FIT INDICES IN STRUCTURAL EQUATION MODELING AND CONFIRMATORY FACTOR ANALYSIS: REPORTING GUIDELINES

ABSTRACT

This research explores the essential aspects of reporting fit indices in Structural Equation Modeling (SEM), focusing on their significance, methodologies for evaluation, and implications for model validity. The aim is to provide a comprehensive understanding of how fit indices contribute to the rigor and reliability of SEM studies. Methodologically, the study reviews prominent fit indices such as Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA), Standardized Root Mean Residual (SRMR), and Chi-Square Test of Model Fit (χ^2). Each index is defined, and specific threshold values are discussed to guide researchers in interpreting their findings effectively. Originality in this study lies in synthesizing current literature to emphasize the importance of transparent reporting practices in SEM, enhancing methodological clarity and promoting replicable research outcomes. Contributions include a structured approach to understanding fit indices' roles in model assessment and validation, aiding researchers in advancing theoretical frameworks with robust empirical support.

JEL Classification: C52, C38, C87

Keywords: Structural Equation Modeling, Fit Indices, Model Evaluation, Reliability, Validity, HTMT

I. INTRODUCTION

Structural Equation Modeling (SEM) is a comprehensive statistical method that integrates factor analysis and multiple regression to examine complex relationships among observed and latent variables, aiding in the testing of theoretical models involving multiple dependent relationships across fields like psychology, sociology, and marketing (Byrne, B. M. (2010); Kline, R. B. (2015); Schumacker, R. E., & Lomax, R. G. (2016)). Fit indices in SEM are essential as they provide quantitative measures to assess model adequacy, guiding the refinement of models to better capture theoretical constructs (Hu, L. T., & Bentler, P. M. (1999); Marsh et al., (2004)). Confirmatory Factor Analysis (CFA) tests whether observed variables represent hypothesized latent constructs, requiring a predefined number and nature of factors (Brown, T. A. (2015)). CFA is used to validate measurement instruments and assess construct validity, with fit indices evaluating model-data fit to guide necessary modifications. CFA is a key component of the broader family of methods known as structural equation modeling (SEM) and is crucial for validating measurement models in path or structural analyses (Brown, 2006; MacCallum & Austin, 2000). When conducting SEM, researchers typically begin by evaluating the measurement model to ensure that the measured variables accurately represent the intended constructs or factors before proceeding to assess the structural model. There has been a positive trend in the use of CFA, with most applications focusing on scale development and construct validation (Brown & Russell, 2002). Mediation analysis explores whether the effect of an independent variable on a dependent variable is mediated by another variable, used extensively in psychological and social research to understand causal pathways (Preacher, K. J., & Hayes, A. F. (2008); Baron, R. M., & Kenny, D. A. (1986)). In SEM-based mediation, fit indices are crucial for evaluating overall model fit and validating mediation effects. Path analysis and latent growth modeling (LGM) are related techniques; path analysis examines direct and indirect relationships among observed variables, while LGM estimates growth trajectories over time for latent constructs (Ximénez, et al., (2022)). Both rely on fit indices to ensure model adequacy and accurate representation of developmental processes. Fit indices are vital across these techniques as they validate hypothesized models, enable model comparison, identify misspecifications, and enhance measurement reliability and validity, thus ensuring models accurately reflect theoretical constructs and relationships (Curran, P. J., Obeidat, K., & Losardo, D. (2010); Kyriazos, T. A. (2018); Shi et al., (2022)).

Many researchers advocate for the use of a range of fit indices when assessing model fit in structural equation modeling (SEM). Marsh, Balla, and Hau (1996) emphasize the importance of considering multiple indices to gain a comprehensive understanding of model fit. Similarly, Jaccard and Wan (1996) recommend using indices from different classes to address the limitations inherent in any single index. This strategy ensures a more robust evaluation of the model. Hu and Bentler (1999) suggest combining multiple fit indices, including RMSEA, SRMR, CFI, and TLI, to make more informed decisions about model fit. Kline (2015) recommends using a variety of fit indices and emphasizes the need for a nuanced approach in interpreting these indices. Byrne (2016) advises researchers to report multiple fit indices to provide a comprehensive evaluation of model fit, as different indices can offer different insights into the model's performance (Shi et al., 2020).

COMMONLY USED AND REPORTED VARIOUS FIT INDICES

There are numerous excellent practice guidelines available for both SEM and CFA (Hoyle & Panter, 1995; Boomsma, 2000; Byrne, 2001; Thompson, 2004; Brown, 2006; Kline, 2005). An issue of particular importance across these guidelines is what should be reported in SEM/CFA studies. Reporting standards emphasize the need to include details on model specification, data preparation, estimation methods, fit indices, and model modification processes (Schumacker & Lomax, 2010; Bentler, 2007; Hu & Bentler, 1999; Marsh, Hau, & Wen, 2004). Additionally, for researchers aiming to contribute original insights to the field, it's crucial to provide comprehensive and transparent reporting to facilitate replication and validation of findings. This includes detailing the theoretical rationale for the model, the selection and justification of indicators, and the steps taken to ensure the robustness of the findings (Muthén & Muthén, 2010; Little, 2013; Bollen, 1989). This thoroughness helps to build a solid foundation for advancing knowledge and promoting further research in SEM and CFA. The state of reporting results from SEM analyses has indeed been critiqued for its variability and occasional inadequacy in providing comprehensive details. Empirical findings support this assertion, highlighting issues such as inconsistent reporting practices and insufficient documentation of key methodological details. For instance, Boomsma (2000) discussed the need for clearer reporting standards to enhance transparency and reproducibility in SEM studies. MacCallum and Austin (2000) examined the variability in reporting fit indices and model fit assessment criteria, emphasizing the importance of standardizing these practices. McDonald and Ho (2002) pointed out deficiencies in reporting measurement model specifications and estimation procedures, which can affect the reliability and validity of study findings. Steiger (2001) highlighted challenges related to the reporting of model modification indices and the implications for model refinement and interpretation (MacCallum & Austin (2000); McDonald and Ho (2002)). Existing reviews of SEM studies often focus on methodological aspects and model fit criteria rather than specifically addressing reporting practices in the CFA literature. This gap in the literature suggests a need for more comprehensive examinations of how CFA studies report their methods and findings. For example, Byrne (2001) discussed the importance of transparent reporting in CFA studies to enhance the credibility and replicability of research findings. Kline (2005) emphasized the need for detailed documentation of measurement model specifications and fit indices in CFA reports. Thompson (2004) highlighted challenges in assessing construct validity and the importance of clearly articulating measurement and structural model details. These citations emphasize the call for empirical investigations or systematic reviews that explicitly examine and critique the state of reporting practices in the CFA literature. Such studies could provide valuable insights into current practices, identify areas needing improvement, and offer recommendations for enhancing the quality of reporting in CFA research. Furthermore, reporting guidelines in SEM/CFA are not universally standardized; however, there is substantial consensus among scholars who have examined this issue (Medsker et al., 1994; Chin, 1998; Hoyle & Panter, 1995; MacCallum & Austin, 2000; McDonald & Ho, 2002; Barrett, 2007; Thompson, 2004). Scholars emphasize the importance of consistent reporting practices to ensure transparency and reproducibility. Key aspects often highlighted include the thorough documentation of model specifications, details of estimation procedures, justification for model fit indices, handling of missing data, and sensitivity analyses to assess model robustness (Bollen, 1989; Schreiber et al., 2006; Kline, 2015; Wolf et al., 2013). These guidelines aim to improve the quality and rigor of SEM/CFA research, facilitating clearer interpretation and comparison of findings across studies (Perry et al., 2015; Marsh et al., 2020).

