Smart Choices for Measurement Models

Executable Manuscript Tutorial for your Confirmatory Factor Analysis in R Environment

Pablo Rogers

March 7, 2024

Abstract

This article aims to accomplish three objectives: first, to compile guidelines for the application of Confirmatory Factor Analysis (CFA), a widely utilized technique in applied social sciences; second, to demonstrate how these guidelines can be practically implemented through a real-world example; and third, to structure this narrative using tools that promote reproducibility, replicability, and transparency of results. To this end, we propose a solution in the form of a tutorial article wherein the key decisions made in conducting a CFA are validated through recent literature and presented within a dynamic document framework. This framework enables readers to access the article's source code, utilized data, analytical execution codes, and various reading media. We anticipate that by employing this pedagogical approach, developed entirely within an open environment (utilizing Git/Github/RStudio/Quarto/packages R + lavaan/Docker), researchers proficient in specific statistical techniques relevant to their domains will adopt and disseminate this proposal, thereby benefiting their colleagues.

Keywords: Confirmatory Factor Analysis, Structural Equation Modeling, Internal Structure Validity, *lavaan*.

1 Introduction

Confirmatory Factor Analysis (CFA) is a key method for assessing the validity of a measurement instrument through its internal structure (Bandalos 2018; Hughes 2018; Sireci and Sukin 2013). Validity is arguably the most crucial characteristic of a measurement model (Furr 2021), as it addresses the essential question of what measuring instruments truly assess (Bandalos 2018). This concern is closely linked with the classical definition of validity: the degree to which a test measures what it claims to measure (Bandalos 2018; Furr 2021; Sireci and Sukin 2013; Urbina 2014), aligning with the tripartite model still embraced by numerous scholars (Widodo 2018).

The tripartite model of validity frames the concept using three categories of evidence: content, criterion, and construct (Bandalos 2018). Content validity pertains to the adequacy and representativeness of test items relative to the domain or objective under investigation (Cohen, Schneider, and Tobin 2022). Criterion validity is the correlation between test outcomes and a significant external criterion, such as performance on

another measure or future occurrences (Cohen, Schneider, and Tobin 2022). Construct validity evaluates the test's capacity to measure the theoretical construct it is intended to assess, taking into account related hypotheses and empirical data (Cohen, Schneider, and Tobin 2022).

Introduced in the American Psychological Association (APA) "Standards for Educational and Psychological Testing" in 1966, the tripartite concept of validity has been a cornerstone in the social sciences for decades (Bandalos 2018). However, its fragmented and confusing nature has led to widespread criticism, prompting a shift towards a more holistic view of validity (Sireci and Sukin 2013). This evolution was signified by the publication of the 1999 standards (AERA, APA, and NCME 1999), and further by the 2014 standards (AERA, APA, and NCME 2014), which redefined test validity in terms of the interpretations and uses of test scores (Furr 2021). Under this new paradigm, validation requires diverse theoretical and empirical evidence, recognizing validity as a unified concept – construct validity – encompassing various evidence sources for evaluating potential interpretations of test scores for specific purposes (Furr 2021; Urbina 2014).

Thus, key authorities in psychological assessment now define validity as the degree to which evidence and theory support the interpretations of test scores for their intended purposes (AERA, APA, and NCME 2014). Validity involves a comprehensive evaluation of how well empirical evidence and theoretical rationales uphold the conclusions and actions derived from test scores or other assessment types (Bandalos 2018; Furr 2021; Urbina 2014).

According to APA guidelines (AERA, APA, and NCME 2014), five types of validity evidence are critical: content, response process, association with external variables, consequences of test use, and internal structure. Content validity examines the extent to which test content accurately represents the domain of interest exclusively (Furr 2021). The response process refers to the link between the construct and the specifics of the examinees' responses (Sireci and Sukin 2013). Validity based on external variables concerns the test's correlation with other measures or constructs expected to be related or unrelated to the evaluated construct (Furr 2021). The implications of test use focus on the positive or negative effects on the individuals or groups assessed (Bandalos 2018).

Evidence based on internal structure assesses how well the interactions among test items and their components align with the theoretical framework used to explain the outcomes of the measurement instrument (AERA, APA, and NCME 2014; Rios and Wells 2014). Sources of internal structural validity evidence may include analyses of reliability, dimensionality, and measurement invariance.

Reliability is gauged by internal consistency, reflecting i) the reproducibility of test scores under consistent conditions and ii) the ratio of true score variance to observed score variance (Rios and Wells 2014). Dimensionality analysis aims to verify if item interrelations support the inferences made by the measurement model's scores, which are assumed to be unidimensional (Rios and Wells 2014). Measurement invariance confirms that item properties remain consistent across specified groups, such as gender or ethnicity.

CFA facilitates the integration of these diverse sources to substantiate the validity of the internal structure (Bandalos 2018; Flora and Flake 2017; Hughes 2018; Reeves and Marbach-Ad 2016; Rios and Wells 2014). In the applied social sciences, researchers

often have a theoretical dimensional structure in mind (Sireci and Sukin 2013), and CFA is employed to align the structure of the hypothesized measurement model with the observed data (Rios and Wells 2014).

CFA constitutes a fundamental aspect of the covariance-based Structural Equation Modeling (SEM) framework (CB-SEM) (Brown 2023; Harrington 2009; Jackson, Gillaspy, and Purc-Stephenson 2009; Kline 2023; Nye 2022). SEM is a prevalent statistical approach in the applied social sciences (Hoyle 2023; Kline 2023), serving as a generalization of multiple regression and factor analysis (Hoyle 2023). This methodology facilitates the examination of complex relationships between variables and the consideration of measurement error, aligning with the requirements for measurement model validation (Hoyle 2023).

Applications of CFA present significant complexities (Crede and Harms 2019; Jessica K. Flake, Pek, and Hehman 2017; Jessica Kay Flake and Fried 2020; Jackson, Gillaspy, and Purc-Stephenson 2009; Nye 2022; Rogers 2023), influenced by data structure, measurement level of items, research goals, and other factors. CFA can proceed smoothly in scenarios involving unidimensional measurement models with continuous items and large samples, but may encounter challenges, such as diminished SEM flexibility, when dealing with multidimensional models with ordinal items and small sample sizes (Rogers 2023).

This leads to an important question: Can certain strategies within CFA applications simplify the process for social scientists seeking evidence of validity in the internal structure of a measurement model? This inquiry does not suggest that research objectives should conform to quantitative methods. Rather, research aims guide scientific inquiry, defining our learning targets and priorities. Quantitative methods serve as tools towards these ends, not as objectives themselves. They represent one among many tools available to researchers, with the study's purpose dictating method selection (Pilcher and Cortazzi 2023).

However, as the scientific method is an ongoing journey of discovery, many questions, especially in Psychometrics concerning measurement model validation, remain openended. The lack of consensus on complex and varied topics suggests researchers should opt for paths offering maximal analytical flexibility, enabling exploration of diverse methodologies and solutions while keeping research objectives forefront (Price 2017).

