Journal of Modern Applied Statistical Methods
https://jmasm.com

Article

Cut-Off Points for RMSEA and SRMR in Structural Equation Modeling Using ULS and RULS

Francisco Pablo Holgado-Tello [1], Julia Sánchez-García [2,*], José Mena Raposo [3] and Juan C. Suárez-Falcón [1]

1 Department of Methodology of Behavioral Science, Universidad Nacional de Educación a Distancia (UNED), 28040 Madrid, Spain
2 Zaragoza University Centre of Defense, University of Zaragoza, 50090 Zaragoza, Spain
3 Department of Experimental Psychology, University of Seville, 41018 Sevilla, Spain
* Correspondence: juliasanchezg@unizar.es

How To Cite: Holgado-Tello, F.P.; Sánchez-García, J.; Raposo, J.M.; et al. Cut-Off Points for RMSEA and SRMR in Structural Equation Modeling Using ULS and RULS. *Journal of Modern Applied Statistical Methods* **2026**, *25*(1), 1. https://doi.org/10.53941/jmasm.2026.100001

Abstract: The use of Likert scales in the field of social research is becoming increasingly common; therefore, it is necessary to investigate which is the most appropriate methodology to carry out the analysis of the ordinal data obtained. If they are ordinal, they should be treated as such, however, they are frequently analyzed considering them as continuous variables. One of the most widely used techniques to obtain construct validity evidence based on the internal structure within nomological measurement models is Confirmatory Factor Analysis (CFA). Using simulation studies in which four factors were manipulated (number of factors, number of items response categories, skewness and sample size) our objective is twofold: (1) to examine, under ordinal measurement, the Type I error rate and statistical power associated with common global fit indices, specifically the Root Mean Square Error of Approximation (RMSEA) and the Standardized Root Mean Square Residual (SRMR) when computed under ULS and robust ULS (RULS) estimation; and (2) to evaluate RMSEA and SRMR cut-off values using Receiver Operating Characteristic (ROC) analysis. It is found that, depending on the estimation method chosen, the Type I error and power vary, as well as the values reported by RMSEA and SRMR. RULS seems to obtain better results regardless of experimental factors manipulated. Finally, it is found that it would be convenient to review the cut-off points for these global fit indices recommended by the literature.

Keywords: ULS; RULS; RMSEA; SRMR; Type I error rate; statistical power; ROC

1. Introduction

Measurement is one of the most relevant aspects of science, and validity is a quality closely related to this. Measurement can be understood as the systematic process of assigning numbers or other categories to the characteristics of people, objects, or phenomena according to explicit and consistent rules [1,2]. In other words, measurement goes beyond simply assigning values to units of observation. According to [3], measurement theory is a branch of applied statistics that seeks to describe, categorize, and evaluate the quality of the measurement, improve the usefulness and meaning of measurements, and propose methods for developing new and better measurement instruments.

In the context of latent variables, such as psychological constructs, there is sometimes confusion about the purpose of validity. Validity does not refer to the property of a measurement, it is concerned with the interpretations of the scores for the intended use [4]. The 5th edition of the Standards for Educational and Psychological Testing states that "Validity refers to the degree to which evidence and theory support interpretations of test scores for the proposed uses of a test" [5] (p. 11). Inadequate understanding of the theoretical concept or poor measurement of constructs can threaten construct validity. Construct validity, therefore, requires empirical evidence of hypothesized relationships between constructs and their observable manifestations, always grounded in a coherent theoretical framework [6].

Confirmatory Factor Analysis (CFA) is an advanced statistical technique used to assess the factor structure of a set of observed variables. It enables researchers to test the hypothesis that there is a relationship between the observed variables and their underlying latent constructs [7]. Unlike Exploratory Factor Analysis (EFA), which identifies the structure of the data without a predefined model, CFA is based on theories and/or previous analytical research to confirm the validity of specific constructs [8–10].

In recent years, the literature has highlighted both the strengths and limitations of CFA. For instance, Refs. [11,12] have noted problems related to the routine use of fit indices and cutoff criteria, while Ref. [13] emphasized the importance of considering the robustness of CFA in non-normal data contexts. Additionally, various authors point to the need to take into account the characteristics of the variables and the robustness of the estimators. Ref. [14] discuss the issue of treating ordinal items as continuous, whereas Ref. [15] analyze the impact of sample size and data distribution on fit indices. Overall, this literature highlights the need for a cautious and contextualized interpretation of CFA results.

In the context of CFA, a measurement model specifies how observable indicators reflect underlying latent constructs. This model is formulated by comparing the observed variance-covariance matrix with the variance-covariance matrix implied by the proposed model. It is important to note that the implied matrix does not contain the parameters directly but rather reflects the predicted relationships between the indicators according to the model equations once the parameters have been estimated [16,17].

The choice of the parameter estimation method directly affects how well the theoretical model and of the measured construct fit the empirical data. The most commonly used method is Maximum Likelihood (ML) due to its high probability of convergence and its asymptotic normal distribution [18]. ML, however, requires that the observed variables be continuous and multivariate normally distributed. Therefore, ML is not appropriate for ordinal variables, as treating them as continuous can result in biased parameters estimates, particularly when the number of categories is small or the category values are asymmetrically distributed [12].

When Likert scales have five or more points, most traditional parametric methods that assume normally distributed and continuous latent variables can be applied. In these cases, Likert scales can be treated as continuous if certain distributional conditions are met, such as symmetry or balanced response distributions [19]. Nevertheless, a normal or continuous latent variable should not be automatically assumed, even with more than five response points; distributional checks are necessary. For scales with fewer than five categories, traditional methods like ML are generally not robust due to violations of assumptions, including non-normality, lack of continuity, and skewness [14].

In psychology, measurement instruments such as Likert-type scales are commonly used, but their items do not always produce distributions that can be treated as interval scales [13]. Additionally, multivariate normality is rarely met, often due to skewness in the distribution of item responses [20].

In such cases, it is preferable to use polychoric correlations [21] to obtain the working matrices. These matrices are generally more robust to the violation of the multivariate normality assumption and allow the use of estimation methods better suited for ordinal data [22,23]. Robust methods can estimate model fit more accurately by employing an asymptotic covariance matrix [24]. Some of the most employed methods in these cases are Weighted Least Squares (WLS) or Generalized Least Squares [25]. While GLS assumes normal theory, WLS is asymptotically distribution free, which is advantageous for ordinal scales with non-normal distributions. As a result, WLS has become the most frequently used estimation method under these conditions [8,26,27]. However, WLS has limitations, including the need for adequate sample size, potential non-positive definite matrices, and imprecise solutions [28].

Selecting the appropriate estimation method is, therefore, crucial for accurately evaluating model fit. This paper focuses on the Unweighted Least Squares (ULS), and its robust correction, Robust Unweighted Least Squares (RULS). Although ML was initially considered, it was excluded from detailed reporting to maintain the focus on ULS and RULS, which are recommended for ordinal data and polychoric correlations. Briefly, supplementary analyses (not shown) indicated that, under ML, RMSEA often fell below 0.08 when skewness was severe, potentially leading to poorly specified models being accepted. SRMR frequently exceeded 0.05 in small samples (inflating Type I error). Type I error also increased as the number of response categories decreased, remaining above 0.05 in all conditions, a pattern consistent with the known behavior of normal-theory ML when its assumptions are violated in highly non-normal ordinal data with few categories.

ULS does not assume any specific distribution of the variables and is recommended for ordinal data [29]. Despite being less efficient, it is valued for its simplicity and ease of interpretation [26]. ULS is applicable in situations where other estimation methods are not feasible, such as when some variables are linearly dependent in the data matrix [30]. ULS overcomes problems like non-positive definite matrices or imprecise solutions and remains robust when data are not multivariate normal [31]. RULS adds advantages for complex models or when the minimum required sample size is not met, using the asymptotic covariance (AC) matrix in its calculations [32].

Both methods are recommended for categorical variables based on polychoric correlation matrices [33,34]. They reduced the influence of extreme observations, preserving Type I error rates within nominal values, and maintained adequate statistical power [32]. However, few simulation studies have evaluated the fit indices obtained under ULS and RULS, examining Type I error, statistical power, and the adequacy of recommended cut-off point using ROC analysis, despite being addressed in several studies [28,35].

