ORIGINAL ARTICLE

The Performance of Regularization Technique in Assessing Reliability and Validity of the Constructs in Structural Equation Modeling: Application in Breast Cancer Awareness Research

Nurul Raudhah Zulkifli[1], Hezlin Md Zin[2]

[1] School of Mathematical Science, College of Computing, Informatics and Mathematics, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia
[2] Department of Family Medicine, Kuliyyah of Medicine, International Islamic University Malaysia, Jalan Sultan Ahmad Shah, Bandar Indera Mahkota, 25200, Kuantan, Pahang, Malaysia

ABSTRACT

Introduction: In structural equation modeling (SEM), among the desirable requirements in the measurement model are the reliability and validity of the constructs. The average variance extracted (AVE) provides a numerical measure of the overall validity of each construct in the model. Meanwhile, composite reliability (CR) reflects the internal consistency reliability of the items under each construct. **Materials and methods:** In this study, the existing estimator in SEM namely unweighted least squares (ULS) has been used for nonnormal data in SEM. However, the method is seen less efficient as the method leads to improper solutions like unique variance which introduces some level of bias, hence affecting the reliability and validity of the constructs. Therefore, the regularized unweighted least squares (ULS), a new approach of regularization is introduced in this study. Utilizing 300 samples of breast cancer awareness data, the analysis was carried out using “lavaan” package in R programming Environment. **Results:** Regularized ULS consistently yields higher CR and AVE values, enhancing the reliability and validity of measurement instruments. **Conclusion:** Besides assisting researchers in achieving the reliability and validity of a construct, the findings of this study can aid survey-based researchers to generate a more reliable model. The findings indicate that employing regularized ULS estimation allows for the retention of a greater number of items or questions within the respective construct in the Malay Version of the Breast Cancer Awareness instrument. This proves to be invaluable in validating the factor structure through confirmatory factor analysis.
Malaysian Journal of Medicine and Health Sciences (2024) 20(SUPP8): 112-118. doi:10.47836/mjmhs20.s8.15

Keywords: Regularization, Structural equation modeling, Breast cancer awareness, Unweighted least squares, Non-normal data

Corresponding Author:
Nurul Raudhah binti Zulkifli, PhD
Email: nurulraudhahzulkifli@gmail.com
Tel : 013-7126620

INTRODUCTION

Covariance-based Structural Equation Modeling (CB-SEM) is a widely utilized method in fields like social science, healthcare, and management studies (1–5) and others in recent years. In CB-SEM, it is typically essential for data distributions to adhere to the normality assumption when employing various estimators (6). However, an alternative is available in the form of the Unweighted Least Squares (ULS) estimator, which can handle non-normally distributed data (7). Nonetheless, there are drawbacks associated with the ULS estimator. ULS often yields improper solutions, such as negative or boundary estimates of unique variances (8). In the context of SEM, unique variance is synonymous with disturbance, representing random errors arising from measurement inaccuracies or unreliability, alongside the reliable variation associated with latent variables that are not directly observed. Furthermore, ULS tends to yield less precise and more biased parameter estimates.

The imprecision introduced by ULS can impact the accuracy of loadings, potentially leading to unreliable and invalid construct, particularly when dealing with non-normally distributed data. To address this issue, an extensive body of literature suggests the application of regularization techniques (9,10). Regularization, according to Arruda et al. (2) , it involves making a process more regular. It can be seen as a mathematical strategy to enhance solutions in challenging scenarios, which aligns with the concept articulated by Bickel et al. (11), involving the modification of methods to provide suitable solutions under adverse conditions. Recent studies and applications of some regularization techniques related to the SEM approach have been made (9,12,13). A number of techniques have been incorporated into SEM (10,14). In addition, regularized structural equation modeling, or RegSEM, was suggested by Jacobucci et al. (10) as a technique for penalizing parameters in order to mitigate model complexity and increase the generalizability of models. However, one weakness of RegSEM, as pointed out by Li et al. (15), is that regularization can lead to biased estimates of the model parameters, particularly when the regularization parameter (λ) is too large. This is because regularization shrinks the estimates towards zero, which can pull them away from their true values. In extreme cases, regularization can even drive some estimates to exactly zero, leading to a less interpretable model. Another potential weakness of RegSEM is that it may not always improve model fit. In some cases, it may even result in a worse fit if the regularization parameter is not chosen carefully. Furthermore, RegSEM can be computationally intensive, particularly when dealing with large datasets and complex models. The addition of a regularization term can significantly increase the computational burden of estimation and may require specialized software and hardware to implement efficiently (16). Thus, this study proposes to incorporate a regularization parameter for the ULS estimator in each element of the sample variance-covariance matrix. By adding small positive values to the covariance matrix elements, this method appears for the best resolution between assessing bias and assessing variability. Therefore, a bias-to-variance ratio can be modified to produce the best estimation results.