A recurrent topic in Factor Analysis (FA) is how to handle the measurement level of scale items. Empirical studies (Rhemtulla, Brosseau-Liard, and Savalei 2012; Robitzsch 2022, 2020) advocating for the treatment of scales with five or more response options as continuous variables have shown to enhance CFA flexibility and address validity evidence for the internal structure. The FA literature acknowledges methodological dilemmas faced when dealing with binary and/or ordinal response items with fewer than five options (Rogers 2023, 2022).

For continuous scale items, the maximum likelihood (ML) estimator and its robust variations are applicable. For non-continuous items, estimators from the Least Squares (cat-LS) family are recommended (Nye 2022; Rogers 2023, 2022). Though cat-LS estimators impose fewer assumptions on data, they require larger sample sizes, more computational power, and greater researcher expertise (Robitzsch 2020).

Assessing model fit is more challenging with cat-LS estimated models compared to those estimated by ML, which are better established and more familiar to researchers

(Rhemtulla, Brosseau-Liard, and Savalei 2012). Despite their increasing popularity, cat-LS models are newer, less recognized, and seldom available in software (Rhemtulla, Brosseau-Liard, and Savalei 2012). Handling missing data remains straightforward with ML models using the Full Information ML (FIML) method but is problematic with ordinal data (Rogers 2023).

Thus, we can optimize the potential of some of the available software (Arbuckle 2019; Bentler and Wu 2020; Fox 2022; JASP Team 2023; Jöreskog and Sörbom 2022; Muthén and Muthén 2023; Neale et al. 2016; Ringle, Wende, and Becker 2022; Rosseel 2012; The jamovi project 2023) and overcome many of the limitations for ordinal and nominal data, which are still present in some of them (Arbuckle 2019; Bentler and Wu 2020; Neale et al. 2016; Ringle, Wende, and Becker 2022).

This discussion does not intend to oversimplify, digress, or claim superiority of one software over another. Rather, it underscores a fundamental statistical principle: transitioning from nominal to ordinal and then to scalar measurement levels increases the flexibility of statistical methods. Empirical studies in CFA support these clarifications (Rhemtulla, Brosseau-Liard, and Savalei 2012; Robitzsch 2022, 2020).

This article assists applied social scientists in decision-making from selecting a measurement model to comparing and updating models for enhanced CFA flexibility. It addresses power analysis, data preprocessing, estimation procedures, and model modification from three angles: smart choices or recommended practices (Jessica K. Flake, Pek, and Hehman 2017; Nye 2022; Rogers 2023), pitfalls to avoid (Crede and Harms 2019; Rogers 2023), and essential reporting elements (Jessica Kay Flake and Fried 2020; Jackson, Gillaspy, and Purc-Stephenson 2009; Rogers 2023).

The aim is to guide researchers through CFA to access the underlying structure of measurement models without falling into common traps at any stage of the validation process. Early-stage decisions can preempt later limitations, while missteps may necessitate exploratory research or additional efforts in subsequent phases.

Practically, this includes an R tutorial utilizing the lavaan package (Rosseel 2012), adhering to reproducibility, replicability, and transparency standards of the Open Science movement (Gilroy and Kaplan 2019; Kathawalla, Silverstein, and Syed 2021; Klein et al. 2018).

Tutorial articles, following the FAIR principles (Findable, Accessible, Interoperable, and Reusable) (Wilkinson et al. 2016), play a vital role in promoting open science (Martins 2021; Mendes-Da-Silva 2023), by detailing significant methods or application areas in an accessible yet comprehensive manner. This encourages adherence to best practices among researchers, minimizing the impact of positive publication bias.

This tutorial is structured into three sections, beyond the introductory discussion. It includes a thorough review of CFA recommended practices, an example of real-world research application in the R ecosystem, and final considerations, following Martins' (2021) format for tutorial articles. This approach, combined with workflow recommendations for reproducibility, aims to support the applied social sciences community in effectively utilizing CFA (Martins 2021; Mendes-Da-Silva 2023).

2 Smart Choices in CFA

This paper presents a comprehensive approach to conducting a standard CFA within the applied social sciences, following the guidelines outlined by Rogers (2023). According to Rogers (2023), a typical CFA study seeks to fit a reflective common factor model with a predefined multifactor structure, established psychometric properties, and a maximum of five Likert-type response options. This scenario frequently occurs in research endeavors where the measurement model facilitates the examination of hypotheses derived from the structural model.

The initial phase in such research involves data preprocessing. Specifically, for categorical data, Rogers (2023) advises employing multiple imputation to handle missing data, taking into consideration the limitations posed by available software and methodologies (Rogers 2023). When a measurement model allows for the treatment of items as continuous variables, addressing this challenge can be deferred to the estimation process stage through the selection of an appropriate estimator (Robitzsch 2022).

This paper reinterprets the insights from Rogers (2023) for CFAs that accommodate continuous item treatment. Thus, a strategic choice involves opting for measurement models that permit this approach, thereby circumventing methodological hurdles (Robitzsch 2022, 2020) associated with binary and/or ordinal response items with up to four or five gradations. Such a decision influences various aspects of the research process, including the choice of software, power analysis, estimation techniques, criteria for model adjustment, and model comparisons. These choices, in turn, affect requirements concerning sample size, computational resources, and the researcher's expertise (Robitzsch 2020).

Subsequent sections delve into themes previously summarized by Rogers (2023), specifically concerning CFAs with ordinal items. These themes are explored in terms of recommended practices (Jessica K. Flake, Pek, and Hehman 2017; Nye 2022; Rogers 2023), pitfalls to avoid (Crede and Harms 2019; Rogers 2023), and reporting guidelines (Jessica Kay Flake and Fried 2020; Jackson, Gillaspy, and Purc-Stephenson 2009; Rogers 2023), all within the context of selecting measurement models that accommodate continuous data interpretation.

Assuming that readers possess a foundational understanding of the topic, this paper omits certain technical details, directing readers to authoritative texts (Brown 2015; Kline 2023) and scholarly articles that provide an introduction to Covariance-Based Structural Equation Modeling (CB-SEM) (Davvetas et al. 2020; **shek2014a?**). The discussion is framed within the CB-SEM paradigm (Brown 2015; Jackson, Gillaspy, and Purc-Stephenson 2009; Kline 2023; Nye 2022), with a focus on CFA. The paper explicitly excludes discussions on measurement model modifications in Variance-Based SEM (VB-SEM), which are predominantly addressed in the literature on Partial Least Squares SEM (PLS-SEM) (Hair et al. 2022, 2017; Henseler 2021).

2.1 Measurement Model Selection

Selecting an appropriate measurement model is a critical initial step in the research process. For robust analysis, it is advisable to prioritize models that provide five or more ordinal response options. Research has shown that a higher number of response gradations enhances the ability to detect inaccurately defined models (Green et al. 1997; Maydeu-Olivares, Fairchild, and Hall 2017), even when using estimators designed for

ordinal items (Xia and Yang 2018). This strategy also mitigates some of the methodological challenges associated with the analysis of ordinal data in CFA (Rhemtulla, Brosseau-Liard, and Savalei 2012; Robitzsch 2022, 2020).

