluated through the Average Variance Extracted ($AVE \geq 0.50$) and the Heterotrait-Monotrait Ratio ($HTMT \leq 0.85$). Model fit is assessed using indices such as CFI (≥ 0.90), TLI (≥ 0.90), NFI (≥ 0.90), NNFI (≥ 0.90), RFI (≥ 0.90), IFI (≥ 0.90), GFI (≥ 0.90), AGFI (≥ 0.90), PGFI (≥ 0.50), PNFI (≥ 0.50), PCFI (≥ 0.50), RMSEA (≤ 0.08 acceptable, ≤ 0.05), and SRMR (≤ 0.08). Confirmatory Factor Analysis (CFA) is also employed to investigate measurement invariance, ensuring the stability of the instrument across different groups.

Keywords: Adaptation; Psychometric Validation; Reliability; Factor Analysis; Psychometrics

1. Introduction

Psychological assessment enables the understanding and evaluation of qualities, attributes, constructs, and behavior patterns (Cherry et al., 2021; Kapur et al., 2022). Participants' responses allow for inferences about psychological constructs or attributes, which are systematically organized into standardized questionnaires. These instruments must be effective and precise to ensure accurate measurement of the constructs or attributes (Hunsley & Meyer, 2003; Loewenthal & Lewis, 2020; Meyer et al., 2001; Reynolds et al., 2021; Rönkkö & Cho, 2022; Schneider et al., 2024; Strauss & Smith, 2009; Wilson, 2023). The quality of these questionnaires depends on the rigor of the methodology used during their development (Einola & Alvesson, 2021).

Exploratory Factor Analysis (EFA) is the most commonly used statistical method for test development (Goretzko et al., 2021), while Confirmatory Factor Analysis (CFA) is widely applied to validate the evidence obtained from EFA (Marsh et al., 2020; Tavakol & Wetzel, 2020; Widaman & Helm, 2023). However, depending on the context, the development of assessment instruments across various areas of psychology often remains in its psychometric infancy. This complex issue has led to numerous proposals for adapting psychometric questionnaires to ensure rigorous measurement by

demonstrating reliability and validity evidence—essential aspects for research and clinical practice (Muñiz, 2018; Shroff et al., 2023).

The adaptation process begins with translation or linguistic adaptation, but this is merely the starting point of an exhaustive procedure. Content adaptation must accurately reflect the cultural and contextual reality of the target group, ensuring that the items are understandable and culturally appropriate for the new population (Abad et al., 2011; Muñiz et al., 2013).

The objective of this tutorial is to provide a guide for the adaptation and validation of psychometric instruments in psychological assessment, focusing on evidence of validity and reliability. The article outlines procedures and criteria for calculating reliability, evaluating validity evidence using current psychometric standards, and guiding researchers in ensuring that the instruments they use to evaluate psychological attributes and constructs are accurate and reliable (Elosua & Egaña, 2020; Haladyna & Rodriguez, 2013).

Adaptation of the Psychological Questionnaires

Test adaptation has advanced significantly over the past 25 years. This procedure, applied across various fields, follows several guidelines, such as those proposed by the International Test Commission (2017). These guidelines, outlined in 18 directives, represent a comprehensive effort to ensure a rigorous test adaptation process.

The first section, preconditions, involves researchers making initial decisions prior to the translation and adaptation process. The second section, test development, focuses on adapting or translating the instrument. The third phase, confirmation, involves analyzing equivalence, reliability, and validity evidence based on empirical data collected from the new context to which the questionnaire is being adapted. Subsequent phases include administration, scoring scales and interpretation, and documentation (International Test Commission, 2017).

This adaptation process is a complex effort that goes beyond mere item translation. Strict adherence to these guidelines is essential to ensure the subsequent validation analyses of the questionnaire within the target population.

Confirmatory Factor Analysis

Confirmatory Factor Analysis (CFA) is conducted after a theoretical structure for the test has been proposed. CFA evaluates whether the empirical data fit the predefined factorial structure and confirms if the proposed model aligns with the observed data (Luo et al., 2019). Through CFA, expected factor loadings, factor correlations, and other model parameters can be verified.

This analysis is crucial for validating the proposed factorial structure and ensuring that the test accurately and reliably measures the intended constructs (Muñiz, 2018).

Table 1. Key psychometric indicators in the validation of Psychometric Properties through Model Fit indices.

Indicators	Abbreviation	Acceptable range
Average Validity Extracted	AVE	≥ 0.50
Comparative Fit Index	CFI	≥ 0.90
Tucker-Lewis Index	TLI	≥ 0.90
Quadratic Approximation Error	RMSEA	≤ 0.08
Root Mean Square Residual	SRMR	≤ 0.08
Heterotrait-Monotrait Ratio	HTMT	≤ 0.85

Confirmatory Factor Analysis (CFA) is a widely used tool for questionnaire validation (Marsh et al., 2014; Montoya & Edwards, 2021; Taasobshirazi & Wang, 2016). This analysis assesses model fit using various indices, including the Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA), and Standardized Root Mean Square Residual (SRMR).

Each index provides insights into different aspects of model fit. The recommended thresholds for a well-fitting model are typically CFI ≥ 0.90, TLI ≥ 0.90, RMSEA ≤ 0.06, and SRMR < 0.08 (Hu &

Bentler, 1999; Tucker & Lewis, 1973; Browne & Cudeck, 1993; Hu & Bentler, 1999) (see Table 1 and Supplementary Table 1).

Relative fit index

The Relative Fit Index (RFI) is a measure that compares the fit of the proposed model to a null model (a model in which all variables are uncorrelated). The RFI aims to evaluate how well the proposed model fits compared to a reference model, balancing the simplicity of the model with its explanatory power (Hutchinson & Olmos, 1998).

The RFI ranges from 0 to 1, with values close to 1 indicating a good model fit, while values below 0.90 generally suggest that the model may not fit the data adequately. The RFI accounts for the model's degrees of freedom, penalizing unnecessarily complex models.

In summary, the RFI is useful for assessing whether a simpler model might be preferable to one that, while fitting well, introduces complexity that is not proportional to the improvement in fit. The formula is presented below.

$$RFI = \frac{x^2_{model} / df_{model}}{x^2_{nullo} / df_{model}}$$

Comparative Adjustment Index (CFI)

The Comparative Fit Index (CFI) is a widely used metric in structural equation modeling to evaluate the quality of the proposed model's fit. The CFI compares the evaluated model to a null model, which assumes no relationships between variables. This index adjusts for differences in sample size and model complexity, providing a measure of how much the proposed model improves fit compared to the null model (Marsh et al., 2014).

A CFI value of ≥ 0.90 is considered indicative of good model fit. This cutoff suggests that the proposed model significantly improves the fit compared to the null model, though it may not be optimal. Values above 0.95 indicate excellent fit, suggesting that the proposed model aligns very well with the observed data and offers a robust representation of the relationships among variables.

These recommendations are based on the work of Hu and Bentler (1999), who established these thresholds to guide researchers in interpreting CFI in the context of model evaluation. The formula is presented below.

$$CFI = \frac{(TLI_{model} - TLI_{nullo})}{(TLI_{perfecto} - TLI_{nullo})}$$

Tucker-Lewis Index (TLI)

The Tucker-Lewis Index (TLI), also known as the Non-Normed Fit Index (NNFI), is a key metric in Confirmatory Factor Analysis (CFA). Like the Comparative Fit Index (CFI), the TLI assesses the fit of the proposed model by comparing it to a null model. However, the TLI has the distinct feature of penalizing overly complex models, making it particularly useful for promoting parsimony in model evaluation (Cai et al., 2023; Shi et al., 2019).

A TLI value of ≥ 0.90 is considered acceptable, indicating that the proposed model provides an adequate fit relative to the null model, adjusted for model complexity. Higher values suggest better fit, with scores above 0.95 generally interpreted as indicative of excellent fit. This penalty for complexity helps avoid overfitting and encourages models that are both parsimonious and representative of the data. The formula is presented below.

$$TLI = \frac{x^2_{nullo} / df_{nullo} - x^2_{model} / df_{model}}{x^2_{nullo} / df_{nullo} - 1}$$

Normed Fit Index (NFI)

The Normed Fit Index (NFI) evaluates the fit of a proposed model in comparison to a null model, where all variables are assumed to be unrelated. The NFI value ranges from 0 to 1, with values closer to 1 indicating a better fit. A value of 0.90 or higher is generally accepted as indicative of good fit.