MATERIALS AND METHODS

Data

Secondary data were obtained from the prior study by Md Zin et al. (17). The breast cancer awareness dataset includes 300 samples of female transgender in Pahang, Kedah, and Kelantan. This dataset encompasses a series of questions related to breast cancer awareness, which are categorized into five domains: knowledge of symptoms and clinical features, confidence, skills, and behavior concerning breast self-examination, barriers in seeking medical help, knowledge of breast screening programs and knowledge of risk factors for breast cancer. Each domain contains between three to nine items. Respondents provided answers as either yes, no, or not sure. To confirm the multivariate non-normality of the data, a multivariate normality test was performed. The Mardia test for multivariate normality was applied using the QuantPsyc package in the R programming language. Furthermore, k-fold cross validation approach is used to improve the generalizability of the estimation methods. It is a robust method used in machine learning and statistics to evaluate the performance of a model. It involves dividing the dataset into k subsets, or folds, and using these subsets to train and validate the model in a systematic way. This study utilizes 20-fold validation in "caret" package to ensure more reliable outcomes (18,19).

Method of estimation

All data were analysed using R Programming. To assess the appropriateness of fit and coefficients in CB-SEM, the non-normally distributed data was analyzed through the application of the Unweighted Least Squares (ULS) estimation technique. ULS, as described in reference (16), reduces the fit function by utilizing derivatives. For the purpose of evaluating fit and coefficients in CB-SEM, nonnormal data were analysed using the Unweighted Least Squares (ULS) estimation method. ULS, according to, minimises the fit function by employing derivatives:

$$F_{ULS} = \frac{1}{2}\, tr\{[S - \Sigma(\theta)]\}^2 \quad (1)$$

where tr is the trace of the matrix, S is the sample covariance matrix, $\Sigma(\theta)$ is the model-implied covariance matrix and θ is the vector of parameters. This method specifies which parameters to penalize, and, in this study, the factor loading matrix was penalized. The penalization of factor loadings from one matrix in a sense pushes the parameter values into the other matrix, therefore, estimation problems could occur if dependent parameters were both penalized. It quantifies the influence of the penalty, with larger λ incurring larger penalty, thus resulting in greater shrinkage of the factor loadings. This leads to estimates of factor loadings which are closer to zero. However, one weakness of RegSEM, as pointed out by Jacobucci et al. (10) is that regularization can lead to biased estimates of the model parameters, particularly when the regularization parameter is too large. This is because regularization shrinks the estimates towards zero, which can pull them away from their true values. In extreme cases, regularization can even drive some estimates to exactly zero, leading to a less interpretable model. Another potential weakness of RegSEM is that it may not always improve model fit. Regularization is typically used to improve the stability of parameter estimates and prevent overfitting, but it may not necessarily lead to a better fit to the data. In some cases, it may even result in a worse fit if the regularization parameter is not chosen carefully. Next, the regularized ULS method introduced by Zulkifli et al. (6) is as follows:

