

Implementing alternative estimation methods to evaluate the reliability of Likert-scale instruments

Chang Gi Park

Department of Population Health Nursing Science, University of Illinois Chicago, Chicago, IL, USA

Introduction

For the unbiased evaluation of the construct validity of Likert-scale instruments by nursing researchers, polychoric correlation is strongly recommended over Pearson correlation or covariance-based confirmatory factor analysis (CFA) [1]. Similarly, when estimating the reliability of these instruments, the use of polychoric correlation should be considered.

Cronbach's alpha has been observed to underestimate the inter-item reliability of Likert-type ordinal item scales. As a result, the ordinal alpha coefficient, based on the polychoric correlation coefficient, has been introduced for a more accurate estimation of reliability [2,3]. Since the introduction of ordinal alpha, it has gained acceptance as a reliability measure for ordinal categorical item scales [4-9]. Alternative reliability measures for ordinal items have also been proposed [10,11]. However, the application of these reliability estimation methods in nursing studies remains limited, as does the available guidance for nursing researchers regarding how to apply these methods.

Consequently, the purpose of this manuscript is to evaluate the current practice of relying solely on the Cronbach's alpha for the assessment of ordinal Likert-scale instruments. Additionally, this article aims to introduce alternative methods for estimating reliability coefficients, address practical issues associated with the use of software for these estimation approaches, and demonstrate the application of these programs. Although methods other than Cronbach's alpha may be suitable for estimating reliability, nursing researchers may lack sufficient information regarding their implementation.

Alternative reliability estimation methods for Likert scales

Polychoric correlation coefficient as a replacement for Pearson correlation or covariance matrix as an alpha input

The standard formula for the reliability alpha coefficient is not appropriate for ordinal item Likert scales, as Pearson correlation and covariance tend to underestimate associations for ordinal variables. The ordinal alpha coefficient, which estimates reliability based on a polychoric correlation matrix, is the most commonly used measure of reliability for Likert scales [2].

Gadermann et al. [3] used a polychoric correlation matrix to demonstrate the use of polychoric correlation coefficients and factor loading coefficients derived from factor analysis. The average factor loading coefficient was utilized to calculate ordinal alpha. This calculation was demonstrated us-

Received: March 5, 2024

Accepted: March 12, 2024

Corresponding author:

Chang Gi Park
Department of Population Health
Nursing Science, University of
Illinois Chicago, 845 S. Damen Ave.,
MC 802, Chicago 60612, IL, USA
E-mail: parkcg@uic.edu

ing R software (R Foundation for Statistical Computing, Vienna, Austria). However, the ordinal alpha coefficient has been regarded as a consistent measure of theoretical reliability [3]. Since the estimation of the polychoric correlation coefficient relies on the latent interval scale (rather than the actual raw item scores) of the underlying ordered item scores, the ordinal alpha reliability coefficient may be overestimated [12].

Structural equation modeling-based reliability coefficient estimation

Because ordinal alpha is defined based on the polychoric correlations of latent items, it is considered to represent theoretical reliability [2,11,13]. To estimate the association of the actual item-score covariance, researchers have introduced a nonlinear structural equation modeling (SEM)-based reliability estimation method [2,3] for ordered categorical items. This method utilizes the estimated item thresholds, factor loading coefficients, and polychoric correlation matrix derived from the CFA model [6,11]. To compute the ordinal alpha, it is necessary to use model-based population covariances for the categorical scale score.

Other reliability estimation methods

The primary limitation of the ordinal alpha method, which is based on the polychoric correlation coefficient, is that polychoric correlation measures the association between underlying latent variables rather than the actual scale. Consequently, ordinal alpha is an unsuitable measure of reliability for Likert scales with ordered items [11,13]. As alternatives, nonparametric correlations such as Spearman correlation and the Cramer V have been recommended for assessing relationships between ordered categorical variables [13].

Deflation-corrected estimators of reliability (DCER) methods offer another alternative for reliability estimation by applying nonparametric item-score associations to the reliability estimation formula [14-16]. A primary cause of reliability coefficient underestimation is the biased Pearson item-score correlation used by traditional reliability estimators. By employing nonparametric item-score correlations, this bias can be corrected. DCER methods can be used to calculate alpha and omega by incorporating nonparametric association estimation techniques, such as the Goodman adjusted item-score correlation coefficient, Kruskal gamma, and Sommer D.

