ATATÜRK ÜNİVERSİTESİ YAYINLARI
ATATURK UNIVERSITY PUBLICATIONS

Derleme *Review*   DOI: 10.62425/jopres.1789821

Estimating Internal Consistency: Did We Choose the Right Coefficient?

İç Tutarlılığın Tahmini: Doğru Katsayıyı Seçtik mi?

Néstor MONTORO-PÉREZ[1,2]

[1]Department of Nursing, Faculty of Health Sciences, Person-Centred Care and Health Outcomes Innovation Group, University of Alicante, San Vicente del Raspeig, Alicante, Spain.

[2]GREIACC Research Group. La Fe Health Research Institute. Valencia, Spain

Silvia ESCRIBANO[2]

[1]Department of Nursing, Faculty of Health Sciences, Person-Centred Care and Health Outcomes Innovation Group, University of Alicante, San Vicente del Raspeig, Alicante, Spain

ABSTRACT

Internal consistency is a concept extensively used in academic discourse, yet its definition remains debated. In the context of validation studies, it is noteworthy that, although internal consistency is commonly assessed, some studies could benefit from employing more accurate estimators that are better suited to the underlying factorial structure. These editorial addresses the various recommended estimators for calculating internal consistency based on the characteristics of the studied model. We explore one-dimensional measures, identifying when estimators such as α are suitable, particularly for tau-equivalent models. For congeneric measurement models, coefficient ω is recommended. We also discuss complex models incorporating multidimensional structures, including essential unidimensionality, scales with multiple correlated or uncorrelated factors, and higher-order models. Researchers should avoid reporting the total internal consistency of the instrument unless unidimensionality or a higher-order factor structure has been demonstrated. When data are approximately unidimensional, measures are congeneric with moderate factor loadings, and sample sizes are large, it is reasonable to report both α and ω.

Keywords: Internal consistency, reliability, psychometrics

ÖZ

İç tutarlılık, akademik yazında yaygın olarak kullanılan bir kavram olmasına rağmen, tanımı hâlâ tartışmalıdır. Geçerlilik çalışmaları bağlamında, iç tutarlılığın sıklıkla değerlendirildiği fakat bazı çalışmaların altta yatan faktör yapısına daha uygun ve daha doğru kestiricileri kullanmaktan fayda görebileceği dikkate değerdir. Bu editoryal, çalışılan modelin özelliklerine göre iç tutarlılığın hesaplanmasında önerilen çeşitli kestiricileri ele almaktadır. Tek boyutlu ölçümleri inceliyor, α gibi kestiricilerin özellikle tau-eşdeğer modellerde ne zaman uygun olduğunu belirtiyoruz. Konjenerik ölçüm modelleri için ω katsayısı önerilmektedir. Ayrıca, zorunlu tek boyutluluk, birden fazla ilişkili veya ilişkisiz faktöre sahip ölçekler ve üst düzey modeller dahil olmak üzere çok boyutlu yapıları içeren karmaşık modelleri de tartışıyoruz. Araştırmacılar, tek boyutluluk veya üst düzey faktör yapısı gösterilmeden aracın toplam iç tutarlılığını raporlamaktan kaçınmalıdır. Veriler yaklaşık olarak tek boyutlu olduğunda, ölçümler orta düzey faktör yüklerine sahip konjenerik ise ve örneklem büyüklüğü büyükse, hem α hem de ω'nın raporlanması makuldür.

Anahtar Kelimeler: İç tutarlılık, güvenirlik, psikometri

Geliş Tarihi/Received	23.09.2025
Kabul Tarihi/Accepted	07.01.2025
Yayın Tarihi/Publication Date	28.02.2026

Sorumlu Yazar/Corresponding author: Néstor MONTORO-PÉREZ
E-mail: nestor.montoro@ua.es

Cite this article: Montoro-Pérez, N., & Escribano, S. (2026). Estimating internal consistency: Did we choose the right coefficient? A discussion. *Journal of Psychometric Research, 4*(1), 70-76.