The NFI directly compares the proposed model to the null model and can be sensitive to sample size, particularly in studies with small samples (Goretzko et al., 2024; Zheng & Bentler, 2024). The formula is presented below.

$$NFI = \frac{x^2_{\text{null}} - x^2_{\text{model}}}{x^2_{\text{null}} / df_{\text{null}}}$$

Relative Fit Index (RFI)

The Relative Fit Index (RFI) compares the fit of a model with a null model, similar to the NFI, but takes into account the complexity of the model and the covariance structure between variables. The RFI is interpreted on a scale from 0 to 1, where values closer to 1 indicate a better relative fit of the proposed model compared to the null model. This index can be useful for avoiding overfitting and considering more parsimonious models (Van Laar & Braeken, 2022). The formula is presented below.

$$RFI = \frac{x^2_{\text{null}} / df_{\text{null}} - x^2_{\text{model}} / df_{\text{model}}}{x^2_{\text{null}} / df_{\text{null}} - 1}$$

Absolute fit index

The Absolute Fit Index refers to a category of indices that directly evaluate how well a proposed model fits the observed data, without making comparisons to other models, as relative fit indices would. Absolute fit indices measure the quality of the model's fit based on the discrepancy between the observed covariance matrix and the covariance matrix estimated by the model. These indices are essential because they provide a direct assessment of the fit without involving model complexity or comparison with other models (Goretzko et al., 2024).

Goodness of fit index (GFI)

It is an absolute fit measure that evaluates the proportion of variance and covariance explained by the proposed model in relation to the observed data. It is similar to a coefficient of determination (R^2) in regression, and its value ranges from 0 to 1, where values close to 1 indicate a good model fit. A value of 0.90 or higher is generally considered acceptable. The GFI is sensitive to sample size, and in studies with small samples, it may underestimate the model fit (Wang et al., 2020). The formula is presented below.

$$GFI = 1 - \frac{(S_{\text{model}} - S_{\text{observed}})^2}{S_{\text{model}}}$$

Adjusted goodness of fit index (AGFI)

It is an adjusted version of the GFI that takes into account the complexity of the model, penalizing models that include more parameters than necessary. The AGFI adjusts the GFI value based on the model's degrees of freedom, so models with more parameters (more complex) receive a penalty. Like the GFI, the AGFI also ranges from 0 to 1, with values close to 0.90 considered indicative of a good fit. The AGFI is more useful when there is a need to balance model fit with its parsimony, that is, achieving a good fit with the fewest possible parameters (Maassen et al., 2023). The formula is presented below.

$$AGFI = 1 - \frac{(1 - GFI) \cdot (df_{\text{model}} - 1)}{df_{\text{total}} - 1}$$

Root Mean Square Error of Approximation (RMSEA)

The Root Mean Square Error of Approximation (RMSEA) is an index used to evaluate the discrepancy between the proposed model and the observed data in confirmatory factor analysis (CFA). Unlike other indices that assess model fit in comparative terms, RMSEA measures the absolute fit error and provides an estimate of how well the model fits the data in terms of average error per degree of freedom (Browne et al., 1993; Hu & Bentler, 1995).

An RMSEA value ≤ 0.06 is considered indicative of a good model fit. This threshold suggests that the discrepancy between the model and the data is small, implying that the model adequately represents the relationships among the observed variables. RMSEA values less than 0.08 are generally acceptable, indicating that the model has a reasonably good fit, though not as precise as lower values. These criteria help researchers interpret the quality of model fit and decide whether the proposed model is suitable for the data. The formula is presented below.

$$RMSEA = \sqrt{\frac{\chi^2 - df}{df \cdot N}}$$

Standard Square Root Mean Residual (SRMR)

The Standardized Root Mean Square Residual (SRMR) is an index that evaluates the discrepancies between the observed covariances and the covariances predicted by the model in confirmatory factor analysis (CFA). Unlike other indices that compare model fit with a null model or adjust for model complexity, SRMR provides a direct measure of absolute fit by quantifying the average standardized difference between the observed and estimated covariances (Hu & Bentler, 1999).

An SRMR value < 0.08 is considered indicative of a good fit. This threshold suggests that the model has a relatively low discrepancy between the observed and predicted covariances, indicating that the model fits the data well. Higher values may suggest issues with model fit, and it may be necessary to review and adjust the model to improve its ability to represent the relationships in the data. The formula is presented below.

$$SRMR = \sqrt{\frac{\sum_{ij} (S_{ij} - \hat{S}_{ij})^2}{m}}$$

Parsimony fit index

Parsimony Fit Indices focus on evaluating the balance between a model's fit and its simplicity. A parsimonious model is one that achieves a good fit with the least number of parameters, avoiding overfitting by including unnecessary parameters. These indices penalize models that, while they may fit the data well, are overly complex.

Parsimony Goodness of Fit Index (PGFI)

The index is an extension of the GFI, which adjusts according to the parsimony of the model. This exponent incorporates a penalty for model complexity, so that a model with more parameters, even if it has a good fit, will receive a lower score if it is unnecessarily complex. PGFI values are generally lower than GFI values. A PGFI value ≥ 0.50 is considered indicative of a good balance between fit and parsimony (Thakkar, 2020). The formula is presented below.

$$PGFI = 1 - \frac{(1 - GFI) \cdot \frac{df_{model}}{df_{total}}}{1}$$

Parsimony Normed Fit Index (PNFI)

The Parsimony Normed Fit Index (PNFI) is an adjusted version of the NFI that penalizes model complexity. The PNFI introduces a correction that takes into account the degrees of freedom of the model, encouraging simpler models that achieve a good fit. Like the NFI, PNFI values vary between 0 and 1. A PNFI value ≥ 0.50 is considered acceptable and reflects a good fit with parsimony (Marsh et al., 2005). The formula is presented below.

$$PNFI = NFI \cdot \frac{df_{model}}{df_{nulo}}$$

Parsimony Comparative Fit Index (PCFI)

Parsimony Comparative Fit Index (PCFI) is an adjusted version of the CFI, which also penalizes model complexity. Like the PNFI, the PCFI considers the degrees of freedom of the model and favors those that achieve a good fit with fewer parameters. A PCFI value ≥ 0.50 is considered adequate and reflects a good balance between model fit and simplicity (Maassen et al., 2023). The formula is presented below.

$$PCFI = CFI \cdot \frac{df_{model}}{df_{nulo}}$$

Likelihood-ratio χ^2 /degree of freedom

Chi-Square Ratio (χ^2) and the degrees of freedom (χ^2/df) is an index commonly used to evaluate the parsimony of the model. It is calculated by dividing the value of the chi-square test by the degrees of freedom of the model. A lower value indicates a more parsimonious model. Generally, a value of (χ^2/df) less than 3 is considered indicative of good fit and parsimony, although some authors allow values up to 5 depending on the context (Van Laar & Braeken, 2022).

Reliability Indicators

The reliability of a test refers to the consistency of scores obtained on different occasions, or under different conditions. A reliable test produces stable and consistent results, meaning that scores should be similar if the test is administered again under equivalent conditions (Kalkbrenner, 2024, Viladrich et al., 2017; Zijlmans et al., 2018; Zimmerman, 2007). The main reliability indicators used to assess this consistency include (see Table 2).

Table 2. Key psychometric indicators in the validation of the Psychometric Properties through Reliability Indexes.

Indicator	Abbreviation	Acceptable range
Cronbach's Alpha	α	≥ 0.70
McDonald's Omega Coefficient	ω	≥ 0.70
Guttman's Lambda	λ	≥ 0.70
Composite Reliability	CR	≥ 0.70

Alfa de Cronbach (α)

It is one of the most commonly used indexes to measure the internal consistency of a test. Cronbach's alpha evaluates the average correlation between the items of a test and provides a measure of the homogeneity of the items. A value of $\alpha \geq 0.70$ is considered acceptable, although higher values (≥ 0.80 or ≥ 0.90) are preferable, indicating higher internal consistency (American Psychological Association, 2022; Doval et al., 2023; McNeish, 2017; Viladrich et al., 2017). Cronbach's alpha is calculated using the following formula:

$$\alpha = \frac{k}{k - 1} \left(1 - \frac{\sum_{i=1}^k \sigma_i^2}{\sigma_t^2} \right)$$

Where k is the number of items in the test, σ_i^2 is the variance of the item i , y σ_t^2 is the total variance of the test. While the ordinal Cronbach's alpha is calculated using the following formula:

$$\alpha_{ordinal} = \frac{N \cdot \bar{r}_{ij}}{-1 + (N - 1) \cdot \bar{r}_{ij}}$$

Where k is the number of items in the test, σ_i^2 is the variance of the item i , y σ_t^2 is the total variance of the test. While the ordinal Cronbach's alpha is calculated using the following formula:

Where N is the number of items y \bar{r}_{ij} is the average polychoric correlation between items.