In structural equation modeling (SEM), confirmatory factor analysis (CFA), and mediation analysis, a variety of fit indices are employed to evaluate model fit. These include the Chi-Square Test (χ^2), which assesses the discrepancy between observed and expected covariance matrices, and the Chi-Square to Degrees of Freedom Ratio (χ^2/df), which adjusts for model complexity (Schumacker & Lomax, 2016). The Goodness-of-Fit Index (GFI) and the Adjusted Goodness-of-Fit Index (AGFI) measure the fit between the observed and model-implied covariance matrices, with AGFI adjusting for degrees of freedom (Kline, 2015). Comparative Fit Index (CFI) and Tucker-Lewis Index (TLI), also known as the Non-Normed Fit Index (NNFI), compare the fit of the target model to an independent baseline model, with higher values indicating better fit (Hu & Bentler, 1999). The Normed Fit Index (NFI) also compares the target model to a null model, while the Root Mean Square Error of Approximation (RMSEA) and the Standardized Root Mean Square Residual (SRMR) measure model fit per degree of freedom and standardized differences between observed and predicted correlations, respectively (Byrne, 2010). Additionally, the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) are used for model comparison, with lower values indicating better fit, and the Parsimony Goodness-of-Fit Index (PGFI) and Parsimony Normed Fit Index (PNFI) adjust other indices for model complexity (Marsh, Hau, & Wen, 2004). Collectively, these fit indices provide a comprehensive assessment of model adequacy, guiding researchers in validating, comparing, and refining models to ensure accurate representation of theoretical constructs and relationships (Brown, 2015; Gallagher, M. W., & Brown, T. A. (2013)).

II. PROMINENT FIT INDICES FOR REPORTING

In the domain of Structural Equation Modeling (SEM), fit indices serve as critical metrics for assessing the alignment between theoretical models and empirical data. These indices provide quantitative measures that help researchers evaluate the adequacy of their models in capturing the relationships among latent constructs and observed variables. Understanding these fit indices is essential for ensuring the reliability, validity, and generalizability of SEM findings. This introduction sets the stage for exploring how these indices, such as Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA), Standardized Root Mean Residual (SRMR), and Chi-Square Test of Model Fit (χ^2), contribute to the rigorous evaluation and reporting of SEM models.The following are the most prominent fit indices while reporting the SEM and CFA models:

1. ABSOLUTE FIT INDICES

Absolute fit indices provide a baseline assessment of how well the hypothesized model fits the observed data. They are essential for initial model evaluation and can guide researchers in determining whether the model adequately represents the data before considering more complex evaluations such as incremental or comparative fit indices. These indices, particularly the chi-square test and its derivatives, serve as fundamental measures in SEM, providing a direct test of the fit between the theoretical model and the observed data. While no single fit index can provide a complete picture of model fit, collectively they offer valuable insights into the adequacy of the specified model.

1.1 Chi-Square Test (χ^2)

When reporting Structural Equation Modeling (SEM), Confirmatory Factor Analysis (CFA), or mediation models, the first fit index typically reported is the Chi-Square Test (χ^2). This index assesses the overall fit of the model by comparing the observed covariance matrix with the model-implied covariance matrix. It provides a fundamental indication of whether the model adequately fits the data. Chi-Square test is a statistical test that evaluates the discrepancy between the observed data and the model-implied covariance matrices. It is reported along with degrees of freedom (df) and the associated p-value. In AMOS it is also known as the Chi-Square Minimum (CMIN). A non-significant chi-square ($p > 0.05$) indicates a good fit, but it can be sensitive to sample size.

Chi-Square Test (χ^2): χ^2 (102) = 154.50, p = 0.25

Rationale for Reporting Chi-Square Test First: The chi-square test is the most fundamental measure of model fit and provides a direct test of the model's overall goodness-of-fit. Many other fit indices are derived from or adjusted based on the chi-square statistic, making it a logical starting point for model evaluation. Traditionally, the chi-square test has been the first reported fit index in SEM, CFA, and mediation analysis, setting the context for additional fit indices.

1.2 The Chi-Square to Degrees of Freedom Ratio (χ^2/df)

This is a normalized version of the chi-square statistic used in structural equation modeling (SEM) and confirmatory factor analysis (CFA). It adjusts the chi-square statistic for the complexity of the model by dividing it by the degrees of freedom, providing a more interpretable measure of model fit (Ullman, 2001; Schumacker & Lomax, 2004; Kline, R. B. (2015)).

$$\chi^2/df = (\text{Chi-Square}) / (\text{Degrees of Freedom})$$

It indicates how well the model fits the data relative to the number of estimated parameters. A lower χ^2/df ratio suggests a better fit. Generally, a χ^2/df ratio less than 2 is considered indicative of a good fit. A ratio less than 3 is often considered acceptable (Hu & Bentler, (1999); Kline, 2015); Schumacker, R. E., & Lomax, R. G. (2016)). However, in a seminal work by Hu and Bentler (1999) suggest that values less than 5 indicate a reasonable fit, though this can be somewhat flexible depending on the complexity of the model and sample size. Bentler (1990), discusses that χ^2/df values around 2 to 5 may indicate an acceptable fit, with lower values preferred for better fit.

Advantages: By dividing chi-square by degrees of freedom, χ^2/df accounts for the complexity of the model, making it more robust and less sensitive to sample size compared to the raw chi-square statistic. χ^2/df provides a straightforward interpretation where lower values indicate better fit, allowing for easier comparison across different models or datasets (Clark, D. A., & Bowles, R. P. (2018)).

Limitations: Like the chi-square statistic, χ^2/df can still be sensitive to sample size, potentially leading to significant results in large samples even when the model fit is reasonably good. Extremely complex models or those with small degrees of freedom may artificially inflate the χ^2/df ratio, potentially leading to misinterpretation of model fit.

Note: These references provide general guidelines rather than strict rules, as interpretations can vary based on factors such as sample size, model complexity, and research context. Researchers often consider multiple fit indices alongside χ^2/df, such as Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), and Root Mean Square Error of Approximation (RMSEA), for a comprehensive assessment of model fit.

1.3 Goodness-of-Fit Index (GFI)

The GFI compares the fit of the hypothesized model to the fit of a baseline model, often the independence model where variables are assumed to be unrelated. It quantifies the proportion of variance and covariance in the observed data that is explained by the hypothesized model. The GFI ranges from 0 to 1, with higher values indicating better fit. The formula for the GFI is:

$$GFI = 1 - \frac{tr(S - \Sigma(\hat{\theta}))^2}{tr(S^2)}$$

Where: S is the sample covariance matrix

$\Sigma(\hat{\theta})$is the model-implied covariance matrix.

tr denotes the trace of a matrix, which is the sum of the diagonal elements (Jöreskog, K. G., & Sörbom, D. (1981); Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003)).

Interpretation: A GFI value close to 1 suggests that the model fits the data well, indicating a good fit. Conversely, values closer to 0 indicate poor fit, suggesting that the model does not adequately represent the observed data. Hu and Bentler (1999) suggested that for the GFI, values above 0.90 are generally considered indicative of good fit (Byrne, 1994). However, this threshold can vary depending on the complexity of the model and the specific research context. Bentler and Bonett (1980) proposed that GFI values greater than 0.85 can be considered acceptable, though this guideline has been updated over time with stricter criteria.

1.4 Adjusted Goodness-of-Fit Index (AGFI)

The Adjusted Goodness-of-Fit Index (AGFI) is a modification of the Goodness-of-Fit Index (GFI) in structural equation modeling (SEM). It adjusts for the complexity of the model by penalizing models with more parameters, aiming to provide a more conservative measure of model fit. AGFI adjusts the GFI by considering the degrees of freedom in the model, penalizing for model complexity. Like GFI, AGFI ranges from 0 to 1, with higher values indicating better fit. The formula for the AGFI is: (Jöreskog, K. G., & Sörbom, D. (1981); Tanaka, J. S., & Huba, G. J. (1985)).

$$AGFI = 1 - \frac{\frac{(p \times (p+1)}{2} - df}{\frac{(p \times (p+1)}{2} \times (1 - GFI)}$$

Where:

p is the number of observed variables.

df is the degrees of freedom of the model.