Introduction

The concept of internal consistency has been extensively utilised in academic discourse, yet its definition remains a subject of considerable debate (Sijtsma, 2008; Tang et al., 2014). Cronbach (1951) initially employed the terms "internal consistency" and "homogeneity" interchangeably, asserting that "an internally consistent or homogeneous test should be independent of test length". In contrast, Revelle (1979) conceptualised the term as “the degree to which all items in a test measure the same construct, or the general factor saturation”. Other scholars applied the term to denote the interrelatedness of items, distinguishing it from homogeneity, which they posited refers to the unidimensionality of a set of test items (Green et al., 1977; McDonald, 1981; Miller, 1995; Mokkink et al., 2010; Viladrich et al., 2017). This multiplicity of meanings underscores the need for greater clarity and consensus in the conceptualisation of internal consistency within the academic community (Tang et al., 2014).

Internal consistency is commonly expressed through reliability coefficients such as alpha (α) or omega (ω), among others, within the framework of Classical Test Theory (CTT) (Doval et al., 2023; Viladrich et al., 2017). In this context, the calculation of internal consistency is applicable only to measures grounded in a reflective measurement model, wherein all items are manifestations of the same underlying construct. Such items, known as effect indicators, are characterised by their anticipated high correlation and interchangeability (Jarvis et al., 2003; Mokkink et al., 2024). Consequently, it is widely recommended to consistently fit a measurement model and derive reliability coefficients from the different parameter estimates (Crutzen & Peters, 2015; Graham, 2006; Viladrich et al., 2017).

Depending on the dimensionality and measurement model, different estimators can either understate or overstate the true reliability of the scores (Doval et al., 2023; Gu et al., 2012; Sijtsma, 2008; Viladrich et al., 2017). In fact, misapplication of internal consistency coefficients is not merely a technical concern but can have substantive and clinical consequences. When a single α or ω is reported for a clearly multidimensional measure, a high coefficient can create unwarranted confidence in total scores and encourage their use for comparing groups or evaluating treatment effects, even though they do not represent a single construct (Doval et al., 2023; Viladrich et al., 2017). Conversely, failing to recognise essential unidimensionality and discarding a reliable total score may lead researchers to work with fragmented constructs and to lose unnecessary statistical power, with potential repercussions for patient classification, screening decisions and the monitoring of change over time (Crutzen & Peters, 2015; Doval et al., 2023; Prinsen et al., 2018; Viladrich et al., 2017).

Internal consistency, as defined by COSMIN (COnsensus-based Standards for the selection of health Measurement INstruments), is a concept within one of the measurement properties (reliability) (Mokkink et al., 2010) and can be operationalised using various statistical parameters, with numerical values ranging from 0 to 1. The scientific field generally concur that when developing novel measures, coefficients exceeding 0.70 are deemed satisfactory. However, in scenarios involving high-stakes individual decision-making processes, a more stringent threshold of 0.90 or higher is necessitated to ensure robust reliability (Nunnally & Bernstein, 1994; Thorndike, 1995). Nevertheless, it is also important to emphasise that when there is high internal consistency, scientific evidence suggests that there may be redundant items (Streiner, 2003). In this discussion, we aim to present a range of recommended estimators for calculating internal consistency. These recommendations are based on the characteristics of the studied model and are aligned with the latest advances in psychometrics.

Internal Consistency Coefficient for One-dimensional Measures

A one-dimensional measure represents a methodological tool aimed at consolidating several related items into a single numerical value, which captures a singular underlying construct (Lord & Novick, 1968; Thorndike, 1995).

In this sense, researchers can identify essentially tau-equivalent models and models that fulfil tau-equivalent criteria. Tau-equivalent models are those in which all items exhibit the same factor loadings. Additionally, we can consider models that fulfil tau-equivalent criteria, specifically, when factor loadings average approximately 0.70 and their variability remains within ±0.20 (Berge, 2014; Viladrich et al., 2017).

When these criteria are met, the recommended estimator would be α (see Table 1), as it has been demonstrated to be an unbiased estimator of internal consistency reliability (Viladrich et al., 2017). Under these circumstances, α remains a viable point estimator of internal consistency reliability, potentially preferred for its computational simplicity compared with ω (Doval et al., 2023; Viladrich et al., 2017).