Coefficiente de Fiabilidad de Congenericidad (ρ)

The coefficient is a generalization of Cronbach's alpha coefficient and is used when items are not tau-equivalent (i.e., they do not have the same variance or the same relationship to the construct). The formula is as follows:

$$\rho = \frac{(\sum \lambda_i)^2}{(\sum \lambda_i)^2 + \sum \theta_{ii}}$$

Where λ_i are the factor loadings of the items and θ_{ii} are the variances of the errors.

McDonald's Omega coefficient (ω)

This index is an alternative to Cronbach's alpha and provides a more accurate estimate of reliability when items are assumed to measure several underlying factors. The omega coefficient is based on factor analysis models and takes into account the dimensional structure of the test, providing a more robust measure of internal consistency. Values of $\omega \geq 0.70$ indicate good reliability (McDonald, 1999; Viladrich et al., 2017). The omega coefficient is calculated using the following formula:

$$\omega = \frac{\sum_{i=1}^k \lambda_i^2}{\sum_{i=1}^k \lambda_i^2 + \sum_{i=1}^k \theta_i}$$

Where λ_i^2 is the variance explained by item i (squared factor loadings) y θ_i is the residual variance of the item i . While the ordinal omega coefficient is calculated using the following formula:

$$\omega_{ordinal} = \frac{\sum_{i=1}^k \lambda_i^2}{\sum_{i=1}^k \lambda_i^2 + \sum_{j=1}^k \psi_j}$$

Where λ_i^2 are the standardized factor loadings of the items on factor, ψ_j are the error variances of the items, and k is the total number of items.

Coefficiente de fiabilidad Omega Jerárquico (ω_h)

The Hierarchical Omega reliability coefficient (ω_h), which is used in hierarchical factor models. This coefficient assesses the internal consistency of a test when items are grouped into subscales or first-order factors. Unlike the alpha coefficient, which assumes that all items are tau-equivalent, the Hierarchical Omega allows a more accurate assessment in contexts where items have different factor loadings. Its formula is:

$$\omega_h = \frac{(\sum \lambda_g)^2}{(\sum \lambda_g)^2 + \sum \theta_{ii}}$$

Where λ_g are the factor loadings of the items on the overall factor and θ_{ii} are the variances of the item errors.

Lambda de Guttman (λ)

Guttman's lambda is an index that assesses reliability in terms of the stability of scores on different administrations of the test. Unlike Cronbach's alpha, which assumes a unidimensional model, Guttman's lambda can be adapted to more complex models. Values of $\lambda \geq 0.70$ suggest good reliability (Benton, 2015; Guttman, 1945; Motallebzadeh, 2023). The Guttman lambda is calculated with the formula:

$$\lambda = \frac{\sum_{i=1}^k \sigma_i^2}{\sigma_t^2}$$

Where σ_i^2 is the variance of the item i y σ_t^2 is the total variance of the test.

Each of these indicators offers a different perspective on the consistency of test scores and may be useful in different contexts and for different types of instruments. Using a combination of these indices provides a more complete and robust assessment of test reliability.

Evidence of Validity

Validity refers to how well a test measures what it purports to measure. Among the main validity indicators are the average variance extracted (AVE), the Heterotrait-Monotrait index (HTMT) (Salessi & Omar, 2019).

Convergent Validity: Average Extracted Validity (AVE)

Average Extracted Validity (AVE) is an index used to assess the convergent validity of a construct in a measurement model. AVE quantifies the amount of variance in the items that is explained by the construct in question. A high AVE value indicates that a large proportion of the variance in the items is attributable to the construct it is intended to measure, suggesting that the construct is being measured effectively and accurately (Fornell & Larcker, 1981; Rönkkö & Cho, 2022). The formula for calculating the AVE is as follows:

$$AVE = \frac{\sum_{i=1}^n \lambda_i^2}{\sum_{i=1}^n \lambda_i^2 + \theta_i}$$

Where λ_i^2 is the square of the loading of item i on the construct, θ_i is the variance of the item error i and n is the total number of items in the construct.

For a construct to have adequate convergent validity, the variance must explain at least 50% of its component items (Fornell & Larcker, 1981). Therefore, the cut-off point for the AVE is ≥ 0.50 . Due to the difficulty of certain constructs to capture such variance, it has been proposed to include as admissible those constructs that explain at least 37.5% of the variance, therefore, the lax cut-off point would be located at ≥ 0.375 (Moral de la Rubia, 2019).

Discriminant Validity: Index HTMT

The Heterotrait-Monotrait Ratio (HTMT) is a measure used to assess the discriminant validity between constructs in a measurement model. The HTMT examines the relationship between items of different constructs (heterotrait) compared to the relationship between items within the same construct (monotrait). In other words, the HTMT assesses the difference in correlation between items measuring different constructs versus items measuring the same construct.

An HTMT value of less than 0.85 suggests that the constructs are sufficiently differentiated from each other, indicating good discriminant validity. If the HTMT value is greater than 0.85, it could indicate a lack of differentiation between the constructs, suggesting that the constructs may be overlapping (Henseler et al., 2015). The formula for calculating the HTMT is:

$$HTMT = \frac{\text{Mean Heterotrait Correlations}}{\text{Average of Monotrait Correlations}}$$

Where the mean of the heterotrait correlations is the mean of the correlations between items of different constructs and the mean of the monotrait correlations is the mean of the correlations between items of the same construct. A HTMT value of less than 0.85, as a cutoff point, indicates good discriminant validity, suggesting that the constructs are sufficiently differentiated from each other (Henseler et al., 2015).

Measurement Invariance

Measurement invariance refers to the consistency in the factor structure of an instrument when applied to different groups (Van De Schoot et al., 2015). This aspect ensures that the number and type of factors, as well as the relationships between items and factors, are equivalent in the groups compared. To assess measurement invariance, confirmatory factor analysis (CFA) is performed on each group to confirm that the items load on the same factors in all samples (Marsh et al., 2014). This type of invariance is crucial to ensure that the instrument measures the same construct in different populations (Maassen et al., 2023; Schmitt & Kuljanin, 2008).

The classic levels of factorial invariance have been widely used to assess the equivalence of measurements across groups. Configural invariance tested whether the basic factorial structure, i.e., the pattern of relationships between items and factors, remained consistent across groups without imposing additional parameter constraints. Subsequently, metric invariance evaluated whether factor loadings were equivalent across groups, ensuring that items consistently reflected their

relationship with latent factors. At a more advanced level, scalar invariance examined the equality of intercepts, guaranteeing that scores reflected differences in latent factors rather than measurement biases. Finally, strict invariance, the most rigorous stage, assessed whether measurement errors were homogeneous across groups, enabling precise and noise-free comparisons in estimates.

In recent research, new forms of invariance were explored to address the limitations of traditional approaches. Rutkowski and Svetina (2017) proposed invariance in hierarchical models, which allowed the evaluation of factorial equivalence in nested data, such as students within schools. They also introduced partial invariance as a solution for scenarios where full invariance was unattainable, permitting some parameters to vary across groups without compromising overall comparability. On the other hand, Xu and Soland (2024) investigated predictive invariance, which assessed whether relationships between latent factors and external variables were consistent across groups, thus integrating predictive validity into the analysis. Approximate invariance was also adopted, tolerating minor deviations in parameters and offering greater flexibility in culturally or linguistically diverse contexts.

Load invariance

Loading invariance examines whether the factor loadings of items are equal across different groups. This type of invariance verifies that each item has the same impact on the underlying factor, regardless of the group to which the test is applied (Kim & Yoon, 2011; Rutkowski & Svetina, 2017; Xu & Soland, 2024). To assess this invariance, we impose restrictions on the factor loadings in the model and compare the fit of the restricted model with an unrestricted model. The loading invariance ensures that the items have an equivalent influence on the factors across the samples studied (Kim et al., 2017). The formula is presented below.

$$X_i = \lambda_{ij} F_j + \varepsilon_i$$

Where X_i is the response observed in the item i , λ_{ij} is the factor loading of the item i in the factor j , F_j is the latent value of the factor j and ε_i is the residual error of the item i . The criteria of loading invariance, part of the stepwise factor analysis, ensures that the relationships between items and factors are equivalent across groups. It is assessed after establishing a configural model, which verifies the same factor structure. (CFI > 0.90, RMSEA < 0.08, SRMR < 0.08) without restrictions. In the metric model, the factor loadings (λ_{ij}) are constrained to be equal between groups, and is considered achieved if the change in fit indices is acceptable ($\Delta\text{CFI} \leq 0.01$, $\Delta\text{RMSEA} \leq 0.015$, $\Delta\text{SRMR} \leq 0.030$). This step is crucial to validate the comparability of item-factor relationships before moving on to stricter invariance such as scalar (includes intercepts) and strict invariance (includes residuals).