GFI is the Goodness-of-Fit Index.

Interpretation: A high AGFI value (close to 1) indicates that the model explains a large proportion of the variance and covariance in the observed data, adjusted for its complexity. Lower AGFI values suggest poorer fit, indicating that the model may not adequately represent the data. Hu and Bentler (1999) suggested that AGFI values above 0.90 are typically considered indicative of good fit. This guideline reflects the adjustment made by AGFI for model complexity, providing a more conservative assessment compared to GFI alone.

1.5 Root Mean Square Error of Approximation (RMSEA)

The Root Mean Square Error of Approximation (RMSEA) is a widely used absolute fit index in structural equation modeling (SEM) that measures the discrepancy between the hypothesized model and the observed data per degree of freedom. RMSEA is crucial in SEM because it considers both model complexity and data fit, providing a balanced assessment of model adequacy. RMSEA assesses how well the model fits the data, considering the complexity of the model and the sample size (Savalei, V. (2018)). It quantifies the amount of error in the model's predictions per degree of freedom, considering the lack of fit relative to the number of parameters estimated. Unlike simpler indices like chi-square, RMSEA adjusts for the number of parameters in the model, making it particularly useful for evaluating complex models with many variables and relationships. The formula to compute RMSEA is (Steiger, J. H. (1990); Browne, M. W., & Cudeck, R. (1993); Lai, K., & Green, S. B. (2016)):

$$RMSEA = \sqrt{\frac{\frac{x^2}{df} - 1}{N - 1}}$$

Where:

x^2 is the chi − square statistic of the model.

df are the degrees of freedom of the model.

N is the sample size.

Interpretation: Lower RMSEA values indicate better fit, with values typically less than 0.05 considered indicative of good fit. Values up to 0.08 are often considered reasonable in some contexts (Browne & Cudeck, 1993; Kim, K. H. (2005); (Preacher et al., 2013)). However, Steiger (1990), proposed that RMSEA values below 0.05 indicate a close fit, with values up to 0.08 acceptable and values greater than 0.10 indicating a poor fit. RMSEA values greater than 0.10 generally suggest poor fit.

Significance Testing: In addition to the point estimate of RMSEA, significance testing can be conducted to assess whether the RMSEA value significantly deviates from an ideal fit (close to 0). This is typically done by examining the 90% or 95% confidence interval around the RMSEA estimate. A non-significant p-value (e.g., p > 0.05) suggests that the model fits the data adequately within the specified range.

At the bottom line, RMSEA can be sensitive to sample size, with smaller samples potentially leading to higher RMSEA values (Kenny, Kaniskan, & McCoach, 2015). Highly complex models with many parameters may yield lower RMSEA values even when the fit is not optimal. Conversely, simpler models may show higher RMSEA values due to the lack of parameter flexibility (Marsh, Hau, & Wen, 2004). Models with fewer degrees of freedom tend to produce lower RMSEA values. Therefore, it is essential to consider the model's complexity and degrees of freedom when interpreting RMSEA (Chen, Curran, Bollen, Kirby, & Paxton, 2008).

1.6 Standardized Root Mean Square Residual (SRMR)

SRMR assesses the discrepancy between the observed covariance matrix and the model-predicted covariance matrix. It provides a standardized measure of the average absolute standardized residual, representing the average discrepancy per degree of freedom in the model.

Interpretation: Lower SRMR values indicate better fit, with guidelines typically suggesting values less than 0.08 as indicative of good fit. This threshold indicates that the model adequately reproduces the observed covariance structure (Hu, L. T., & Bentler, P. M. (1999)). SRMR is important in SEM and CFA as it directly measures the model's ability to reproduce the observed data covariance structure. Unlike incremental fit indices (e.g., CFI, TLI) that compare the fit of the hypothesized model to a baseline model, SRMR focuses on the absolute fit of the model to the observed data without considering model complexity.

$$SRMR = \Sigma_{ij}\sqrt{\frac{\left(\frac{\hat{\Sigma}_{ij}-\Sigma_{ij}}{s.e(\hat{\Sigma}_{ij})}\right)}{p(p+1)/2}}$$

Where $s.e(\hat{\Sigma}_{ij})$ is the estimated standard error of $\hat{\Sigma}_{ij}$

1.7 Root Mean Square Residual (RMR)

Root Mean Square Residual (RMR) is a statistical measure used in structural equation modeling (SEM) to quantify the discrepancy between the observed covariance matrix and the model-predicted covariance matrix. RMR calculates the square root of the average squared residuals between the observed (Σ) and predicted $(\widehat{\Sigma})$ covariance matrices (Bollen, K. A. (1989); Byrne, B. M. (2016)). Mathematically, it is expressed as:

$$SRMR = \sqrt{\frac{\Sigma_{ij}\left(\hat{\Sigma}_{ij} - \Sigma_{ij}\right)^2}{p(p+1)/2}}$$

Where: $\hat{\Sigma}_{ij}$ is the predicted covariance between *i* and *j*

Σ_{ij} is the observed covariance between *i* and *j*

P is the number of observed variables

Interpretation: RMR provides a direct measure of the average discrepancy per covariance element between the model-predicted and observed covariance matrices. Lower values of RMR indicate better fit, suggesting that the model adequately reproduces the covariance structure observed in the data.

Difference

Scale: RMR provides a raw measure of the average residual, while SRMR standardizes the residuals by their standard errors.

Interpretation: Lower values of both RMR and SRMR indicate better fit, with SRMR being more commonly used due to its standardized nature.

2. INCREMENTAL FIT INDICES

Incremental fit indices in structural equation modeling (SEM) assess how much better the hypothesized model fits the data compared to a baseline model, typically the null model or a more restricted model. These indices provide additional insights beyond absolute fit indices like chi-square and RMSEA, focusing on the improvement in fit achieved by the model of interest.

2.1 Comparative Fit Index (CFI)

The Comparative Fit Index (CFI) is an incremental fit index used in structural equation modeling (SEM) to compare the fit of the hypothesized model to a baseline model, typically the null model or a more restricted model. CFI compares how much better the hypothesized model fits the data compared to a baseline model, adjusting for sample size and model complexity. CFI values range from 0 to 1, with values closer to 1 indicating better fit. The formula to compute CFI is (Bentler, P. M. (1990); Hu, L. T., & Bentler, P. M. (1999)):

$$CFI = 1 - \frac{x^2_{model} - df_{model}}{x^2_{baseline} - df_{baseline}}$$

$\chi 2_{\text{model}}$ is the chi-square statistic for the user-specified model.

df_{model} is the degrees of freedom for the user-specified model.

$\chi2_{\text{baseline}}$ is the chi-square statistic for the baseline (null) model.

df_{baseline} is the degrees of freedom for the baseline (null) model.

Interpretation: Bentler (1990) originally proposed that CFI values above 0.90 or even 0.95 are indicative of good fit. This guideline suggests that the hypothesized model provides a significantly better fit to the data compared to the baseline model. However, thresholds can vary based on factors such as model complexity and research context. Some researchers have suggested slightly different thresholds for CFI. Marsh et al. (2004) proposed that CFI values above 0.95 indicate excellent fit, reflecting a more stringent criterion for model adequacy. Hu and Bentler (1999) discussed that CFI values above 0.90 can be considered acceptable, especially in models with more complexity or smaller sample sizes.

Importance: CFI is important in SEM because it provides a robust measure of incremental fit improvement over baseline models. It helps researchers evaluate whether the hypothesized relationships between variables explain the data adequately, considering both model fit and parsimony. According to Fan, Thompson, and Wang (1999), CFI is less affected by sample size compared to other fit indices. This attribute makes it a reliable choice in various research scenarios, as its values remain relatively stable even with changes in sample size. However, there are concerns about the potential bias in CFI. Raykov (2000, 2005) points out that CFI may be biased due to its reliance on non-centrality parameters. This bias arises because the CFI evaluates the fit of a model relative to a null model, which assumes no relationship among variables. The measure of non-centrality, which reflects the discrepancy between the hypothesized model and the null model, can introduce bias into the CFI, leading to overestimation or underestimation of the model fit (Lai, M. H., & Yoon, M. (2015)).