Table 1

Criteria for using the two types of α

Criterion	Cronbach's α	Ordinal α
Type of response scale*	Continuous, interval/ratio scales	Ordinal (e.g. Likert), categorical
Correlation matrix⸸	Pearson	Polychoric
Model assumption	Normality, linearity, tau-equivalence	Ordinal latent variables, non-lineear relationships

Note. Based on the scientific evidence provided by Cho and Béland (2025), Doval et al. (2023), Gadermann et al. (2012), Sijtsma and Pfadt (2021), Sijtsma (2008), Viladrich et al. (2017), and Zumbo and Kroc (2019)

* According to Doval et al. (2023), a scale with five or more response categories can be considered as a continuous item,

⸸ Using an inappropriate matrix can lead to incorrect factor solutions, poor model fit, and invalid interpretations. For example, Pearson correlations applied to ordinal data might fail to capture the true associations, while polychoric correlations applied to continuous data could overfit the model.

On the other hand, congeneric measurement models operationalise the premise that item factor loadings inherently exhibit variability rather than uniformity (Berge, 2014; Jöreskog, 1971). Unlike restrictive tau-equivalent assumptions, congeneric models acknowledge the empirically observed heterogeneity in items, which is often more frequently encountered in the real-world context of measurement instruments (Viladrich et al., 2017). When the congeneric measurement model provides a good fit to the data, but the more stringent essentially tau-equivalent model does not, it is recommended to estimate the internal consistency of the total or mean score of the items using coefficient ω (see Table 2) (Doval et al., 2023; Viladrich et al., 2017). Notably, research has demonstrated that α tends to yield lower values compared to ω (Doval et al., 2023; Raykov, 1997; Viladrich et al., 2017). This suggests that α can be considered the lower limit of reliability (Raykov, 1997).

Table 2

Criteria for Using α and ω

Criterion	α	ω
Model Assumptions	Tau-equivalence criteria	Congeneric
Factor Loadings	Homogeneous (±0.20)	Heterogeneous
Error Structure	Uncorrelated	Uncorrelated or correlated*
Recommended When	Factor loadings ~0.70 ±0.20	Significant factor loading differences or model misfit

Note. Based on the scientific evidence provided by Doval et al. (2023) and Viladrich et al. (2017);

* ω remains vulnerable to correlated errors, taking into account the Structural Equation Model (SEM); Gu et al. (2012) found that α remains largely unbiased regardless of the number of non-tau-equivalent items or the variability in factor loadings, the presence of correlated errors degrades the performance of coefficient alpha; package in R, documented by Zhang and Yuan (2015), estimates α and ω coefficients and their confidence intervals, handling missing data and outliers.

Advancing this discussion, we find the correlated errors models. Correlated errors models refer to systematic relationships between measurement errors across different items, challenging the fundamental assumption that errors are random and uncorrelated (Viladrich et al., 2017). Correlated errors can significantly impact reliability estimates, potentially leading to under- or overestimation of reliability coefficients. These correlations may arise from various sources, including measurement methods, shared environmental factors, or response biases in surveys (Doval et al., 2023; Green & Hershberger, 2000; Viladrich et al., 2017).

When models with correlated errors are identified, it is recommended to estimate internal consistency using coefficient ωcorrected for correlated errors (Doval et al., 2023; Green & Hershberger, 2000; Viladrich et al., 2017). However, to address the bias in correlated errors models, it is essential to incorporate the covariance between errors in both the model parameter estimation and the ω formula (Doval et al., 2023; Raykov, 2004; Viladrich et al., 2017). This approach builds upon Bollen's (1980) original unstandardised factor formulation. By implementing these modifications, researchers can enhance the accuracy of their analyses and mitigate potential distortions in their results (see also Table 2) (Doval et al., 2023; Viladrich et al., 2017).

Internal Consistency Coefficient for Multi-dimensional Measures

This section discusses complex models that incorporate multidimensional structures, including the concept of essential unidimensionality. It also examines scales with multiple factors, which may be either correlated or uncorrelated (Reise et al., 2012a; Viladrich et al., 2017), as well as higher-order models (Flora, 2020; Reise, 2012b; Viladrich et al., 2017).

Essential unidimensionality measures refer to whether a scale functions predominantly as a single construct despite potential multidimensionality arising from minor secondary factors. A pivotal methodological framework for evaluating this concept involves bifactor measurement models (Reise, 2012b). In such models, the correlations between items can be explained by a general factor that represents the shared variance among all items and/or a set of group factors in which the variance, beyond that accounted for by the general factor, is shared among subsets of items presumed to be highly similar in content (Doval et al., 2023; Reise, 2012b; Rodriguez et al., 2015; Viladrich et al., 2017). These group factors, which may reflect spurious associations (e.g., method effects or item wording artefacts), are explicitly modelled to disentangle their influence from the general factor (Reise, 2012b; Rodriguez et al., 2015).