Once an estimation method has been applied, the goodness-of-fit indices help determine whether the hypothesized model is well suited to the study variables and the results obtained. Each index interpreted relative to a threshold of values that determine whether the proposed model is valid [36]. The χ^2 measure is one of the most prominent for providing a test of statistical significance [26,37]. In Confirmatory Factor Analysis (CFA), the model Chi-square (χ^2) is a test statistic that applied researchers frequently interpret as a measure of goodness-of-fit. Larger values suggest a poorer fit, while smaller values indicate a better fit [26]. The Chi-square statistic, however, is sensitive to sample size when the model is misspecified. In large samples, a significant result indicates that the model is imperfect, but this does not necessarily mean the discrepancies are trivial [38]. A lack of fit test may signal substantial model misspecification regardless of sample size [39]. In such cases, it is highly advisable to inspect the residuals rather than attributing the statistical significance of the Chi-square test solely to large samples. [38] also notes that sampled cases are assumed to be causally homogeneous. This assumption weakens as sample size increases, since some cases are more likely to be embedded in different causal processes. As a result, models in large samples are prone to misspecifications, and significant χ^2 values do not rule out strong discrepancies. Researchers should therefore examine potential misspecifications carefully through residuals, regardless of sample size.

χ^2 assumes that the data follow a multivariate normal distribution. If this assumption is not met, the validity of χ^2 as a fit index may be compromised [40]. To mitigate these limitations, it is recommended to use χ^2 alongside other fit indices, such as RMSEA and SRMR, which provide a more robust assessment of model fit [8,12,41,42].

RMSEA estimates the degree of model misfit in the population based on degrees of freedom [43], reflecting the discrepancy due to approximation by degrees of freedom [40]. Values close to 0.05 are interpreted as a good fit [44] and values near 0.08 are considered acceptable [45–47]. These thresholds are approximate heuristics rather than absolute rules, as interpretation may vary depending on the context and complexity of the model [40,48]. RMSEA is sensitive to the number of model parameters. Ref. [49] reported a loss of power with higher numbers of data categories. Models with more degrees of freedom tend to produce lower RMSEA values, while simpler models may show higher values even when they fit reasonably well [48,50]. Therefore, RMSEA is more informative when interpreted in conjunction with other fit indices. A critical aspect of using RMSEA concerns which chi-square statistic should be used for its calculation: scaled or unscaled. Ref. [27] caution that RMSEA derived from scaled chi-square may underestimate the misfit. Similarly, Ref. [42] emphasize that RMSEA is sensitive to the estimation method and distribution, supporting the recommendation to report multiple fit indices rather than relying solely on RMSEA.

SRMR represents the average of the standardized discrepancies between the observed correlations and the correlations predicted by the model [47]. A perfect fit corresponds to 0 [51], values below 0.05 indicate good fit [52] and values below 0.08 are considered acceptable [46,47]. However, as with RMSEA, these cut-offs are rough heuristics rather than strict thresholds, as their applicability depends on sample size, model complexity, and other conditions [49]. With sufficiently large samples, SRMR is expected to be robust with respect to the selection of the estimation method [53]. However, empirical support for this is limited.

Global fit indices such as SRMR or RMSEA are influenced by multiple interacting factors, including sample size, data normality, model parsimony, estimation method, amount and type of misspecification, and characteristics of the data. Therefore, proposed cut-off values, mainly from [47,54], should be considered heuristic guidelines rather than strict rules, informing decisions regarding model fit.

Overall goodness-of-fit indices assess the validity of a model and its ability to represent the observed data. Determining whether a model is correctly specified relies on cut-off values. However, established cut-off values may sometimes classify a well-specified model as misspecifed, or vice versa, leading to Type I or Type II errors. To address this, the present study examines Type I error and statistical power, and evaluates the cut-off points of RMSEA and SRMR using Receiver Operating Characteristic (ROC) analysis, which considers all possible cut-off values and provides a comprehensive representation of diagnostic accuracy.

To address these aims, the Method section describes the simulation study design, and the Data Analysis section details how Type I error and power are computed and how ROC analysis is used to evaluate RMSEA and SRMR across cut-off values, under both ULS and robust ULS (RULS) estimation. The Results section summarizes the main patterns in Type I error and power and reports the ROC-curve findings for both indices under each estimation method. Finally, the Conclusions section discusses the implications of these results and provides recommendations for interpreting and selecting cut-off values.

2. Method

2.1. Simulation Study

Simulation studies are the main tool for the researcher to exhaustively study the fit of different models under different study conditions [11,55]. To obtain the simulated data, different scenarios were used in which four conditions were manipulated: (a) number of factors or latent variables; (b) number of response categories; (c) degree of asymmetry; and (d) sample size.

The number of factors had five levels (2, 3, 4, 5 and 6 factors). Each factor was simulated by 3 items, which is the minimum recommended for identification and numerical stability in CFA with categorical indicators and freely estimated loadings. This decision does not establish construct validity, which depends on theoretical and empirical evidence in context [7,56]. The factor loadings were 0.9, 0.8, and 0.7 for the first, second, and third items of each factor. The items followed a normal distribution $N(0, 1)$.

The number of response categories had four levels (3, 4, 5, and 6 response categories in a Likert scale). For the degree of asymmetry (0 = asymmetric; 1 = moderate and 2 = severe), the scales were categorized so that (a) the distribution of responses was symmetric; (b) all items had moderate asymmetry; or (c) the items had severe asymmetry. Thresholds defining the skewness conditions were calculated according to [57]. The sample size had six levels (100, 150, 250, 450, 650, and 850), since, according to [58], simulation studies usually contain samples of between 100 and 1000 subjects.

By crossing levels of the manipulated factors, a total of 360 experimental conditions were obtained. For each condition, 500 replications were conducted. Data sets were generated using R and PRELIS according to the specification of each experimental condition. For each generated data set, two alternative models were tested using CFA with the ULS and RULS methods in LISREL: (1) correctly specified models, in which the items loaded according to the theoretical model used to generate the data and; (2) misspecified models, in which one item from each theoretical factor was loaded on a different factor.

Polychoric correlation matrices and asymptotic covariance matrices were stored for each replication, and CFA was then performed. For ordinal CFA, models were estimated based on polychoric correlation matrices. The asymptotic covariance matrix of the sample statistics was computed to ensure robustness, yielding robust (scaled) test statistics and standard errors. The RMSEA value was derived directly from the robust (scaled) model chi-square statistic.

In the CFA for correctly specified models, the LISREL syntax corresponded to the data generating model. For example, if the data were generated from a three-factor model, the syntax specified a three-factor model in which each item loaded on its theoretical factor. In contrast, for misspecified models, the LISREL syntax did not correspond to the generating model. For instance, when the data were generated from a two-factor model, one item from each factor was incorrectly loaded on a different factor. The reassigned items were those with higher loading (0.9), whereas the uniquenesses (error variances) were not adjusted.

A total of 720,000 outputs (360,000 for ULS and 360,000 for RULS) were obtained, of which 36 were missing values (NA) and were discarded from the analyses.

2.2. Data Analysis

2.2.1. Type I Error and Power

To determine whether the proposed model is correct or incorrect based on the adequacy of the data to the theoretical model, the values obtained from the fit indices are examined. Four possible outcomes may occur. If the fit index recommends accepting the tested model and the model from which the data were generated is correct, this outcome is classified as a True Positive (TP). Conversely, if the tested model is incorrect but the fit index suggests a good fit (i.e., the model is accepted), the result is a False Positive (FP). Similarly, a True Negative (TN) occurs when the fit index indicates poor fit and the tested model is indeed incorrect. Finally, a False Negative (FN) arises when the fit indices indicate poor fit even though the model tested is correctly specified (Fuentes, 2013).

In general, if the null hypothesis is rejected when it is true, a Type I error is made. The probability of committing a Type I error is α. In this case, it coincides with the false positive rate (FPR). When the null hypothesis is false and it is not rejected, a Type II error is committed. The probability of committing a Type II error is β and the probability of rejecting the null hypothesis when it is false is equal to 1-β. This value is the statistical power of the test and will coincide with the true positive rate (TPR) in the study [59].

For the analysis of Type I error, the proportion of cases classified as "incorrect" was calculated, i.e., cases in which the fit indices recommended rejecting the null hypothesis in correctly specified models. For statistical power, the proportion of cases classified as "incorrect" was analyzed, i.e., cases in which the fit indices recommended rejecting the null hypothesis in misspecified models. All calculations were performed for each manipulated factor (factors, categories, symmetry, and sample size), as well as for the aggregated data, based on the values provided by the goodness-of-fit indices for the ULS and RULS estimation methods. The decision thresholds used were 0.05 for RMSEA and SRMR and 0.08 for RMSEA.

2.2.2. ROC Curves

The ROC curve is used to assess the diagnostic accuracy of a model. Accuracy refers to (1) sensitivity or fraction of TPR, defined as the probability of correctly accepting a well-specified model; and (2) specificity or fraction of true negative rate (TNR), defined as the probability of correctly rejecting a poorly specified model. The optimal value is reached when both percentages are 100%, corresponding to the point (1, 1) in Cartesian coordinates.