2.2 Tucker-Lewis Index (TLI)

The Tucker-Lewis Index (TLI), also known as the Non-Normed Fit Index (NNFI), is an incremental fit index used in structural equation modeling (SEM) to compare the fit of a hypothesized model to a baseline model. TLI assesses the relative fit of the hypothesized model by comparing it with a more restricted baseline model, typically the null model. TLI values range from 0 to 1, with values closer to 1 indicating better fit. TLI is important in SEM because it provides a complementary perspective to other fit indices like CFI and RMSEA, focusing on incremental fit improvement. It helps researchers assess whether the hypothesized relationships among variables explain the observed data patterns adequately. The formula to compute TLI is (Tucker, L. R., & Lewis, C. (1973); Bentler, P. M. (1990)):

$$TLI = \frac{\frac{x^2_{baseline}}{df_{baseline}} - \frac{x^2_{model}}{df_{model}}}{\left(\frac{x^2_{baseline}}{df_{baseline}}\right) - 1}$$

$\chi2_{\text{model}}$ is the chi-square statistic for the user-specified model.

df_{model} is the degrees of freedom for the user-specified model.

$\chi2_{\text{baseline}}$ is the chi-square statistic for the baseline (null) model.

df_{baseline} is the degrees of freedom for the baseline (null) model.

Interpretation: Hu and Bentler (1999) recommended TLI values above 0.90 as indicative of good fit. This guideline suggests that the hypothesized model provides a significantly better fit to the data compared to the baseline model. TLI values closer to 1 indicate that the model is a good representation of the observed data patterns. Some researchers have suggested slightly different thresholds for TLI. Marsh et al., (2004) proposed that TLI values above 0.95 indicate excellent fit, reflecting a stricter criterion for model adequacy. Hu and Bentler (1999) discussed that TLI values above 0.90 can be considered acceptable, especially in models with moderate complexity or smaller sample sizes.

2.3 Incremental Fit Index (IFI)

The Incremental Fit Index (IFI) is an incremental fit index used in structural equation modeling (SEM) to quantify the improvement in fit achieved by the hypothesized model compared to a null or more restricted model. IFI compares how much better the hypothesized model fits the data compared to a null or more restricted model, while adjusting for model complexity. IFI values range from 0 to 1, with values closer to 1 indicating better fit. Bentler (1990) introduced IFI as a measure to enhance the assessment of incremental fit improvement, aiming to provide a more detailed evaluation of model adequacy beyond absolute fit indices such as chi-square and RMSEA. Higher IFI values indicate that the hypothesized model significantly improves the fit compared to the baseline model. The formula to compute IFI is (Bollen, 1989; Bentler & Bonnet, 1980)

$$IFI = \frac{x^2_{baseline} - x^2_{model}}{x^2_{baseline} - df^2_{model}}$$

$\chi 2_{\text{model}}$ is the chi-square statistic for the user-specified model.

df_{model} is the degrees of freedom for the user-specified model.

$\chi 2_{\text{baseline}}$ is the chi-square statistic for the baseline (null) model.

Interpretation: IFI values closer to 1 suggest that the hypothesized model explains a substantial proportion of the variance and covariance in the observed data, adjusted for its complexity. Lower IFI values indicate less improvement in fit compared to the baseline model.IFI is important in SEM because it helps researchers evaluate whether the hypothesized relationships among variables contribute significantly to explaining the observed data patterns. It complements other fit indices such as CFI and TLI, providing a comprehensive assessment of model fit.

2.4 Normed Fit Index (NFI)

The Normed Fit Index (NFI) is an incremental fit index used in structural equation modeling (SEM) to compare the fit of a target model to a baseline model, typically the null model. NFI compares the relative improvement in fit of the hypothesized model over a null model, considering model complexity and sample size. NFI values range from 0 to 1, with values closer to 1 indicating better fit. The formula to compute NFI (Bentler, P. M., & Bonett, D. G. (1980); Bentler, P. M. (1990)):

$$NFI = \frac{x^2_{baseline} - x^2_{model}}{x^2_{baseline}}$$

$\chi 2_{model}$ is the chi-square statistic for the user-specified model.

$\chi 2_{baseline}$ is the chi-square statistic for the baseline (null) model.

Interpretation: The original guideline for NFI values was proposed by Bentler and Bonett (1980), who suggested that NFI values above 0.90 indicate good fit. This criterion implies that the hypothesized model provides a substantial improvement in fit compared to the null model. NFI is important in SEM because it focuses on the improvement in fit achieved by the hypothesized model, offering insights into whether the model adequately explains the observed data patterns. It complements other fit indices like CFI, TLI, and RMSEA, providing a comprehensive assessment of model adequacy. According to Ullman (2001), small sample sizes frequently lead to an underestimation of model fit. Moreover, while the TLI can overestimate fit when the number of parameters in the model increases, the NNFI (Non-Normed Fit Index) effectively mitigates this problem, providing a more accurate assessment.

2.5 The Normed Fit Index (NFI)

The Normed Fit Index (NFI) has been a useful incremental fit index in structural equation modeling (SEM), but its usage and interpretation have evolved over time. Initially, Bentler and Bonett suggested that NFI values above 0.90 indicate good fit. This guideline was influential in establishing a benchmark for evaluating model adequacy. In their work on fit indices in SEM, Hu and Bentler suggested that NFI values above 0.95 may be considered indicative of excellent fit, reflecting a stricter criterion than the original recommendation by Bentler and Bonett. Marsh et al., (2004) discussed that NFI values above 0.90 are generally acceptable, aligning with the earlier guideline by Bentler and Bonett. They emphasized that while higher values are desirable, the interpretation should consider the complexity of the model and the sample size. The formula for NFI is (Bentler, P. M., & Bonett, D. G. (1980); Bentler, P. M. (1990))

$$NFI = \frac{x^2_{baseline} - x^2_{model}}{x^2_{baseline} - x^2_{model}}$$

$\chi 2_{model}$ is the chi-square statistic for the user-specified model.

$\chi 2_{baseline}$ is the chi-square statistic for the baseline (independence) model.

2.6 RELATIVE FIT INDEX (RFI)

The Relative Fit Index (RFI) is an incremental fit index used in structural equation modeling (SEM) that, like NFI, compares the fit of a target model to a baseline model, adjusting for degrees of freedom. RFI measures the relative improvement in fit of the hypothesized model compared to a null or baseline model, considering the degrees of freedom used. It provides a measure of how well the hypothesized model fits the data relative to a more restricted model. Formula to compute RFI is (Bentler, P. M., & Bonett, D. G. (1980); Bentler, P. M. (1990))

$$RFI = \frac{x^2_{baseline} - x^2_{model}}{x^2_{baseline}}$$

$\chi 2_{model}$ is the chi-square statistic for the user-specified model.

$\chi 2_{\text{baseline}}$ is the chi-square statistic for the baseline (null) model.

Interpretation: As an incremental fit index, RFI values range from 0 to 1, with values closer to 1 indicating better fit. The specific threshold values for RFI may vary depending on the study and the complexity of the model. Guidelines for RFI are less standardized compared to other fit indices like CFI or RMSEA. Unfortunately, specific threshold values for RFI are not as extensively documented in the literature compared to other fit indices like NFI or CFI. Researchers often interpret RFI values in conjunction with other fit indices to assess model adequacy. Bentler and Bonett (1980) initially discussed RFI in the context of structural equation models, emphasizing its role in comparing fit between the hypothesized model and the null model, but did not provide specific threshold values for RFI. Due to the lack of standardized threshold values, researchers may use their judgment based on the specific context of their study, including model complexity, sample size, and theoretical considerations.