In this sense, is recommended ωhierarchical, as conceptualised by Zinbarg et al. (2005), quantifies reliability by exclusively attributing true score variance to a general factor while relegating variance from specific factors—such as method effects or item-specific variance—to the error term in the denominator. This formulation deliberately excludes spurious variance sources (e.g., response biases or measurement artifacts) from the numerator, ensuring reliability estimates reflect only systematic variance attributable to the target construct (Doval et al., 2023; Reise, 2012b; Rodriguez et al., 2015; Viladrich et al., 2017). Green and Yang (2015) advocate for concurrently reporting ωhierarchical and ωtotal. Whereas ωhierarchical isolates the general factor's contribution, ωtotal incorporates variance from both general and specific factors, providing complementary perspectives on scale performance. The degree of congruence between these coefficients offers empirical insight into unidimensionality (Doval et al., 2023; Reise, 2012b; Rodriguez et al., 2015; Viladrich et al., 2017). Also, the scientific literature recommends calculating ωhierarchical subscale (an index reflecting the reliability of a subscale score after controlling for the variance due to the general factor). This can be consulted in more detail in Reise et al. (2012a; 2012b) and Rodriguez et al. (2015).

Correlated factors measures represent structures in which latent variables (factors) within a measurement model are interrelated (Bentler, 1990). In this context, the coefficients αtotal and ωtotal are not suitable for multidimensional tests measuring different constructs that do not share one general factor (Doval et al., 2023). Nevertheless, once the different factors have been identified, these coefficients can be calculated for each subscale separately (Bentler, 2021; Doval et al., 2023; Prinsen et al., 2018; Sijtsma & Pfadt, 2021). Therefore, it is recommended to calculate the internal consistency for each subscale following the recommendations previously outlined in Table 2. Researchers should avoid reporting the total internal consistency of the instrument unless unidimensionality or a higher-order factor structure has been demonstrated (Doval et al., 2023; Prinsen et al., 2018).

Higher-order models represent a hierarchical structure of latent variables in factor analysis, comprising lower-order factors that directly influence observed variables and a higher-order factor that affects these lower-order factors (Flora, 2020; Stout, 1987; Viladrich et al., 2017).

In such models, the reliability measure omega-higher-order (ωho) becomes relevant. ωho quantifies the proportion of total-score variance attributable to the higher-order factor, calculated using parameter estimates from the model. The significance of ωho lies in its ability to represent the reliability of a total score in measuring a single construct that influences all items, despite the test's multidimensional nature (Flora, 2020; Yung et al., 1999).

Conclusion and Recommendations

The choice of internal consistency coefficient should be explicitly guided by the dimensionality and the type of measurement model. For tau equivalent scales (i.e. homogeneous factor loadings around 0.70 ± 0.20), α is an appropriate and computationally simple estimator (Doval et al., 2023; Viladrich et al., 2017). When the scale is clearly congeneric (heterogeneous loadings), ω is preferable because it incorporates the actual loading pattern and usually provides less biased estimates than α (Doval et al., 2023; Viladrich et al., 2017). For essentially unidimensional instruments modelled with bifactor structures, ωhierarchical and ωtotal are recommended to quantify internal consistency (Doval et al., 2023; Green & Yang, 2015; Reise, 2012b; Rodriguez et al., 2015; Viladrich et al., 2017; Zinbarg et al., 2005). Also, it is recommended to calculate ωhierarchical subscale, an index that reflects the internal consistency of a subscale score after controlling for the variance attributable to the general factor. Finally, in higher order models with several first order factors and a second order factor, ωho can be used to evaluate how much of the total score variance is attributable to that higher order construct, whereas internal consistency for each first order dimension should be reported separately (Flora, 2020; Yung et al., 1999).

Although all these coefficients are intended to estimate the ratio of true score variance to observed score variance, they rest on different assumptions and can therefore yield different values (Doval et al., 2023; Viladrich et al., 2017). Using an inappropriate coefficient may lead to overestimating precision and to misleading substantive or clinical decisions; conversely, failing to recognise essential unidimensionality and reporting only subscale indices can underestimate the internal consistency of a well defined general construct (Doval et al., 2023; Sijtsma, 2008; Viladrich et al., 2017).

