Ordinal Factor Analysis with Robust Estimation and Predictive Models

Abuelgasim Osman, Zahayu binti Md Yusof
School of Quantitative Sciences, Universiti Utara Malaysia (UUM), Kedah, Malaysia

***Abstract*—Ordinal survey indicators are common in behavioral and social science research. Still, they violate continuous normal assumptions when category spacing is unequal, response distributions are skewed, or categories are sparse. This study evaluates a practical workflow for ordinal factor analysis that combines robust confirmatory factor analysis (CFA) estimation with complementary predictive modeling. Simulated five-category ordinal datasets and an empirical Malaysian green consumption dataset were analyzed using WLS, WLSMV, and DWLS estimators based on polychoric correlations. CFA performance was examined through conventional fit indices, parameter recovery, and bootstrap stability of standardized loadings. Predictive models, including random forests, support vector machines, gradient boosting, dense neural networks, convolutional neural networks, and recurrent neural networks, were assessed using matched 5-fold cross-validation. The revised comparison separates measurement validation from prediction, reports estimator stability through resampling, and clarifies the conditions under which robust ordinal estimators and nonlinear predictive models are useful. The study contributes a reproducible benchmark and reporting framework for researchers analyzing Likert-type ordinal data in psychometric and behavioral applications.**

Keywords—Ordinal data; confirmatory factor analysis; polychoric correlation; WLSMV; DWLS; bootstrap; machine learning; deep learning; cross-validation

I. Introduction

Ordinal Likert-type indicators are widely used in behavioral, educational, social science, and psychometric measurement. These indicators record ordered categories rather than continuous measurements with equal intervals. When they are analyzed as if they were continuous and normally distributed, Pearson correlations and normal-theory estimation can attenuate associations, bias factor loadings, and distort global model fit, especially under skewed category usage and floor or ceiling effects [1].

A standard approach to ordinal measurement assumes an unobserved continuous response propensity that is discretized into observed categories using thresholds. Under this threshold framework, polychoric correlations estimate associations among the latent response variables and provide a more appropriate input structure for ordinal factor models than Pearson correlations [2], [3]. Prior simulation and empirical studies have shown that the consequences of ignoring ordinality are most evident when sample sizes are modest, response categories are imbalanced, or indicators contain sparse cells [4], [5].

Robust ordinal confirmatory factor analysis (CFA) commonly relies on weighted least squares estimators. Full weighted least squares (WLS) is theoretically attractive but requires estimating a large asymptotic weight matrix that can become unstable in applied samples. Diagonally weighted least squares (DWLS) and mean- and variance-adjusted WLS (WLSMV) reduce this computational burden and are widely recommended for ordinal CFA [6]–[9]. Nevertheless, estimator performance remains condition-dependent; therefore, applied studies should report not only fit indices but also indicators of numerical stability and sampling uncertainty.

At the same time, machine learning and deep learning models are increasingly used to predict behavioral outcomes from item-level data. These models can capture nonlinearities and interactions, but they do not directly provide confirmatory measurement outputs such as factor loadings, thresholds, and fit indices. This study therefore distinguishes between two related but distinct objectives: measurement validation through robust ordinal CFA and predictive evaluation through machine learning and deep learning using matched validation procedures.

The contribution of this study is a compact benchmarking and reporting workflow for ordinal survey data. It compares WLS, WLSMV, and DWLS for ordinal CFA, assesses loading stability through nonparametric bootstrap resampling, and evaluates selected predictive models using consistent cross-validation. The intention is not to replace theory-driven CFA with prediction models but to clarify how robust estimation and predictive modeling can be used together without conflating their inferential purposes.

II. Related Work

A. Ordinal Measurement and Polychoric CFA

Ordinal response scales encode rank order but not equal distances between adjacent categories. This distinction matters because normal-theory CFA estimates model-implied covariances for continuous variables, whereas Likert-type items are generated by thresholded response processes. Polychoric correlations address this issue by estimating the association between two unobserved continuous response variables underlying the observed categories [2], [3]. However, polychoric estimation is not a universal remedy: sparse cells, extreme thresholds, and small sample sizes can yield unstable or non-positive-definite matrices, which, in turn, affect the downstream CFA solution.