The ROC curve is obtained by plotting the TPR (sensitivity) on the y-axis and the FPR (1-specificity) on the x-axis for each possible cut-off value. The ROC curve is necessarily non-decreasing, reflecting the trade-off between sensitivity and specificity: increasing sensitivity can only occur at the expense of decreasing specificity. Each prediction result or confusion matrix instance represents a point in the ROC space [60]. If the analysis cannot discriminate between groups, the ROC curve lies along the diagonal from the lower-left to the upper right vertex. As the curve moves towards the upper-left vertex accuracy increases; the optimal point is (0, 1), where there are no false positives and the true positive rate is 100%.

In the study, ROC curves were constructed and analyzed using nonparametric methods. All possible threshold values of the fit indices were generated for each study case, and each of them determines a point in the ROC space.

Finally, to find the optimal value for each of the indices according to the experimental conditions and their levels, we proceed to calculate the Euclidean distance (*d-index*) in the ROC plane between sensitivity and specificity. The objective of the d-index function is to measure the minimum distance between the pair [1−specificity(d), sensitivity(d)] and the optimal point in the analysis of the ROC curves [61] for each value of the goodness-of-fit indices. The operating point is called the closest point to (0, 1), and is obtained by applying the Euclidean distance as follows:

$$d = \sqrt{(1 - Sensitivity(d))^2 + (1 - Specificity(d))^2} \quad (1)$$

For the aggregated data, the Area Under the Curve (AUC) was calculated as a global measure of index quality since it does not depend on a specific threshold and summarizes results across all possible values [62,63]. The threshold that minimizes the distance to the optimal sensitivity-specificity point was then identified. This distance was calculated, and the corresponding ROC curves were plotted for analysis.

3. Results

3.1. RMSEA and SRMR

We evaluated whether RMSEA and SRMR behave as expected under ULS and robust ULS (RULS) by computing Type I error and power at each level of the experimental factors. The mean and standard deviation (SD) of the values obtained for both fit indices in each estimation method are presented for the levels of the experimental factors. In this context, "acceptable fit" was operationalized using the decision thresholds RMSEA/SRMR < 0.05 (and, for RMSEA, also < 0.08).

3.1.1. Number of Factors

Table 1 summarizes the effect of the number of factors on RMSEA and SRMR under ULS and RULS. Under ULS, RMSEA shows elevated Type I error at the 0.05 cut-off that increases slightly with model complexity (from 0.591 with two factors to 0.658 with six factors), whereas at the 0.08 cut-off it remains broadly stable (≈0.36 across conditions). Power, in turn, increases for both 0.05 and 0.08 as the number of factors increases. SRMR Type I error remains comparatively high and tends to increase with additional factors. Mean values are consistent with these patterns: for misspecified models, RMSEA and SRMR means exceed 0.05 across methods, whereas for correctly specified models values are more consistently below 0.05 under RULS than under ULS.

Table 1. Number of factors for RMSEA and SRMR in ULS and RULS.

Index	Number of factors	Threshold	Type I Error Rate ULS[RULS]	Power ULS[RULS]	Mean (SD)	
					Misspecified Models	Correctly Specified Models
RMSEA	2	0.05	0.591[0.159]	0.995[0.903]	0.162(0.049)	0.078(0.071)
		0.08	0.376[0.084]	0.983[0.669]	[0.094(0.036)]	[0.022(0.03)]
	3	0.05	0.631[0.048]	1.00[1.00]	0.188(0.034)	0.075(0.058)
		0.08	0.355[0.004]	1.00[0.910]	[0.117(0.029)]	[0.013(0.018)]
	4	0.05	0.657[0.026]	1.00[1.00]	0.159(0.030)	0.077(0.056)
		0.08	0.362[0.001]	1.00[0.732]	[0.096(0.023)]	[0.011(0.015)]
	5	0.05	0.665[0.023]	1.00[1.00]	0.151(0.027)	0.079(0.056)
		0.08	0.365[0.001]	1.00[0.662]	[0.091(0.020)]	[0.010(0.014)]
	6	0.05	0.658[0.000]	1.00[1.00]	0.141(0.028)	0.078(0.056)
		0.08	0.359[0.000]	1.00[0.960]	[0.100(0.011)]	[0.008(0.010)]
SRMR	2	0.05	0.261[0.193]	0.951[0.938]	0.080(0.033) [0.080(0.033)]	0.046(0.041) [0.042(0.038)]
	3	0.05	0.329[0.256]	1.00[1.00]	0.101(0.018) [0.099(0.016)]	0.044(0.021) [0.040(0.017)]
	4	0.05	0.383[0.313]	1.00[1.00]	0.097(0.016) [0.095(0.014)]	0.048(0.023) [0.043(0.018)]
	5	0.05	0.415[0.326]	1.00[1.00]	0.098(0.017) [0.097(0.016)]	0.050(0.024) [0.044(0.018)]
	6	0.05	0.440[0.142]	1.00[1.00]	0.090(0.018) [0.079(0.008)]	0.052(0.026) [0.035(0.012)]

Note. In brackets results for RULS.

Using RULS, Type I error for RMSEA decreases as the number of factors increases, both for the 0.05 and 0.08 threshold. The power for the 0.05 value increases as the number of factors increases, while there is variability for the 0.08 value. Under RULS, SRMR Type I error increases as the number of factors increases, except in the six-factor condition where it decreases, whereas power increases across factor levels; overall, mean RMSEA and SRMR values are largely stable, with the lowest values observed for six factors.

For incorrectly specified models, the mean RMSEA and SRMR values for both ULS and RULS consistently exceed 0.05. However, as the number of factors increases, the mean RMSEA values for ULS decrease, from 0.162 for two factors to 0.141 for six factors, whereas, SRMR shows no clear variation. For correctly specified models, the mean values are below 0.05 with SRMR, except for 6 factors; under RULS, both SRMR and RMSEA are below 0.05 and systematically lower than under ULS.

The complexity of the model could be related to the increase in Type I error. Thus, the behavior of these indices is consistent with previous findings.

3.1.2. Number of Categories

Table 2 summarizes the effect of the number of response categories on RMSEA and SRMR under ULS and RULS. Under ULS, RMSEA shows higher Type I error and power at the 0.05 cut-off than at 0.08, and Type I error decreases as the number of categories increases (e.g., 0.771 at 3 categories vs. 0.512 at 6 categories for the 0.05 cut-off; 0.475 vs. 0.258 for the 0.08 cut-off). Under RULS, RMSEA Type I error performs well and remains low—especially at the 0.08 cut-off—and power stays near ceiling; SRMR shows comparatively high Type I error but similarly decreases as categories increase. SRMR follows the same general pattern, with Type I error decreasing with additional categories (from 0.413 to 0.278), while power remains largely stable.

Mean values are consistent with these trends: misspecified models yield means above 0.05, whereas correctly specified models are more consistently below 0.05 under RULS than under ULS. Across methods, increasing the number of categories tends to reduce Type I error, with the reduction being most pronounced under ULS.

Table 2. Number of categories for RMSEA and SRMR in ULS and RULS.

Index	Categories	Threshold	Type I Error Rate ULS[RULS]	Power ULS[RULS]	Mean (sd)	
					Misspecified Models	Correctly Specified Models
RMSEA	3	0.05	0.771[0.049]	0.999[0.953]	0.168(0.044)	0.095(0.069)
		0.08	0.475[0.006]	0.996[0.644]	[0.089(0.025)]	[0.012(0.018)]
	4	0.05	0.664[0.103]	0.999[0.976]	0.161(0.038)	0.084(0.062)
		0.08	0.388[0.065]	0.997[0.765]	[0.099(0.027)]	[0.018(0.029)]
	5	0.05	0.571[0.045]	0.999[0.975]	0.156(0.032)	0.069(0.052)
		0.08	0.303[0.007]	0.996[0.794]	[0.103(0.029)]	[0.012(0.018)]
	6	0.05	0.512[0.042]	0.998[0.973]	0.153(0.032)	0.061(0.048)
		0.08	0.258[0.006]	0.995[0.820]	[0.106(0.030)]	[0.012(0.018)]
SRMR	3	0.05	0.413[0.304]	0.988[0.985]	0.095(0.023) [0.092(0.020)]	0.051(0.025) [0.044(0.019)]
	4	0.05	0.397[0.312]	0.990[0.989]	0.098(0.029) [0.098(0.029)]	0.054(0.037) [0.049(0.036)]
	5	0.05	0.322[0.231]	0.988[0.983]	0.090(0.019) [0.089(0.019)]	0.045(0.024) [0.038(0.017)]
	6	0.05	0.278[0.206]	0.987[0.983]	0.089(0.018) [0.089(0.019)]	0.041(0.020) [0.012(0.019)]

Note. in brackets results for RULS.