Nevertheless, considering the findings of Doval et al. (2023), it is equally reasonable to use α or ω when data are approximately unidimensional, measures are congeneric with moderate factor loadings, and sample sizes are large, as both coefficients perform similarly under these conditions. A conservative approach, as recommended by Doval et al. (2023) and Revelle and Condon (2019), involves reporting both α and ω based on these criteria. Reporting α allows comparison across studies, given its widespread use, while ω provides a model based reliability estimate. If there is a notable discrepancy between α and ω, it is important to explore potential causes.

Finally, although this discussion offers a necessarily concise and technical overview, readers interested in more extensive mathematical details and formal formulations are encouraged to consult the cited sources.

Etik Komite Onayı: Bu makale derleme niteliğinde olup insan deneklerden orijinal veri toplama içermediğinden, etik kurul onayının bildirilmesi gerekli değildir.

Hakem Değerlendirmesi: Dış bağımsız.

Yazar Katkıları: Fikir- N.M.P., S.E.; Tasarım- N.M.P., S.E.; Denetleme- N.M.P., S.E.; Kaynaklar- N.M.P., S.E.; Veri Toplanması ve/veya İşlemesi- N.M.P., S.E.; Analiz ve/ veya Yorum- N.M.P., S.E.; Literatür Taraması- N.M.P., S.E.; Yazıyı Yazan- N.M.P., S.E.; Eleştirel İnceleme- N.M.P., S.E.

Teşekkür: Amsterdam UMC, Vrije Universiteit Amsterdam ve Amsterdam Halk Sağlığı Araştırma Enstitüsü Epidemiyoloji ve Veri Bilimi Bölümü'nden Dr. Lidwine B. Mokkink'e, bu başyazıyı incelemedeki değerli katkılarından dolayı en içten şükranlarımızı sunuyoruz. Ölçme metodolojisi ve psikometri alanındaki uzmanlığı, çalışmamızın kalitesini ve titizliğini önemli ölçüde artırmıştır.

Çıkar Çatışması: Yazarlar, çıkar çatışması olmadığını beyan etmiştir.

Finansal Destek: Yazarlar, bu çalışma için finansal destek almadığını beyan etmiştir.

Yapay Zeka Kullanımı: Yapay zekâ araçları yalnızca metnin yazım ve dilbilgisi düzeltmesi için kullanılmıştır.

Ethics Committee Approval: This article is a review and does not involve original data collection from human subjects; therefore, reporting ethical committee approval is not required.

Peer-review: Externally peer-reviewed.

Author Contributions: Concept -N.M.P., S.E.; Design- N.M.P., S.E.; Supervision- N.M.P., S.E.; Resources- N.M.P., S.E.; Data Collection and/or Processing- N.M.P., S.E.; Analysis and/or Interpretation- N.M.P., S.E.; Literature Search- N.M.P., S.E.; Writing Manuscript- N.M.P., S.E.; Critical Review- N.M.P., S.E.

Acknowledgements: We extend our sincere gratitude to Lidwine B. Mokkink, PhD, from the Department of Epidemiology and Data Science at Amsterdam UMC, Vrije Universiteit Amsterdam, and the Amsterdam Public Health Research Institute, for her valuable contribution in reviewing this editorial. Her expertise in measurement methodology and psychometrics has significantly enhanced the quality and rigour of our work.

Conflict of Interest: The authors have no conflicts of interest to declare.

Financial Disclosure: The authors declared that this study has received no financial support.

Use of Artificial Intelligence: Artificial intelligence tools were used only for grammar and language revision of the manuscript.