The literature, therefore, supports a conditional rather than automatic preference for ordinal methods. Flora and Curran [4] and Holgado-Tello et al. [5] show that treating data as ordinal often improves parameter recovery compared with treating them as continuous. However, they also highlight the sensitivity of ordinal methods to response distributions and sample size. For applied researchers, this means that the choice of estimator should be evaluated alongside empirical diagnostics rather than justified solely by the presence of Likert-type items.

B. *Robust Ordinal CFA Estimators*

Full WLS uses the complete asymptotic covariance matrix of sample thresholds and correlations. In principle, this approach can be efficient; in practice, the weight matrix can be high-dimensional and poorly estimated when the number of items is moderate, and the sample size is limited. DWLS simplifies the weight matrix by using only its diagonal elements, while WLSMV adds robust mean and variance adjustments to improve test statistics and standard errors [6]–[8]. These estimators are now commonly used in ordinal CFA because they often balance computational stability with acceptable parameter recovery.

Nevertheless, robust ordinal estimators are not identical. Comparative studies indicate that WLSMV and DWLS can behave differently under conditions of threshold imbalance, small sample size, and model misspecification [7]–[9]. The practical implication is that an estimator comparison should not report only the best-fitting index. It should also assess convergence behavior, confidence interval width, and the substantive meaningfulness of differences in fit indices.

C. *Stability Assessment and Reporting*

Model fit indices such as CFI, TLI, RMSEA, and SRMR are useful but provide incomplete summaries of the adequacy of an ordinal CFA model. Established guidance recommends interpreting multiple indices together rather than relying on a single cutoff [10]. In ordinal settings, uncertainty arising from thresholds and polychoric correlations can also propagate to loading estimates, factor correlations, and convergence behavior. Nonparametric bootstrap resampling is a practical method for quantifying this uncertainty because it repeatedly resamples cases, refits the model, and summarizes the empirical variability of key parameters [11].

Bootstrap reporting is particularly useful when the same model appears acceptable according to conventional fit indices, but some estimators yield wider confidence intervals for factor loadings or lower convergence rates. Therefore, this study treats bootstrap convergence and the width of loading confidence intervals as complementary diagnostics rather than secondary details.

D. *Predictive Modeling for Ordinal Survey Data*

Machine learning serves a different analytical objective from CFA. Random forests, support vector machines, and gradient boosting models can capture nonlinear patterns and interactions in tabular data [12]–[14]. Deep neural networks can learn flexible representations, although their advantage is not guaranteed for small tabular survey datasets and depends heavily on hyperparameter tuning, regularization, and validation design [15]. These methods may be valuable when the researcher aims to predict an observed outcome, but they do not provide the same confirmatory measurement evidence as CFA.

The main gap addressed here is the integration of these methods without conceptual confusion. Prior studies on ordinal CFA emphasize robust estimation, whereas predictive modeling studies emphasize cross-validated accuracy. Fewer applied workflows incorporate both components into a single reporting framework while preserving their distinct purposes. This study addresses that gap by pairing robust ordinal CFA with resampling-based stability checks and evaluating predictive models separately using matched cross-validation.

III. METHODOLOGY

A. *Data Sources*

Two data sources were used. First, simulated ordinal datasets were generated to examine estimator behavior under controlled conditions. For each sample size n = 250, 500, and 1000, a one-factor latent response model with six indicators was generated. The continuous latent response was defined as y*_ij = λ_j F_i + ε_ij, where F_i followed a standard normal distribution, ε_ij represented normally distributed measurement error, and the target standardized loadings were approximately 0.70. The continuous responses were discretized into five ordered categories using both balanced and uneven thresholds, with the latter producing floor and ceiling effects. Each condition was replicated 500 times.