Variance Inflation Factor (VIF)

VIF is a measure used in regression analysis to detect multicollinearity between predictor variables. It indicates how much the variance of a regression coefficient increases due to correlation between independent variables. The formula is presented below.

$$\text{VIF}_i = \frac{1}{1 - R_i^2}$$

Where R_i^2 is the coefficient of determination of the regression of the variable i on the other predictor variables. The Variance Inflation Factor (VIF) is an indicator used to assess the presence of multicollinearity among predictor variables in a regression model.

A VIF close to 1 indicates no multicollinearity, meaning the independent variable is not highly correlated with the others. This suggests that the regression coefficients can be estimated accurately and without significant distortion.

On the other hand, a VIF greater than 5 or 10 signals the presence of problematic multicollinearity, implying that a variable is strongly correlated with others in the model. In such cases, coefficient estimation becomes unstable, increasing variance and reducing the reliability of results. To mitigate this issue, techniques such as removing redundant variables, combining correlated variables, or using regularization methods like ridge regression can be applied.

Intercept Invariance

Intercept invariance focuses on whether the starting points of the items in the measurement scale are the same in different groups. That is, it assesses whether the relationship between item and factor scores is consistent across groups. To assess this invariance, restrictions are imposed on the intercepts of the items and the fitted model is compared with one that does not have these restrictions. Intercept invariance is essential to ensure that differences in scores between groups are not due to biases in item starting points (De Roover, 2021). The formula is presented below.

$$X_i = \lambda_{ij} F_j + \tau_i + \varepsilon_i$$

Where X_i is the response observed in the item i , λ_{ij} is the factor loading of the item i in the factor j , F_j is the latent value of the factor j , τ_i is the intercept of the item i and ε_i is the residual error of the item i .

The intercept or scalar invariance criteria is assessed by comparing the metric model, which constrains factor loadings (λ_{ij}), with a scalar model that also constrains the intercepts (τ_i) so that they are equal between groups. It is considered achieved if the change in the overall fit indices is acceptable. ($\Delta CFI \leq 0.01$, $\Delta RMSEA \leq 0.015$, $\Delta SRMR \leq 0.010$). The process involves fitting both models and assessing the comparability of the fit indices to ensure that differences between groups reflect only variations in the latent scores and not in the items themselves.

Invariance of Residue

Residual invariance determines whether the residual variances (measurement errors) are equal across groups. This type of invariance ensures that the amount of variance in the items that is not explained by the factor is consistent across groups. To assess residual invariance, the model is adjusted to equalise measurement errors across groups and compared to models that do not have these restrictions. Residual invariance is crucial to ensure that differences in results are not due to differences in residual variability between groups (Putnick & Bornstein, 2016). The formula is presented below.

$$X_i = \lambda_{ij} F_j + \tau_i + \varepsilon_i$$

Where X_i is the response observed in the item i , λ_{ij} is the factor loading of the item i in the factor j , F_j is the latent value of the factor j , τ_i is the intercept of the item i and ε_i is the residual error of the item i , whose variance ($\text{Var}(\varepsilon_i)$) is restricted in this step.

The criteria for assessing residual invariance include the analysis of the global fit indices, where an acceptable change in the $\Delta CFI \leq 0.01$, $\Delta RMSEA \leq 0.015$, y $\Delta SRMR \leq 0.010$. For the model comparison, the scalar model, which restricts factor loadings and intercepts, is evaluated against the strict model, which further restricts the variances of the residual errors. If the changes in the fit indices are minimal, it is concluded that the residual invariance is achieved, indicating that the measurements are fully equivalent between the groups.

Temporal Invariance

Time invariance refers to the stability of the instrument's properties over time. A test should maintain the same structure and psychometric properties when administered at different points in time. This ensures that changes in observed scores over time reflect changes in the measured construct and not variations in the properties of the instrument (Li et al., 2018). The formula for time invariance can be expressed similarly to factorial invariance, but in a longitudinal context, considering the different points in time $t1, t2, \dots, tn$ in which the same constructs are measured.

$$X_{it} = \lambda_{ij(t)} F_{jt} + \tau_{it} + \varepsilon_{it}$$

Where X_{it} is the observed score on the item i in time t , $\lambda_{ij(t)}$ is the factor loading of the item i in the factor j in time t , F_{jt} is the latent factor in time t , τ_{it} is the intercept of the item i in time t , y ε_{it} is the residual error over time t .

Criteria for assessing time invariance include the analysis of global adjustment indices, where acceptable changes are considered in $\Delta CFI \leq 0.01$, $\Delta RMSEA \leq 0.015$, y $\Delta SRMR \leq 0.010$. For model comparison, an initial temporal model, which may have minor restrictions, is compared with more restrictive models that impose equality of factor loadings, intercepts and residual errors over time. If

the changes in the fit indices are minimal, it is concluded that time invariance has been achieved, indicating that the measurements are equivalent over time.

Structural invariance

The observed differences in the correlations between factors indicate that structural invariance does not hold, suggesting that the relationships between factors vary significantly across the different groups analysed (Sass & Schmitt, 2013). Structural invariance is essential to ensure that the latent relationships between the underlying constructs, represented by these factors, are consistent across groups. When this condition is not met, it implies that the underlying associations between the assessed dimensions or latent variables vary across groups, which may compromise the validity of intergroup comparisons. In a robust analysis, structural invariance is essential to assert that factors are comparable in terms of their relationships and magnitudes across different populations or subgroups (Kang et al., 2016; Rogers, 2024). The formula for structural invariance is expressed by structural equation modelling, similar to factorial invariance, but focusing on the relationships between latent factors (structural paths) and not only on loadings or intercepts.

$$F_1 = \beta_1 F_2 + \varepsilon$$

Where F_1, F_2 are the latent factors, β_1 is the regression coefficient representing the structural relationship between the factors, and ε is the specification error.

Model Interpretation

Differences in correlations between factors could reflect that different groups interpret or respond to the items that make up the scale in distinct ways (Loewenthal & Lewis, 2020; Reise et al., 1993; Streiner et al., 2024). This may signal that the factors not only have different meanings for each group but also that their relevance or weight within the model varies. For example, in the context of mental health, two factors that are highly correlated in one population may not show the same correlation in another, indicating differences in how psychological constructs manifest or interact within those groups. Such discrepancies could suggest that the underlying psychological or conceptual structure of the factors differs between groups, potentially influenced by variables such as cultural context, personal experiences, or socioeconomic differences.

Modeling and Comparisons

When differences in correlations between factors are observed, making direct comparisons of factor scores between different groups can be problematic and potentially misleading (Christensen & Golino, 2021). If the correlations between factors are not homogeneous, it indicates that the underlying factor model is not structurally invariant. In such a case, this means that comparisons between groups may not be valid, as the relationships between the constructs are not equivalent. Such discrepancies can lead to misinterpretations of the observed differences in factor scores, as they may reflect structural differences rather than true differences in the measured characteristics. Furthermore, the lack of structural invariance complicates the ability to conduct more advanced analyses, such as multi-group or causal path analyses, as the interpretations may be biased (Morin et al., 2020).

The existence of significant differences in correlations between factors suggests the need for a critical reevaluation of the proposed model. One option could be to explore alternative models that allow for a more flexible structure adapted to the observed differences between groups (Leitgöb et al., 2023). It may also be necessary to conduct further research to identify the underlying reasons for these differences in correlations, which could include a deeper analysis of the characteristics of the subgroups and their contexts. In some cases, it might be useful to adjust the model to reflect the lack of structural invariance, which could involve introducing specific parameters for each group or identifying subgroups that exhibit particular correlation patterns. These modifications would not only improve the model's accuracy but also provide a better understanding of the dynamics influencing different groups, facilitating a more appropriate and contextually relevant interpretation of the results.

Methods for Evaluating Invariance

To evaluate invariance, advanced statistical techniques such as Confirmatory Factor Analysis (CFA) (Kenny, 1976; Marsh et al., 2014; Rogers, 2024) and Structural Equation Modeling (SEM) (Jorgensen et al., 2020; Kline, 2011; Oberski & Satorra, 2013; Rosseel, 2012; 2020; Vispoel et al., 2024) are used. In the context of CFA, different hierarchical invariance models (configural, metric, scalar, and residual) are compared to determine if there are significant differences between groups. The process includes configuring the model, assessing metric invariance, scalar invariance, and residual invariance, thereby ensuring that the instrument is valid and reliable across different contexts and time points.