When choosing a measurement scale, it is crucial to select ones that have been validated in the language of application and with the study's target audience (Jessica K. Flake, Pek, and Hehman 2017). Avoid scales that are proprietary or specific to certain professions. An examination of your country's Psychological Test Assessment System can be an effective starting point. If the desired scale is not found within these resources, consider looking into scales developed with the support of public institutions, non-governmental organizations, research centers, or universities, as these entities often invest significant resources in validating measurement models for broader public policy purposes.

An extensive literature review is essential for selecting a suitable measurement model. This should include consulting specialized journals, books, technical reports, and academic dissertations or theses. Schumacker, Wind, and Holmes (2021) provide a detailed guide for initiating this search. Consideration should also be given to systematic reviews or meta-analyses focusing on measurement models related to your topic of interest. It is important to review both the original articles on the scales and subsequent applications. Kline (2016) offers a useful checklist for assessing various measurement methods.

Incorporate control questions, such as requiring respondents to select "strongly agree" on specific items, and monitor survey response times to gauge participant engagement (Collier 2020).

Avoid adopting measurement models designed for narrow purposes or those lacking rigorous psychometric validation (Jessica Kay Flake and Fried 2020; Kline 2016). The mere existence of a scale does not ensure its validity (Jessica K. Flake, Pek, and Hehman 2017). Also, steer clear of seldom-used or outdated scales, as they may have compromised psychometric properties. Translating a scale from another language for immediate use without thorough translation and retranslation processes is inadvisable. Be cautious of overlooking alternative factorial structures (e.g., higher-order or bifactor models) that could potentially salvage the research if considered thoroughly (Crede and Harms 2019).

When selecting a scale, justify its choice by highlighting its strong psychometric properties, including previous empirical evidence of its application within the target population and its reliability and validity metrics (Jessica Kay Flake and Fried 2020; Jackson, Gillaspy, and Purc-Stephenson 2009; Kline 2016). If the scale has multiple potential factorial structures, provide a rationale for the chosen model to prevent the misuse of CFA for exploratory purposes (Jackson, Gillaspy, and Purc-Stephenson 2009).

Clearly specify the selected model and rationalize your choice by detailing its advantages over other theoretical models. Illustrating the models under consideration can further clarify your research approach (Jackson, Gillaspy, and Purc-Stephenson 2009). Finally, identify and explain any potential cross-loadings based on prior empirical evidence (Brown 2023; Nye 2022), ensuring a comprehensive and well-justified methodological foundation for your study.

2.2 Power Analysis

When addressing Power Analysis (PA) in CFA and SEM, it's essential to move beyond general rules of thumb for determining sample sizes. Commonly cited guidelines sug-

gesting minimum sizes or specific ratios of observations to parameters (e.g., 50, 100, 200, 300, 400, 500, 1000 for sample sizes or 20/1, 10/1, 5/1 for observation/parameter ratios) (Kline 2023; Kyriazos 2018) are based on controlled conditions that may not directly transfer to your study's context.

Reliance on lower-bound sample sizes as a substitute for thorough PA risks inadequate power for detecting meaningful effects in your model (Westland 2010; Yilin Andre Wang 2023). Tools like Soper's calculator (https://www.danielsoper.com/statcalc/), while popular and frequently cited (as of 02/20/2024, with almost four years of existence, it had collected more than 1,000 citations on Google Scholar), should not replace a tailored PA approach. Such calculators, despite their utility, may not fully accommodate the complexities and specific requirements of your research design (Kyriazos 2018; Feng and Hancock 2023; Moshagen and Bader 2023).

A modern perspective on sample size determination emphasizes customizing power calculations to fit the unique aspects of each study, incorporating specific research settings and questions (Feng and Hancock 2023; Moshagen and Bader 2023). This approach underscores that there is no universal sample size or minimum that applies across all research scenarios (Kline 2023).

Planning for PA should ideally precede data collection, enhancing the researcher's understanding of the study and facilitating informed decisions regarding the measurement model based on existing literature and known population characteristics (Feng and Hancock 2023; Leite, Bandalos, and Shen 2023). A priori PA not only ensures adequate sample size for detecting the intended effects, minimizing Type II errors, but also aids in budgeting for data collection and enhancing overall research design (Feng and Hancock 2023).

PA in SEM can be approached analytically, using asymptotic theory, or through simulation methods. Analytical methods require specifying the effect size in relation to the non-centrality parameter, while simulated PA leverages a population model to empirically estimate power (Moshagen and Bader 2023; Feng and Hancock 2023). These approaches are applicable to assessing both global model fit and specific model parameters.

For CFA, evaluating the power related to the global fit of the measurement model is recommended (Nye 2022). Although analytical solutions have their limitations, they can serve as preliminary steps, complemented by simulation techniques for a more comprehensive PA (Feng and Hancock 2023; Moshagen and Bader 2023).

Several resources offer analytical solutions for global fit PA, including ShinyApps by Jak et al. (2021), Moshagen and Bader (2023), Y. Andre Wang and Rhemtulla (2021), and Zhang and Yuan (2018), with the last application providing a comprehensive suite for Monte Carlo Simulation (SMC) that accommodates missing data, non-normal distributions, and facilitates model testing without extensive coding (Y. Andre Wang and Rhemtulla 2021). For an overview of these solutions and a discussion of analytical approaches, see Feng and Hancock (2023), Jak et al. (2021), Nye (2022), and Yilin Andre Wang (2023).

However, it is a smart decision to run an SMC for the PA of your CFA model using solutions that are consistent with the results' reproducible and replicability. In this way, even analytical solutions that the researcher may use as a starting point are recommended in the R environment via the semTools packages (Jak et al. 2021) and

semPower 2 (Jobst, Bader, and Moshagen 2023; Moshagen and Bader 2023). The first option is compatible with the lavaan syntax and looks to be enough. The second, albeit including SMC in some cases, has a more difficult syntax.

For detailed and tailored PA, especially in complex models or unique study designs, the simsem package offers a robust solution, allowing for the relaxation of traditional assumptions and supporting the use of robust estimators. This package, which utilizes the familiar lavaan syntax, simplifies the learning curve for researchers already accustomed to SEM analyses, providing a user-friendly interface for conducting SMC (Pornprasertmanit et al. 2022).

Publishing the sampling design and methodology enhances the reproducibility and replicability of research, contributing to the scientific community's collective understanding and validation of measurement models (Jessica K. Flake, Pek, and Hehman 2017; Jessica Kay Flake et al. 2022; Jessica Kay Flake and Fried 2020; Leite, Bandalos, and Shen 2023). In the context of CFA, acknowledging the power limitations of your study can signal potential concerns for the broader inferences drawn from your research, emphasizing the importance of external validity and the relevance of the outcomes over mere precision (Leite, Bandalos, and Shen 2023).

2.3 Pre-processing

Upon gathering and tabulating original data, ideally in non-binary formats such as CSV, TXT, or JSON, the first step in data preprocessing should be to eliminate responses from participants who have abandoned the study. This identification often occurs at the end of preprocessing, where these incomplete responses can offer insights into handling missing data, outliers, and multicollinearity.