Second, an empirical dataset on Malaysian green consumption behavior was used to validate the applied model. The dataset contains 375 respondents and measures of environmental concern, social influence, perceived behavioral control, consumer novelty seeking, and green consumption behavior [16]. Demographic variables were excluded to focus the analysis on ordinal measurement and item-based prediction.

B. *CFA Estimators and Model Fit*

Ordinal CFA models were estimated using WLS, WLSMV, and DWLS. All CFA estimators were applied to polychoric correlation matrices, consistent with the threshold view of ordinal indicators. Model adequacy was evaluated using CFI, TLI, RMSEA, and SRMR. The analysis emphasizes comparative interpretation of indices rather than rigid reliance on a single cutoff [10]. Estimation was implemented in R using the lavaan package [17].

C. *Predictive Models*

Predictive modeling was conducted on the empirical dataset to evaluate item-based prediction of green consumption behavior. The machine learning models included random forest (RF), support vector machine with radial basis function kernel (SVM), and gradient boosting machine (GBM). RF models used 500 trees, with mtry tuned via a grid search. SVM hyperparameters C and γ were tuned on logarithmic grids. GBM models were tuned for learning rate, tree depth, and number of boosting iterations, with early stopping to reduce overfitting.

The deep learning baselines included a dense neural network (DNN), a convolutional neural network (CNN), and a recurrent neural network (RNN). The DNN used two hidden layers, dropout regularization, and early stopping based on validation loss. CNN and RNN models were treated as simple baselines because the dataset is tabular rather than image- or sequence-based. Predictive models were evaluated using R^2, RMSE, and MAE.

D. Validation Procedures

The CFA and prediction components used different validation designs because they answer different questions. For CFA estimator stability, nonparametric bootstrap resampling with B = 500 resamples was applied to the real dataset. The bootstrap analysis summarized convergence rates and the average width of 95% confidence intervals for standardized loadings. For predictive models, matched 5-fold cross-validation was used so that all algorithms were evaluated on consistent data splits. Fig. 1 presents an overview of the revised methodological workflow.

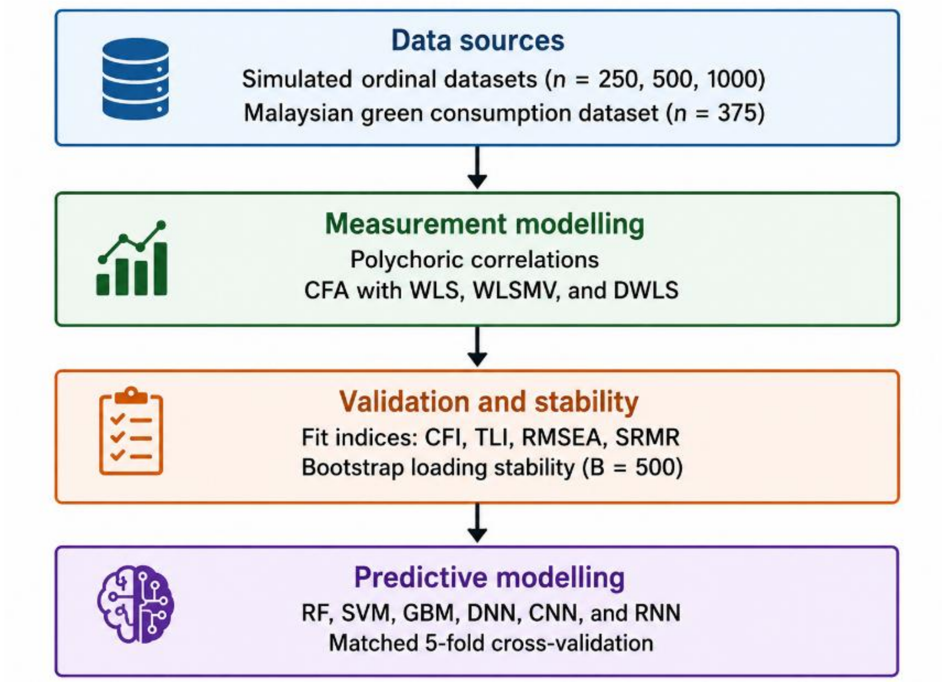

Fig. 1. Overview of the revised methodological workflow.