$$F_{regULS} = \frac{1}{2}\, tr\{[\hat{S} - \Sigma(\theta)]\}^2 \quad (2)$$

where $\hat{S}$ is the regularized sample covariance matrix, $\hat{S} = (S + \lambda)$. λ is the regularization parameter. The amount of weight that should be given to the variance-covariance matrix depends critically on the value of lambda (2). Several lambda values ($\lambda > 0$) are tested and the lambda with the minimum Root Mean Square Error of Approximation (RMSEA) is selected. RMSEA quantifies the discrepancy between the observed and predicted covariance matrices, accounting for the model's complexity. It takes into account both the number of estimated parameters in the model and the sample size. Therefore, it is considered to be a relatively more robust fit index for large sample sizes and complex models compared to other fit indices such as the chi-square statistic. The optimal regularization parameter value that corresponds to the lowest RMSEA value is selected. This optimal regularization parameter value represents the optimal balance between model complexity and fit to the data. Instead of continuing to increase the value until all regularized parameters become zero, the testing process can be stopped once estimation problems arise such as model is unfit. This is particularly important when regularizing parameters since reaching a sufficiently high regularization parameter may result in over shrinkage of the estimates (10). Accordingly, the procedure begins with finding the optimal regularization parameter by defining a range of potential values for the regularization parameter. This study varied across 100 values, ranging from 0 to 1 in equal increments, 0.01. The increment of 0.01 is applied in this study based on the existence of vast literature that set 0.01 as initial values (10,16). For each value of the regularization parameter, the model is fitted and RMSEA is computed. Next, the optimal regularization parameter value that corresponds to the lowest RMSEA value is selected. Finally, using the optimal regularization parameter, the parameters are estimated.

Performance indicators

Composite reliability (CR) is used to measure the internal consistency of the constructs. The CR formula is as follows (17)

$$CR = \frac{\left[\sum_{i=1}^{p}\lambda_i\right]^2}{\left[\sum_{i=1}^{p}\lambda_i\right]^2 + \sum_{i=1}^{p}\theta_{\delta ii}} \quad (3)$$

where the number of observed indicators is denoted as p (i =1 though p), λ_i are the item loadings and $\theta_{\delta ii}$ are the measurement error variances. Acceptable values for CR in exploratory study are 0.60 to 0.70, whereas values of 0.70 to 0.90 can be considered satisfactory in more advanced research (18). Meanwhile, the average amount of variance justified by a construct in its indicator variables in relation to the overall variance of its indicators is denoted by average variance extracted, AVE (19). According to Fornell et al. (20)

$$AVE = \frac{\sum_{i=1}^{p}\lambda_i^2}{\sum_{i=1}^{p}\lambda_i^2 + \sum_{i=1}^{p}\theta_{\delta ii}} \quad (4)$$

where the number of observed indicators is denoted as p (i = 1 though p), λ_i are the indicator loadings and $\theta_{\delta ii}$ are the measurement error variances. AVE value of 0.50 or greater, determined using the same reasoning as the individual indicators, implies that the concept, on average, accounts for more than 50% of the variance in its indicators.

RESULTS

Validity and Reliability of The Breast Cancer Awareness Domains

As the preliminary step, the multivariate normality test to validate the multivariate nonnormality of the data was conducted. The findings reveal significant skewness in the data, as evidenced by Mardia's coefficient which is 7263.23 and the p-value <0.001. This indicates that the data is deemed multivariate nonnormal. Table I presents the Composite Reliability (CR) and Average Variance Extracted (AVE) values before item deletion, while Table II shows the values after item deletion for different domains. In the domain of Knowledge of symptoms and clinical features, it is worth noting that prior to the removal of items, the ULS method yielded a CR of 0.85, which was considered acceptable for CR, but the AVE of 0.39 fell below the threshold. However, following the removal of items, the ULS CR showed improvement, increasing to 0.86, and the AVE also increased to 0.41. Meanwhile, the regularized ULS values consistently remained high for both CR and AVE. In the context of Knowledge of cancer risk factors domain, the ULS method initially produced a CR of 0.68, which was above the acceptable level, and an AVE of 0.21, falling short of the required standard. After removing items, the ULS CR remained at 0.68, while the AVE improved to 0.34. Nevertheless, the regularized ULS values improves the CR and AVE values.