Incorporating control questions and measuring response time allows researchers to further refine their dataset by excluding participants who fail control items or complete the survey unusually quickly (Collier 2020). Calculating individual response variability (standard deviation) can identify respondents who may not have engaged meaningfully with the survey, indicated by minimal variation in their responses.

These preliminary data cleaning steps are fundamental yet frequently overlooked in empirical research. They can significantly enhance data quality before engaging in more complex statistical analyses. Visual and descriptive examination of measurement model items is implicitly beneficial for any statistical investigation and should be considered standard practice.

While data transformation methods like linearization or normalization are available, they are generally not necessary given the robust estimation processes that can handle non-normal data (Brown 2015). Parceling items is also discouraged due to its potential to obscure underlying multidimensional structures (Brown 2015; Crede and Harms 2019).

Addressing missing data, outliers, and multicollinearity is critical. Single imputation methods should be avoided as they underestimate error variance and can lead to identification problems in your model (Enders 2023). For missing data under 5%, the impact may be minimal, but for higher rates, Full Information ML (FIML) or Multiple Imputation (MI) should be utilized, with FIML often being the most straightforward and effective choice for CFA (Brown 2015; Kline 2023).

FIML and MI are preferred for handling missing data due to their ability to produce consistent and efficient parameter estimates under conditions similar to MI (Enders 2023; Kline 2023). FIML it can be adapted for non-normal data using robust estimators (Brown 2015).

Calculating the Variance Inflation Factor (VIF) helps identify items with problematic multicollinearity (VIF > 10), which should be addressed to prevent model convergence issues and misinterpretations (Kline 2016; Whittaker and Schumacker 2022). Reflective constructs in CFA require some level of item correlation but not to the extent that it causes statistical or validity concerns.

Consider multivariate outliers rather than univariate ones, identifying and assessing their exclusion based on sample characteristics. Reporting all data cleaning processes, including any loss of items and strategies for assessing respondent engagement, is crucial for transparency. Additionally, documenting signs of multicollinearity and the software or packages used (with versions) enhances the reproducibility and credibility of the research (Jessica Kay Flake and Fried 2020; Jackson, Gillaspy, and Purc-Stephenson 2009).

Finally, making raw data public adheres to the principles of open science, promoting transparency and allowing for independent validation of research findings (Crede and Harms 2019; Jessica Kay Flake et al. 2022; Jessica Kay Flake and Fried 2020). This practice not only contributes to the scientific community's collective knowledge base but also reinforces the integrity and reliability of the research conducted.

2.4 Estimation Process

In CFA with ordinal items, such as those involving Likert-type scales with up to five points, Rogers (2023) advocates for the use of estimators from the Ordinary Least Squares (OLS) family. Specifically, for smaller samples, the recommendation is to utilize the Unweighted Least Squares (ULS) in its robust form (RULS), and for larger samples, the Diagonally Weighted Least Squares (DWLS) in its robust version (RDWLS), citing substantial supporting research.

Despite this, empirical evidence (Rhemtulla, Brosseau-Liard, and Savalei 2012; Robitzsch 2022) and theoretical considerations (Robitzsch 2020) suggest that treating ordinal data as continuous can yield acceptable outcomes when the response options number five or more. Particularly with 6-7 categories, comparisons between methods under various conditions reveal little difference, and it is recommended to use a greater number of response alternatives (5) to enhance the power for detecting model misspecifications (Maydeu-Olivares, Fairchild, and Hall 2017).

The ML estimator, noted for its robustness to minor deviations from normality (Brown 2015), is further improved by using robust versions like MLR (employing Huber-White standard errors and Yuan-Bentler scaled ^2). This adjustment allows for generating robust standard errors and adjusted test statistics, with MLR offering extensive applicability including in scenarios of missing data (RFIML) or where data breaches the independence of observations assumption (Brown 2015; Rosseel 2012). Comparative empirical studies have supported the effectiveness of MLR against alternative estimators (Bandalos 2014; Holgado-Tello, Morata-Ramirez, and García 2016; Li 2016; Nalbantoğlu-Yılmaz 2019; Yang and Liang 2013; Yang-Wallentin, Jöreskog, and Luo 2010).

Researchers are advised to carefully describe and justify the chosen estimation method

based on the data characteristics and the specific model being evaluated (Crede and Harms 2019). It is also critical to report any estimation challenges encountered, such as algorithm non-convergence or model misidentification (Nye 2022). In case of estimation difficulties, alternative approaches like MLM estimators (employing robust standard errors and Satorra-Bentler scaled χ^2) or the default ML with non-parametric bootstrapping, as proposed by Bollen-Stine, can be considered. This latter approach is also capable of accommodating missing data (Brown 2015; Kline 2023).

Additionally, it is important to clarify whether the variance of a marker variable was fixed (=1) to scale the latent variables (Jackson, Gillaspy, and Purc-Stephenson 2009), and to provide both standardized and unstandardized parameter estimates (Nye 2022). These steps are crucial for ensuring transparency, reproducibility, and the ability to critically assess the validity of the CFA results.

2.5 Model Fit

In conducting CFA with ordinal items, such as Likert-type scales, it's crucial to approach model evaluation with nuance and avoid reliance on rigid cutoff values for fit indices. Adhering strictly to traditional cutoffs – whether more conservative (e.g., SRMR .06, RMSEA .06, CFI .95) or less conservative (e.g., RMSEA .08, CFI .90, SRMR .08) – should not be the sole criterion for model acceptance (Xia and Yang 2019). The origins of these thresholds are in simulation studies with specific configurations (up to three factors, fifteen items, factor loadings between 0.7 and 0.8) (West et al. 2023), and may not universally apply due to the variance in the number of items, factors, model degrees of freedom, misfit types, and presence of missing data (Groskurth, Bluemke, and Lechner 2023; Niemand and Mai 2018; West et al. 2023).

Evaluation of global fit indices (SRMR, RMSEA, CFI) should be done in a collective manner, rather than fixating on any single index. A deviation from traditional cutoffs warrants further investigation into whether the discrepancy is attributable to data characteristics or limitations of the index, rather than indicating a fundamental model misspecification (Nye 2022). Interpreting fit indices as effect sizes can offer a more meaningful assessment of model fit, aligning with their original conceptualization (McNeish and Wolf 2023a; McNeish 2023b).

The SRMR is noted for its robustness across various conditions, including non-normality and different measurement levels of items. Pairing SRMR with CFI can help balance Type I and Type II errors, but reliance on alternative indices may increase the risk of Type I error (Mai, Niemand, and Kraus 2021; Niemand and Mai 2018).

Emerging methods like the Dynamic Fit Index (DFI) and Flexible Cutoffs (FCO) offer tailored approaches to evaluating global fit. DFI, based on simulation, provides model-specific cutoff points, adjusting simulations to match the empirical model's characteristics (McNeish 2023a; McNeish and Wolf 2023b; McNeish and Wolf 2023a). FCO, while not requiring identification of a misspecified model like DFI, conservatively defines misfit, shifting focus from approximate to accurate fit (McNeish and Wolf 2023b).