IV. Results

A. CFA Model Fit Across Estimators

Table I reports model fit across estimators and sample conditions. In the simulated datasets, the largest differences appeared at n = 250. WLS produced weaker fit than WLSMV and DWLS, while the differences decreased as the simulated sample size increased. In the empirical dataset, WLSMV and DWLS yielded a better fit than WLS, although the fit was weaker than in the large simulated conditions.

TABLE I. CFA Model Fit Indices by Estimator and Sample Size

Sample	Estimator	CFI	TLI	RMSEA	SRMR
250 (sim)	WLS	0.94	0.92	0.08	0.07
250 (sim)	WLSMV	0.97	0.96	0.05	0.04
250 (sim)	DWLS	0.96	0.95	0.06	0.05
500 (sim)	WLS	0.96	0.95	0.05	0.05
500 (sim)	WLSMV	0.99	0.99	0.03	0.03
500 (sim)	DWLS	0.98	0.98	0.04	0.03
1000 (sim)	WLS	0.99	0.99	0.02	0.02
1000 (sim)	WLSMV	1.00	1.00	0.01	0.01
1000 (sim)	DWLS	0.99	0.99	0.01	0.01
375 (real)	WLS	0.92	0.90	0.08	0.07
375 (real)	WLSMV	0.95	0.93	0.06	0.05
375 (real)	DWLS	0.93	0.91	0.07	0.06

Note: Fit indices were computed from polychoric correlation-based CFA models. Simulated and empirical conditions are reported together for compact comparison.

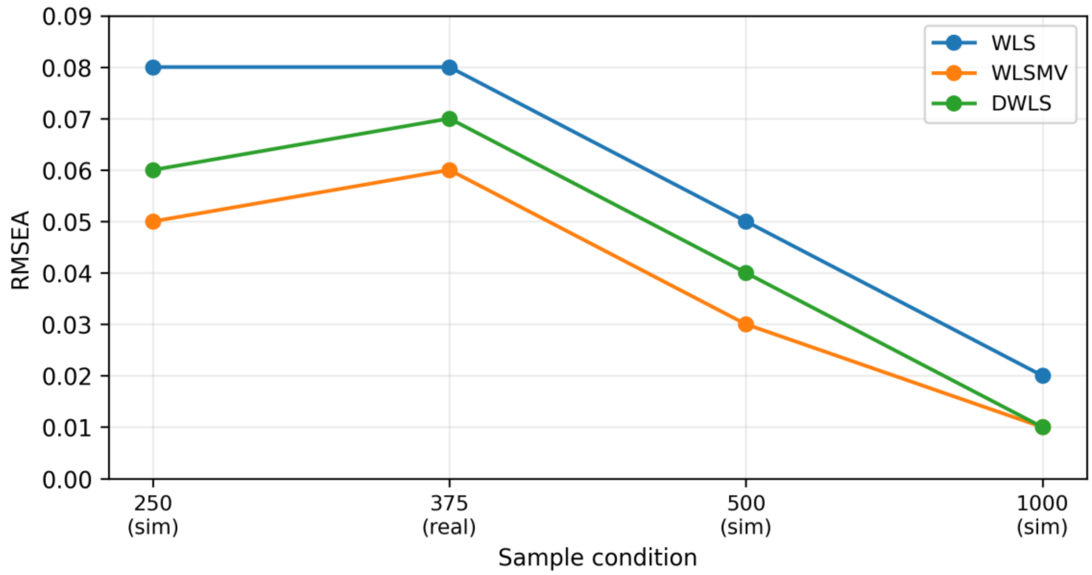

Fig. 2. RMSEA across estimators and sample conditions.