Item response theory (IRT)-based reliability estimation [6] is another approach for estimating the reliability of ordinal item Likert scales. The ordinal alpha estimation was demonstrated in a recent Kuder-Richardson formula 20 (KR-20) and 21 (KR-21)

in the context of Likert scale data [17].

Application of reliability coefficient for ordinal item scales in nursing research

Since the first application of ordinal alpha in a nursing study [18], several articles have reported the use of ordinal alpha [19-22] and ordinal theta [23]. For the model-based reliability assessment of ordinal item scales, both the estimated ordinal alpha and composite reliability of Likert-scale instruments have been reported, utilizing coefficients from CFA [24]. Additionally, the application of a nonlinear SEM reliability method employing the semTools R library has been described [25-27]. The limited use of ordinal alpha and nonlinear SEM reliability coefficients in the evaluation of Likert scales within nursing research may stem from an absence of detailed guidance regarding why and how nurse researchers should implement these alternative approaches.

Software

A straightforward method for estimating ordinal alpha is available through an online program accessible at <https://r-apps.shinyapps.io/shinyapps/>. Users simply upload a data file, with multiple supported formats that include text, .txt, .csv, .sav, .xls, .xslm, and .xlsx. The output includes the estimated polychoric correlation matrix and the ordinal alpha reliability coefficient, as well as both the raw and standardized alpha and the reliability of each item when dropped.

A second easily implementable method for assessing ordinal reliability involves the Jamovi PPDA library. The R interface provided by Jamovi enables the use of R libraries without the need to directly operate the R program. Jamovi's PPDA library — which stands for Psychometrics & Post-Data Analysis [28] — can estimate ordinal reliability coefficients such as alpha and omega, as well as yielding the polychoric correlation matrix.

Other programs for estimating ordinal alpha are R-based and thus require a basic understanding of R to operate.

- `R_ufs`: This program is the updated version of `R_userfriendlyscience`. Its `scaleStructure()` function can automatically detect ordinal categorical variables and calculate reliability coefficients under both interval and ordinal scale assumptions. When invoked without data, `scaleStructure()` triggers a pop-up window that allows the user to select a data file for reliability coefficient estimation. Depending on the version of R being used, the results may not be displayed on the screen; if this occurs, executing a `plot.new()` command before running

scaleStructure() can resolve the issue. When data are loaded in the R session, entering scaleStructure(dataset name) will present the ordinal alpha estimation, which includes the alpha coefficient based on the interval scale assumption along with omega coefficients.

- R_misty: The item.alpha (dataset name, ordered = TRUE) function is specifically used to determine the ordinal alpha coefficient. The function's output also includes the estimated polychoric correlation matrix.
- R_psych: When the omega function is used with the poly = TRUE option, the ordinal alpha coefficient can be estimated. The output also contains estimated results for the ordinal version of G.6, omega, omega hierarchical (omega H), omega H asymptotic, and omega total. The alpha function can estimate ordinal alpha when given a polychoric correlation matrix as an input. To generate this matrix, the polychoric(dataset) function must first be applied to the dataset. Once the estimated polychoric correlation matrix is obtained, it can then be used as input for the alpha function.
- R_semTools: The reliability function within the semTools library provides estimates of the alpha and omega coefficients [29]. When used along with the CFA function from the lavaan R library cfa (using the option ordered = TRUE), this function yields ordinal alpha and omega coefficients. The compReISEM() function represents an updated version of the reliability function, designed for specific types of estimation. In particular, it allows for the estimation of ordinal reliability, with or without correction for score variance, through the use of the ord.scale and tau.eq options. When ord.scale = TRUE and tau.eq = TRUE, the resulting coefficient is equivalent to the alpha coefficient. However, when ord.scale = FALSE, the reliability coefficient corresponds to the ordinal alpha value. The lrv2ord function can generate a model-estimated population covariance and correlation matrix. This, in turn, facilitates the estimation of the alpha coefficient using the alpha formula.
- DCER approach: This set of calculation methods has been introduced through an Excel worksheet, with detailed step-by-step instructions presented in the corresponding papers [14-16].
- R_irtreliability: IRT-based reliability programs can be utilized for Likert scales [30]. The R_irtreliability program leverages the results from the R mirt package. The options "mrc" and "trc" are used to estimate the marginal and test reliability coefficients, respectively. The mirt package also includes the empirical_rxx function, designed for the calculation