For those hesitant to delve into simulation-based methods, Equivalence Testing (EQT) presents an alternative. EQT aligns with the analytical mindset of PA and incorporates DFI principles, challenging the conventional hypothesis testing framework by considering model specification and misspecification size control (Yuan et al. 2016).

When addressing reliability, Cronbach's Alpha should not be the default measure due

to its limitations. Instead, consider McDonald's Omega or the Greatest Lower Bound (GLB) for a more accurate reliability assessment within the CFA context (Bell, Chalmers, and Flora 2023; Cho 2022; Dunn, Baguley, and Brunsden 2014; Flora 2020; Goodboy and Martin 2020; Green and Yang 2015; Hayes and Coutts 2020; Kalkbrenner 2023; McNeish 2018; Trizano-Hermosilla and Alvarado 2016).

Before modifying the model, first check for Heywood instances, which are standardized factor loadings greater than one or negative variances (Nye 2022) and document the chosen cutoffs for evaluation. Tools and resources like ShinyApp for DFI and the FCO package in R can facilitate the application of these advanced methodologies (McNeish and Wolf 2023a; Mai, Niemand, and Kraus 2021; Niemand and Mai 2018). Always report corrected chi-square and degrees of freedom, alongside a minimum of three global fit indices (RMSEA, CFI, SRMR) and local fit measures to provide a comprehensive view of model fit and adjustment decisions (Crede and Harms 2019; Jessica Kay Flake and Fried 2020).

2.6 Model Comparisons and Modifications

Researchers embarking on CFA should avoid prematurely committing to a specific factor structure without thoroughly evaluating and comparing alternate configurations. It's advisable to consider various potential structures early in the study design, ensuring the selected model is based on its merits relative to competing theories (Jackson, Gillaspy, and Purc-Stephenson 2009). Since models are inherently approximations of reality, adopting the most effective "working hypothesis" is a dynamic process, contingent on ongoing assessments against emerging alternatives (Preacher and Yaremych 2023).

Good models are characterized not only by their interpretability, simplicity, and generalizability but notably by their capacity to surpass competing models in critical aspects. This competitive advantage frames the selected theory as the prevailing hypothesis until a more compelling alternative is identified (Preacher and Yaremych 2023).

The evaluation of model fit should extend beyond isolated assessments using fit indices. A comprehensive approach involves comparing multiple models, each grounded in substantiated theories, to discern the most accurate representation of the underlying structure. This comparative analysis is preferred over singular model evaluations, fostering a more holistic understanding of the phenomena under study (Preacher and Yaremych 2023).

Uniform application of models across the same dataset, utilizing identical software and sample size, ensures consistency in the researcher's analytical freedom, mitigating the risk of results manipulation. This standardized approach underpins a more rigorous and transparent investigative process (Preacher and Yaremych 2023).

Model selection is instrumental in pinpointing the most effective explanatory framework for the observed phenomena, enabling the dismissal of less performant models while retaining promising ones for further exploration. This methodological flexibility enhances the depth of analysis, contributing to the advancement of knowledge within the social sciences (Preacher and Yaremych 2023).

Adjustments to a model, particularly in response to unsatisfactory fit indices, should be theoretically grounded and reflective of findings from prior research. Blind adherence to a pre-established model or making hasty modifications can adversely affect the structural model's integrity. Thoughtful adjustments, potentially revisiting exploratory

factor analysis (EFA) or considering Exploratory SEM (ESEM) for cross-loadings representation, are preferable to drastic changes that might shift the study from confirmatory to exploratory research (Brown 2023; Jessica K. Flake, Pek, and Hehman 2017; Jackson, Gillaspy, and Purc-Stephenson 2009; Crede and Harms 2019).

All modifications to the measurement model, especially those enhancing model fit, must be meticulously documented to maintain transparency and support reproducibility (Jessica Kay Flake and Fried 2020). Openly reporting these adjustments, including item exclusions and inter-item correlations, is vital for the scientific integrity of the research (Nye 2022; Jessica Kay Flake et al. 2022).

Regarding model comparison and selection, traditional fit indices (SRMR, RMSEA, CFI) have limitations for direct model comparisons. Adjusted chi-square tests and information criteria like AIC and BIC are more suitable for this purpose, balancing model fit and parsimony. These criteria, however, should be applied with an understanding of their constraints and complemented by theoretical judgment to inform model selection decisions (Preacher and Yaremych 2023; Brown 2015; Huang 2017; Lai 2020, 2021).

Ultimately, model selection in SEM is a nuanced process, blending empirical evidence with theoretical insights. Researchers are encouraged to leverage a range of models based on theoretical foundations, ensuring that the eventual model selection is not solely determined by statistical criteria but is also informed by substantive theory and expertise (Preacher and Yaremych 2023). This balanced approach underscores the importance of theory-driven research in the social sciences, guiding the interpretation and application of findings derived from chosen models.

3 Executable Manuscript

3.1 Measurement Model Selection

3.2 Power Analysis

3.3 Pre-processing

3.4 Estimation Process

3.5 Model Fit

3.6 Model Comparisons and Modifications

4 Final Considerations

References

AERA, APA, and NCME. 1999. "Standards for Educational and Psychological Testing." Washington: American Educational Research Association, American Psychological Association, & National Council on Measurement in Education.

———. 2014. "Standards for Educational and Pshychological Testing." Washington: American Educational Research Association, American Psychological Association & National Council on Measurement in Education.

Arbuckle, J. L. 2019. "Amos." Chicago: IBM Corp.

Bandalos, Deborah L. 2014. "Relative Performance of Categorical Diagonally Weighted Least Squares and Robust Maximum Likelihood Estimation." *Structural Equation Modeling* 21 (1): 102–16. https://doi.org/10.1080/10705511.2014.859510.

———. 2018. *Measurement Theory and Applications for the Social Sciences.* New York: Guilford Press.

Bell, Stephanie M., R. Philip Chalmers, and David B. Flora. 2023. "The Impact of Measurement Model Misspecification on Coefficient Omega Estimates of Composite Reliability." *Educational and Psychological Measurement*, 1–36. https://doi.org/10.1177/00131644231155804.

Bentler, Peter M., and Erik Wu. 2020. "EQS 6.4 for Windows." Multivariate Software, Inc. https://mvsoft.com.

Brown, Timothy A. 2015. *Confirmatory Factor Analysis for Applied Research.* New York: The Guilford Press.

———. 2023. "Confirmatory Factor Analysis." In *Handbook of Structural Equation Modeling*, edited by Rick H. Hoyle. New York: The Guilford Press.

Cho, Eunseong. 2022. "Reliability and Omega Hierarchical in Multidimensional Data: A Comparison of Various Estimators." *Psychological Methods.* https://doi.org/10.1037/met0000525.

Cohen, Ronald Jay, Joel W. Schneider, and Renée M. Tobin. 2022. *Psychological Testing and Assessment: An Introduction to Test and Measurement.* New York: McGraw Hill LLC.

Collier, Joel E. 2020. *Applied Structural Equation Modeling Using AMOS: Basic to Advanced Techniques.* New York: Routledge.