Fig. 2 summarizes RMSEA trends. The figure shows that RMSEA decreases as the simulated sample size increases. WLSMV provides the lowest RMSEA in the small and moderate simulated samples, whereas DWLS reports RMSEA values of 0.06 and 0.04 at n = 250 and n = 500, respectively. Therefore, DWLS should be described as improving over WLS under these conditions rather than as producing a negligible RMSEA across all sample sizes.

B. Bootstrap Stability of Loadings

Table II reports the bootstrap stability results for the empirical dataset. WLS converged in 94% of bootstrap resamples, while WLSMV and DWLS converged in all resamples. The average 95% confidence interval width for standardized loadings was wider under WLS than under WLSMV and DWLS, indicating greater sampling variability for the full WLS solution in this dataset. Fig. 3 details the bootstrap convergence rate by estimator.

TABLE II. BOOTSTRAP CONVERGENCE AND LOADING CONFIDENCE INTERVAL WIDTH.

Estimator	Bootstrap Conv.	Mean CI Width
WLS	94%	0.20
WLSMV	100%	0.14
DWLS	100%	0.15

Note: CI width refers to the average width of the 95% bootstrap interval for standardized loadings across 500 resamples.

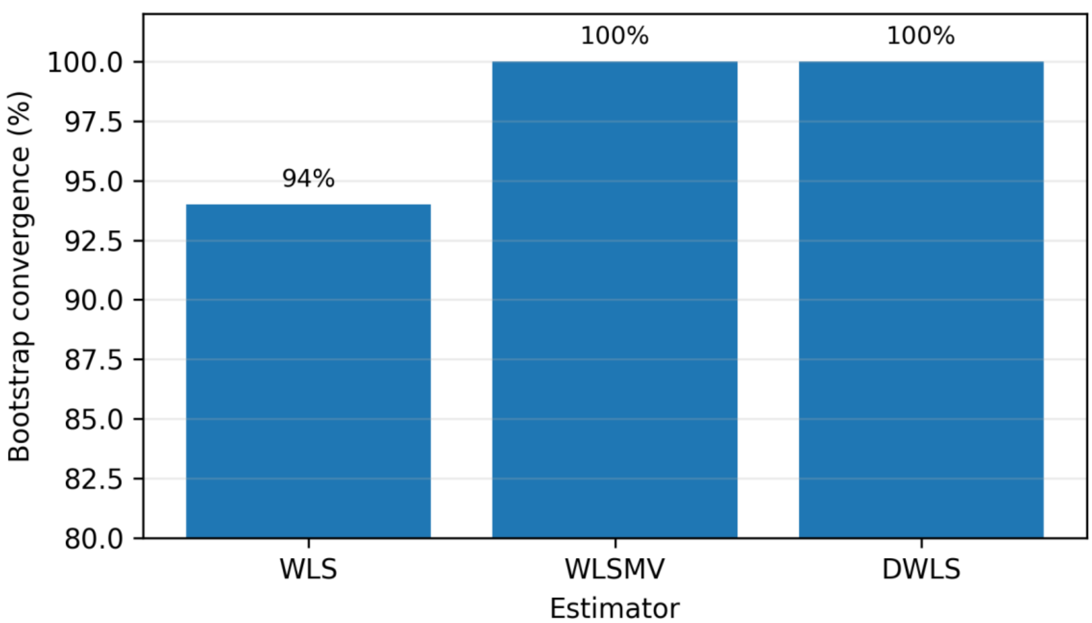

Fig. 3. Bootstrap convergence rate by estimator.

C. Predictive Modeling Results

Table III reports 5-fold cross-validated predictive performance on the empirical dataset. Among the evaluated models, DNN produced the highest R^2 and the lowest RMSE and MAE, followed by GBM. CNN and RNN baselines did not improve prediction in this tabular setting. These results should be interpreted as predictive evidence only; they do not replace the measurement evidence provided by ordinal CFA. Fig. 4 presents the cross-validated R^2 by the predictive model.