3. PARSIMONY FIT INDICES

Parsimony Fit Indices refer to a category of fit indices used in structural equation modeling (SEM) that specifically assess the balance between model fit and model complexity. These indices aim to reward models that explain the data well while using fewer parameters, thereby promoting parsimony. The rationale behind parsimony fit indices is to avoid overfitting, where a model may fit the sample data well but generalize poorly to new data or populations due to excessive complexity. The following are the reasons for using Parsimony Fit Indices:

Promoting Generalizability: One of the primary goals in SEM is to develop models that not only fit the observed data well but also generalize to new data or populations. Parsimony fit indices, such as the Parsimonious Normed Fit Index (PNFI) or Parsimonious Comparative Fit Index (PCFI), encourage the selection of models that achieve good fit while using fewer parameters. This approach reduces the risk of overfitting, where a model may excessively tailor itself to noise or idiosyncrasies in the sample data, leading to poor performance with new data.

Enhancing Model Interpretability: Simpler models are often easier to interpret and communicate, making the relationships between variables clearer and more actionable. By penalizing overly complex models, parsimony fit indices help researchers prioritize theoretical clarity and practical relevance in their SEM analyses. This ensures that the model's structure and parameters are more meaningful and reflective of underlying theoretical constructs.

Optimizing Model Selection: SEM typically involves comparing multiple competing models to determine which one best represents the relationships in the data. Parsimony fit indices provide a quantitative measure to guide this selection process, aiding researchers in choosing models that strike an appropriate balance between explanatory power and simplicity. This selection process is crucial for advancing scientific understanding and theory building in various fields that utilize SEM.

Statistical Rigor: Incorporating parsimony fit indices adds a layer of statistical rigor to SEM analyses by formalizing the trade-off between model fit and complexity. This helps in avoiding the pitfalls of overly complex models that may lead to inflated fit indices but lack practical utility or generalizability.

3.1 The Parsimony Goodness-of-Fit Index (PGFI)

The Parsimony Goodness-of-Fit Index (PGFI) is a fit index used in structural equation modeling (SEM) to adjust the Goodness-of-Fit Index (GFI) for the complexity of the model. PGFI adjusts the traditional GFI by penalizing the model for its complexity, typically by incorporating a correction factor that accounts for the number of estimated parameters. This adjustment aims to promote parsimony, rewarding models that achieve good fit while using fewer parameters. Formula to compute PGFI is (Jöreskog, K. G., & Sörbom, D. (1981))

$$PGFI = \frac{x^2_{model}}{df_{model}}$$

$\chi2_{\text{model}}$ is the chi-square statistic for the user-specified model.

df_{model} is the degrees of freedom for the user-specified model.

Interpretation: PGFI values range from 0 to 1, like GFI and other fit indices, with values closer to 1 indicating better fit. The interpretation of PGFI involves comparing its value against established threshold criteria to assess model adequacy. Specific threshold values for PGFI may vary in the literature. However, guidelines generally suggest that higher PGFI values indicate better model fit while considering model complexity. As with other fit indices, thresholds are context-dependent and can vary based on the specific research area, sample size, and complexity of the model.

3.2 The Parsimony Normed Fit Index (PNFI)

The Parsimony Normed Fit Index (PNFI) is a fit index used in structural equation modeling (SEM) to adjust the Normed Fit Index (NFI) for the complexity of the model. PNFI adjusts the traditional NFI by penalizing the model for its complexity, typically by incorporating a correction factor that accounts for the number of estimated parameters. This adjustment aims to promote parsimony, rewarding models that achieve good fit while using fewer parameters. Essentially, PNFI provides a more balanced assessment of model fit by considering both the goodness of fit and the model's simplicity. Formula to compute PNFI is (Bentler, P. M., & Bonett, D. G. (1980); Bentler, P. M. (1990))

$$PNFI = \frac{NFI \text{ x} df_{baseline}}{df_{baseline}}$$

NFI is the Normed Fit Index.

df_{baseline} is the degrees of freedom for the baseline (null) model.

df_{model} is the degrees of freedom for the user-specified model.

Interpretation: Like NFI, PNFI values range from 0 to 1, with values closer to 1 indicating better fit. Higher PNFI values suggest that the model fits the data well while maintaining simplicity, thereby enhancing its generalizability and interpretability.

3.3 The Akaike Information Criterion (AIC)

The Akaike Information Criterion (AIC), along with the Bayesian Information Criterion (BIC), Browne-Cudeck Criterion (BCC), and Consistent Akaike Information Criterion (CAIC), is recognized as a measure of model fit based on information theory. These indices are particularly relevant when maximum likelihood estimation is utilized (Burnham & Anderson, 1998). These indices serve the primary purpose of comparing different models to determine which one offers the best fit. The models yielding the lowest values for these criteria are considered the most optimal. It is important to note that the absolute value of the AIC, BIC, BCC, or CAIC is not inherently meaningful; instead, the focus should be on the relative values between models. Specifically, the model with the lower AIC value is preferred. Although values closer to zero are generally considered ideal, it is the comparison between models that provides meaningful insights (Burnham & Anderson, 1998; Akaike, 1974; Schwarz, 1978; Sawa, 1978).

The Akaike Information Criterion (AIC) is a statistical measure used to compare the relative quality of different models based on their fit to the data. AIC balances the goodness of fit of a model with its complexity, penalizing models that are more complex. The principle behind AIC is to select a model that adequately explains the data with the fewest parameters, thereby promoting parsimony and generalizability (Akaike, H. (1974)). AIC is calculated using the formula:

$$AIC = -2.\ln(\mathcal{L}) + 2.k$$

$\mathcal{L}$ is the maximized value of the likelihood function of the model.

k is the number of estimated parameters in the model.

Interpretation: Lower Values Better Fit: AIC values are relative, meaning lower values indicate a better fit of the model to the data. Models with lower AIC scores are considered to be more parsimonious and better fitting to the observed data. AIC is widely used in model selection across various fields, including statistics, econometrics, and machine learning. It allows researchers to compare multiple models and determine which one strikes the best balance between goodness of fit and model complexity.

CONFIRMATORY FACTOR ANALYSIS

In addition, in Confirmatory Factor Analysis (CFA), assessing reliability (Fornell, C., & Larcker, D. F. (1981); Bagozzi, R. P., & Yi, Y. (1988)), convergent validity (Fornell, C., & Larcker, D. F. (1981); Hair et al., (2010)), and discriminant validity (Fornell, C., & Larcker, D. F. (1981); Kline, R. B. (2015)) is crucial to ensure the measurement model's adequacy (Jackson et al., 2009)).

Reliability

Reliability refers to the consistency of a measure. A reliable measurement produces the same results under consistent conditions. In the context of psychological and educational testing, reliability ensures that the instrument measures a construct consistently across different occasions, items, and raters. High reliability indicates that the measurement tool yields stable and consistent results over repeated applications (Nunnally, J. C., & Bernstein, I. H. (1994)).

Composite Reliability (CR)

Composite Reliability (CR), also known as construct reliability, is a measure used to assess the internal consistency of a set of latent construct indicators in structural equation modeling (SEM). Unlike Cronbach's alpha, which assumes equal weight for all items, CR takes into account the varying loadings of the items, providing a more accurate estimation of the reliability of the construct. Composite Reliability is crucial for evaluating the reliability of a measurement model because it provides insight into the degree to which the indicators reflect the latent construct. High CR values indicate that the indicators consistently represent the underlying construct, enhancing the validity of the measurement model (Fornell, C., & Larcker, D. F. (1981); Bagozzi, R. P., & Yi, Y. (1988); Beauducel, A., & Wittmann, W. W. (2005)).

$$CR = \frac{(\Sigma\lambda_i)^2}{(\Sigma\lambda_i)^2 + \Sigma\theta_i}$$

Where λ_i are the factor loadings and θ_i are the error variances.