Psychometric validation is essential to guarantee the accuracy and relevance of assessment instruments in psychology. This process must begin with translation and cultural adaptation to ensure the instrument fits the target population as closely as possible. Psychometric property validation assesses the reliability and validity of the instrument, as well as the model fit, to achieve the effective application of psychometric tools in psychological research.

Conclusions

This article has reviewed the main theoretical and methodological aspects of adapting and validating psychometric instruments in the field of psychology, emphasizing the importance of rigorous evaluation at each phase of the process. The use of exploratory and confirmatory factor analysis, along with internal consistency and validity indicators such as Cronbach's alpha, omega, AVE, HTMT, CFI, TLI, RMSEA, and SRMR, is crucial to ensure that a test is not only reliable but also valid. Simultaneously, considering these indicators provides a comprehensive view of the instrument's structure and quality, allowing for the identification of possible areas for improvement and ensuring accurate assessment.

It is assumed that the correct linguistic and cultural adaptation of the instrument has been carried out to ensure that the items are relevant and understandable in the new cultural context. This process involves not only accurate translation but also the adaptation of the assessed concepts to reflect the realities and experiences of the target cultural group, minimizing the risk of biases or misinterpretations. In this way, the process of cross-cultural validation, along with psychometric validation, will enhance the utility and applicability of the instruments in diverse populations, promoting equity in psychological assessment.

In summary, applied psychometrics enables psychological instruments to be relevant and accurate for assessing different constructs in diverse populations. By ensuring the reliability and validity of measurements, the quality of the collected data is assured, which in turn strengthens the validity of the results obtained, both in research and clinical evaluation. This methodological approach is essential for advancing evidence-based psychology, where measurement accuracy is key for making well-founded clinical or educational decisions and for developing public policies that respond to the real needs of the populations assessed.

Recommendations

The use of multiple psychometric indicators is essential for a comprehensive validation of psychological assessment instruments. Reliability and validity are complementary aspects that should be addressed with a multidimensional approach. The use of Cronbach's alpha and omega as reliability indicators provides an overview of a test's internal consistency, while validity should be evaluated with measures such as AVE (Average Variance Extracted) and HTMT (Heterotrait-Monotrait Ratio), which assess convergent and discriminant validity. Furthermore, fit indices such as CFI (Comparative Fit Index), TLI (Tucker-Lewis Index), RMSEA (Root Mean Square Error of Approximation), and SRMR (Standardized Root Mean Square Residual) are crucial to evaluate whether the proposed model fits the data appropriately. By employing these multiple indicators, a completer and more robust picture of the psychometric properties of the instrument is obtained, avoiding the limitations of relying solely on one measure.

Confirmatory Factor Analysis (CFA) is an indispensable tool in the psychometric validation of instruments. While Exploratory Factor Analysis (EFA) is useful in the early stages to identify possible underlying structures, CFA allows testing more specific hypotheses about the factorial structure of

the data. The recommendation to use both approaches provides stronger validation, as EFA allows for open exploration of factors, while CFA validates the structure found with additional data or different samples. Statistical programs such as *SPSS*, *R*, *Mplus*, *Jasp*, or *Jamovi* offer robust tools that enable modeling complex data and conducting the necessary tests to ensure factorial validity.

It is essential to consider the recommended values for model fit indicator cut-offs. For CFI and TLI, values above 0.90 are considered indicative of a good fit, although ideally, they should exceed 0.95. For RMSEA, values below 0.08 indicate an acceptable fit, while values below 0.05 reflect an excellent fit. For SRMR, a value below 0.08 is interpreted as a good fit. Ensuring that models meet these criteria is crucial for rigorously validating the structure of instruments, guaranteeing that the proposed theoretical model fits the sample data optimally.

It is recommended to consider the International Test Commission (ITC) when adapting and using psychological tests in different cultural and linguistic contexts. The ITC promotes the ethical and scientific use of assessments globally, ensuring that tests maintain their validity and reliability. The organization has developed guidelines that address key issues such as cultural equivalence, ensuring that assessments are fair and appropriate for the populations they are applied to. For professionals working with international tests, following the ITC's recommendations is essential.

It is also advisable to follow the guidelines of the American Psychological Association (APA), a globally influential organization in psychology and psychometrics. The APA provides ethical and scientific standards that are used by psychologists worldwide, not just in the United States. Its committees and divisions offer guidelines on research and test adaptation, helping ensure that assessment instruments are culturally sensitive and valid. The APA guidelines are widely used in the adaptation of psychological and educational tests.

It is also recommended to consider the standards promoted by the European Federation of Psychologists' Associations (EFPA). This organization groups psychology associations in Europe and works on harmonizing psychological assessment standards in the region. The EFPA's psychological assessment committee focuses on the proper adaptation of tests in different European countries, promoting the development of common guidelines to ensure the validity and reliability of assessments in various cultural contexts. For professionals working in Europe or with European populations, following the EFPA's recommendations is essential.

Another relevant organization is the International Association for Educational Assessment (IAEA), particularly recommended for those involved in educational assessment. Although its primary focus is education, the IAEA also works on the adaptation of psychometric tests to be used in various educational systems and cultures. The organization promotes research to ensure that assessment instruments are culturally appropriate and valid. For those involved in educational assessment in an international context, it is crucial to consider the guidelines and studies promoted by the IAEA.

Finally, it is recommended to consider the products and guidelines of The Psychological Corporation, part of Pearson, one of the leading publishers in the creation and adaptation of psychological and educational tests. Pearson has a long history in the standardization and adaptation of tests for use in different cultural contexts. When working with assessment tools internationally, it is important to ensure that the instruments used meet validity and reliability standards in each country, which Pearson guarantees through its rigorous adaptation and standardization process.

Limitations

One of the main limitations of this tutorial is its scope, as it is designed as an introductory guide and does not cover all possible methodological variations or contexts in which confirmatory factor analysis (CFA) can be applied. This could limit its applicability in studies requiring advanced analyses or in very specific cultural and linguistic contexts, where methodological adaptations need to be more detailed.

Additionally, the tutorial focuses on the validation of psychological instruments through CFA, but it does not address other complementary techniques that could also be relevant for evaluating psychometric properties, such as exploratory factor analysis (EFA), structural equation modeling

(SEM), or methods for handling non-normal data or small samples. This might lead researchers to overlook other options that could be more suitable for their particular studies.

Another limitation is that the tutorial does not use specific, simplified examples to illustrate the steps of the analysis. Finally, the tutorial does not delve deeply into ethical or quality-related aspects of adapting and validating questionnaires in different populations, such as ensuring appropriate sample representation or the cultural sensitivity of items. This could be a challenge for researchers working with diverse or vulnerable populations.

Future Studies

Since this tutorial aims to be a practical guide, future studies should focus on the creation of more specialized complementary materials, such as advanced tutorials that include multigroup analysis, cross-validation, and handling of complex data. It would also be useful to develop guides tailored to specific contexts, such as the validation of instruments in minority populations or in longitudinal studies.

Additionally, practical guides addressing the use of specific software for performing confirmatory factor analysis would be valuable, comparing the advantages and limitations of platforms like JASP, AMOS, Mplus, R (with packages such as lavaan), or Jamovi. This could help researchers choose the most suitable tool according to their needs.

Finally, future studies could explore the integration of emerging technologies, such as artificial intelligence, to automate parts of the psychometric validation process, optimizing time and precision in analysis. This would be particularly relevant for researchers with limited resources or those working in contexts with high demands for data analysis.

Cross-cultural validation remains a critical area of research in psychometrics. Instruments developed in a specific cultural context may not have the same relevance or meaning in other cultures, posing the need to conduct studies that compare the underlying psychometric structures of tests across different populations. This not only helps ensure the validity of instruments in diverse contexts but also enables the establishment of norms that are globally applicable. Cross-cultural studies provide valuable information on the universality or cultural specificity of psychological constructs, which in turn promotes the creation of more inclusive measures adapted to cultural diversity.

The exploration of new psychometric techniques is another area that should continue to evolve. Traditionally, classical test theory has dominated the field of psychometric evaluation, but more advanced techniques such as Item Response Theory (IRT) and Structural Equation Modeling (SEM) have proven to be powerful tools for improving the accuracy and validity of instruments. These techniques allow, for example, the identification of items that function differently in distinct subgroups and a more accurate evaluation of the latent structure of an instrument. Additionally, these techniques offer greater flexibility in data analysis, especially in situations where the assumptions of classical theory are not fully met.

Longitudinal research is another area that has been overlooked in many psychometric studies. Most studies rely on cross-sectional designs, which limit our understanding of the stability and consistency of instruments over time. Future research should focus on conducting longitudinal validations (at least longitudinal cross-sectional) to assess how measurements behave at different time points and under various circumstances. This not only provides data on the temporal reliability of the instrument but also allows for the examination of changes in the constructs measured as individuals or populations evolve.