TABLE III. CROSS-VALIDATED PREDICTION PERFORMANCE

Model	R^2	RMSE	MAE
Random Forest (RF)	0.30	6.3	4.8
Support Vector Machine (SVM)	0.28	6.5	5.0
Gradient Boosting Machine (GBM)	0.33	6.1	4.5
Deep Neural Network (DNN)	0.35	5.9	4.4
Convolutional Neural Network (CNN)	0.26	6.7	5.1
Recurrent Neural Network (RNN)	0.27	6.6	5.0

Note: Metrics were averaged across matched 5-fold cross-validation splits on the empirical dataset.

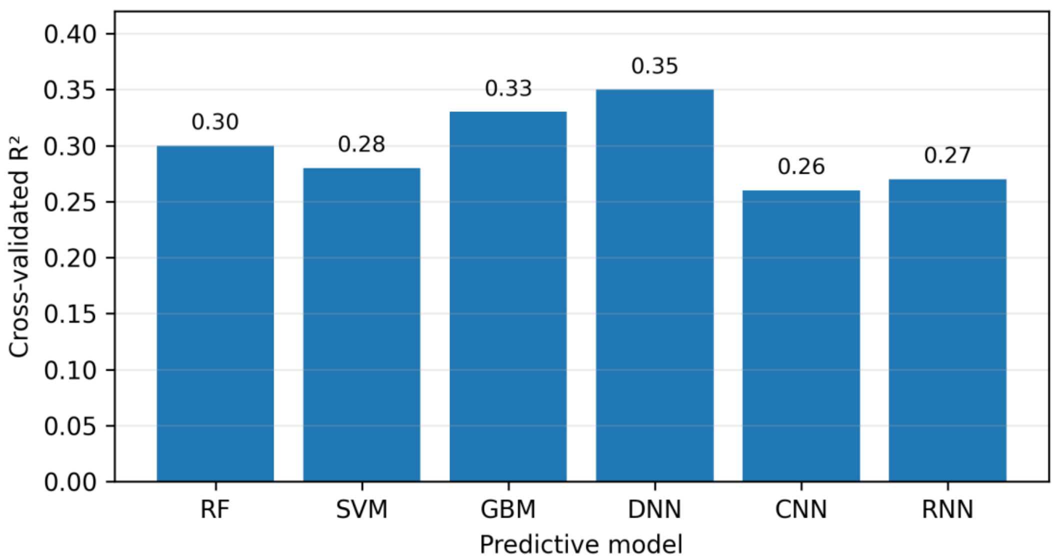

Fig. 4. Cross-validated R^2 by the predictive model.

V. DISCUSSION

The results support the use of robust ordinal estimators for CFA with Likert-type indicators. WLSMV and DWLS provided a better fit than WLS under small-to-moderate sample conditions and demonstrated greater bootstrap stability in the empirical dataset. This pattern is consistent with the methodological argument that full WLS can become sensitive when estimating a large weight matrix. In contrast, WLSMV and DWLS use reduced or adjusted weighting strategies that are more stable in typical applied samples.

The bootstrap analysis is important because it demonstrates that fit indices alone do not fully describe estimator performance. Two estimators may produce acceptable global fit yet differ in convergence rates and the uncertainty of factor loadings. Reporting bootstrap convergence and the width of loading confidence intervals therefore strengthens ordinal CFA reporting by providing an empirical assessment of estimator stability.

The predictive results show that nonlinear models can be useful for predicting outcomes from ordinal survey indicators. Among the evaluated models, DNN and GBM achieved the strongest cross-validated performance. However, these models address a different objective from CFA. They can improve prediction, but they do not estimate thresholds, standardized loadings, latent factor correlations, or global measures of model fit. For this reason, machine learning and deep learning should be treated as complementary predictive tools rather than substitutes for confirmatory measurement modeling.

A practical workflow follows from these findings. Researchers analyzing ordinal survey data should first validate the measurement structure using a robust ordinal CFA estimator, such as WLSMV or DWLS, report multiple fit indices, and, whenever possible, provide resampling-based evidence of estimator stability. If prediction is also a research objective, machine learning and deep learning models can then be evaluated using cross-validation on the same dataset without interpreting their predictive accuracy as evidence of measurement validity.

