Tehran University of Medical Sciences

Nursing Practice Today

2024; Volume 11, No 1, pp. 16-21

Commentary

Statistical concerns, invalid construct validity, and future recommendations

Hamid Sharif-Nia[1,2], Long She[3], Jason Osborne[4], Ozkan Gorgulu[5], Fatemeh Khoshnavay Fomani*[6], Amir Hossein Goudarzian[7]

[1]Psychosomatic Research Center, Mazandaran University of Medical Sciences, Sari, Iran
[2]Department of Nursing, Amol Faculty of Nursing and Midwifery, Mazandaran University of Medical Sciences, Sari, Iran
[3]Sunway Business School, Sunway University, Sunway, Malaysia
[4]Department of Statistics, Miami University, Oxford, Ohio, USA
[5]Department of Biostatistics and Medical Informatics, Faculty of Medicine, Kirsehir Ahi Evran University, Kirsehir, Turkey
[6]Nursing and Midwifery Care Research Center, School of Nursing and Midwifery, Tehran University of Medical Sciences, Tehran, Iran
[7]School of Nursing and Midwifery, Tehran University of Medical Sciences, Tehran, Iran

Introduction

Psychometric studies account for a significant portion of the existing academic knowledge in many disciplines. The need for evidence of high-quality measurement, particularly in the health and social sciences, has prompted many researchers to investigate the validity and reliability of new or existing measures. However, not all studies are executed with the highest level of rigor or quality, leading researchers to develop and make available various tools such as QUADAS, STARD, and COSMIN (1, 2). Peer review processes are valuable but not foolproof, and occasionally, published studies can be criticized on process, conceptualization (such as construct validity), sampling, data cleaning, analysis, bias, or even misconduct, as scholars have access to better tools for detecting these issues. In this article, we discussed one example of a problematic study and presented some recommendations.

Overview

One of our concerns about some psychometric studies related to claiming variance is explained. Authors can make the common mistake of reporting cumulative variance accounted for (communality) for each factor extracted rather than the unique (incremental) communality. This is evident in one study (3) in which the authors claimed that the three factors of their newly developed instrument could explain 56.85%, 60.54%, and 64.05% of the variance, respectively, which is, of course, impossible as it sums to over 100%. More concerning, however, is that the reported communalities appear to substantially inflate the variance one would expect from the factor loadings. Additional concerns in the field include using outdated methods for determining the number of factors to extract, such as the outdated "Little Jiffy" criteria (also referred to as the Kaiser Criterion, (4), which has been superseded by modern guidelines such as parallel analysis and/or bootstrap analysis (5). When reporting the results of factor analysis, it is best practice to report eigenvalues, communalities, and a scree plot for an independent examination of the results and a better understanding of the factor structure. In this example study (3), the authors failed to report eigenvalues or a scree plot, leaving readers unable to independently judge whether the right interpretation was made. Finally, we remind readers that exploratory factor analysis is an exploratory technique, and therefore, results can be volatile and challenging to replicate (5, 6), meaning that other tools, such as confirmatory factor analysis, should be used to

Corresponding Author: Fatemeh Khoshnavay Fomani, Nursing and Midwifery Care Research Center, School of Nursing and Midwifery, Tehran University of Medical Sciences, Tehran, Iran. Email: f.khoshnava@gmail.com

DOI: 10.18502/npt.v11i1.14938

Please cite this article as: Sharif-Nia H, She L, Osborne J, Gorgulu O, Khoshnavay Fomani F, Goudarzian A.H. Statistical concerns, invalid construct validity and future recommendations. Nursing Practice Today. 2024; 11(1):16-21

confirm or validate and confirm the suspected psychometric properties of an instrument.