Threshold Values: The threshold values for Composite Reliability are generally similar to those for Cronbach's alpha:

CR ≥ 0.70: Indicates adequate reliability

CR ≥ 0.80: Indicates good reliability

CR ≥ 0.90: Indicates excellent reliability

Values below 0.70 suggest that the construct indicators may not be consistently measuring the same underlying concept, thus questioning the reliability of the construct (Nunnally, J. C., & Bernstein, I. H. (1994)).

Cronbach's Alpha

Cronbach's alpha (α) is a widely used measure of internal consistency, indicating how closely related a set of items are as a group. It is essential for determining the reliability of a multi-item scale. Cronbach's alpha values range from 0 to 1, with higher values indicating greater internal consistency (Cronbach, L. J. (1951)). A high Cronbach's alpha suggests that the items measure the same underlying construct, making the scale reliable. Cronbach's alpha is crucial because it helps validate the consistency of the items in a test or survey, ensuring that they collectively measure the intended construct. This is particularly important in research and assessment contexts where the accuracy and dependability of measurement tools directly impact the validity of conclusions drawn (Tavakol, M., & Dennick, R. (2011)).

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

Where: N is the number of items.

σ_i^2 is the variance of the ith item.

σ_t^2 is the variance of the total score formed by summing all the items.

Threshold Values for Cronbach's Alpha

The acceptable level of Cronbach's alpha depends on the context and purpose of the measurement. General guidelines for interpreting Cronbach's alpha values are as follows:

$\alpha \geq 0.9$: Excellent
$0.8 \leq \alpha < 0.9$: Good
$0.7 \leq \alpha < 0.8$: Acceptable
$0.6 \leq \alpha < 0.7$: Questionable
$0.5 \leq \alpha < 0.6$: Poor
$\alpha < 0.5$: Unacceptable
Note: These thresholds serve as benchmarks for evaluating the reliability of a scale. However, the interpretation of Cronbach's alpha should also consider the specific research context, including the nature of the construct being measured and the intended use of the instrument (George, D., & Mallery, P. (2003)).

Average Variance Extracted (AVE)
Average Variance Extracted (AVE) is a measure used in structural equation modeling (SEM) to assess the amount of variance that is captured by a latent construct in relation to the amount of variance due to measurement error. It is an indicator of convergent validity, which ensures that the items of a construct are truly representative of the intended latent variable (Fornell, C., & Larcker, D. F. (1981)). AVE is crucial because it helps researchers determine whether the indicators of a construct are adequately capturing the underlying factor they are intended to measure. High AVE values indicate that the construct explains a substantial portion of the variance in its indicators, thus demonstrating good convergent validity. AVE also aids in assessing discriminant validity, which ensures that constructs are distinct and not excessively correlated with each other. The formula for calculating AVE is:

$$AVE = \frac{\Sigma_{i=1}^{N} \lambda_i^2}{N}$$

Where: λ_i represents the standardized factor loading of each indicator on the latent construct.
N is the number of indicators.

Threshold Values for AVE
The commonly accepted threshold for AVE is 0.50. This means that at least 50% of the variance in the indicators should be accounted for by the latent construct. Values below this threshold suggest that the construct may not be adequately measured by its indicators (Hair et al., 2010).

Discriminant Validity

1. Fornell-Larcker Criterion

The Fornell-Larcker criterion is a method used to assess discriminant validity in structural equation modeling (SEM). It evaluates whether the constructs in a model are distinct from each other by comparing the square root of the Average Variance Extracted (AVE) for each construct with the correlations between constructs (Fornell, C., & Larcker, D. F. (1981)). The criterion is important because it helps researchers ensure that the constructs they are studying are distinct and not highly correlated with each other. This is crucial for avoiding multicollinearity issues and for accurately interpreting the relationships between constructs in a model. If the Fornell-Larcker criterion is met, it provides confidence that the measurement model is valid and that the constructs are adequately differentiated (Hair et al., 2010).

How It Works?

The Fornell-Larcker criterion is applied as follows:

Step1: Calculate AVE: Compute the Average Variance Extracted (AVE) for each construct. AVE measures the amount of variance explained by the construct's indicators relative to measurement error.

Step 2: Calculate Correlations: Calculate the correlations between all pairs of constructs in the model.

Step3: Compare AVE and Correlations: Compare the square root of the AVE (for each construct) ($\sqrt{AVE_i}$) with the correlations between that construct and all other constructs. According to the criterion, the square root of the AVE for a construct should be greater than the correlations between that construct and any other construct in the model to confirm discriminant validity.

2. Heterotrait-Monotrait Ratio (HTMT)

The Heterotrait-Monotrait Ratio (HTMT) is a measure used to assess discriminant validity in structural equation modeling (SEM). It compares the correlations between constructs (heterotrait correlations) to the average correlations within constructs (monotrait correlations). It is defined as the average of the heterotrait correlations divided by the average of the monotrait correlations. HTMT is important because it provides a direct and quantitative assessment of discriminant validity by comparing how much more constructs correlate with themselves (monotrait correlations) compared to how much they correlate with other constructs (heterotrait correlations). This helps researchers ensure that the measures used in their models are distinct and do not overlap substantially with each other.

How It Works?

Step1: Calculate Correlations: Compute the correlations between all pairs of constructs in the model.

Step2: Calculate HTMT: For each pair of constructs i and j

$$HTMT_{ij} = \frac{Averageof\ heterotraitcorrelations_{ij}}{Averageofmonotraitcorrelations_{ij}}$$

Step3: Interpretation: A HTMT value close to or less than 1 indicates that the constructs are sufficiently distinct (good discriminant validity). Values significantly greater than 1 suggest potential issues with discriminant validity, indicating that the constructs may not be adequately differentiated (Henseler, J., Ringle, C. M., & Sarstedt, M. (2015)).

The following is the comprehensive table summarizing the key criteria used in Confirmatory Factor Analysis (CFA), including Cronbach's alpha, factor loadings, Average Variance Extracted (AVE), Composite Reliability (CR), and discriminant validity. Each criterion is defined along with its threshold value and relevant citations for further reading:

TABLE 1
THE KEY CRITERIA USED IN CONFIRMATORY FACTOR ANALYSIS (CFA)

Criterion	Definition	Threshold Value	Citations
Cronbach's Alpha	Measure of internal consistency reliability of a scale.	≥ 0.70	Cronbach (1951), Nunnally & Bernstein (1994)
Factor Loadings	Strength and direction of the relationship between each item and its construct.	≥ 0.50	Hair et al. (2010), Byrne (2016)
Average Variance Extracted (AVE)	Amount of variance explained by the construct's indicators relative to measurement error.	≥ 0.50	Fornell & Larcker (1981), Hair et al. (2010)
Composite Reliability (CR)	Measure of internal consistency reliability considering factor loadings.	≥ 0.70	Fornell & Larcker (1981), Hair et al. (2010)
Discriminant Validity	Ensures that constructs are distinct from each other.	HTMT ≤ 1.00	Fornell & Larcker (1981), Henseler et al., (2015)

Note: In confirmatory factor analysis (CFA), loadings of 0.70 or higher are often desired (Kline, R. B. (2016)) to ensure that the latent constructs are well-represented by the observed variables. This is particularly important in developing and validating measurement instruments.

TABLE 2
THE KEY INDICES FOR REPORTING PURPOSES

Fit Index	**Definition**	**Threshold Values**	**Citations**
χ^2	Measures overall fit; sensitive to sample size.	Non-significant (p > 0.05)	Bollen, K. A. (1989)
χ^2/df	Chi-square divided by degrees of freedom.	< 2: Good fit 2-3: Acceptable fit > 3: Poor fit	Kline, R. B. (2015)
CFI	Compares fit to a null model; adjusts for sample size.	> 0.90: Acceptable fit > 0.95: Good fit	Bentler, P. M. (1990)
TLI	Compares fit to a null model; penalizes complexity.	> 0.90: Acceptable fit > 0.95: Good fit	Hu, L. T., & Bentler, P. M. (1999).
RMSEA	Measures fit per degree of freedom; penalizes complexity.	< 0.05: Good fit 0.05-0.08: Acceptable fit > 0.10: Poor fit	Browne, M. W., & Cudeck, R. (1993).