Table I: CR and AVE values before deletion of items

Domain	Composite reliability (CR>0.6)		Average variance extracted (AVE >0.5)	
	ULS	**Regularized ULS**	**ULS**	**Regularized ULS**
Knowledge of symptoms and clinical features	**0.85**	**0.87**	0.39	0.43
Knowledge of cancer risk factors	**0.68**	**0.77**	0.21	0.30
Knowledge of breast screening programs	**0.68**	**0.69**	0.36	0.40
Barriers in seeking medical help	0.44	0.54	0.21	0.31
Confidence, skills and behaviour concerning breast self-examination	0.27	0.33	0.21	0.21

Note: Bold values indicate that the CR and AVE values are acceptable.

Table II: CR and AVE values after deletion of items

Domain	Composite reliability (CR>0.6)		Average variance extracted (AVE >0.5)	
	ULS	**Regularized ULS**	**ULS**	**Regularized ULS**
Knowledge of symptoms and clinical features	**0.86**	**0.87**	0.41	0.45
Knowledge of cancer risk factors	**0.68**	**0.75**	0.34	0.36
Knowledge of breast screening programs	**0.71**	**0.76**	**0.50**	**0.62**
Barriers in seeking medical help	0.55	**0.62**	0.38	**0.50**
Confidence, skills and behaviour concerning breast self-examination	0.41	0.50	0.26	0.31

Note: Bold values indicate that the CR and AVE values are acceptable.

Next, before item deletion in the Knowledge of breast screening programs domain, the regularized ULS values for CR and AVE were 0.69 and 0.40, respectively, and they further improved to 0.76 and 0.62 after item removal, while the ULS AVE was 0.40, below the 0.5 threshold, suggesting that the items in this domain might not converge as strongly as desired. The improvement in regularized ULS suggests that the domain Knowledge of breast screening programs became more reliable and better measured the intended construct after the removal of items, particularly when using the regularized ULS estimation method. Furthermore, the ULS CR in the Barriers in seeking medical help domain was 0.44, indicating weak internal consistency. The ULS AVE was 0.21, falling below the typical threshold of 0.5 for AVE. After deletion of item, the ULS CR improved to 0.55, but it still remained below the 0.6 threshold for CR and the ULS AVE increased to 0.38, which is a step toward meeting the acceptable AVE threshold.

However, regularized ULS values consistently displayed higher CR and AVE values both before and after item deletion. Before item deletion, the regularized ULS values for CR and AVE were 0.54 and 0.31, respectively. After item deletion, they improved to 0.62 for CR and 0.50 for AVE. These improvements suggest that the domain Barriers in seeking medical help became more reliable and better measured the intended construct after the removal of items, particularly when utilizing the regularized ULS estimation method. Finally, after item deletion, the Confidence, skills, and behavior concerning breast self-examination domain has higher ULS CR (0.41) and AVE (0.26), but they are still below the threshold. Regularized ULS significantly enhances both CR and AVE. In summary, item deletion improves CR and AVE in some domains. Regularized ULS consistently yields higher CR and AVE values, enhancing the reliability and validity of measurement instruments. This finding is consistent with several findings that highlights the effectiveness of the regularization methods (2,6,21).

Although multiple AVE values were observed below the 0.50 threshold, and CR values remained under 0.70, eliminating items that are conceptually irrelevant can strengthen the validity and reliability of the construct by focusing on items that best represent the construct of interest (26). In structural equation modeling, eliminating items with low factor loadings can enhance fit indices, resulting in a more parsimonious model that better represents the underlying constructs. Following this, the results are validated through the application of a 20-fold validation approach. We found that there exists only a marginal difference in the AVE and CR values before and after item removal, which does not alter the overall conclusion drawn as presented in Table III and Table IV. Consequently, this substantiates that the cross-validation approach improved generalizability of regularized ULS over ULS.

Table III: Mean of CR and AVE values before deletion of items (using 20-fold validation)

Domain	Composite reliability (CR>0.6)		Average variance extracted (AVE >0.5)	
	ULS	**Regularized ULS**	**ULS**	**Regularized ULS**
Knowledge of symptoms and clinical features	**0.80**	**0.85**	0.37	0.45
Knowledge of cancer risk factors	**0.65**	**0.74**	0.21	0.32
Knowledge of breast screening programs	**0.62**	**0.67**	0.34	0.40
Barriers in seeking medical help	0.43	0.56	0.22	0.33
Confidence, skills and behaviour concerning breast self-examination	0.25	0.35	0.20	0.23

Note: Bold values indicate that the CR and AVE values are acceptable.