Choosing an appropriate decision rule

In exploratory factor analysis, extraction method and decision rules are important, as is the rotation methodology used and information about communalities- which helps the reader understand how well the analysis is capturing (or not capturing) the variance in the data. While many statistical computing packages use Varimax as the default rotation, it is an orthogonal rotation method, meaning that it extracts uncorrelated factors. This can cause mis-estimation of the factor structure and loadings. When we expect that factors will be correlated, it is a common practice to utilize oblique rotations, which are widely available and easy to interpret. Regardless, authors should be clear in terms of what extraction and rotation techniques were used to aid the reader in understanding the results (for a thorough explanation of extraction and rotation options, see (7). In this example paper, the authors report that the components extracted had correlations, suggesting the use of an oblique rotation method that allows for correlated components or factors, yet the methodology is unclear. It is also unclear whether the structure or pattern matrix loadings were reported, which is important to understand in oblique rotation as structure matrix coefficients are simple correlations while pattern matrix coefficients are more akin to standardized regression coefficients, reporting the unique relationship of the item to the latent factor holding other factors constant.

Reporting communalities accurately is important

Communalities are essentially "variance accounted for" and, as such, cannot exceed 100% (and rarely approaches it). In our example paper, authors reported communalities of 56.85, 60.54, and 64.05. Of course, this could be a simple error of reporting cumulative rather than unique communalities, which is borne out by the fact that the first factor should always have the largest communality. Yet these numbers do not seem to match what one would expect from the reported factor loadings reported.

It is possible to calculate each subscale's variance ratios using the reported factor loadings of the three sub-constructs. Eigenvalues (λ) are also the sum of squared factor loadings (SSL) component (or factor) loadings across all items (k) for each component (or factor), which represents the amount of variance in each item that can be explained by the analysis (6). Then, to get the percent of total variance explained by the factor (or principal component), we divided the eigenvalue by the total number of items $\frac{\lambda}{k}$: where λ represents the eigenvalue of the sub-construct, F is the standardized factor loading of item number I in a sub-construct, and k is the total number of the items in the main construct (three-factor construct). In other words, the ratio of variance is estimated by the formula:

$$\text{Proportion of variance} = \sum_{i=1}^{N} l_i^2 \big/ N = [(l_1^2 + l_2^2 + \cdots l_N^2)]/N \quad (8).$$

N is the number of variables (Items), and l_i is the factor load of the i^{th} variable.

Unfortunately, one complication with this fairly basic approach to verifying the results of the analysis is that in our example article, factor loadings less than 0.40 were suppressed, leaving us without the ability to complete these calculations exactly and also leaving us without specific information about potential cross-loadings (as some of the retained loadings were close to 0.40).

Although we were left without an accurate way to calculate eigenvalues and variance accounted for, we can adopt an approach of calculating a minimum and maximum, where we assumed unreported loadings were 0.00 (minimum) or 0.30 (a reasonable maximum close to the stated upper limit for suppressed loadings) to estimate a reasonable range for eigenvalue and variance.

As shown in Table 1, our computed variance ratio for all sub-constructs and total variance explained by the three-factor construct is markedly different from the results reported by the authors of the mentioned paper in their original article. We estimate that the total variance explained by the three-factor construct should range from 25.54% (minimum) to 44.34% (maximum) based on the factor loadings that the study provided and the estimated minimum or maximum values for suppressed loadings. Although there are no well-established filed-specific thresholds in the exploratory factor analysis literature, Plonsky suggests that the minimum cumulative percentage of explained variance should be around 55-65% (9) in a factor analysis. Regardless, even the upper bound "maximum" we calculated based on reported results lies far below that reported by the authors. The same is true for each component extracted.