of the reliability coefficient. The R source syntax for estimating reliability using the KR-20 and KR-21 methods, tailored for Likert scale reliability [17], is available at <https://osf.io/rk5e2/>.

Demonstration of reliability coefficient estimation methods

To illustrate the ordinal alpha approach and the other alternative reliability estimation methods, the first five items from the dataset bfi (the Big Five Inventory 25 personality items) were utilized as a single factor. This represents the most frequently accessed dataset available from the 'psych' R library.

In the R session:

```
> library(psych)
> library(ufs)
> data (bfi)
> data <- bfi[,1:5]
> data <- invertitems(data,1) for reverse coding for the first
```

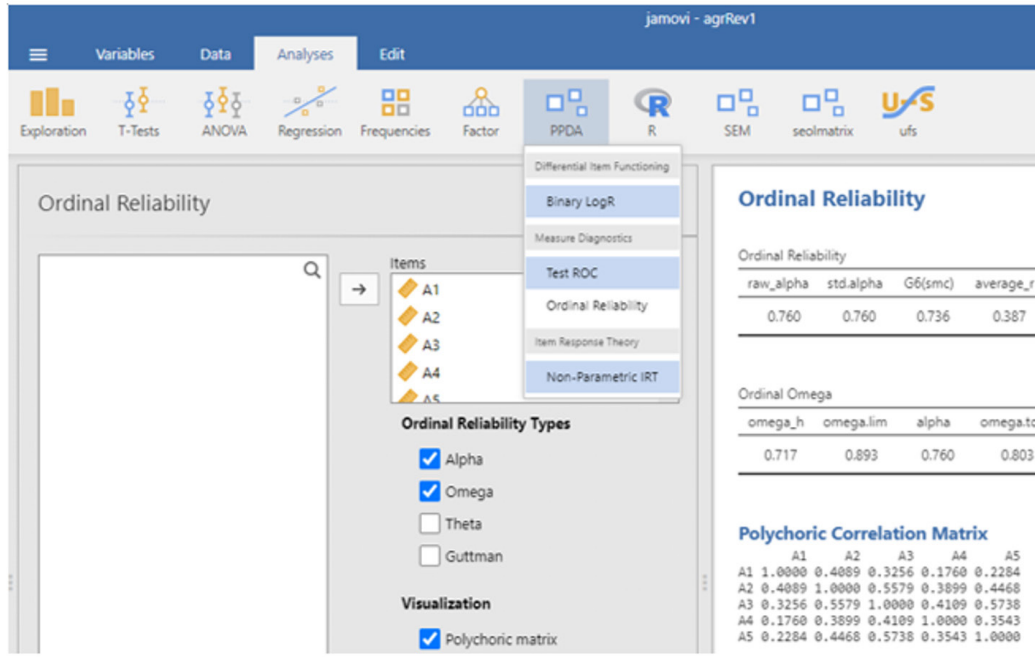
item.

The application of ordinal alpha estimation can be performed using either an online program or Jamovi software. Two examples of ordinal alpha estimation are presented in Figure 1. The process is straightforward and readily accessible for most nursing researchers.

Table 1 shows the reliability coefficient estimations for the Likert scale, as demonstrated with the sample data. Previous studies have demonstrated methods for ordinal reliability estimation using the R libraries psych, ufs, and semTools [3,7,8].