Crede, Marcus, and Peter Harms. 2019. "Questionable Research Practices When Using Confirmatory Factor Analysis." *Journal of Managerial Psychology* 34 (1): 18–30. https://doi.org/10.1108/JMP-06-2018-0272.

Davvetas, Vasileios, Adamantios Diamantopoulos, Ghasem Zaefarian, and Christina Sichtmann. 2020. "Ten Basic Questions about Structural Equations Modeling You Should Know the Answers to – But Perhaps You Don't." *Industrial Marketing Management* 90 (October): 252–63. https://doi.org/10.1016/j.indmarman.2020.07.016.

Dunn, Thomas J., Thom Baguley, and Vivienne Brunsden. 2014. "From Alpha to Omega: A Practical Solution to the Pervasive Problem of Internal Consistency Estimation." *British Journal of Psychology* 105 (3): 399–412. https://doi.org/10.1111/bjop.12046.

Enders, Craig. 2023. "Fitting Structural Equation Models with Missing Data." In *Handbook of Structural Equation Modeling*, edited by Rick H. Hoyle. New York: The Guilford Press.

Feng, Yi, and Gregory R. Hancock. 2023. "Power Analysis Within a Structural Equation Modeling Framework." In *Handbook of Structural Equation Modeling*, edited by Rick H. Hoyle. New York: The Guilford Press.

Flake, Jessica Kay, Ian J. Davidson, Octavia Wong, and Jolynn Pek. 2022. "Construct Validity and the Validity of Replication Studies: A Systematic Review." *American Psychologist* 77 (4): 576–88. https://doi.org/10.1037/amp0001006.

Flake, Jessica Kay, and Eiko I. Fried. 2020. "Measurement Schmeasurement: Questionable Measurement Practices and How to Avoid Them." *Advances in Methods and Practices in Psychological Science* 3 (4): 456–65. https://doi.org/10.1177/2515245920952393.

Flake, Jessica K., Jolynn Pek, and Eric Hehman. 2017. "Construct Validation in

Social and Personality Research: Current Practices and Recommendations." *Social Psychological and Personality Science* 8 (4): 370–78. https://doi.org/10.1177/1948550617693063.

Flora, David B. 2020. "Your Coefficient Alpha Is Probably Wrong, but Which Coefficient Omega Is Right? A Tutorial on Using R to Obtain Better Reliability Estimates." *Advances in Methods and Practices in Psychological Science* 3 (4): 484–501. https://doi.org/10.1177/2515245920951747.

Flora, David B., and Jessica K. Flake. 2017. "The Purpose and Practice of Exploratory and Confirmatory Factor Analysis in Psychological Research: Decisions for Scale Development and Validation." *Canadian Journal of Behavioural Science* 49 (2): 78–88. https://doi.org/10.1037/cbs0000069.

Fox, John. 2022. "Sem: Structural Equation Modeling." R package. https://cran.r-project.org/web/packages/sem/.

Furr, Michael R. 2021. *Psychometrics: An Introduction*. SAGE Publications.

Gilroy, Shawn P., and Brent A. Kaplan. 2019. "Furthering Open Science in Behavior Analysis: An Introduction and Tutorial for Using GitHub in Research." *Perspectives on Behavior Science* 42 (3): 565–81. https://doi.org/10.1007/s40614-019-00202-5.

Goodboy, Alan K., and Matthew M. Martin. 2020. "Omega over Alpha for Reliability Estimation of Unidimensional Communication Measures." *Annals of the International Communication Association* 44 (4): 422–39. https://doi.org/10.1080/23808985.2020.1846135.

Green, Samuel B., Theresa M. Akey, Kandace K. Fleming, Scott L. Hershberger, and Janet G. Marquis. 1997. "Effect of the Number of Scale Points on Chi-Square Fit Indices in Confirmatory Factor Analysis." *Structural Equation Modeling: A Multidisciplinary Journal* 4 (2): 108–20. https://doi.org/10.1080/10705519709540064.

Green, Samuel B., and Yanyun Yang. 2015. "Evaluation of Dimensionality in the Assessment of Internal Consistency Reliability: Coefficient Alpha and Omega Coefficients." *Educational Measurement: Issues and Practice* 34 (4): 14–20. https://doi.org/10.1111/emip.12100.

Groskurth, Katharina, Matthias Bluemke, and Clemens M. Lechner. 2023. "Why We Need to Abandon Fixed Cutoffs for Goodness-of-Fit Indices: An Extensive Simulation and Possible Solutions." *Behavior Research Methods*, August. https://doi.org/10.3758/s13428-023-02193-3.

Hair, Joseph F., Tomas M. G. Hult, Christian M. Ringle, and Marko Sarstedt. 2022. *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*. Thousand Oaks: Sage Publications.

Hair, Joseph F., Marko Sarstedt, Christian Ringle, and Siegfried P. Gudergan. 2017. *Advanced Issues in Partial Least Squares Structural Equation Modeling*. London: SAGE Publications, Inc.

Harrington, Donna. 2009. *Confirmatory Factor Analysis*. New York: Oxford University Press.

Hayes, Andrew F., and Jacob J. Coutts. 2020. "Use Omega Rather Than Cronbach's Alpha for Estimating Reliability. But…." *Communication Methods and Measures* 14 (1): 1–24. https://doi.org/10.1080/19312458.2020.1718629.

Henseler, Jörg. 2021. *Composite-Based Structural Equation Modeling: Analyzing Latent and Emergent Variables*. New York: The Guilford Press.

Holgado-Tello, F., M. Morata-Ramirez, and M. García. 2016. "Robust Estimation Methods in Confirmatory Factor Analysis of Likert Scales: A Simulation Study." *International Review of Social Sciences and Humanities* 11 (2): 80–96.

Hoyle, Rick H. 2023. "Structural Equation Modeling: An Overview." In *Handbook of Structural Equation Modeling*, edited by Rick H. Hoyle. New Yoirk: Guilford Press.

Huang, Po-Hsien. 2017. "Asymptotics of AIC, BIC, and RMSEA for Model Selection in Structural Equation Modeling." *Psychometrika* 82 (2): 407–26. https://doi.org/10.1007/s11336-017-9572-y.

Hughes, David J. 2018. "Psychometric Validity: Establishing the Accuracy and Appropriateness of Psychometric Measures." In *The Wiley Handbook of Psychometric Testing: A Multidisciplinary Reference on Survey, Scale and Test Development*, edited by Paul Irwing, Tom Booth, and David J. Hughes. John Wiley & Sons Ltd.

Jackson, Dennis L., J. Arthur Gillaspy, and Rebecca Purc-Stephenson. 2009. "Reporting Practices in Confirmatory Factor Analysis: An Overview and Some Recommendations." *Psychological Methods* 14 (1). https://doi.org/10.1037/a0014694.

Jak, Suzanne, Terrence D Jorgensen, Mathilde G E Verdam, Frans J Oort, and Louise Elffers. 2021. "Analytical Power Calculations for Structural Equation Modeling: A Tutorial and Shiny App." *Behavior Research Mehods* 53: 1385–1406. https://doi.org/10.3758/s13428-020-01479-0/Published.