Table 1. The comparison of variance ratio and total variance explained between our re-calculated results and the results reported in Shamsalinia et al. article

Domains	Items No.	Factor loading	Eigenvalues (min/ max)*	Variance ratio (min/ max)*	The variance ratio reported in Shamsalinia et al.'s paper
Adherence efficacy	1	.550	3.647/5.079	12.57/17.51	56.85
	3	.432			
	11	.554			
	12	.633			
	17	.524			
	18	.476			
	20	.570			
	21	.444			
	22	.408			
	26	.486			
	4	.595			
	5	.458			
	9	.432			
	25	.523			
Preventive behaviors	2	.507	2.456/4.166	8.47/14.37	60.54
	7	.448			
	8	.423			
	14	.517			
	15	.490			
	16	.449			
	19	.531			
	23	.600			
	27	.545			
	28	.414			
Information effectiveness	6	.606	1.453/3.613	5.01/12.46	64.05
	10	.469			
	13	.409			
	24	.709			
	29	.443			
Total variance explained (%)				**26.06/44.34**	**64.05**

Note: Eigenvalues were calculated by $\sum$ (Factor loading)^{2,} and the variance ratio was calculated by dividing eigenvalues by several items. Because the authors did not report the full table of component loadings, we cannot compute the correct eigenvalue exactly. However, we can calculate a minimum (assuming all loadings were a minimum of 0.00) and a maximum (assuming all non-reported loadings were 0.30, modestly below the cut-off of 0.40).

In sum, we are left more confused than ever by the reported results, possibly because of incomplete reporting or perhaps because the results were not reported accurately or honestly. It is not entirely clear which might be true, but the lessons from this example are that authors should report clearly what methods are being used, what coefficients are being reported, double-check communalities and summary statistics, and if suppressing small coefficients for clarity, be sure to note the level at which coefficients are suppressed.

Cross-loadings must be addressed

In exploratory factor analysis, as in this example paper, we can often see an item load on multiple factors, and it is often the case that authors assign the item to the theoretically appropriate factor or the one in which there was the strongest loading. However, when loadings are very close, it might be the case that the item is unclear or in need of revision. Authors should carefully examine items that load on multiple factors and consider the best way forward rather than ignoring this issue or setting high suppression levels for reporting coefficients.

The importance of replication and multiple samples

Replication and multiple samples are important as factor analysis, like many statistical procedures, can be impacted by the idiosyncratic nature of a particular sample, and therefore replication analysis or bootstrap methodology is important to ensure that results and conclusions are likely to generalize to other samples (10). In our example study, numerous items were deleted using the same data or sample as the final analysis, leaving us with an increased risk that these results may not replicate.

Validation should use multiple measures clearly appropriate to the task

Construct validity exercises often include predictive or convergent validity, discriminant validity, confirmatory analyses, or Rasch analyses, for example. It is often the case that authors will use behavioral measures of predictive validity or other indicators of validity to ensure the evaluation of the psychometric properties of an instrument is clear. In our example study, the authors make a passing mention of three indicators of validity but provide limited information on how a reader should understand these statistics as there is no information in the methods or results section as to how these numbers were calculated, and another examination by Pahlevan Sharif, Naghavi and Sharif-Nia (11) suggest that they have been miscalculated.

Principal Components Analysis (PCA) is an antiquated technique

One final recommendation concerns the continued use of PCA in many fields, which is a technique similar to factor analysis but with simplified mathematics appropriate for when computing power is rare and limited. PCA should not be used in the modern era, and exploratory analyses should use EFA with appropriate extraction, rotation, and replication techniques. Osborne (7) and other modern scholars have extensive discussions of the appropriate applications of these exploratory techniques, how to use them to maximum advantage and with rigor, and also when to use other, more appropriate confirmatory techniques. Our example study used PCA- which produces orthogonal components despite asserting the factors should be correlated.

Leave exploratory analyses to exploration

As mentioned previously, there are many exploratory techniques, like EFA, that are valuable when appropriately deployed. However, in the modern era, validation and drawing conclusions about measures requires confirmatory techniques, like Confirmatory Factor Analysis, which is more rigorous and produces, when used appropriately, more replicable and generalizable results. Exploratory techniques, in short, should not be used for validation. Our example paper asserts conclusions appropriate for confirmatory techniques using exploratory analyses, which is undesirable.