3.1.3. Skewness

Table 3 summarizes the impact of skewness on RMSEA and SRMR under ULS and RULS. For both estimation methods, Type I error increases as skewness becomes more pronounced (e.g., RMSEA under ULS rises from 0.304 at skewness = 0 to 0.938 at skewness 2; SRMR from 0.200 to 0.507), while power remains near ceiling at the 0.05 cut-off. In contrast, at the 0.08 RMSEA cut-off, power decreases as skewness increases (from 0.957 to 0.425 under ULS).

Overall, RULS produces substantially lower Type I error for RMSEA than ULS across skewness levels and maintains the mean values for correctly specified models below 0.05, whereas under ULS the mean values for correctly specified models exceed 0.05 under moderate and severe skewness (e.g., RMSEA 0.074 and 0.117; SRMR 0.061 at skewness 2).

Table 3. Degree of skewness for RMSEA and SRMR in ULS and RULS.

Index	Skewness	Threshold	Type I Error Rate ULS[RULS]	Power ULS[RULS]	Mean (SD)	
					Misspecified Models	Correctly Specified Models
RMSEA	0	0.05	0.304[0.044]	0.998[0.991]	0.143(0.025)	0.041(0.029)
		0.08	0.106[0.007]	0.993[0.957]	[0.117(0.025)]	[0.012(0.018)]
	1	0.05	0.648[0.085]	0.988[0.976]	0.156(0.030)	0.074(0.042)
		0.08	0.356[0.038]	0.996[0.827]	[0.098(0.023)]	[0.016(0.026)]
	2	0.05	0.938[0.051]	0.996[0.934]	0.178(0.046)	0.117(0.072)
		0.08	0.610[0.022]	0.998[0.425]	[0.079(0.022)]	[0.013(0.020)]
SRMR	0	0.05	0.200[0.174]	0.985[0.983]	0.085(0.015) [0.085(0.015)]	0.036(0.016) [0.034(0.015)]
	1	0.05	0.350[0.339]	0.988[0.985]	0.092(0.022) [0.093(0.023)]	0.047(0.027) [0.047(0.029)]
	2	0.05	0.507[0.304]	0.990[0.988]	0.101(0.026) [0.100(0.026)]	0.061(0.032) [0.047(0.025)]

Note. in brackets results for RULS.

3.1.4. Sample Size

Table 4 summarizes the effect of sample size on RMSEA and SRMR under ULS and RULS. Under ULS, power remains near ceiling across sample sizes but Type I error is strongly inflated in small samples and decreases as n increases (e.g., RMSEA Type I error drops from 0.888 at n = 100 to 0.339 at n = 850 at the 0.05 cut-off; at the 0.08 cut-off it decreases to 0.041 by n = 850). Mean values for misspecified models also decline with larger samples (e.g., RMSEA from 0.187 at n = 100 to 0.145 at n = 850), while correctly specified models approach values below 0.05 only at moderate to large n (RMSEA ≈ 0.059 by n = 450; SRMR ≈ 0.051 by n = 250). Under RULS, RMSEA shows substantially lower Type I error and stable mean values across sample sizes, with correctly specified models consistently below 0.05 and misspecified models above 0.05; SRMR shows a similar pattern, although its Type I error remains comparatively higher than RMSEA in small samples. Overall, increasing sample size primarily improves Type I error control, particularly for RMSEA at the 0.08 cut-off.

Table 4. Sample size for RMSEA and SRMR in ULS and RULS.

Index	Sample	Threshold	Type I Error Rate ULS[RULS]	Power ULS[RULS]	Mean (SD)	
					Misspecified Models	Correctly Specified Models
RMSEA	100	0.05 0.08	0.888[0.175] 0.739[0.037]	0.995[0.917] 0.998[0.732]	0.187(0.058) [0.100(0.038)]	0.133(0.087) [0.022(0.028)]
	150	0.05 0.08	0.840[0.125] 0.649[0.033]	0.998[0.944] 0.992[0.734]	0.171(0.040) [0.098(0.032)]	0.106(0.058) [0.020(0.027)]
	250	0.05 0.08	0.744[0.068] 0.455[0.033]	0.999[0.974] 0.997[0.755]	0.158(0.030) [0.099(0.028)]	0.079(0.042) [0.016(0.024)]
	450	0.05 0.08	0.568[0.044] 0.221[0.032]	0.999[0.987] 0.999[0.776]	0.151(0.025) [0.100(0.025)]	0.059(0.035) [0.013(0.022)]
	650	0.05 0.08	0.422[0.002] 0.059[0.000]	1.00[0.991] 0.999[0.769]	0.147(0.022) [0.100(0.024)]	0.046(0.023) [0.008(0.011)]
	850	0.05 0.08	0.339[0.000] 0.041[0.000]	1.00[0.994] 0.999[0.773]	0.145(0.021) [0.100(0.024)]	0.043(0.029) [0.007(0.009)]
SRMR	100	0.05	0.883[0.806]	0.990[0.987]	0.112(0.025) [0.106(0.024)]	0.075(0.022) [0.064(0.016)]
	150	0.05	0.720[0.620]	0.987[0.986]	0.104(0.024) [0.103(0.025)]	0.065(0.026) [0.058(0.022)]
	250	0.05	0.404[0.336]	0.990[0.984]	0.094(0.021) [0.094(0.023)]	0.051(0.025) [0.050(0.027)]
	450	0.05	0.093[0.083]	0.987[0.985]	0.087(0.020) [0.088(0.021)]	0.039(0.025) [0.039(0.026)]
	650	0.05	0.002[0.002]	0.985[0.984]	0.082(0.012) [0.082(0.013)]	0.029(0.008) [0.028(0.008)]
	850	0.05	0.048[0.000]	0.987[0.985]	0.080(0.012) [0.081(0.012)]	0.029(0.019) [0.024(0.007)]

Note. in brackets results for RULS.

3.1.5. Comparison of RMSEA and SRMR between Estimation Methods for Aggregated Data

Table 5 compares ULS and RULS using aggregated results. Overall, Type I error is higher under ULS than under RULS, particularly for RMSEA at the 0.05 cut-off (ULS 0.630 vs. RULS 0.060). The power for both estimation methods and goodness-of-fit indices is high, although slightly lower with RMSEA in RULS when the 0.08 cutoff point is taken into consideration. For misspecified models, mean RMSEA and SRMR values exceed 0.05 (and RMSEA exceeds 0.08), as expected. For correctly specified models, ULS yields an inflated RMSEA mean (0.077), consistent with its elevated Type I error, while RULS produces substantially lower values (RMSEA 0.014; SRMR 0.042–0.048), indicating better control of false rejections.

Table 5. Influence of estimation methods on RMSEA and SRMR for aggregated data.

Index	Threshold	Type I Error Rate ULS[RULS]	Power ULS[RULS]	Mean ULS[RULS]	
				Misspecified Models	Correctly Specified Models
RMSEA	0.05 0.08	0.630[0.060] 0.364[0.021]	0.999[0.970] 0.996[0.757]	0.160[0.100]	0.077[0.014]
SRMR	0.05	0.366[0.263]	0.990[0.985]	0.093[0.092]	0.048[0.042]

Note. in brackets results for RULS.

3.2. ROC Curve Analysis on RMSEA and SRMR

Figure 1 presents smoothed histograms for the four cases under study. Before conducting a detailed analysis using ROC curves, these plots already provide an idea of how the RMSEA and SRMR indices behave for both ULS and RULS. They also highlight the range within which our optimal cutoff value is likely to be found.