The use of emerging technologies also offers new possibilities for psychometrics. The growing prevalence of digital platforms and online psychological assessment apps necessitates the adaptation and validation of instruments for use in virtual environments. Although these technologies offer opportunities for large-scale test administration, they also introduce challenges related to accessibility, data security, and comparability with traditional methods. However, the implementation of computerized adaptive tests and the use of artificial intelligence to analyze

response patterns represents a significant opportunity to optimize the precision and efficiency of psychological assessment in digital settings.

Author Contributions: Contribution to the conception and design: A.R.; Contribution to data collection: A.R., L.B.-B., H.S., J.H.D. and F.J.R.-D.; Contribution to data analysis and interpretation: A.R., J.V.Q.-C., and L.B.-B.; Drafting and/or revising the article: A.R., L.B.-B., F.J.R.-D. and J.V.Q.-C.; Approval of the final version for publication: A.R., L.B.-B., J.H.D. and F.J.R.-D.; Obtaining authorization for the scale: J.H.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research was financed by the Universidad Politécnica Salesiana Sede Cuenca, Ecuador.

Institutional Review Board Statement: All procedures conducted in this study involving human participants adhered to the ethical standards established by the ethics committee. This study was approved by the Human Research Ethics Committee of the Pontifical Catholic University of Ecuador (code 046-UIO-2022). This research is derived from the research project entitled "Prevalence and risk factors of dating violence among Ecuadorian adolescents and university students and evaluation of the effectiveness of psychological intervention with virtual reality in reducing anxiety, depression and post-traumatic stress derived from violence", under the direction of Dr. Andres Ramirez, with the support of the Research Group in Psychology (GIPSI-SIB) of the Salesian Polytechnic University (Universidad Politécnica Salesiana), Cuenca, Ecuador.

Informed Consent Statement: Informed consent was obtained from all subjects involved in this study.

Data Availability Statement: The data supporting this research are publicly available and can be obtained by emailing the first author of this article.

Acknowledgments: The authors would like to thank the Universidad Politécnica Salesiana, Ecuador, and especially to Juan Cárdenas Tapia.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Abad, F. J., Olea, J., Ponsoda, V. & García, C. (2011). *Medición en ciencias sociales y de la salud*. Síntesis.
- Alamer, A., Schubert, F., & Henseler, J. (2024). When and how to use confirmatory composite analysis (CCA) in second language research. *Studies in Second Language Acquisition*, 46(2), 597-616. <https://doi.org/10.1017/S0272263124000020>
- American Psychological Association. (2022). *Publication manual of the American psychological association* (7a ed.). APA.
- American Educational Research Association, American Psychological Association y National Council on Measurement in Education. (2018). *Estándares para pruebas educativas y psicológicas*. American Educational Research Association.
- American Educational Research Association, American Psychological Association y National Council on Measurement in Education (2014). *Standards for educational and psychological testing*. APA.
- Bell, S. M., Chalmers, R. P., & Flora, D. B. (2024). The Impact of Measurement Model Misspecification on Coefficient Omega Estimates of Composite Reliability. *Educational and Psychological Measurement*, 84(1), 5-39. <https://doi.org/10.1177/00131644231155804>
- Benton, T. (2015). *An empirical assessment of Guttman's Lambda 4 reliability coefficient*. In *Quantitative psychology research: The 78th annual meeting of the Psychometric Society* (pp. 301-310). Springer International Publishing.
- Browne, M. W., & Cudeck, R. (1993). *Alternative ways of assessing model fit*. En K. A. Bollen & J. S. Long (Eds.), *Testing Structural Equation Models* (pp. 136-162). Sage Publications.
- Caldwell, A. R. (2022). SimplyAgree: an R package and jamovi module for simplifying agreement and reliability analyses. *Journal of Open Source Software*, 7(71), 4148. <https://doi.org/10.21105/joss.04148>
- Cahigas, M. M. L., Prasetyo, Y. T., Nadlifatin, R., Persada, S. F., & Gumasing, M. J. J. (2023). Determinants of continuous visiting behavior to Palawan, Philippines: Integrating Uncertainty Reduction Theory and Expectation Confirmation Theory. *Plos one*, 18(10), e0291694. <https://doi.org/10.1371/journal.pone.0291694>

- Cai, L., Chung, S. W., & Lee, T. (2023). Incremental model fit assessment in the case of categorical data: Tucker–Lewis index for item response theory modeling. *Prevention Science*, 24(3), 455–466. <https://doi.org/10.1007/s11121-021-01253-4>
- Cherry, K. M., Vander Hoeven, E., Patterson, T. S., & Lumley, M. N. (2021). Defining and measuring “psychological flexibility”: A narrative scoping review of diverse flexibility and rigidity constructs and perspectives. *Clinical psychology review*, 84, 101973. <https://doi.org/10.1016/j.cpr.2021.101973>
- Christensen, A. P., & Golino, H. (2021). On the equivalency of factor and network loadings. *Behavior research methods*, 53(4), 1563–1580. <https://doi.org/10.3758/s13428-020-01500-6>
- De Roover, K. (2021). Finding clusters of groups with measurement invariance: Unraveling intercept non-invariance with mixture multigroup factor analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, 28(5), 663–683. <https://doi.org/10.1080/10705511.2020.1866577>
- Doval, E., Viladrich, C., & Angulo-Brunet, A. (2023). Coefficient alpha: the resistance of a classic. *Psicothema*, 35(1), 5. <https://doi.org/10.7334/psicothema2022.321>
- Elosua Oliden, P. y Egaña, M. (2020). *Psicometría aplicada. Guía para el análisis de datos y escalas con jamovi*. Universidad del País Vasco / Euskal Herriko Unibertsitatea.
- Embretson, S. E. y Reise, S. P. (2000). *Item response theory for psychologists*. Lawrence Erlbaum Associates.
- Ferrando, P. J., Lorenzo-Seva, U., Hernández-Dorado, A. y Muñiz, J. (2022). Decálogo para el análisis factorial de los ítems de un test. *Psicothema*, 34(1), 7–17. <https://doi.org/10.7334/psicothema2021.456>
- Field, A. (2024). *Discovering statistics using IBM SPSS statistics*. Sage publications limited.
- Geiser, C. (2023). Structural equation modeling with the Mplus and lavaan programs. *Handbook of structural equation modeling*, 241–258.
- Goretzko, D., Pham, T. T. H., & Bühner, M. (2021). Exploratory factor analysis: Current use, methodological developments and recommendations for good practice. *Current psychology*, 40, 3510–3521. <https://doi.org/10.1007/s12144-019-00300-2>
- Goretzko, D., Siemund, K., & Sterner, P. (2024). Evaluating Model Fit of Measurement Models in Confirmatory Factor Analysis. *Educational and Psychological Measurement*, 84(1), 123–144. <https://doi.org/10.1177/00131644231163813>
- Haladyna, T. M. & Rodriguez, M.C. (2013). *Developing and validating test items*. Routledge.
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55.
- Hunsley, J., & Meyer, G. J. (2003). The Incremental Validity of Psychological Testing and Assessment: Conceptual, Methodological, and Statistical Issues. *Psychological Assessment*, 15(4), 446–455. <https://doi.org/10.1037/1040-3590.15.4.446>
- Hutchinson, S. R., & Olmos, A. (1998). Behavior of descriptive fit indexes in confirmatory factor analysis using ordered categorical data. *Structural Equation Modeling: A Multidisciplinary Journal*, 5(4), 344–364. <https://doi.org/10.1080/10705519809540111>
- International Test Commission. (2017). *The ITC Guidelines for Translating and Adapting Tests* (Second edition).
- Jorgensen T. D., Pornprasertmanit S., Schoemann A. M., Rosseel Y. (2020). *semTools: Useful tools for structural equation modeling*. <https://CRAN.R-project.org/package=semTools>
- Kang, Y., McNeish, D. M., & Hancock, G. R. (2016). The Role of Measurement Quality on Practical Guidelines for Assessing Measurement and Structural Invariance. *Educational and Psychological Measurement*, 76(4), 533–561. <https://doi.org/10.1177/0013164415603764>
- Kalkbrenner, M. T. (2024). Choosing Between Cronbach’s Coefficient Alpha, McDonald’s Coefficient Omega, and Coefficient H: Confidence Intervals and the Advantages and Drawbacks of Interpretive Guidelines. *Measurement and Evaluation in Counseling and Development*, 57(2), 93–105. <https://doi.org/10.1080/07481756.2023.2283637>
- Kapur, P., Misra, G., & K. Verma, N. (2022). Methodological Approach. In *Psychological Perspectives on Identity, Religion and Well-Being* (pp. 61–72). Springer Nature Singapore. https://doi.org/10.1007/978-981-19-2844-4_4
- Kenny D. A. (1976). An empirical application of confirmatory factor analysis to the multitrait-multimethod matrix. *Journal of Experimental Social Psychology*, 12(3), 247–252. [https://doi.org/10.1016/0022-1031\(76\)90055](https://doi.org/10.1016/0022-1031(76)90055)