Table IV: Mean of CR and AVE values after deletion of items (using 20-fold validation)

Domain	Composite reliability (CR>0.6)		Average variance extracted (AVE >0.5)	
	ULS	**Regularized ULS**	**ULS**	**Regularized ULS**
Knowledge of symptoms and clinical features	**0.82**	**0.86**	0.40	0.46
Knowledge of cancer risk factors	**0.65**	**0.74**	0.35	0.38
Knowledge of breast screening programs	**0.70**	**0.75**	**0.51**	**0.60**
Barriers in seeking medical help	0.53	**0.60**	0.36	**0.51**
Confidence, skills and behaviour concerning breast self-examination	0.42	0.51	0.27	0.35

Note: Bold values indicate that the CR and AVE values are acceptable.

Next, the discriminant validity in Table V and Table VI is also computed using regularized ULS and ULS to enhance the overall evaluation of the measurement model. This would aid in achieving a more thorough understanding of the differentiation among the constructs within the model. The bold diagonal values represent the square root of the AVE for each construct. The off-diagonal values signify correlations between pairs of constructs, indicating the extent of shared variance. Generally, off-diagonal values should be smaller than diagonal ones, suggesting stronger correlations within constructs than between them. Overall, the results before and after deletion of items for both methods affirm the discriminant validity of the measurement model, implying that each construct is unique and differentiated from others within the framework.

Table V: Discriminant validity after deletion of items for ULS

	Knowledge of symptoms and clinical features	Confidence, skills and behaviour concerning breast self-examination	Barriers in seeking medical help	Knowledge of breast screening programs	Knowledge of cancer risk factors
Knowledge of symptoms and clinical features	**0.64**				
Confidence, skills and behaviour concerning breast self-examination	0.32	**0.51**			
Barriers in seeking medical help	0.04	0.03	**0.62**		
Knowledge of breast screening programs	0.45	0.15	0.04	**0.71**	
Knowledge of cancer risk factors	0.30	0.36	0.15	0.25	**0.58**

Note: Bold values are the square root of AVE.

Table VI: Discriminant validity after deletion of items for Regularized ULS

	Knowledge of symptoms and clinical features	Confidence, skills and behaviour concerning breast self-examination	Barriers in seeking medical help	Knowledge of breast screening programs	Knowledge of cancer risk factors
Knowledge of symptoms and clinical features	**0.67**				
Confidence, skills and behaviour concerning breast self-examination	0.32	**0.56**			
Barriers in seeking medical help	0.04	0.03	**0.71**		
Knowledge of breast screening programs	0.45	0.15	0.04	**0.79**	
Knowledge of cancer risk factors	0.30	0.36	0.15	0.25	**0.60**

CONTINUE

Table VI: Discriminant validity after deletion of items for Regularized ULS (CONT.)

Note: Bold values are the square root of AVE.

DISCUSSION

From the findings of the analyses, the regularized ULS method is effective in obtaining more reliable and valid constructs in breast cancer awareness instruments. This revelation aligns seamlessly with the scholarly contribution of (27), illuminating that in instances characterized by low factor loadings, regularization methodologies tend to surpass ULS in adeptly recovering loadings, enhancing the reliability and validity of the construct. The stability of the covariance matrix is improved by adding a positive constant value to the sample variance-covariance matrix, reducing the likelihood of negative error variances. During confirmatory factor analysis, the values provided by the sample variance-covariance matrix are significant in establishing the parameters of the model. The CR and AVE values before and after item deletion within each domain reveals that, in general, item deletion led to improvements in CR and AVE, enhancing the reliability and validity of the constructs as evidenced in (26,28). Furthermore, it is noteworthy that the regularized ULS estimation method consistently yielded better results, especially in AVE, underlining the significance of method choice in structural equation modeling. Researchers should carefully consider these findings when determining which items to retain to improve the overall quality of their research model. In some situations, the relationships between construct and its respective items in the model may be strong and robust, regardless of the sample size. The finding resonates with the conclusions drawn by Mabel et al. (29), stating that the attainment of accurate parameter estimates remains achievable when each factor boasts three to four measured variables, and the items' factor loadings surpass the commendable threshold of 0.7. To conclude, this study has confirmed the factor structure suggested by Md Zin (17) through confirmatory factor analysis by testing whether the data fit this model well in terms of RMSEA. This underscores the significance of selecting optimal regularization parameter in the regularized ULS approach (30). Next, it is important to acknowledge several limitations of this study. Relying on secondary data from breast cancer awareness studies may restrict the generalizability of findings, given potential differences in data collection methods, sample characteristics, and measurement tools compared to other contexts. Additionally, the implementation of cross-validation, like k-fold cross-validation, proves time-intensive, particularly with large datasets, as each iteration necessitates training and evaluating the model multiple times.