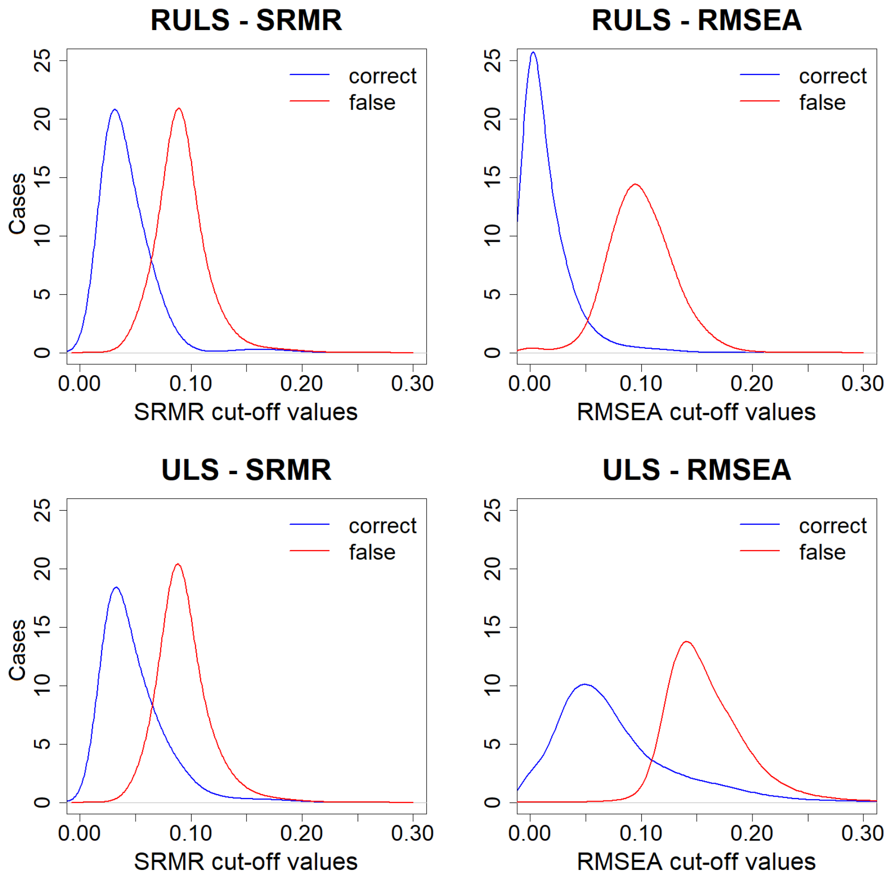

Figure 1. Distribution for ULS and RULS based on cutoff values for SRMR and RMSEA.

Specifically, a cut-off point study was carried out to analyze the sensitivity and specificity of the values reported by RMSEA and SRMR for each estimation method. A critical interpretation of the cut-off points is also made based on the *d*-index. To this end, and for parsimony, aggregate data were used, since analyzing this aspect in each experimental condition would yield an enormous number of results that, although they must be examined, nevertheless exceed the purpose of this work.

Table 6 presents the main results obtained from the ROC curve analysis for both methods in the general case, including both the commonly used values and the optimal ones, obtained by selecting the optimal point *d* that balances sensitivity and specificity when choosing the threshold. This applies to the data obtained in both ULS and RULS.

Table 6. Optimal cut-off points in ULS and RULS for the general case.

Index	Method	AUC	Cut-Off Value	Sensitivity	Specificity	*d* Value
RMSEA	ULS	0.8945	0.1150 *	0.8045	0.9633	0.1989
			0.0500	0.3697	0.9989	0.6303
			0.0800	0.6436	0.9961	0.3564
	RULS	0.9775	0.0545 *	0.9490	0.9615	0.0639
			0.0500	0.9401	0.9697	0.0671
			0.0800	0.9788	0.7575	0.2434
SRMR	ULS	0.9084	0.0715 *	0.8409	0.8766	0.2013
			0.0500	0.6470	0.9880	0.3532
	RULS	0.9534	0.0665 *	0.8950	0.9014	0.1440
			0.0500	0.7373	0.9853	0.2631

Note. * optimal cut-off value.

Figure 2 shows the four ROC curves corresponding to the four analysis conditions in our study. At first glance, it can be observed that the RULS curves have a larger area and a operating point closer to the upper left corner. In contrast, the ULS curves exhibit a smaller area under the curve and an optimal decision point that is farther from the upper left corner. Under this ROC analysis, for ULS the optimal RMSEA cutoff value according to the d-index is 0.115, while it is common to use 0.05 which achieves only a 36.97% of true-positive rate. Regarding RULS, the optimal RMSEA cutoff value is 0.0545, which is coherent with the general recommendations of the literature.

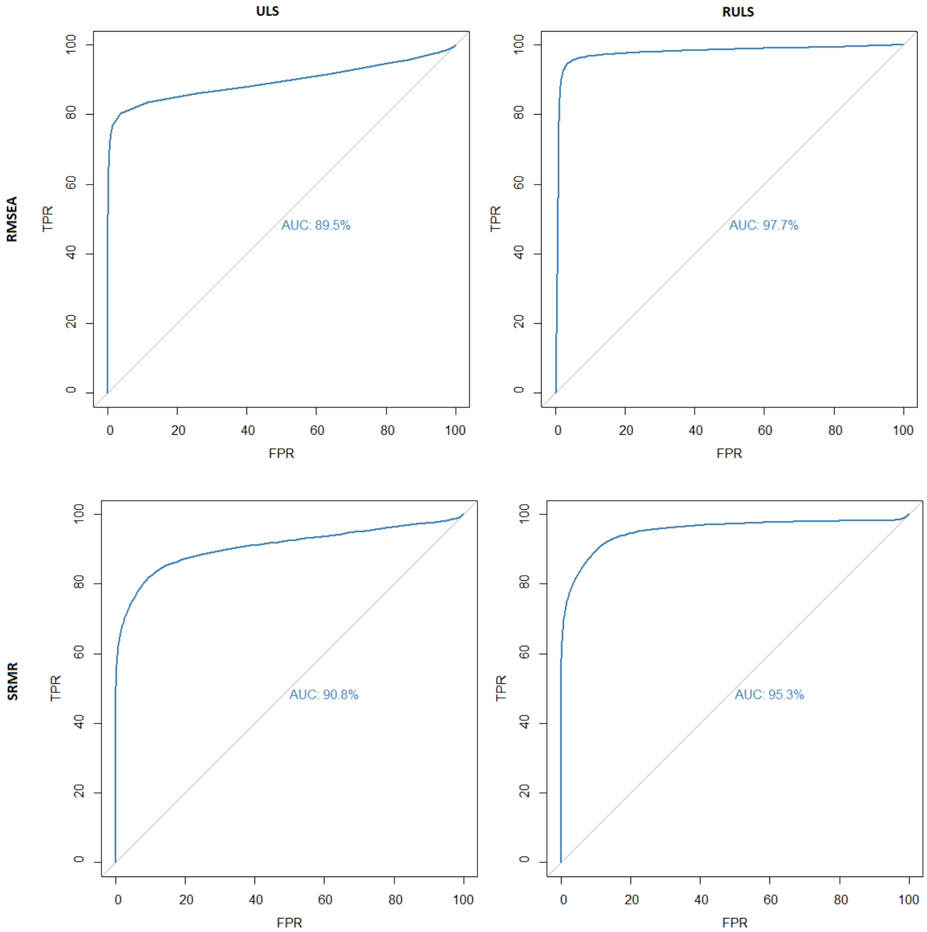

Figure 2. ROC curves for the ULS and RULS methods with the cutoff values for SRMR and RMSEA.

The different pattern is found for SRMR, where the optimal cutoff point is 0.0715 for ULS, achieving a sensitivity of 84.1% and a specificity of 87.7%, while with 0.05 the sensitivity is 64.70%. However, using RULS the optimal SRMR cutoff point should be higher than 0.05 because with this value only 73.73% of true positives would be detected. Our model finds an optimal value of 0.0665, achieving a sensitivity of 89.5% and a specificity of 90.14%. That is, using ULS the optimal cutoff point for RMSEA and SRMR would be different from the values recommended. In general, RULS achieves better results in Type I and Type II errors compared to ULS around its optimal value.

4. Conclusions

The first objective of this study was to examine the influence of several variables (number of factors, categories, skewness and sample size) on the Type I error rate and power of two parameter estimation methods in CFA models, namely ULS and RULS. Specifically, the study tested the ability of both methods to detect correctly specified and misspecified models using two commonly used fit indices in structural equation modelling: RMSEA and SRMR. The second objective was to review the RMSEA and SRMR cut-off points by analyzing ROC curves, given that inappropriate cut-off points may lead to the rejection of correctly specified models or the acceptance of incorrectly specified models.

The results of this research highlight the importance of using estimation methods correctly to ensure that inferences drawn from the results are coherent and consistent. Below, the outcomes of comparing ULS and RULS in terms of Type I error and power, as a function of the manipulated variables, are discussed.

With respect to the number of factors in the theoretical model, ULS may be less suitable because, although it exhibits high power, it also shows high Type I error, leading to the rejection of correctly specified models. RULS, in contrast, demonstrates lower Type I error while maintaining high power, making it the preferable method.