SRMR	Measures average standardized residuals.	< 0.08: Good fit	Hu, L. T., & Bentler, P. M. (1999).
RMR	Measures average residuals.	Lower values indicate better fit (specific thresholds not widely cited)	Byrne, B. M. (2016).
GFI	Compares model to a null model; adjusts for sample size.	> 0.90: Good fit	Jöreskog, K. G., & Sörbom, D. (1989).
AGFI	Adjusts GFI for model complexity.	> 0.90: Good fit	Jöreskog, K. G., & Sörbom, D. (1989).
PGFI	Adjusts GFI for model parsimony.	No specific threshold (used comparatively)	Mulaik, S. A., et al., (1989).
NFI	Compares fit to a null model.	> 0.90: Good fit	Bentler, P. M., & Bonett, D. G. (1980).
PNFI	Adjusts NFI for model parsimony.	No specific threshold (used comparatively)	Mulaik, S. A., et al., (1989).
IFI	Adjusts for sample size and model complexity.	> 0.90: Good fit	Bollen, K. A. (1989).
RFI	Adjusts NFI for degrees of freedom.	> 0.90: Good fit	Bollen, K. A. (1986).
AIC	Compares relative quality of models; penalizes complexity.	Lower values indicate better fit (no absolute threshold)	Akaike, H. (1974).
BIC	Similar to AIC but with a larger penalty for complexity.	Lower values indicate better fit (no absolute threshold)	Schwarz, G. (1978).

III. DISCUSSION AND CONCLUSION

Proper reporting of Structural Equation Modeling (SEM) and Confirmatory Factor Analysis (CFA) findings is crucial for transparency, reproducibility, and the validity of research conclusions. This section discusses the importance of reporting fit indices and criteria, outlines key aspects to consider, and emphasizes the significance of clear and thorough reporting practices. Reporting fit indices such as Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), and Root Mean Square Error of Approximation (RMSEA) provides readers with an objective assessment of how well the hypothesized models fit the data. This transparency allows others to evaluate the robustness of the findings and the validity of the underlying theoretical constructs. Criteria like Average Variance Extracted (AVE), Composite Reliability (CR), and discriminant validity measures (e.g., Fornell-Larcker criterion, Heterotrait-Monotrait Ratio) are essential for assessing the quality of measurement models. Reporting these measures ensures that the constructs under investigation are adequately defined, reliable, and distinct from each other. Proper citation of methodological references and established thresholds (e.g., guidelines for AVE, CR, discriminant validity) enhances the credibility of the research. It demonstrates adherence to best practices and allows for comparisons with other studies in the field. Clearly state the values of fit indices (CFI, TLI, RMSEA) obtained from the analysis. Interpret these values in relation to accepted thresholds (e.g., Hu & Bentler, 1999; Browne & Cudeck, 1993) to assess model fit. Provide AVE and CR values for each construct to demonstrate convergent validity and reliability. Discuss these values in relation to established benchmarks (e.g., Fornell & Larcker, 1981; Hair et al., 2010) to evaluate the adequacy of the measurement models. Describe how discriminant validity was assessed (e.g., Fornell-Larcker criterion, HTMT ratio) and report the findings. Highlight measures taken to ensure that constructs are distinct and not highly correlated with each other. Include citations to methodological references and seminal works that justify the chosen criteria and interpretations. This supports the methodological rigor of the study and provides a basis for readers to further explore the theoretical framework.

Therefore, the meticulous reporting of fit indices and criteria in SEM and CFA studies serves as a cornerstone for ensuring methodological rigor, credibility, and the advancement of theoretical knowledge. By adhering to best practices in reporting, researchers not only bolster the validity of their findings but also foster a conducive environment for scholarly discourse and the cumulative growth of knowledge in the field of structural equation modeling and confirmatory factor analysis. By emphasizing these points, researchers can effectively communicate the significance of their findings and contribute meaningfully to the ongoing dialogue within their academic communities.

Disclaimer (Artificial intelligence)

We hereby declare that NO generative AI technologies such as Large Language Models (ChatGPT, COPILOT, etc.) and text-to-image generators have been used during writing or editing of manuscripts except for conceptual clarity.

REFERENCES

1. Akaike, H. (1974). A new look at the statistical model identification. IEEE Transactions on Automatic Control, 19(6), 716-723.
2. Bagozzi, R. P., & Yi, Y. (1988). On the evaluation of structural equation models. Journal of the Academy of Marketing Science, 16(1), 74-94. doi:10.1007/BF02723327
3. Baron, R. M., & Kenny, D. A. (1986). The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations. Journal of Personality and Social Psychology, 51(6), 1173-1182.
4. Barrett, P. (2007). Structural equation modeling: Adjudging model fit. Personality and
5. Beauducel, A., & Wittmann, W. W. (2005). Simulation study on fit indexes in CFA based on data with slightly distorted simple structure. Structural equation modeling, 12(1), 41-75.
6. Bentler, P. M. (1990). Comparative fit indexes in structural models. Psychological Bulletin, 107(2), 238-246.
7. Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. Psychological Bulletin, 88(3), 588-606.
8. Bollen, K. A. (1989). Structural equations with latent variables. Wiley series in probability and mathematical statistics. John Wiley & Sons, Inc.
9. Boomsma, A. (2000). Reporting analyses of covariance structures. Structural Equation Modeling, 7(3), 461-483.
10. Brown, T. A. (2006). Confirmatory factor analysis for applied research. New York: Guilford.

11. Brown, T. A. (2015). Confirmatory factor analysis for applied research. Guilford publications.
12. Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. Sociological Methods & Research, 21(2), 230-258.
13. Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), Testing Structural Equation Models (136-162). Sage.
14. Burnham, K. P., & Anderson, D. R. (1998). Model Selection and Inference: A Practical Information-Theoretic Approach. Springer.
15. Byrne, B. M. (2010). Structural Equation Modeling with AMOS: Basic Concepts, Applications, and Programming. Routledge.
16. Byrne, B. M. (2016). Structural equation modeling with AMOS: Basic concepts, applications, and programming (3rd ed.). Routledge.
17. Chen, F. F., Curran, P. J., Bollen, K. A., Kirby, J., & Paxton, P. (2008). An empirical evaluation of the use of fixed cutoff points in RMSEA test statistic in structural equation models. Sociological Methods & Research, 36(4), 462-494.
18. Chin, W. W. (1998). Commentary: Issues and opinion on structural equation modeling. MIS quarterly, vii-xvi.
19. Clark, D. A., & Bowles, R. P. (2018). Model fit and item factor analysis: Overfactoring, underfactoring, and a program to guide interpretation. Multivariate behavioral research, 53(4), 544-558.
20. Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. Psychometrika, 16(3), 297-334. doi:10.1007/BF02310555
21. Curran, P. J., Obeidat, K., & Losardo, D. (2010). Twelve Frequently Asked Questions About Growth Curve Modeling. Journal of Cognition and Development, 11(2), 121-136.
22. Fan, X., & Sivo, S. A. (2007). Sensitivity of fit indices to model misspecification and model types. Multivariate Behavioral Research, 42(3), 509-529.
23. Fan, X., B. Thompson, and L. Wang (1999). Effects of sample size, estimation method, and model specification on structural equation modeling fit indexes. Structural Equation Modeling, 6, 56-83.
24. Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. Journal of Marketing Research, 18(1), 39-50. doi:10.2307/3151312
25. Gallagher, M. W., & Brown, T. A. (2013). Introduction to confirmatory factor analysis and structural equation modeling. In Handbook of quantitative methods for educational research (pp. 287-314). Brill.
26. George, D., & Mallery, P. (2003). SPSS for Windows Step by Step: A Simple Guide and Reference (4th ed.). Allyn & Bacon.
27. Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). Multivariate Data Analysis (7th ed.). Pearson.