How SPSS users can estimate ordinal alpha

The only method available for estimating ordinal alpha within SPSS is to utilize the R program. However, users can also employ the polychoric correlation function HETCOR without directly using R. This function can be added to the SPSS menu by installing SPSSINC_HETCOR.spe, which is available for download at https://github.com/IBMPredictiveAnalytics/SPSSINC_HETCOR. This installation enables the estimation of polychoric correlation coefficients. Once installed, a new menu item, labeled "heterogeneous correlation," will appear under the "correlation" menu. The polychoric correlation matrix generated by running this function can be copied into an Excel worksheet. In Excel, the average correlation coefficient can be calculated. This average polychoric correlation coefficient is then inserted into the alpha formula: $a = k\bar{r} / (1 + (k-1)\bar{r})$, where k represents the number of items and \bar{r} denotes the average correlation coefficient. For the



Polychoric Matrix and Ordinal Alpha

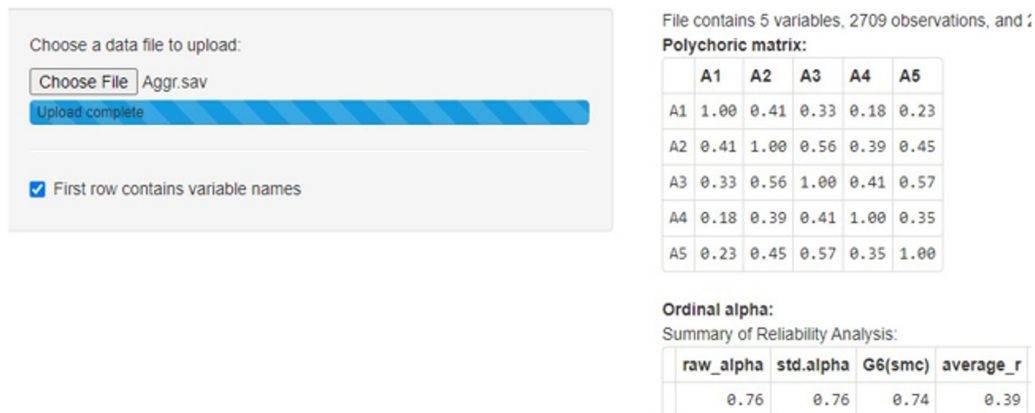


Figure 1. Estimation of ordinal alpha using Jamovi software and an online program.

sample data, the average bivariate polychoric correlation coefficient was 0.387, resulting in a reliability calculation of 0.7594 based on the alpha coefficient formula. Alternatively, the alpha coefficient can be calculated using other bivariate correlation coefficients, such as the Spearman rho and the Kendall tau_b. These methods may enable calculation of the reliability coefficient for Likert-ordered items. For example, the average Spearman inter-item correlation coefficient was 0.35, leading to an estimated alpha coefficient of 0.729. However, due to the multiple steps required to estimate ordinal reliability coefficients with SPSS, the use of alternative programs is strongly recommended.

The estimated ordinal alpha values obtained from various programs (omega, alpha with polychoric correlation matrix, structuralScale, and item.alpha) were consistent, each yielding a value of 0.76. This uniformity arises because all programs employ the same formula and polychoric correlation coefficient. The manually calculated value, using the estimated polychoric correlation coefficients, corroborated this result.

The results of the nonlinear SEM reliability estimation [10,11] included both the ordinal alpha computed using the method of Zumbo et al. [2] and compReSEM, with the tau.eq=TRUE option applied with and without the ord.scale option. To facilitate

Table 1. Demonstration of ordinal reliability estimation methods with sample data

Category	Function	Description	Alpha
Package			
jamovi PPDA	ordinal reliability		0.76
psych	omega	> omega(data, poly = TRUE)	0.76
	alpha	> poly < -polychoric(data) > alpha(poly\$rho)	0.76
ufs	structuralScale	> structuralScale(data, ordered = TRUE)	0.76
misty	item.alpha	> item.alpha(data, ordered = TRUE)	0.76
SEM methods			
lavaan	reliability	> fit < -cfa("F = ~A1+A2+A3+A4+A5", data, ordered = T)	
semTools	compReSEM	> reliability(fit)	0.76
		> compReSEM(fit, tau.eq = T, ord.scale = F)	0.76
		> compReSEM(fit, tau.eq = T, ord.scale = T)	0.70
Nonparametric methods			
R psych & stats	cor	> cor < -cor(data, method = "spearman") > alpha(cor)	0.73
	alpha	> cor < -cor(data, method = "Kendall") > alpha(cor)	0.68
DCER [†]		Rit (max. item-score correlation ratio)	0.72
		gamma	0.78
		Sommer D	0.72
IRT methods			
mirt	mirt	> re < -mirt(data, 1, "gpcm", SE = TRUE)	
	fscores	> theta < -fscores(re, fullscores, SE = TRUE)	0.74
	empirical_rxx	> empirical_rxx(theta)	
irtreliability	irtreliability	> irtreliability(re, "GPCM", rep(6, 5), relcoef = "mrc")	0.74