Regarding number of categories, ULS produces a higher proportion of Type I errors than RULS. Although both methods achieve high power, RULS is recommended because ULS tends to reject correctly specified models more frequently. Across the indices, RMSEA is more likely than SRMR to reject incorrectly specified models, consistent with the pattern observed for the number of factors. SRMR, however, generally performs better than RMSEA for both ULS and RULS in detecting correctly specified models, except in the case of six categories with RULS.

When considering the degree of skewness, both ULS and RULS exhibit high Type I error for distributions with moderate and severe skewness for both RMSEA (particularly at the 0.05 cut-off point) and SRMR. The probability of committing a Type I error, however, increases more with ULS than with RULS as skewness rises. The rejection rate of correctly specified models is lower with RULS, especially when using RMSEA. Consistent with [32], RULS is recommended for estimating model parameters when the response distribution is skewed.

Consistent with the previous sections, ULS produces the highest Type I error rates regardless of sample size. RULS maintains low rejection rates of correctly specified models, particularly with RMSEA, and achieves slightly higher power. Following the criteria proposed by [32] regarding the use of estimation methods with polychoric correlations for Likert-type scales, RULS is preferred due to its advantages for parameter estimation, especially when sample size is considered.

Based on the results for Type I error and power, RULS appears to offer more advantages than ULS. This conclusion holds both when considering each factor individually (number of factors, the number of categories, the degree of skewness and sample size) and when all experimental conditions are considered in aggregate.

It is important to note that goodness-of-fit indices do not provide a dichotomous decision regarding the acceptance or rejection of a proposed model. Most indices, including RMSEA and SRMR, evaluate fit on a continuous scale where lower values indicate better fit, as they measure error.

Traditionally, cut-off points for certain fit indices have followed the recommendations of [47,54]. These authors, however, caution that the estimation method, sample size, and data distribution can affect goodness-of-fit values, making a universal cut-off point inappropriate [64]. In the study by [65] the cut-off values recommended by [47] were not confirmed; the authors emphasize that the appropriate cut-off depends on the type and degree of model misspecification one is willing to tolerate. With respect to RMSEA and SRMR, which should accompany interpretations of the chi-squared test, general conclusions for each estimation method could be summarized as follows.

The ROC curve analysis and the optimal cut-off point indicate that the robust version of ULS achieves superior performance in both sensitivity and specificity (i.e., a better trade-off between Type I and Type II errors). By contrast, ULS involves a trade-off: reducing Type I error leads to an increase in Type II error.

Based on the optimal values obtained in the general case, it can be concluded that the values that optimize both Type I and Type II errors differ for RMSEA and SRMR. For RULS, the calculated optimal values are similar to those commonly used, whereas for ULS, they deviate considerably from accepted values in the literature. When using ULS, a higher false positive rate is observed, particularly for RMSEA: the aggregated rate was 0.63, compared to 0.06 with RULS. Interpreted in terms of the ROC curve, a higher critical value (0.115) is required to adequately balance sensitivity and specificity.

The review of the cut-off points provides empirical evidence regarding the behavior of goodness-of-fit indices. RMSEA, in particular, is highly sensitive to very high and very low factor loadings, leading to elevated Type I and Type II error rates, respectively [66–68]. This sensitivity can result in the rejection of well-specified models or the acceptance of poorly specified models. The critical evaluation of RMSEA and SRMR confirms, as noted by [47,54], that universal cut-off points are not justified regardless of data characteristics or estimation method. In this sense, at least, in this research work it has been shown by means of a classic technique (ROC) whose application in this context is novel, that the choice of estimation method must be carefully reviewed jointly with the interpretation of the fit indices based on the measures of error. In any case, one of the most relevant conclusions of this work is that the commonly used RMSEA cutoff point may not be directly applicable when using ULS with ordinal data.

Considering all the results regarding the dependent variables used in this study, i.e., Type I error, power, RMSEA and SRMR, we could venture that when working with ordinal data, RULS seems to have certain advantages over ULS. These advantages are given insofar as RULS, regardless of the number of factors, categories, degree of skewness and sample size, controls reasonably well the probability of rejecting correctly specified models (Type I error) and increases the probability of rejecting incorrectly specified models (power). RMSEA and SRMR values under RULS align more closely with expected cut-offs for both correctly specified and misspecified models. Consequently, global error rates are more robust under RULS, and no alternative cut-off points substantially improve sensitivity and specificity, in contrast with ULS. Furthermore, RULS exhibits higher AUC values for RMSEA and SRMR than ULS (Figure 2).

These findings support the recommendation that RULS should be the preferred estimation method for ordinal Likert-type data. When using ULS, researchers should be aware of its high Type I error rate and interpret RMSEA and SRMR with caution, potentially applying higher thresholds than traditionally recommended.

Finally, this study does not aim to establish new universal cut-offs. Rather, it seeks to empirically quantify, via simulation, the behavior of RMSEA and SRMR with ULS and RULS estimators for ordinal data, providing insight into their properties. ROC curve analysis allows identification of estimator-specific thresholds that balance Type I and Type II errors within this context. These thresholds should be treated as screening tools to be supplemented by local-fit diagnostics and theory-based judgment, offering more nuanced guidance for applied researchers.

5. Limitations

As mentioned above, several factors that could affect the results were not considered in this study, including number of items, reliability, higher-order factors, kurtosis and degree of misspecification, for example. Therefore, the conclusions are only generalizable to the conditions examined in this research. In addition, the results apply specifically to models with the type of misspecification implemented in the simulation. Future research could expand the variety of poorly specified models to assess their influence on the validation of measurement instruments using Likert-type scales. For example, Ref. [64] identified up to 12 types of models according to the degree of misspecification. Furthermore, the use of high and homogeneous factor loadings and a minimum of three items per factor in ordinal CFA represents conditions that are uncommon in real-world applications, highlighting the need for further investigation. Considering all these aspects, there are still many issues to be investigated in future research.

When considering the levels of non-normality in response distribution, both skewness and kurtosis should be considered, as recommended by [69].

Given the striking results in terms of the revision of some cut-off points according to which estimation method, a future line of research could consist of a more in-depth review of the RMSEA and SRMR cut-off points according to the levels of the factors considered in each estimation method studied and not only consider the data in an aggregated manner.

Additionally, the effects of the experimental conditions on fit indices could be formally tested using MANOVA to examine the significance of the main effects and by estimation method.

Finally, one of the future lines of work must necessarily include other estimation methods such as PSL (Partial Least Squares), GLS (Generalized Least Squares), DWLS (Diagonally Weighted Least Squares), as well as other robust variants for ML, DWLS and ULS based on the correction of mean and variance by means of different procedures such as Satterthwaite's (MLMVS, MLMV, MLF, WLSM, WLSMVS, WLSMV, ULSM, ULSMVS, ULSMV). Likewise, the study of the *Satorra-Bentler adjusted* χ^2 would have to be included, as well as other global fit indices such as for example the unbiased SRMR [70], CFI, TLI, GFI and AGFI that are frequently used together.

Author Contributions

Conceptualization: F.P.H.-T. and J.C.S.-F.; Data curation: F.P.H.-T., J.S.-G., J.M.R. and J.C.S.-F.; Formal analysis: F.P.H.-T., J.S.-G., J.M.R. and J.C.S.-F.; Validation: F.P.H.-T. and J.C.S.-F.; Methodology: F.P.H.-T., J.S.-G., J.M.R. and J.C.S.-F.; Software: F.P.H.-T.; Visualisation: F.P.H.-T., J.S.-G., J.M.R. and J.C.S.-F.; Supervision: F.P.H.-T. and J.C.S.-F.; Writing—original draft: F.P.H.-T., J.S.-G., J.M.R. and J.C.S.-F.; Writing—revision and editing: F.P.H.-T., J.S.-G., J.M.R. and J.C.S.-F. All authors have read and agreed to the published version of the manuscript.

Funding

This research received no external funding.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

The data supporting the findings of this study are available from the corresponding author and can be shared upon reasonable request.

Acknowledgments

The authors also gratefully acknowledge the careful work of the reviewers and their contributions, which have significantly improved the paper.

Conflicts of Interest

The authors declare no conflict of interest.

Use of AI and AI-Assisted Technologies

The authors acknowledge the use of ChatGPT (OpenAI, GPT-5, 2025) as a writing assistant. The tool was used exclusively to improve the readability and linguistic accuracy of the manuscript in English, as the authors are non-native speakers. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article. All conceptualization, analyses, and interpretations were entirely the responsibility of the authors.