References

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*(2), 238-246. https://doi.org/10.1037/0033-2909.107.2.238

Bentler, P. M. (2021). Alpha, FACTT, and beyond. *Psychometrika, 86*(4), 861-868. https://doi.org/10.1007/s11336-021-09797-8

Berge, J. M. F. T. (2014). *Tau-equivalent and congeneric measurements.* Wiley StatsRef: Statistics Reference Online. https://doi.org/10.1002/9781118445112.stat06393

Bollen, K. A. (1980). Issues in the comparative measurement of political democracy. *American Sociological Review, 45*(3), 370. https://doi.org/10.2307/2095172

Cho, E., & Béland, S. (2025). Reliability in unidimensional ordinal data: A comparison of continuous and ordinal estimators. *Psychological Methods*. https://doi.org/10.1037/met0000739

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297-334. https://doi.org/10.1007/bf02310555

Crutzen, R., & Peters, G. Y. (2015). Scale quality: Alpha is an inadequate estimate and factor-analytic evidence is needed first of all. *Health Psychology Review, 11*(3), 242-247. https://doi.org/10.1080/17437199.2015.1124240

Doval, E., Viladrich, C., & Angulo-Brunet, A. (2023). Coefficient alpha: The resistance of a classic. *Psicothema, 35*(1), 5-20. https://doi.org/10.7334/psicothema2022.321

Flora, D. B. (2020). Your coefficient alpha is probably wrong, but which coefficient omega is right? A tutorial on using r to obtain better reliability estimates. *Advances in Methods and Practices in Psychological Science, 3*(4), 484-501. https://doi.org/10.1177/2515245920951747

Gadermann, A. M., Guhn, M., & Zumbo, B. D. (2012). Estimating ordinal reliability for likert-type and ordinal item response data: A conceptual, empirical, and practical guide. *Practical Assessment, Research & Evaluation, 17*(3), 1-13. https://doi.org/10.7275/n560-j767

Graham, J. M. (2006). Congeneric and (essentially) tau-equivalent estimates of score reliability. *Educational and Psychological Measurement, 66*(6), 930-944. https://doi.org/10.1177/0013164406288165

Green, S. B., & Hershberger, S. L. (2000). Correlated errors in true score models and their effect on coefficient alpha. *Structural Equation Modeling A Multidisciplinary Journal, 7*(2), 251-270. https://doi.org/10.1207/s15328007sem0702_6

Green, S. B., & Yang, Y. (2015). Evaluation of dimensionality in the assessment of internal consistency reliability: Coefficient alpha and omega coefficients. *Educational Measurement Issues And Practice, 34*(4), 14-20. https://doi.org/10.1111/emip.12100

Green, S. B., Lissitz, R. W., & Mulaik, S. A. (1977). Limitations of coefficient alpha as an index of test unidimensionality1. *Educational And Psychological Measurement, 37*(4), 827-838. https://doi.org/10.1177/001316447703700403

Gu, F., Little, T. D., & Kingston, N. M. (2012). Misestimation of reliability using coefficient alpha and structural equation modeling when assumptions of tau-equivalence and uncorrelated errors are violated. *Methodology, 9*(1), 30-40. https://doi.org/10.1027/1614-2241/a000052

Jarvis, C. B., MacKenzie, S. B., & Podsakoff, P. M. (2003). A critical review of construct indicators and measurement model misspecification in marketing and consumer research. *Journal of Consumer Research, 30*(2), 199-218. https://doi.org/10.1086/376806

Jöreskog, K. G. (1971). Statistical analysis of sets of congeneric tests. *Psychometrika, 36*(2), 109-133. https://doi.org/10.1007/bf02291393

Lord, F. M., & Novick, M. R. (1968). Statistical theories of mental test scores. Addison-Wesley.

McDonald, R. P. (1981). The dimensionality of tests and items. *British Journal of Mathematical and Statistical Psychology, 34*(1), 100-117. https://doi.org/10.1111/j.2044-8317.1981.tb00621.x

Miller, M. B. (1995). Coefficient alpha: A basic introduction from the perspectives of classical test theory and structural equation modeling. *Structural Equation Modeling A Multidisciplinary Journal, 2*(3), 255-273. https://doi.org/10.1080/10705519509540013

Mokkink, L. B., Elsman, E. B., & Terwee, C. B. (2024). COSMIN guideline for systematic reviews of patient-reported outcome measures version 2.0. *Quality of Life Research, 33*(11), 2929-2939. https://doi.org/10.1007/s11136-024-03761-6

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., Bouter, L. M., & de Vet, H. C. (2010). The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of Clinical Epidemiology, 63*(7), 737-745. https://doi.org/10.1016/j.jclinepi.2010.02.006

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.