28. Hoyle, R. H., & Panter, A. T. (1995). Writing about structural equation models. In R. H. Hoyle (Ed.), Structural equation modeling: Concepts, issues, and applications (pp. 158-176). Thousand Oaks, CA: Sage.
29. 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.
30. Jaccard, J., & Wan, C. K. (1996). LISREL approaches to interaction effects in multiple regression (No. 114). sage.
31. Jackson, D. L., Gillaspy Jr, J. A., & Purc-Stephenson, R. (2009). Reporting practices in confirmatory factor analysis: an overview and some recommendations. Psychological methods, 14(1), 6.
32. Jöreskog, K. G., & Sörbom, D. (1981). LISREL V: Analysis of Linear Structural Relationships by Maximum Likelihood, Instrumental Variables, and Least Squares Methods. Scientific Software.
33. Kenny, D. A., Kaniskan, B., & McCoach, D. B. (2015). The performance of RMSEA in models with small degrees of freedom. Sociological Methods & Research, 44(3), 486-507.
34. Kim, K. H. (2005). The relation among fit indexes, power, and sample size in structural equation modeling. Structural equation modeling, 12(3), 368-390.
35. Kline, R. B. (2005). Principles and practice of structural equation modeling (2nd ed.). Guilford Press.
36. Kline, R. B. (2015). Principles and practice of structural equation modeling (4th ed.). Guilford Press.
37. Kyriazos, T. A. (2018). Applied psychometrics: sample size and sample power considerations in factor analysis (EFA, CFA) and SEM in general. Psychology, 9(08), 2207.
38. Lai, K., & Green, S. B. (2016). The problem with having two watches: Assessment of fit when RMSEA and CFI disagree. Multivariate behavioral research, 51(2-3), 220-239.
39. Lai, M. H., & Yoon, M. (2015). A modified comparative fit index for factorial invariance studies. Structural Equation Modeling: A Multidisciplinary Journal, 22(2), 236-248.
40. Little, T. D. (Ed.). (2013). The Oxford handbook of quantitative methods in psychology, Vol. 2: Statistical analysis. Oxford University Press.
41. MacCallum, R. C., & Austin, J. T. (2000). Applications of structural equation modeling in psychological research. Annual Review of Psychology, 51, 201-226.
42. Marsh, H. W., Balla, J. R., & Hau, K. T. (1996). An evaluation of incremental fit indexes: A clarification of mathematical and empirical properties. In G. A. Marcoulides & R. E. Schumacker (Eds.), Advanced structural equation modeling techniques (315-353). Mahwah, NJ: Lawrence Erlbaum.
43. 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.
44. Marsh, H. W., Hau, K. T., & Wen, Z. (2004). In Search of Golden Rules: Comment on Hypothesis-Testing Approaches to Setting Cutoff Values for Fit Indexes and Dangers in Overgeneralizing Hu and Bentler's (1999) Findings. Structural Equation Modeling, 11(3), 320-341.
45. McDonald, R. P., & Ho, M. H. R. (2002). Principles and practice in reporting structural equation analyses. Psychological Methods, 7(1), 64-82.
46. Medsker, G. J., Williams, L. J., & Holahan, P. J. (1994). A review of current practices for evaluating causal models in organizational behavior and human resources management research. Journal of management, 20(2), 439-464.
47. Mulaik, S. A., James, L. R., Van Alstine, J., Bennett, N., Lind, S., & Stilwell, C. D. (1989). Evaluation of Goodness-of-Fit Indices for Structural Equation Models. Psychological Bulletin, 105(3), 430-445.
48. Muthén, L. K., & Muthén, B. O. (2010). Statistical analysis with latent variables (Vol. 1). John Wiley & Sons.
49. Nunnally, J. C., & Bernstein, I. H. (1994). Psychometric Theory (3rd ed.). McGraw-Hill.
50. Perry, J. L., Nicholls, A. R., Clough, P. J., & Crust, L. (2015). Assessing model fit: Caveats and recommendations for confirmatory factor analysis and exploratory structural equation modeling. Measurement in physical education and exercise science, 19(1), 12-21.
51. Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and Resampling Strategies for Assessing and Comparing Indirect Effects in Multiple Mediator Models. Behavior Research Methods, 40(3), 879-891.
52. Preacher, K. J., Zhang, G., Kim, C., & Mels, G. (2013). Choosing the optimal number of factors in exploratory factor analysis: A model selection perspective. Multivariate Behavioral Research, 48(1), 28–56
53. Raykov, T. (2000). On the large-sample bias, variance, and mean squared error of the conventional noncentrality parameter estimator of covariance structure models. Structural Equation Modeling, 7, 431-441.
54. Raykov, T. (2005). Bias-corrected estimation of noncentrality parameters of covariance structure models. Structural Equation Modeling, 12, 120-129.
55. research. *Journal of Management, 20,* 439-464
56. Savalei, V. (2018). On the computation of the RMSEA and CFI from the mean-and-variance corrected test statistic with nonnormal data in SEM. Multivariate behavioral research, 53(3), 419-429.
57. Sawa, T. (1978). Information criteria for discriminating among alternative regression models. Econometrica: Journal of the Econometric Society, 46(6), 1273-1291.
58. Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. Methods of Psychological Research Online, 8(2), 23-74.

59. Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. The Journal of Educational Research, 99(6), 323-338.
60. Schumacker, R. E., & Lomax, R. G. (2004). A beginner's guide to structural equation modeling, Second edition. Mahwah, NJ: Lawrence Erlbaum Associates.
61. Schumacker, R. E., & Lomax, R. G. (2016). A Beginner's Guide to Structural Equation Modeling. Routledge.
62. Schwarz, G. (1978). Estimating the dimension of a model. The Annals of Statistics, 6(2), 461-464.
63. Shi, D., & Maydeu-Olivares, A. (2020). The effect of estimation methods on SEM fit indices. Educational and psychological measurement, 80(3), 421-445.
64. Shi, D., DiStefano, C., Maydeu-Olivares, A., & Lee, T. (2022). Evaluating SEM model fit with small degrees of freedom. Multivariate behavioral research, 57(2-3), 179-207.
65. Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. Multivariate Behavioral Research, 25(2), 173-180.
66. Steiger, J. H. (2001). Understanding the limitations of global fit assessment in structural equation modeling. Personality and Individual Differences, 32(5), 901-912.
67. Tanaka, J. S., & Huba, G. J. (1985). A fit index for covariance structure models under arbitrary GLS estimation. British Journal of Mathematical and Statistical Psychology, 38(2), 197-201.
68. Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. International Journal of Medical Education, 2, 53-55.
69. Thompson, B. (2004). Exploratory and confirmatory factor analysis: Understanding concepts and applications. Washington, DC, 10694(000), 3.
70. Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. Psychometrika, 38(1), 1-10.
71. Ullman, J. B. (2001). Structural equation modeling. In B. G. Tabachnick & L. S. Fidell (2001). Using Multivariate Statistics (4th ed& pp 653- 771). Needham Heights, MA: Allyn & Bacon.
72. Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample size requirements for structural equation models: An evaluation of power, bias, and solution propriety. Educational and Psychological Measurement, 73(6), 913-934.
73. Wright, S. (1934). The Method of Path Coefficients. Annals of Mathematical Statistics, 5(3), 161-215.
74. Ximénez, C., Maydeu-Olivares, A., Shi, D., & Revuelta, J. (2022). Assessing cutoff values of SEM fit indices: Advantages of the unbiased SRMR index and its cutoff criterion based on communality. Structural Equation Modeling: A Multidisciplinary Journal, 29(3), 368-380.