JASP Team. 2023. "JASP." [Computer Software]. https://jasp-stats.org/.

Jobst, Lisa J., Martina Bader, and Morten Moshagen. 2023. "A Tutorial on Assessing Statistical Power and Determining Sample Size for Structural Equation Models." *Psychological Methods* 28 (1): 207–21. https://doi.org/10.1037/met0000423.

Jöreskog, K. G., and D. Sörbom. 2022. "LISREL 12 for Windows." Scientific Software International, Inc. https://ssicentral.com/index.php/products/lisrel/.

Kalkbrenner, Michael T. 2023. "Alpha, Omega, and H Internal Consistency Reliability Estimates: Reviewing These Options and When to Use Them." *Counseling Outcome Research and Evaluation* 14 (1): 77–88. https://doi.org/10.1080/21501378.2021.1940118.

Kathawalla, Ummul-Kiram, Priya Silverstein, and Moin Syed. 2021. "Easing Into Open Science: A Guide for Graduate Students and Their Advisors." *Collabra: Psychology* 7 (1): 18684. https://doi.org/10.1525/collabra.18684.

Klein, Olivier, Tom E. Hardwicke, Frederik Aust, Johannes Breuer, Henrik Danielsson, Alicia Hofelich Mohr, Hans IJzerman, Gustav Nilsonne, Wolf Vanpaemel, and Michael C. Frank. 2018. "A Practical Guide for Transparency in Psychological Science." Edited by Michéle Nuijten and Simine Vazire. *Collabra: Psychology* 4 (1): 20. https://doi.org/10.1525/collabra.158.

Kline, Rex B. 2016. *Principles and Pratice of Structural Equation Modeling.* New York: The Guilford Press.

———. 2023. *Principles and Pratice of Structural Equation Modeling.* Fifth Edition. New York: The Guilford Press.

Kyriazos, Theodoros A. 2018. "Applied Psychometrics: Sample Size and Sample Power Considerations in Factor Analysis (EFA, CFA) and SEM in General." *Psychology* 09 (08): 2207–30. https://doi.org/10.4236/psych.2018.98126.

Lai, Keke. 2020. "Confidence Interval for RMSEA or CFI Difference Between Nonnested Models." *Structural Equation Modeling: A Multidisciplinary Journal* 27 (1): 16–32. https://doi.org/10.1080/10705511.2019.1631704.

———. 2021. "Fit Difference Between Nonnested Models Given Categorical Data: Measures and Estimation." *Structural Equation Modeling: A Multidisciplinary Journal* 28 (1): 99–120. https://doi.org/10.1080/10705511.2020.1763802.

Leite, Walter L., Deborah L. Bandalos, and Zuchao Shen. 2023. "Simulation Methods in Structural Equation Modeling." In *Handbook of Structural Equation Modeling*,

edited by Rick H. Hoyle. New York: The Guilford Press.

Li, Cheng Hsien. 2016. "Confirmatory Factor Analysis with Ordinal Data: Comparing Robust Maximum Likelihood and Diagonally Weighted Least Squares." *Behavior Research Methods* 48 (3): 936–49. https://doi.org/10.3758/s13428-015-0619-7.

Mai, Robert, Thomas Niemand, and Sascha Kraus. 2021. "A Tailored-Fit Model Evaluation Strategy for Better Decisions about Structural Equation Models." *Technological Forecasting and Social Change* 173 (December): 121142. https://doi.org/10.1016/j.techfore.2021.121142.

Martins, Henrique Castro. 2021. "Tutorial-Articles: The Importance of Data and Code Sharing." *Revista de Administração Contemporânea* 25 (1): e200212. https://doi.org/10.1590/1982-7849rac2021200212.

Maydeu-Olivares, Alberto, Amanda J. Fairchild, and Alexander G. Hall. 2017. "Goodness of Fit in Item Factor Analysis: Effect of the Number of Response Alternatives." *Structural Equation Modeling: A Multidisciplinary Journal* 24 (4): 495–505. https://doi.org/10.1080/10705511.2017.1289816.

McNeish, Daniel. 2018. "Thanks Coefficient Alpha, We'll Take It from Here." *Psychological Methods* 23 (3): 412–33. https://doi.org/10.1037/met0000144.

———. 2023a. "Dynamic Fit Index Cutoffs for Factor Analysis with Likert, Ordinal, or Binary Responses." *PsyArXiv Preprints.* https://doi.org/10.31234/osf.io/tp35s.

———. 2023b. "Generalizability of Dynamic Fit Index, Equivalence Testing, and Hu & Bentler Cutoffs for Evaluating Fit in Factor Analysis." *Multivariate Behavioral Research* 58 (1): 195–219. https://doi.org/10.1080/00273171.2022.2163477.

McNeish, Daniel, and Melissa G. Wolf. 2023a. "Dynamic Fit Index Cutoffs for Confirmatory Factor Analysis Models." *Psychological Methods* 28 (1): 61–88. https://doi.org/10.1037/met0000425.

McNeish, Daniel, and Melissa Gordon Wolf. 2023b. "Direct Discrepancy Dynamic Fit Index Cutoffs for Arbitrary Covariance Structure Models." Preprint. PsyArXiv. https://doi.org/10.31234/osf.io/4r9fq.

Mendes-Da-Silva, Wesley. 2023. "What Lectures and Research in Business Management Need to Know About Open Science." *Revista de Administração de Empresas* 63 (4): e0000–0033. https://doi.org/10.1590/s0034-759020230408x.

Moshagen, Morten, and Martina Bader. 2023. "semPower: General Power Analysis for Structural Equation Models." *Behavior Research Methods*, November. https://doi.org/10.3758/s13428-023-02254-7.

Muthén, L. K., and B. O. Muthén. 2023. "Mplus Version 8.9 User's Guide."

Nalbantoğlu-Yılmaz, Funda. 2019. "Comparison of Different Estimation Methods Used in Confirmatory Factor Analyses in Non-Normal Data: A Monte Carlo Study." *International Online Journal of Educational Sciences* 11 (4). https://doi.org/10.15345/iojes.2019.04.010.

Neale, Michael C., Michael D. Hunter, Joshua N. Pritikin, Mahsa Zahery, Timothy R. Brick, Robert M. Kirkpatrick, Ryne Estabrook, Timothy C. Bates, Hermine H. Maes, and Steven M. Boker. 2016. "OpenMx 2.0: Extended Structural Equation and Statistical Modeling." *Psychometrika* 81 (2): 535–49. https://doi.org/10.1007/s11336-014-9435-8.

Niemand, Thomas, and Robert Mai. 2018. "Flexible Cutoff Values for Fit Indices in the Evaluation of Structural Equation Models." *Journal of the Academy of Marketing Science* 46 (6): 1148–72. https://doi.org/10.1007/s11747-018-0602-9.

Nye, Christopher D. 2022. "Reviewer Resources: Confirmatory Factor Analysis." *Organizational Research Methods*, August, 109442812211205. https://doi.org/10.1177/

10944281221120541.