Prinsen, C. A. C., Mokkink, L. B., Bouter, L. M., Alonso, J., Patrick, D. L., De Vet, H. C. W., & Terwee, C. B. (2018). COSMIN guideline for systematic reviews of patient-reported outcome measures. *Quality of Life Research, 27*(5), 1147-1157. https://doi.org/10.1007/s11136-018-1798-3

Raykov, T. (1997). Estimation of composite reliability for congeneric measures. *Applied Psychological Measurement, 21*(2), 173-184. https://doi.org/10.1177/01466216970212006

Raykov, T. (2004). Point and interval estimation of reliability for multiple-component measuring instruments via linear constraint covariance structure modeling. *Structural Equation Modeling A Multidisciplinary Journal, 11*(3), 342-356. https://doi.org/10.1207/s15328007sem1103_3

Reise, S. P., Bonifay, W. E., & Haviland, M. G. (2012a). Scoring and modeling psychological measures in the presence of multidimensionality. *Journal of Personality Assessment, 95*(2), 129-140. https://doi.org/10.1080/00223891.2012.725437

Reise, S. P. (2012b). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research, 47*(5), 667-696. https://doi.org/10.1080/00273171.2012.715555

Revelle, W. (1979). Hierarchical cluster analysis and the internal structure of tests. *Multivariate Behavioral Research, 14*(1), 57-74. https://doi.org/10.1207/s15327906mbr1401_4

Revelle, W., & Condon, D. M. (2019). Reliability from α to ω: A tutorial. *Psychological Assessment, 31*(12), 1395-1411. https://doi.org/10.1037/pas0000754

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2015). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods, 21*(2), 137-150. https://doi.org/10.1037/met0000045

Sijtsma, K. (2008). On the use, the misuse, and the very limited usefulness of cronbach's alpha. *Psychometrika, 74*(1), 107-120. https://doi.org/10.1007/s11336-008-9101-0

Sijtsma, K., & Pfadt, J. M. (2021). Part II: On the use, the misuse, and the very limited usefulness of cronbach's alpha: Discussing lower bounds and correlated errors. *Psychometrika, 86*(4), 843-860. https://doi.org/10.1007/s11336-021-09789-8

Stout, W. (1987). A nonparametric approach for assessing latent trait unidimensionality. *Psychometrika, 52*(4), 589-617. https://doi.org/10.1007/bf02294821

Streiner, D. L. (2003). Starting at the beginning: An introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment, 80*(1), 99-103. https://doi.org/10.1207/s15327752jpa8001_18

Tang, W., Cui, Y., & Babenko, O. (2014). Internal consistency: Do we really know what it is and how to assess it? *Journal of Psychology & Behavioral Science, 2*(2), 205–220. https://jpbs.thebhprj.org/journals/jpbs/Vol_2_No_2_June_2014/13.pdf

Thorndike, R. M. (1995). Book review: Psychometric Theory (3rd ed.) by Jum Nunnally and Ira Bernstein New York: McGraw-Hill, 1994, xxiv + 752 pp. Applied Psychological Measurement, 19(3), 303-305. https://doi.org/10.1177/014662169501900308

Viladrich, C., Angulo-Brunet, A., & Doval, E. (2017). Un viaje alrededor de alfa y omega para estimar la fiabilidad de consistencia interna. *Anales de Psicología, 33*(3), 755 – 782. https://doi.org/10.6018/analesps.33.3.268401

Yung, Y., Thissen, D., & McLeod, L. D. (1999). On the relationship between the higher-order factor model and the hierarchical factor model. *Psychometrika, 64*(2), 113-128. https://doi.org/10.1007/bf02294531

Zhang, Z., & Yuan, K. (2015). Robust coefficients alpha and omega and confidence intervals with outlying observations and missing data. *Educational And Psychological Measurement, 76*(3), 387-411. https://doi.org/10.1177/0013164415594658

Zinbarg, R. E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's α, revelle's β, and mcdonald's ωh: their relations with each other and two alternative conceptualizations of reliability. *Psychometrika, 70*(1), 123-133. https://doi.org/10.1007/s11336-003-0974-7

Zumbo, B. D., & Kroc, E. (2019). A measurement is a choice and stevens' scales of measurement do not help make it: A response to chalmers. *Educational And Psychological Measurement, 79*(6), 1184-1197. https://doi.org/10.1177/0013164419844305