CONCLUSION

In summary, these findings suggest that ULS estimation initially raises concerns about reliability and convergent validity. However, the regularized ULS method often mitigates these issues. Therefore, researchers are encouraged to consider the regularization approach as a means to enhance the overall reliability and validity of their constructs in structural equation modeling. The regularized ULS approach can significantly contribute to the quality and impact of breast cancer awareness research. They help researchers better understand the complex interplay of factors influencing awareness, design more effective interventions, and provide accurate information that can influence healthcare practices, public awareness campaigns, and policy decisions in the context of breast cancer. In the future, utilizing the regularized ULS approach could prove invaluable for verifying the factor structure in various languages within the field of breast cancer awareness studies, thereby facilitating its global applicability.

ACKNOWLEDGEMENT

This work is supported by the secondary data which was obtained and granted from the author's prior research study.

REFERENCES

1. Joel E. C. Applied Structural Equation Modelling Using AMOS. Applied Structural Equation Modelling for Researchers and Practitioners. 2020;113–22. doi: 10.1108/978-1-78635-883-720161025.
2. Arruda EH, Bentler PM. A Regularized GLS for Structural Equation Modeling. Structural Equation Modeling. 2017;24(5):657–65. doi: 10.1080/10705511.2017.1318392.
3. Altikriti S, Anderson CN. Factor Analysis and Structural Equation Modeling. In: The Encyclopedia of Research Methods in Criminology and Criminal Justice. Wiley; 2021. p. 833–8. doi: 10.1002/9781119111931.ch159.
4. Kline RB. Principles and practice of structural equation modeling. Guilford publications; 2023.
5. Anderson SA and CN. Confirmatory Factor Analysis and Structural Equation Modeling. The Palgrave Handbook of Applied Linguistics Research Methodology. 2018;(March):459–500. doi: 10.1057/978-1-137-59900-1_21.
6. Zulkifli NR, Aimran N, Deni SM. The performance of unweighted least squares and regularized unweighted least squares in estimating factor loadings in structural equation modeling. International Journal of Data and Network Science. 2023;7(3). doi: 10.5267/j.ijdns.2023.6.004.
7. Mîndrilă D. Maximum Likelihood (ML) and Diagonally Weighted Least Squares (DWLS) Estimation Procedures: A Comparison of Estimation Bias with Ordinal and Multivariate Non-Normal Data. International Journal for Digital Society. 2010;1(1):60–6. doi: 10.20533/ijds.2040.2570.2010.0010.
8. Jung S, Takane Y. Regularized common factor analysis. New Trends in Psychometrics. 2007;1(1):1–10. Available from: http://takane.brinkster.net/Yoshio/c042.
9. Jacobucci R, Brandmaier AM, Kievit RA. A Practical Guide to Variable Selection in Structural Equation Modeling by Using Regularized Multiple-Indicators, Multiple-Causes Models. Adv Methods Pract Psychol Sci. 2019;2(1):55–76. doi: 10.1177/2515245919826527.
10. Jacobucci R, Grimm KJ, McArdle JJ. Regularized Structural Equation Modeling. Structural Equation Modeling. 2016;23(4):555–66. doi: 10.1080/10705511.2016.1154793.
11. Bickel PJ, Levina E. Covariance regularization by thresholding. Ann Stat. 2008;36(6):2577–604. doi: 10.1214/08-AOS600.
12. Yuan KH, Chan W. Structural Equation Modeling With Unknown Population Distributions: Ridge Generalized Least Squares. Structural Equation Modeling. 2016;23(2):163–79. doi: 10.1080/10705511.2015.1077335.
13. Bentler PM, Yuan KH. Structural equation modeling with small samples: Test statistics. Multivariate Behav Res. 1999;34(2):181–97. doi: 10.1207/S15327906Mb340203.
14. Jung S, Park JH. Consistent partial least squares path modeling via regularization. Vol. 9, Frontiers in Psychology. 2018. doi: 10.3389/fpsyg.2018.00174.