DCER, Deflation-corrected estimators of reliability; IRT, item response theory; PPDA, psychometrics & post-data analysis; SEM, structural equation modeling.

[†]Only an Excel worksheet demonstration is available. The estimated alpha is provided for comparison purposes.

comparison of the estimated ordinal alpha coefficient, the tau.eq option was set to TRUE. The estimated coefficient was 0.76—identical to the ordinal alpha coefficient—when scale variance was unadjusted (ord.scale = FALSE) and 0.7 with the ord.scale = TRUE option was applied. When tau.eq = FALSE, the estimated reliability coefficient represents the omega value.

The application of a nonparametric correlation estimation method, namely Spearman correlation, to the reliability alpha formula yielded an estimated coefficient of 0.73. However, when applying the Kendall tau_b correlation, the coefficient was 0.68.

Results estimated using the DCER method indicate that the reliability coefficient varies based on the method used to estimate item-score correlation. When the ratio of maximum correlation between an item and the total score was employed, the alpha reliability estimate was 0.72. Utilizing gamma correlation statistics, the estimated alpha reliability coefficient was 0.78, compared to

0.72 when using the Sommer D.

The estimated IRT reliability coefficient was 0.74, as determined using the empirical_rxx function from the mirt program and the irtreliability function from the irtreliability program.

The estimated values for Likert scale reliability using the KR-20 and KR-21 methods [17] were 0.7039 and 0.7016, respectively. The results are consistent with those obtained using the reliability alpha coefficient, differing from ordinal alpha and other alternative methods. However, a limitation of this approach is the current unavailability of software for the program.

Conclusion

Despite reported limitations of the reliability alpha and alternative reliability estimation methods, approaches to estimating reliability for the ordinal Likert scale remain limited. Since the intro-

duction of the ordinal alpha method in 2009, its limitations, as well as those of alternative methods like the nonlinear SEM reliability estimation method, have been acknowledged. However, other approaches, such as the nonparametric DCER and the IRT-based reliability estimation method, have not been widely applied to Likert scales.

Until all available estimation methods are fully validated and accessible to nursing researchers, it is recommended to report not only the traditional Cronbach's alpha but also the ordinal alpha. The consequences of biased reliability estimation extend beyond the development of instruments, as the scales from these instruments may also influence the accuracy of statistical evaluations, including even simple bivariate correlations or two-group comparisons. While adopting a single universally accepted reliability coefficient (such as Cronbach's alpha) may be the preferred practice for a given type of scale, the publication of multiple articles discussing the limitations of Cronbach's alpha and alternative reliability coefficients has prompted increased recognition of the complexities involved in measuring reliability.

Likert-scale instruments are commonly employed in nursing research, making their evaluation essential. While reports have described limitations of and alternatives to the application of the reliability alpha, these have not substantially clarified how the distinct attributes of the ordinal categorical scale may influence the reliability evaluation practices of nursing researchers.

The introduction of ordinal alpha by Zumbo et al. [2] has highlighted additional possibilities, prompting nurse researchers to anticipate the development of a new universally accepted reliability indicator. Several studies have recommended decision-making frameworks for identifying suitable reliability estimation methods for Likert scales [6,8,9]. These guidelines could assist nurse researchers in selecting the appropriate reliability estimation method and in reporting their findings, thereby enhancing the comprehension of the instruments used.

Best practices involve reporting a variety of reliability coefficients to gain a comprehensive understanding of the instrument used. This broader perspective can assist nursing researchers in making more informed decisions to improve the instrument. Relying on a single reliability coefficient can lead to misguided development; therefore, a more informed approach to decision-making is necessary and should be applied in further research.