References

1. DeVellis, R.F.; Thorpe, C.T. *Scale Development: Theory and Applications*, 5th ed.; Sage: Los Angeles, CA, USA, 2021.
2. Stevens, S.S. On the Theory of Scales of Measurement. *Science* **1946**, *103*, 677–680. https://doi.org/10.1126/science.103.2684.677.
3. Allen, M.; Yen, W. *Introduction to Measurement Theory*; Waveland Press: Long Grove, IL, USA, 1979.
4. Clark, L.A.; Watson, D. Constructing Validity: New Developments in Creating Objective Measuring Instruments. *Psychol. Assess.* **2019**, *31*, 1412–1427. https://doi.org/10.1037/pas0000626.
5. American Educational Research Association; American Psychological Association; National Council on Measurement in Education. *Standards for Educational and Psychological Testing*; American Educational Research Association: Washington, DC, USA, 2014.
6. Cronbach, L.J.; Meehl, P.E. Construct Validity in Psychological Tests. *Psychol. Bull.* **1955**, *52*, 281–302. https://doi.org/10.1037/h0040957.
7. Van Bork, R. *Interpreting Psychometrics Models*; Riet Van Bork: Utrecht, The Netherlands, 2019. https://doi.org/10.31237/osf.io/x6a7s.
8. Brown, T.A. *Confirmatory Factor Analysis for Applied Research*; Guilford Press: New York, NY, USA, 2015.
9. Catena, A.; Ramos Álvarez, M.M.; Trujillo, H.M. *Análisis Multivariado. Un Manual Para Investigadores*; Biblioteca Nueva: Madrid, Spain, 2003.
10. Jackson, D.L.; Gillaspy, J.A.; Purc-Stephenson, R. Reporting Practices in Confirmatory Factor Analysis. *Psychol. Methods* **2009**, *14*, 6–23. https://doi.org/10.1037/a0014694.
11. Pavlov, G.; Maydeu-Olivares, A.; Shi, D. Using the Standardized Root Mean Squared Residual (SRMR) to Assess Exact Fit in Structural Equation Models. *Educ. Psychol. Meas.* **2021**, *81*, 110–130. https://doi.org/10.1177/0013164420926231.
12. Xia, Y.; Yang, Y. RMSEA, CFI, and TLI in Structural Equation Modeling with Ordered Categorical Data: The Story They Tell Depends on the Estimation Methods. *Behav. Res. Methods* **2019**, *51*, 409–428. https://doi.org/10.3758/s13428-018-1055-2.
13. Wu, H.; Leung, S.O. Can Likert Scales Be Treated as Interval Scales? A Simulation Study. *J. Soc. Serv. Res.* **2017**, *43*, 527–532. https://doi.org/10.1080/01488376.2017.1329775.
14. Rhemtulla, M.; Brosseau-Liard, P.É.; Savalei, V. When Can Categorical Variables Be Treated as Continuous? A Comparison of Robust Continuous and Categorical SEM Estimation Methods under Suboptimal Conditions. *Psychol. Methods* **2012**, *17*, 354–373. https://doi.org/10.1037/a0029315.
15. Shi, D.; Maydeu-Olivares, A.; Rosseel, Y. Assessing Fit in Ordinal Factor Analysis Models: SRMR vs. RMSEA. *Struct. Equ. Modeling* **2020**, *27*, 1–15. https://doi.org/10.1080/10705511.2019.1611434.
16. Kenny, D.A.; Kashy, D.A. Analysis of the Multitrait-Multimethod Matrix Using Confirmatory Factor Analysis. *Psychol. Bull.* **1992**, *112*, 165–172. https://doi.org/10.1037/0033-2909.112.1.165.
17. Millsap, R.E.; Everson, H. Confirmatory Measurement Model Comparisons Using Latent Means. *Multivar. Behav. Res.* **1991**, *26*, 479–497. https://doi.org/10.1207/s15327906mbr2603_6.

18. Byrne, B.M. *Structural Equation Modeling with LISREL, PRELIS, and SIMPLIS: Basic Concepts, Applications, and Programming*; Lawrence Erlbaum Associates: Mahwah, NJ, USA, 1998.
19. Huh, I.; Gim, J. Exploration of Likert Scale in Terms of Continuous Variable with Parametric Statistical Methods. *BMC Med. Res. Methodol.* **2025**, *25*, 218. https://doi.org/10.1186/s12874-025-02668-1.
20. Jöreskog, K.G.; Moustaki, I. Factor Analysis of Ordinal Variables: A Comparison of Three Approaches. *Multivar. Behav. Res.* **2001**, *36*, 347–387. https://doi.org/10.1207/S15327906347-387.
21. Holgado-Tello, F.P.; Chacón-Moscoso, S.; Barbero-García, I.; et al. Polychoric versus Pearson Correlations in Exploratory and Confirmatory Factor Analysis of Ordinal Variables. *Qual. Quant.* **2010**, *44*, 153–166. https://doi.org/10.1007/s11135-008-9190-y.
22. Coenders, G.; Satorra, A.; Saris, W.E. Alternative Approaches to Structural Modeling of Ordinal Data: A Monte Carlo Study. *Struct. Equ. Modeling* **1997**, *4*, 261–282. https://doi.org/10.1080/10705519709540077.
23. Freiberg, A.; Stover, J.B.; De la Iglesia, G.; et al. Correlaciones Policóricas y Tetracóricas en Estudios Factoriales Exploratorios y Confirmatorios. *Cienc. Psicol.* **2013**, *7*, 151–164.
24. Erceg-Hurn, D.M.; Mirosevich, V.M. Modern Robust Statistical Methods: An Easy Way to Maximize the Accuracy and Power of Your Research. *Am. Psychol.* **2008**, *63*, 591–601. https://doi.org/10.1037/0003-066X.63.7.591.
25. Liu, Q.; Okui, R.; Yoshimura, A. Generalized Least Squares Model Averaging. *Econom. Rev.* **2016**, *35*, 1692–1752. https://doi.org/10.1080/07474938.2015.1092817.
26. Jöreskog, K.G.; Sörbom, D. *LISREL 8 User's Reference Guide*; Scientific Software: Chicago, IL, USA, 1996.
27. Savalei, V. Improving Fit Indices in Structural Equation Modeling with Categorical Data. *Multivar. Behav. Res.* **2021**, *56*, 390–407. https://doi.org/10.1080/00273171.2020.1717922.
28. Li, C.H. Confirmatory Factor Analysis with Ordinal Data: Comparing Robust Maximum Likelihood and Diagonally Weighted Least Squares. *Behav. Res. Methods* **2016**, *48*, 936–949. https://doi.org/10.3758/s13428-015-0619-7.
29. MacCallum, R.C. Factor Analysis. In *The SAGE Handbook of Quantitative Methods in Psychology*; Millsap, R.E., Maydeu-Olivares, A., Eds.; Sage: Los Angeles, CA, USA, 2009; pp. 123–147.
30. Saris, W.E.; Stronkhorst, H. *Causal Modelling in Non-Experimental Research: An Introduction to the LISREL Approach*; Sociometric Research Foundation: Amsterdam, The Netherlands, 1984.
31. Morata-Ramírez, M.; Holgado-Tello, F.P.; Barbero-García, I.; et al. Análisis Factorial Confirmatorio: Recomendaciones Sobre Mínimos Cuadrados No Ponderados en Función del Error Tipo I de Ji-Cuadrado y RMSEA. *Acción Psicol.* **2015**, *12*, 79–90. https://doi.org/10.5944/ap.12.1.14362.
32. Yang-Wallentin, F.; Jöreskog, K.G.; Luo, H. Confirmatory Factor Analysis of Ordinal Variables with Misspecified Models. *Struct. Equ. Modeling* **2010**, *17*, 392–423. https://doi.org/10.1080/10705511.2010.489003.
33. Batista, J.M.; Coenders, G. *Modelos de Ecuaciones Estructurales: Modelos Para el Análisis de Relaciones Causales*; Hespérides: Madrid, Spain, 2000.
34. Schumacker, R.E.; Lomax, R.G. *A Beginner's Guide to Structural Equation Modeling*; Lawrence Erlbaum Associates, Inc.: Mahwah, NJ, USA, 1996.
35. Sass, D.A.; Schmitt, T.A.; Marsh, H.W. Evaluating Model Fit with Ordered Categorical Data within a Measurement Invariance Framework: A Comparison of Estimators. *Struct. Equ. Modeling* **2014**, *21*, 167–180. https://doi.org/10.1080/10705511.2014.882658.
36. Jordan, F.M. Valor de Corte de los Índices de Ajuste en el Análisis Factorial Confirmatorio. *Psocial* **2021**, *7*, 66–71.
37. Powell, D.A.; Shafer, W.D. The Robustness of the Likelihood Ratio Chi-Square Test for Structural Equation Models: A Meta-Analysis. *J. Educ. Behav. Stat.* **2001**, *26*, 105–132. https://doi.org/10.3102/10769986026001105.
38. Hayduk, L.A. Shame for Disrespecting Evidence: The Personal Consequences of Insufficient Respect for Structural Equation Model Testing. *BMC Med. Res. Methodol.* **2014**, *14*, 124. https://doi.org/10.1186/1471-2288-14-124.
39. McIntosh, C.N. Improving the evaluation of model fit in confirmatory factor analysis: A commentary on Gundy, C.M., Fayers, P.M., Groenvold, M., Petersen, M. Aa., Scott, N.W., Sprangers, M.A.J., Velikov, G., Aaronson, N.K. (2011). Comparing higher-order models for the EORTC QLQ-C30. *Qual. Life Res.* **2012**, *21*, 1619–1621. https://doi.org/10.1007/s11136-011-0084-4.
40. Shi, D.; Lee, T.; Maydeu-Olivares, A. Understanding the Model Size Effect on SEM Fit Indices. *Educ. Psychol. Meas.* **2018**, *79*, 310–334. https://doi.org/10.1177/0013164418783530.
41. Holgado-Tello, F.P.; Suárez-Falcón, J.C.; Martínez-Huertas, M.A.; et al. *Modelos de Ecuaciones Estructurales con LISREL y Lavaan*; Sanz y Torres: Madrid, Spain, 2024.
42. Shi, D.; Maydeu-Olivares, A. The Effect of Estimation Methods on SEM Fit Indices. *Educ. Psychol. Meas.* **2020**, *80*, 421–445. https://doi.org/10.1177/0013164419885164.
43. Yin, Y.; Fairchild, A.J.; Shi, D.; et al. Estimating the Root Mean Square Error of Approximation (RMSEA) with Multiply Imputed Data Under Non-Normality. *Struct. Equ. Modeling* **2025**, *32*, 832–857. https://doi.org/10.1080/10705511.2025.2515229.