15. Li X, Jacobucci R. Regularized structural equation modeling with stability selection. Psychol Methods. 2022 Aug;27(4):497–518. doi: 10.1037/met0000389.
16. Finch H. Applied regularization methods for the social sciences. CRC Press; 2022.
17. Md Zin H, Draman S, Shalihin SE, Mohd Rus R, Ahmad Alwi AA. Validity and Reliability of Malay Language Breast Cancer Awareness Scale for Male-to-Female Transgender. IIUM Medical Journal Malaysia. 2023 Jan 1;22(1). doi: 10.31436/imjm.v22i1.2136.
18. Zhang P. Model selection via multifold cross validation. The annals of statistics. 1993;299–313. Available from: https://iri.columbia.edu/~tippett/cv_papers/Zhang1993.
19. Nti IK, Nyarko-Boateng O, Aning J. Performance of machine learning algorithms with different K values in K-fold cross-validation. International Journal of Information Technology and Computer Science. 2021;13(6):61–71. doi: 10.5815/ijitcs.2021.06.05.
20. McDonald RP, Bollen KA. Structural Equations with Latent Variables. Vol. 85, Journal of the American Statistical Association. 1990. 1175 p. doi: 10.2307/2289630.
21. Jureskog KG. Simultaneous factor analysis in several populations. Psychometrika. 1971 Dec;36(4):409–26. doi: 10.1007/BF02291366.
22. Thorndike RM. Book Review : Psychometric Theory (3rd ed.) by Jum Nunnally and Ira Bernstein New York: McGraw-Hill, 1994, xxiv + 752 pp. Appl Psychol Meas. 1995 Sep 26;19(3):303–5. doi: 10.1177/014662169501900308.
23. Henseler J, Hubona G, Ray PA. Using PLS path modeling in new technology research: Updated guidelines. Industrial Management and Data Systems. 2016;116(1):2–20. doi: 10.1108/IMDS-09-2015-0382.
24. Fornell C, Larcker DF. Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. Journal of Marketing Research. 1981 Feb;18(1):39. doi: 10.2307/3151312.
25. Finch WH. SEM Estimation in the Context of Small Samples: Comparison of Latent Variable Models, Single Indicator, Regularized 2-Stage Least Squares and Observed Variable Models. General Linear Model Journal. 2022;46(1):1–19. doi: 10.31523/glmj.046001.001.
26. Hair., Matthews LM, Matthews RL, Sarstedt M. PLS-SEM or CB-SEM: updated guidelines on which method to use. International Journal of Multivariate Data Analysis. 2017;1(2):107. doi: 10.1504/ijmda.2017.10008574.
27. Jung S, Seo DG, Park J. Regularized Exploratory Bifactor Analysis With Small Sample Sizes. Front Psychol. 2020 Apr 9;11. doi: 10.3389/fpsyg.2020.00507.
28. Hair, Risher JJ, Sarstedt M, Ringle CM. The Results of PLS-SEM Article information. European Business Review. 2018;31(1):2–24. doi: 10.1108/EBR-11-2018-0203.
29. Mabel OA, Olayemi OS. A comparison of principal component analysis, maximum likelihood and the principal axis in factor analysis. American Journal of Mathematics and Statistics. 2020;10(2):44–54. doi: 10.5923/j.ajms.20201002.03.
30. Zulkifli NR, Aimran N, Mohd Deni S. The performance of unweighted least squares and regularized unweighted least squares in estimating factor loadings in structural equation modeling. International Journal of Data and Network Science. 2023;7(3):1017–24. doi: 10.5267/j.ijdns.2023.6.004.