ORCID

Chang Gi Park, <https://orcid.org/0000-0003-3366-4279>

Authors' contributions

All work was done by Park CG.

Conflict of interest

The author declared no conflict of interest.

Funding

None.

Data availability

Please contact the corresponding author for data availability.

Acknowledgments

None.

References

1. Kiwanuka F, Kopra J, Sak-Dankosky N, Nanyonga RC, Kvist T. Polychoric correlation with ordinal data in nursing research. *Nurs Res.* 2022;71(6):469-476. <https://doi.org/10.1097/NNR.0000000000000614>
2. Zumbo BD, Gadermann AM, Zeisser C. Ordinal versions of coefficients alpha and theta for Likert rating scales. *J Mod Appl Stat Methods.* 2007;6(1):21-29. <https://doi.org/10.56801/10.56801/v6.i.279>
3. Gadermann AM, Guhn M, Zumbo BD. Estimating ordinal reliability for Likert-type and ordinal item response data: a conceptual, empirical, and practical guide. *Pract Assess Res Eval.* 2012;17:1-13. <https://doi.org/10.7275/n560-j767>
4. Peters GJ. The Alpha and the omega of scale reliability and validity: why and how to abandon Cronbach's alpha and the route towards more comprehensive assessment of scale quality. *Eur Health Psychol.* 2014;16(2):56-69. <https://doi.org/10.31234/osf.io/h47fv>
5. Kelley K, Pornprasertmanit S. Confidence intervals for population reliability coefficients: Evaluation of methods, recommendations, and software for composite measures. *Psychol Methods.* 2016;21(1):69-92. <https://doi.org/10.1037/a0040086>
6. Viladrich C, Angulo-Brunet A, Doval E. A journey around alpha and omega to estimate internal consistency reliability.