Pilcher, Nick, and Martin Cortazzi. 2023. "'Qualitative' and 'Quantitative' Methods and Approaches Across Subject Fields: Implications for Research Values, Assumptions, and Practices." *Quality & Quantity*, September. https://doi.org/10.1007/s11135-023-01734-4.

Pornprasertmanit, Sunthud, Patrick Miller, Terrence D. Jorgensen, and Quick Corbin. 2022. "Simsem: SIMulated Structural Equation Modeling." R package. www.simsem.org.

Preacher, Kristopher J., and Haley E. Yaremych. 2023. "Model Selection in Structural Equation Modeling." In *Handbook of Structural Equation Modeling*, edited by Rick H. Hoyle. New York: The Guilford Press.

Price, Larry R. 2017. *Psychometric Methods: Theory into Practice.* 1st Edition. Methodology in the Social Sciences. New York: The Guilford Press.

Reeves, Todd D., and Gili Marbach-Ad. 2016. "Contemporary Test Validity in Theory and Practice: A Primer for Discipline-Based Education Researchers." *CBE Life Sciences Education* 15 (1). https://doi.org/10.1187/cbe.15-08-0183.

Rhemtulla, Mijke, Patricia É Brosseau-Liard, and Victoria Savalei. 2012. "When Can Categorical Variables Be Treated as Continuous? A Comparison of Robust Continuous and Categorical SEM Estimation Methods Under Suboptimal Conditions." *Psychological Methods* 17 (3): 354–73. https://doi.org/10.1037/a0029315.

Ringle, Christian M., Sven Wende, and Jan Michael Becker. 2022. "SmartPLS 4." Oststeinbek: SmartPLS. https://www.smartpls.com.

Rios, Joseph, and Craig Wells. 2014. "Validity Evidence Based on Internal Structure." *Psicothema* 26 (1): 108–16. https://doi.org/10.7334/psicothema2013.260.

Robitzsch, Alexander. 2020. "Why Ordinal Variables Can (Almost) Always Be Treated as Continuous Variables: Clarifying Assumptions of Robust Continuous and Ordinal Factor Analysis Estimation Methods." *Frontiers in Education* 5 (October). https://doi.org/10.3389/feduc.2020.589965.

———. 2022. "On the Bias in Confirmatory Factor Analysis When Treating Discrete Variables as Ordinal Instead of Continuous." *Axioms* 11 (4). https://doi.org/10.3390/axioms11040162.

Rogers, Pablo. 2022. "Best Practices for Your Exploratory Factor Analysis: A Factor Tutorial." *Revista de Administração Contemporânea* 26 (6). https://doi.org/10.1590/1982-7849rac2022210085.en.

———. 2023. "Best Practices for Your Confirmatory Factor Analysis: A JASP and Lavaan Tutorial." Preprint. Open Science Framework. https://doi.org/10.31219/osf.io/57efj.

Rosseel, Yves. 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.

Schumacker, Randall E., Stefanie A. Wind, and Lauren F. Holmes. 2021. "Resources for Identifying Measurement Instruments for Social Science Research." *Measurement: Interdisciplinary Research and Perspectives* 19 (4): 250–57. https://doi.org/10.1080/15366367.2021.1950486.

Sireci, Stephen G., and Tia Sukin. 2013. "Test Validity." In *APA Handbook of Testing and Assessment in Psychology, Vol. 1: Test Theory and Testing and Assessment in Industrial and Organizational Psychology.*, 61–84. Washington: American Psychological Association. https://doi.org/10.1037/14047-004.

The jamovi project. 2023. "Jamovi." [Computer Software]. https://www.jamovi.org.

Trizano-Hermosilla, Italo, and Jesús M. Alvarado. 2016. "Best Alternatives to Cron-

bach's Alpha Reliability in Realistic Conditions: Congeneric and Asymmetrical Measurements." *Frontiers in Psychology* 7 (MAY). https://doi.org/10.3389/fpsyg.2016.00769.

Urbina, Susana. 2014. *Essentials of Psychological Testing.* Hoboken, New Jersey: John Wiley & Sons.

Wang, Y. Andre, and Mijke Rhemtulla. 2021. "Power Analysis for Parameter Estimation in Structural Equation Modeling: A Discussion and Tutorial." *Advances in Methods and Practices in Psychological Science* 4 (1): 1–17. https://doi.org/10.1177/2515245920918253.

Wang, Yilin Andre. 2023. "How to Conduct Power Analysis for Structural Equation Models: A Practical Primer." Preprint. PsyArXiv. https://doi.org/10.31234/osf.io/4n3uk.

West, Stephen G., Wei Wu, Daniel McNeish, and Andrea Savord. 2023. "Model Fit in Structural Equation Modeling." In *Handbook of Structural Equation Modeling*, edited by Rick H. Hoyle. New York: The Guilford Press.

Westland, Christopher J. 2010. "Lower Bounds on Sample Size in Structural Equation Modeling." *Electronic Commerce Research and Applications* 9 (6). https://doi.org/10.1016/j.elerap.2010.07.003.

Whittaker, Tiffany A., and Randall E. Schumacker. 2022. *A Beginner's Guide to Structural Equation Modeling.* Fifth Edition. New York: Routledge Taylor & Francis Group.

Widodo, Estu. 2018. "Some Notes on the Contemporary Views of Validity in Psychological and Educational Assessment." *Advances in Social Science, Education and Humanities Research* 231: 732–34. https://doi.org/10.3968/8877.

Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, et al. 2016. "The FAIR Guiding Principles for Scientific Data Management and Stewardship." *Scientific Data* 3 (1): 160018. https://doi.org/10.1038/sdata.2016.18.

Xia, Yan, and Yanyun Yang. 2018. "The Influence of Number of Categories and Threshold Values on Fit Indices in Structural Equation Modeling with Ordered Categorical Data." *Multivariate Behavioral Research* 53 (5): 731–55. https://doi.org/10.1080/00273171.2018.1480346.

———. 2019. "RMSEA, CFI, and TLI in Structural Equation Modeling with Ordered Categorical Data: The Story They Tell Depends on the Estimation Methods." *Behavior Research Methods* 51 (1): 409–28. https://doi.org/10.3758/s13428-018-1055-2.

Yang, Yanyun, and Xinya Liang. 2013. "Confirmatory Factor Analysis Under Violations of Distributional and Structural Assumptions." *Int. J. Quantitative Research in Education* 1 (1): 61–84.

Yang-Wallentin, Fan, Karl G. Jöreskog, and Hao Luo. 2010. "Confirmatory Factor Analysis of Ordinal Variables with Misspecified Models." *Structural Equation Modeling* 17 (3): 392–423. https://doi.org/10.1080/10705511.2010.489003.

Yuan, Ke Hai, Wai Chan, George A. Marcoulides, and Peter M. Bentler. 2016. "Assessing Structural Equation Models by Equivalence Testing With Adjusted Fit Indexes." *Structural Equation Modeling* 23 (3): 319–30. https://doi.org/10.1080/10705511.2015.1065414.

Zhang, Zhiyong, and Ke-Hai Yuan. 2018. *Practical Statistical Power Analysis Using Webpower and R.* Indiana: ISDSA Press. https://doi.org/10.35566/power.