44. Lai, K. Correct Point Estimator and Confidence Interval for RMSEA Given Categorical Data. *Struct. Equ. Modeling* **2020**, *27*, 678–695. https://doi.org/10.1080/10705511.2019.1687302.
45. Bollen, K.A. *Structural Equations with Latent Variables*; John Wiley & Sons: Hoboken, NJ, USA, 1989. https://doi.org/10.1002/9781118619179.
46. Cho, G.; Hwang, H.; Sarstedt, M.; et al. Cutoff Criteria for Overall Model Fit Indexes in Generalized Structured Component Analysis. *J. Mark. Anal.* **2020**, *8*, 189–202. https://doi.org/10.1057/s41270-020-00089-1.
47. Hu, L.; Bentler, P.M. Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria versus New Alternatives. *Struct. Equ. Modeling* **1999**, *6*, 1–55. https://doi.org/10.1080/10705519909540118.
48. Kenny, D.A.; Kaniskan, B.; McCoach, D.B. The Performance of RMSEA in Models with Small Degrees of Freedom. *Sociol. Methods Res.* **2015**, *44*, 486–507. https://doi.org/10.1177/0049124114543236.
49. Maydeu-Olivares, A.; Fairchild, A.J.; Hall, A.G. Goodness of Fit in Item Factor Analysis: Effect of the Number of Response Alternatives. *Struct. Equ. Modeling* **2017**, *24*, 495–505. https://doi.org/10.1080/10705511.2017.1289816.
50. MacCallum, R.C.; Browne, M.W.; Sugawara, H.M. Power Analysis and Determination of Sample Size for Covariance Structure Modeling. *Psychol. Methods* **1996**, *1*, 130–149. https://doi.org/10.1037/1082-989X.1.2.130.
51. Kline, R.B. *Principles and Practice of Structural Equation Modeling*, 3rd ed.; Guilford Press: New York, NY, USA, 2011.
52. Cole, D.A. Utility of Confirmatory Factor Analysis in Test Validation Research. *J. Consult. Clin. Psychol.* **1987**, *55*, 584–594. https://doi.org/10.1037/0022-006X.55.4.584.
53. Satorra, A. Robustness Issues in Structural Equation Modeling: A Review of Recent Developments. *Qual. Quant.* **1990**, *24*, 367–386. https://doi.org/10.1007/BF00152011.
54. Hu, L.; Bentler, P.M. Fit Indices in Covariance Structure Modeling: Sensitivity to Underparameterized Model Misspecification. *Psychol. Methods* **1998**, *3*, 424–453. https://doi.org/10.1037/1082-989X.3.4.424.
55. Boomsma, A.; Hoogland, J.J. The Robustness of LISREL Modeling Revisited. In *Structural Equation Modeling: Present and Future*; Cudeck, R., du Toit, S., Sörbom, D., Eds.; Scientific Software International: Lincolnwood, IL, USA, 2001; pp. 139–168.
56. Kane, M. The Argument-Based Approach to Validation. *Sch. Psychol. Rev.* **2013**, *42*, 448–457. https://doi.org/10.1080/02796015.2013.12087465.
57. Muthén, B.; Kaplan, D. A Comparison of Some Methodologies for the Factor Analysis of Non-Normal Likert Variables. *Br. J. Math. Stat. Psychol.* **1985**, *38*, 171–189. https://doi.org/10.1111/j.2044-8317.1985.tb00832.x.
58. Newsom, J.T. *Longitudinal Structural Equation Modeling: A Comprehensive Introduction*; Routledge: New York, NY, USA, 2015.
59. Akobeng, A.K. Understanding Type I and Type II Errors, Statistical Power and Sample Size. *Acta Paediatr.* **2016**, *105*, 605–609. https://doi.org/10.1111/apa.13384.
60. Fawcett, T. An Introduction to ROC Analysis. *Pattern Recognit. Lett.* **2006**, *27*, 861–874. https://doi.org/10.1016/j.patrec.2005.10.010.
61. Rota, M.; Antolini, L.; Valsecchi, M.G. Optimal Cut-Point Definition in Biomarkers: The Case of Censored Failure Time Outcome. *BMC Med. Res. Methodol.* **2015**, *15*, 24. https://doi.org/10.1186/s12874-015-0009-y.
62. Krzanowski, W.J.; Hand, D.J. *ROC Curves for Continuous Data*; CRC Press: Boca Raton, FL, USA, 2009.
63. Padgett, R.N.; Morgan, G.B. Multilevel CFA with Ordered Categorical Data: A Simulation Study Comparing Fit Indices across Robust Estimation Methods. *Struct. Equ. Modeling* **2021**, *28*, 51–68. https://doi.org/10.1080/10705511.2020.1759426.
64. Sharma, S.; Mukherjee, S.; Kumar, A.; et al. A Simulation Study to Investigate the Use of Cutoff Values for Assessing Model Fit in Covariance Structure Models. *J. Bus. Res.* **2005**, *58*, 935–943. https://doi.org/10.1016/j.jbusres.2003.10.007.
65. Marsh, H.W.; Hau, K.T.; Wen, Z. 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. *Struct. Equ. Modeling* **2004**, *11,* 320–341. https://doi.org/10.1207/s15328007sem1103_2.
66. Browne, M.W.; MacCallum, R.C.; Kim, C.T.; et al. When Fit Indices and Residuals are Incompatible. *Psychol. Methods* **2002**, *7*, 403–421. https://doi.org/10.1037/1082-989X.7.4.403.
67. Heene, M.; Hilbert, S.; Draxler, C.; et al. Masking Misfit in Confirmatory Factor Analysis by Increasing Unique Variances: A Cautionary Note on the Usefulness of Cutoff Values of Fit Indices. *Psychol. Methods* **2011**, *16*, 319–336. https://doi.org/10.1037/a0024917.
68. Savalei, V.; Rhemtulla, M. The Performance of Robust Test Statistics with Categorical Data. *Br. J. Math. Stat. Psychol.* **2013**, *66*, 201–223. https://doi.org/10.1111/j.2044-8317.2012.02049.x.
69. Flora, D.B.; Curran, P.J. An Empirical Evaluation of Alternative Methods of Estimation for Confirmatory Factor Analysis with Ordinal Data. *Psychol. Methods* **2004**, *9*, 466–491. https://doi.org/10.1037/1082-989X.9.4.466.
70. Ximénez, C.; Maydeu-Olivares, A.; Shi, D.; et al. Assessing Cutoff Values of SEM Fit Indices: Advantages of the Unbiased SRMR Index and Its Cutoff Criterion Based on Communality. *Struct. Equ. Modeling* **2022**, *29*, 368–380. https://doi.org/10.1080/10705511.2021.1992596.