- Ann Psychol. 2017;33(3):755-782. <https://doi.org/10.6018/analesps.33.3.268401>
7. McNeish D. Thanks coefficient alpha, we'll take it from here. *Psychol Methods*. 2018;23(3):412-433. <https://doi.org/10.1037/met0000144>
 8. Flora DB. Your coefficient alpha is probably wrong, but which coefficient omega is right? A tutorial on using R to obtain better reliability estimates. *Adv Methods Pract Psychol Sci*. 2020;3(4):484-501. <https://doi.org/10.1177/2515245920951747>
 9. Doval E, Viladrich C, Angulo-Brunet A. Coefficient alpha: the resistance of a classic. *Psicothema*. 2023;35(1):5-20. <https://doi.org/10.7334/psicothema2022.321>
 10. Green SB, Yang Y. Reliability of summed item scores using structural equation modeling: an alternative to coefficient alpha. *Psychometrika*. 2009;74(1):155-167. <https://doi.org/10.1007/s11336-008-9099-3>
 11. Yang Y, Green SB. Evaluation of structural equation modeling estimates of reliability for scales with ordered categorical items. *Methodology: Eur J Res Methods Behav Soc Sci*. 2015; 11(1):23-34. <https://doi.org/10.1027/1614-2241/a000087>
 12. Revelle W, Condon DM. Reliability from α to ω : a tutorial. *Psychol Assess*. 2019;31(12):1395-1411. <https://doi.org/10.1037/pas0000754>
 13. Chalmers RP. On misconceptions and the limited usefulness of ordinal alpha. *Educ Psychol Meas*. 2018;78(6):1056-1071. <https://doi.org/10.1177/0013164417727036>
 14. Metsämuuronen J. Attenuation-corrected estimators of reliability. *Appl Psychol Meas*. 2022;46(8):720-737. <https://doi.org/10.1177/01466216221108131>
 15. Metsämuuronen J. Typology of deflation-corrected estimators of reliability. *Front Psychol*. 2022;13:891959. <https://doi.org/10.3389/fpsyg.2022.891959>
 16. Metsämuuronen J. Seeking the real reliability: why the traditional estimators of reliability usually fail in achievement testing and why the deflation-corrected coefficients could be better options. *Pract Assess Res Eval*. 2023;28:Article 10. <https://doi.org/10.7275/pare.1264>
 17. Foster RC. KR20 and KR21 for some nondichotomous data (it's not just Cronbach's alpha). *Educ Psychol Meas*. 2021; 81(6):1172-1202. <https://doi.org/10.1177/0013164421992535>
 18. Lee JJ, Clarke CL. Nursing students' attitudes towards information and communication technology: an exploratory and confirmatory factor analytic approach. *J Adv Nurs*. 2015;71(5):1181-1193. <https://doi.org/10.1111/jan.12611>
 19. Rodríguez-Martín B, Martín-Martín R, Suhonen R. Individualized Care Scale-patient: a Spanish validation study. *Nurs Ethics*. 2019;26(6):1791-1804. <https://doi.org/10.1177/0969733018769351>
 20. Säfström E, Nasstrom L, Liljeroos M, Nordgren L, Årestedt K, Jaarsma T, et al. Patient Continuity of Care Questionnaire in a cardiac sample: a confirmatory factor analysis. *BMJ Open*. 2020;10(7):e037129. <https://doi.org/10.1136/bmjopen-2020-037129>
 21. Sterner A, Säfström E, Palmér L, Ramstrand N, Hagiwara MA. Development and initial validation of an instrument to measure novice nurses' perceived ability to provide care in acute situations - PCAS. *BMC Nurs*. 2020;19:13. <https://doi.org/10.1186/s12912-020-0406-3>
 22. Agrazal-García J, Gordon-de Isaacs L, Tuñón R. Design and validation of a questionnaire to measure hantavirus preventive practices in an endemic community. *Rev Peru Med Exp Salud Publica*. 2022;39(1):47-54. <https://doi.org/10.17843/rp-mesp.2022.391.9740>
 23. Schulz E, Gothe RM, Them C, Tufan İ, Mueller G. Self-reported questionnaire of functional health of home-dwelling elderly people living in Austria: based on the patient questionnaire of the standardized assessment of elderly people in primary care. *J Nurs Meas*. 2017;25(2):238-256. <https://doi.org/10.1891/1061-3749.25.2.238>
 24. Barbaranelli C, Lee CS, Vellone E, Riegel B. Dimensionality and reliability of the self-care of heart failure index scales: further evidence from confirmatory factor analysis. *Res Nurs Health*. 2014;37(6):524-537. <https://doi.org/10.1002/nur.21623>
 25. Juliá-Sanchis R, Cabañero-Martínez MJ, Leal-Costa C, Fernández-Alcántara M, Escribano S. Psychometric properties of the health professionals communication skills scale in university students of health sciences. *Int J Environ Res Public Health*. 2020;17(20):7565. <https://doi.org/10.3390/ijerph17207565>
 26. Esteban-Burgos AA, Fernández-Alcántara M, Escribano S, Perpiñá-Galvañ J, Campos-Calderón CP, Cabañero-Martínez MJ. Psychometric properties of the Spanish version of the Fatigue Assessment Scale in caregivers of palliative care patients. *J Clin Med*. 2022;11(14):3999. <https://doi.org/10.3390/jcm11143999>
 27. Maziero VG, Rocha FL, Campos JA, Dias BM. Authentic Leadership Questionnaire applied to Brazilian nurses: evidence of validity. *Rev Lat Am Enfermagem*. 2022;30:e3607. <https://doi.org/10.1590/1518-8345.5868.3607>
 28. Lucasjriesen, Jonathon-love. *Psychometrics & Post-Data*

- Analysis for Jamovi [Internet]. GitHub, Inc.; 2023 [cited 2024 Mar 5]. Available from: <https://github.com/lucasjfriesen/jamoviPsychoPDA>
29. Jorgensen TD, Pornprasertmanit S, Schoemann AM, Rosseel Y. semTools: Useful tools for structural equation modeling (R package Version 0.5-3) [Computer software]. Authors; 2022 [cited 2024 Mar 5]. Available from: <https://CRAN.R-project.org/package=semTools>
30. Andersson B. Irtreliability: item response theory reliability. Version 0.1-1 [Computer software]. Author; 2022 [cited 2024 Mar 5]. Available from: <https://CRAN.R-project.org/package=irtreliability>