- Kim, E. S., Cao, C., Wang, Y., & Nguyen, D. T. (2017). Measurement invariance testing with many groups: A comparison of five approaches. *Structural Equation Modeling: A Multidisciplinary Journal*, 24(4), 524-544. <https://doi.org/10.1080/10705511.2017.1304822>
- Kim, E. S., & Yoon, M. (2011). Testing measurement invariance: A comparison of multiple-group categorical CFA and IRT. *Structural Equation Modeling*, 18(2), 212-228. <https://doi.org/10.1080/10705511.2011.557337>
- Kline R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). Guilford.
- Lakens, D. (2022). Sample size justification. *Collabra: psychology*, 8(1), 33267. <https://doi.org/10.1525/collabra.33267>
- Leitgöb, H., Seddig, D., Asparouhov, T., Behr, D., Davidov, E., De Roover, K., ... & van de Schoot, R. (2023). Measurement invariance in the social sciences: Historical development, methodological challenges, state of the art, and future perspectives. *Social science research*, 110, 102805. <https://doi.org/10.1016/j.ssresearch.2022.102805>
- Li, Z., Shi, H. S., Elis, O., Yang, Z. Y., Wang, Y., Lui, S. S., ... & Chan, R. C. (2018). The structural invariance of the Temporal Experience of Pleasure Scale across time and culture. *PsyCh Journal*, 7(2), 59-67. <https://doi.org/10.1002/pchj.207>
- Loewenthal, K. M., & Lewis, C. A. (2020). *An introduction to psychological tests and scales*. Routledge. <https://doi.org/10.4324/9781315561387>
- López-Aguado, M., & Gutiérrez-Provecho, L. (2019). Com dur a terme i interpretar una anàlisi factorial exploratòria utilitzant SPSS. *REIRE Revista d'Innovació I Recerca En Educació*, 12(2), 1-14. <https://doi.org/10.1344/reire2019.12.227057>
- Luo, L., Arizmendi, C., & Gates, K. M. (2019). Exploratory factor analysis (EFA) programs in R. *Structural Equation Modeling: A Multidisciplinary Journal*, 26(5), 819-826. <https://doi.org/10.1080/10705511.2019.1615835>
- Maassen, E., D'Urso, E. D., Van Assen, M. A., Nuijten, M. B., De Roover, K., & Wicherts, J. M. (2023). The dire disregard of measurement invariance testing in psychological science. *Psychological Methods*. <https://doi.org/10.1037/met0000624>
- Malkewitz, C. P., Schwall, P., Meesters, C., & Hardt, J. (2023). Estimating reliability: A comparison of Cronbach's α , McDonald's ω and the greatest lower bound. *Social Sciences & Humanities Open*, 7(1), 100368. <https://doi.org/10.1016/j.ssaho.2022.100368>
- Martínez-Arias, M. R., Hernández-Lloreda, M. J. & Hernández-Lloreda, M. V. (2006). *Psicometría*. Alianza Editorial.
- Marsh, H., & Alamer, A. (2024). When and how to use set-exploratory structural equation modelling to test structural models: A tutorial using the R package lavaan. *British Journal of Mathematical and Statistical Psychology*. <https://doi.org/10.1111/bmsp.12336>
- Marsh, H. W., Hau, K. T., & Grayson, D. (2005). *Goodness of Fit in Structural Equation*. Contemporary psychometrics.
- Marsh, H. W., Guo, J., Dicke, T., Parker, P. D., & Craven, R. G. (2020). Confirmatory factor analysis (CFA), exploratory structural equation modeling (ESEM), and set-ESEM: Optimal balance between goodness of fit and parsimony. *Multivariate behavioral research*, 55(1), 102-119. <https://doi.org/10.1080/00273171.2019.1602503>
- Marsh H. W., Morin A. J. S., Parker P. D., & Kaur G. (2014). Exploratory structural equation modeling: An integration of the best features of exploratory and confirmatory factor analysis. *Annual Review of Clinical Psychology*, 10, 85-110. <https://doi.org/10.1146/annurev-clinpsy-032813-153700>
- McNeish D. (2017). Thanks coefficient alpha, we'll take it from here. *Psychological Methods*, 23(3), 412-433. <https://doi.org/10.1037/met0000144>
- Meyer, G. J., Finn, S. E., Eyde, L. D., Kay, G. G., Moreland, K. L., Dies, R. R., ... & Reed, G. M. (2001). Psychological testing and psychological assessment: A review of evidence and issues. *American psychologist*, 56(2), 128.
- Motallebzadeh, Z. (2023). A comparison of different methods for investigating the reliability of C-tests. *Educational Methods & Practice*, 1(1).
- Montoya, A. K., & Edwards, M. C. (2021). The Poor Fit of Model Fit for Selecting Number of Factors in Exploratory Factor Analysis for Scale Evaluation. *Educational and Psychological Measurement*, 81(3), 413-440. <https://doi.org/10.1177/0013164420942899>

- Moral de la Rubia, J. (2019). Revisión de los criterios para validez convergente estimada a través de la Varianza Media Extraída. *Psychologia*, 13(2), 25–41. <https://doi.org/10.21500/19002386.4119>
- Morin, A. J., Myers, N. D., & Lee, S. (2020). Modern factor analytic techniques: Bifactor models, exploratory structural equation modeling (ESEM), and bifactor-ESEM. *Handbook of sport psychology*, 1044-1073. <https://doi.org/10.1002/9781119568124.ch51>
- Muñiz, J. (1997). *Introducción a la teoría de respuesta a los ítems*. Pirámide.
- Muñiz, J. (2002). *Teoría clásica de los tests*. Pirámide.
- Muñiz, J. (2018). *Introducción a la psicometría*. Teoría clásica y TRI. Pirámide.
- Muñiz, J. & Fonseca-Pedrero, E. (2019). Diez pasos para la construcción de un test. *Psicothema*, 31(1), 7-16. <https://doi.org/10.7334/psicothema2018.291>
- Oberski D. L., & Satorra A. (2013). Measurement error models with uncertainty about the error variance. *Structural Equation Modeling: A Multidisciplinary Journal*, 20(3), 409–428. <https://doi.org/10.1080/10705511.2013.797820>
- Olea, J. & Ponsoda, V. (2004). *Tests adaptativos informatizados*. Universidad Nacional de Educación a Distancia.
- Olea, J., Ponsoda, V. y Prieto, G. (1998). *Tests informatizados: Fundamentos y aplicaciones*. Pirámide.
- Pfadt, J. M., Bergh, D. V. D., Sijtsma, K., & Wagenmakers, E. J. (2023). A tutorial on Bayesian single-test reliability analysis with JASP. *Behavior Research Methods*, 55(3), 1069-1078. <https://doi.org/10.3758/s13428-021-01778-0>
- Pfadt, J. M., van den Bergh, D., Sijtsma, K., & Wagenmakers, E. J. (2021). A tutorial on Bayesian single-test reliability analysis with JASP.
- Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental review*, 41, 71-90. <https://doi.org/10.1016/j.dr.2016.06.004>
- Reise, S. P., Widaman, K. F., & Pugh, R. H. (1993). Confirmatory factor analysis and item response theory: Two approaches for exploring measurement invariance. *Psychological Bulletin*, 114(3), 552–566. <https://doi.org/10.1037/0033-2909.114.3.552>
- Reynolds, C. R., Altmann, R. A., & Allen, D. N. (2021). *Introduction to Psychological Assessment*. In Mastering Modern Psychological Testing: Theory and Methods (pp. 1-47). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-59455-8_1
- Ridley, C. R., Mollen, D., Console, K., & Yin, C. (2021). Multicultural counseling competence: A construct in search of operationalization. *The Counseling Psychologist*, 49(4), 504-533. <https://doi.org/10.1177/0011000020988110>
- Rodríguez-Rodríguez, J., & Reguant-Álvarez, M. (2020). Calcular la fiabilitat d'un qüestionari o escala mitjançant l'SPSS: el coeficient alfa de Cronbach. *REIRE Revista d'Innovació i Recerca En Educació*, 13(2), 1–13. <https://doi.org/10.1344/reire2020.13.230048>
- Rogers, P. (2024). Best practices for your confirmatory factor analysis: A JASP and lavaan tutorial. *Behavior Research Methods*, 1-21. <https://doi.org/10.3758/s13428-024-02375-7>
- Rönkkö, M., & Cho, E. (2022). An Updated Guideline for Assessing Discriminant Validity. *Organizational Research Methods*, 25(1), 6-14. <https://doi.org/10.1177/1094428120968614>
- Rosseel Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1–36. <https://doi.org/10.18637/jss.v048.i02>
- Rosseel Y. (2020). *Small sample solutions for structural equation modeling*. In van de Schoot R., Miočević M. (Eds.), *Small sample size solutions: A guide for applied researchers and practitioners* (pp. 226–238). Taylor & Francis. <https://doi.org/10.4324/9780429273872-19>
- Rothwell, A., Moorkens, J., Fernández-Parra, M., Drugan, J., & Austermuehl, F. (2023). *Translation tools and technologies*. Routledge. <https://doi.org/10.4324/9781003160793>
- Rutkowski, L., & Svetina, D. (2017). Measurement Invariance in International Surveys: Categorical Indicators and Fit Measure Performance. *Applied Measurement in Education*, 30(1), 39–51. <https://doi.org/10.1080/08957347.2016.1243540>
- Şahin, M., & Aybek, E. (2019). Jamovi: an easy to use statistical software for the social scientists. *International Journal of Assessment Tools in Education*, 6(4), 670-692. <https://doi.org/10.21449/ijate.661803>