This study has several limitations. The simulation conditions employed a compact one-factor structure and a limited range of sample sizes. The empirical validation relied on a single public green consumption dataset. Future research should extend the benchmark to multidimensional CFA models, more severe threshold imbalance, missing-data mechanisms, and larger empirical datasets. Additional work should also evaluate interpretable machine learning methods to connect predictive patterns with substantive item-level interpretation.

VI. CONCLUSION

This study revised and clarified a comparative framework for ordinal factor analysis and predictive modeling. The findings indicate that WLSMV and DWLS are preferable to WLS for ordinal CFA in small-to-moderate applied conditions, particularly when estimator stability is assessed through bootstrap resampling. For prediction, DNN and GBM performed strongest among the evaluated models under matched cross-validation. The central conclusion is that robust ordinal CFA and ML/DL serve complementary purposes: CFA supports theory-driven measurement validation, whereas ML/DL supports outcome prediction. A defensible applied workflow should preserve this distinction while transparently reporting both measurement fit and predictive validation.

VII. Declaration on Generative AI-Assisted Tools

Generative AI tools were used for language editing, formatting assistance, and organization of responses to reviewers during manuscript revision. The authors critically reviewed, edited, and approved the scientific content, analysis, interpretation, references, and final manuscript.

References

[1] A. Agresti, *Analysis of Ordinal Categorical Data*, 2nd ed. Hoboken, NJ, USA: John Wiley & Sons, 2010.

[2] U. Olsson, "Maximum likelihood estimation of the polychoric correlation coefficient," *Psychometrika*, vol. 44, no. 4, pp. 443–460, Dec. 1979.

[3] K. G. Jöreskog, "On the estimation of polychoric correlations and their asymptotic covariance matrix," *Psychometrika*, vol. 59, no. 3, pp. 381–389, Sep. 1994.

[4] D. B. Flora and P. J. Curran, "An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data," *Psychological Methods*, vol. 9, no. 4, pp. 466–491, Dec. 2004.

[5] F. P. Holgado-Tello, S. Chacón-Moscoso, I. Barbero-García, and E. Vila-Abad, "Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables," *Quality & Quantity*, vol. 44, no. 1, pp. 153–166, Jan. 2010.

[6] C. G. Forero, A. Maydeu-Olivares, and D. Gallardo-Pujol, "Factor analysis with ordinal indicators: A Monte Carlo study comparing DWLS and ULS estimation," *Structural Equation Modeling*, vol. 16, no. 4, pp. 625–641, Oct. 2009.

[7] C.-H. Li, "Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares," *Behavior Research Methods*, vol. 48, no. 3, pp. 936–949, Sep. 2016.

[8] D. L. Bandalos, "Relative performance of categorical diagonally weighted least squares and robust maximum likelihood estimation," *Structural Equation Modeling*, vol. 21, no. 1, pp. 102–116, Jan. 2014.

[9] C. DiStefano and G. B. Morgan, "A comparison of diagonally weighted least squares robust estimation techniques for ordinal data," *Structural Equation Modeling*, vol. 21, no. 3, pp. 425–438, Jul. 2014.

[10] L.-T. Hu and P. M. Bentler, "Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives," *Structural Equation Modeling*, vol. 6, no. 1, pp. 1–55, Jan. 1999.

[11] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap*. New York, NY, USA: Chapman & Hall/CRC, 1993.

[12] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. New York, NY, USA: Springer, 2009.

[13] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, Oct. 2001.

[14] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, San Francisco, CA, USA, 2016, pp. 785–794.

[15] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[16] N. A. Mohd Ghani, F. D. Yusop, and Y. Kamarulzaman, "Dataset on the green consumption behaviour amongst Malaysian consumers," *Data in Brief*, vol. 32, Art. no. 106302, Oct. 2020.

[17] Y. Rosseel, "lavaan: An R package for structural equation modeling," *Journal of Statistical Software*, vol. 48, no. 2, pp. 1–36, May 2012.