- Salessi, S., & Omar, A. (2019). Validez discriminante, predictiva e incremental de la escala de com-portamientos laborales proactivos de Belschaky Den Hartog. *Revista Costarricense de Psicología*, 38(1), 75–93. <https://doi.org/10.22544/rcps.v38i01.05>
- Santisteban, C. (2009). *Principios de psicometría*. Síntesis.
- Sass, D. A., & Schmitt, T. A. (2013). *Testing measurement and structural invariance: Implications for practice*. In Handbook of quantitative methods for educational research (pp. 315-345). Brill.
- Schmitt, N., & Kuljanin, G. (2008). Measurement invariance: Review of practice and implications. *Human resource management review*, 18(4), 210-222. <https://doi.org/10.1016/j.hrmr.2008.03.00>
- Schneider, S., Hernandez, R., Junghaenel, D. U., Jin, H., Lee, P. J., Gao, H., ... & Stone, A. A. (2024). Can you tell people's cognitive ability level from their response patterns in questionnaires?. *Behavior Research Methods*, 56: 1-18. <https://doi.org/10.3758/s13428-024-02388-2>
- Shaked, D., Faulkner, L. M. D., Tolle, K., Wendell, C. R., Waldstein, S. R., & Spencer, R. J. (2019). Reliability and validity of the Conners' Continuous Performance Test. *Applied Neuropsychology: Adult*, 27(5), 478–487. <https://doi.org/10.1080/23279095.2019.1570199>
- Shi, D., Lee, T., & Maydeu-Olivares, A. (2019). Understanding the Model Size Effect on SEM Fit Indices. *Educational and Psychological Measurement*, 79(2), 310-334. <https://doi.org/10.1177/0013164418783530>
- Shroff, A., Roulston, C., Fassler, J., Dierschke, N. A., Todd, J. S. P., Ríos-Herrera, Á., Plastino, K. A., & Schleider, J. L. (2023). A Digital Single-Session Intervention Platform for Youth Mental Health: Cultural Adaptation, Evaluation, and Dissemination. *JMIR Mental Health*, 10, e43062. <https://doi.org/10.2196/43062>
- Skadiņa, I., Vasiljevs, A., Pinnis, M., Bērziņš, A., Aranberri, N., Bogaert, J. V. D., ... & Way, A. (2023). *Deep Dive Machine Translation*. In *European Language Equality: A Strategic Agenda for Digital Language Equality* (pp. 263-287). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-28819-7_40
- Strauss M. E., Smith G. T. (2009). Construct validity: Advances in theory and methodology. *Annual Review of Clinical Psychology*, 5, 1–25. <https://doi.org/10.1146/annurev.clinpsy.032408.153639>. Construct
- Streiner, D. L., Norman, G. R., & Cairney, J. (2024). *Health measurement scales: a practical guide to their development and use*. Oxford university press.
- Sürücü, L., Şeşen, H., & Maslakçı, A. (2023). *Regression, mediation/moderation, and structural equation modeling with SPSS, AMOS, and PROCESS Macro*. Livre de Lyon.
- Tavakol, M., & Wetzel, A. (2020). Factor Analysis: a means for theory and instrument development in support of construct validity. *International journal of medical education*, 11, 245. <https://doi.org/10.5116/ijme.5f96.0f4a>
- Taasobshirazi, G., & Wang, S. (2016). The performance of the SRMR, RMSEA, CFI, and TLI: An examination of sample size, path size, and degrees of freedom. *Journal of Applied Quantitative Methods*, 11(3), 31-39.
- Thakkar, J. J. (2020). *Structural equation modelling*. Application for Research and Practice.
- Thériault, R. (2023). lavaanExtra: Convenience Functions for Package lavaan. *Journal of Open Source Software*, 8(90), 5701. <https://doi.org/10.21105/joss.05701>
- Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, 38(1), 1-10.
- Van De Schoot, R., Schmidt, P., De Beuckelaer, A., Lek, K., & Zondervan-Zwijnenburg, M. (2015). Measurement invariance. *Frontiers in psychology*, 6, 1064. <https://doi.org/10.3389/fpsyg.2015.01064>
- Van Laar, S., & Braeken, J. (2022). Caught off base: A note on the interpretation of incremental fit indices. *Structural Equation Modeling: A Multidisciplinary Journal*, 29(6), 935-943. <https://doi.org/10.1080/10705511.2022.2050730>
- Viladrich, C., Angulo-Brunet, A., & Doval, E. (2017). Un viaje alrededor de alfa y omega para estimar la fiabilidad de consistencia interna. *Anales de psicología*, 33(3), 755–782. <https://doi.org/10.6018/analesps.33.3.268401>
- Vispoel, W. P., Lee, H., & Hong, H. (2024). Analyzing multivariate generalizability theory designs within structural equation modeling frameworks. *Structural Equation Modeling: A Multidisciplinary Journal*, 31(3), 552-570. <https://doi.org/10.1080/10705511.2023.2222913>
- Wang, K., Xu, Y., Wang, C., Tan, M., & Chen, P. (2020). A corrected goodness-of-fit index (CGFI) for model evaluation in structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 27(5), 735-749. <https://doi.org/10.1080/10705511.2019.1695213>

- Watkins, M. W. (2018). Exploratory Factor Analysis: A Guide to Best Practice. *Journal of Black Psychology*, 44(3), 219–246. <https://doi.org/10.1177/0095798418771807>
- Widaman, K. F., & Helm, J. L. (2023). *Exploratory factor analysis and confirmatory factor analysis*. In H. Cooper, M. N. Coutanche, L. M. McMullen, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *APA handbook of research methods in psychology: Data analysis and research publication* (2nd ed., pp. 379–410). American Psychological Association. <https://doi.org/10.1037/0000320-017>
- Wilson, M. (2023). *Constructing measures: An item response modeling approach*. Routledge. <https://doi.org/10.4324/9781003286929>
- Xu, R., & Soland, J. (2024). Beyond group comparisons: Accounting for intersectional sources of bias in international survey measures. *International Journal of Testing*, 24(3), 230–258. <https://doi.org/10.1080/15305058.2024.2364168>
- Zheng, B. Q., & Bentler, P. M. (2024). Enhancing Model Fit Evaluation in SEM: Practical Tips for Optimizing Chi-Square Tests. *Structural Equation Modeling: A Multidisciplinary Journal*, 1–6. <https://doi.org/10.1080/10705511.2024.2354802>
- Zijlman E. A. O., van der Ark L. A., Tijmstra J., Sijtsma K. (2018). Methods for estimating item-score reliability. *Applied Psychological Measurement*, 42(7), 553–570. <https://doi.org/10.1177/0146621618758290>
- Zimmerman D. W. (2007). Correction for attenuation with biased reliability estimates and correlated errors in populations and samples. *Educational and Psychological Measurement*, 67(6), 920–939. <https://doi.org/10.1177/0013164406299132>
- Zinbarg R. E., Revelle W., Yovel I., Li W. (2005). Cronbach's α , Revelle's β and McDonald's ω H: Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika*, 70(1), 123–133. <https://doi.org/10.1007/s11336-003-0974-7>

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.