Summary

The paper in question presented results that seem invalid, miscalculated, or impossible; there is not enough information to objectively evaluate whether a three-factor solution is ideal for this measure; we have no expectation that these results would replicate in a different sample; and we have little information as to whether the measure is valid and reliable in this population given the reported data. We outline these shortcomings to help future research utilize best practices to produce strong and defensible results.

Discussion

The goal of this brief Letter to the Editor was to highlight some common methodological practices that can improve your analyses. We used a paper that contains what we believe to be serious methodological and conceptual issues as a contrary example to guide discussion.

In brief, this example paper (3) uses Principal Components Analysis (PCA), which is not considered an appropriate analysis in the modern era. The results report correlated factors, but PCA produces uncorrelated factors. In many fields, it is reasonable to expect factors to be correlated with each other, so it is recommended to use oblique rotations unless there is clear evidence for non-correlation between factors.

The communalities may or may not be accurate, but even if they are, they are reported in such a manner as to create the impression that over 100% of the variance is being accounted for. The extraction and rotation methodology is unclear, and authors should clearly identify what techniques were used and why. Several original items were discarded based on a single analysis from a single sample, cross-loadings may not have been addressed adequately, and conclusions appropriate to confirmatory techniques were drawn using exploratory analyses.

Exploratory techniques like EFA are notoriously challenging to replicate (e.g., Osborne, (7)), leading many to recommend using EFA and confirmatory factor analysis (CFA) sequentially (on different and independent data sets) to provide more conclusive evidence of a stable and replicable structure.

CFA, on the other hand, is useful for testing a pre-specified model of underlying factors. However, it can be overly restrictive and may not allow for new factors to emerge. Using both techniques, researchers can first use EFA to identify potential underlying factors and then use CFA to test a pre-specified model that incorporates those factors using an independent data set. This approach allows for greater flexibility while maintaining rigor in the analysis (12).

Finally, it is recommended that researchers report effect sizes (ideally with confidence intervals) along with their results. Effect sizes provide information about the strength of relationships between variables and can help readers interpret the practical significance of findings (13). Osborne (7) has examples of how to create and use confidence intervals in EFA easily.

Of more concern is the possibility that a paper reports false statistical results, intentionally or unintentionally, in a reputable journal. False results can have serious negative consequences on the reputation of an author, a journal, and even a field. This has been a serious challenge to the integrity and reputation of Semitic research in general (14), although the issue seems to permeate many fields and many cultures.

To avoid such issues in future research, there have been many recommendations for improving peer review, including providing complete results using modern best practices and archiving original data for further evaluation later by interested colleagues.

Conflict of interests

The authors declare that they have no competing interests.

Acknowledgments

We would like to express our sincere gratitude to all those who have supported us in writing this commentary. Without their help, this article would not have been possible.

References

1. Mokkink LB, Terwee CB, Knol DL, Stratford PW, Alonso J, Patrick DL, et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: A clarification of its content. BMC Medical Research Methodology. 2010;10(1):22.
2. Oliveira MR, Gomes Ade C, Toscano CM. QUADAS and STARD: evaluating the quality of diagnostic accuracy studies. Rev Saude Publica. 2011;45(2):416-22.
3. Shamsalinia A, Ghadimi R, Chafjiri RT, Norouzinejad F, Pourhabib A, Ghaffari F. Nutrition self-efficacy assessment: designing and psychometric evaluation in a community-dwelling elderly population. Journal of Health, Population and Nutrition. 2019;38(1):1-10.
4. Kaiser HF. A second generation little jiffy. 1970.
5. Osborne JW, Banjanovic ES. Exploratory factor analysis with SAS: Sas Institute; 2016.
6. Costello AB, Osborne J. Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. Practical assessment, research, and evaluation. 2019;10(1):7.
7. Osborne JW. Best practices in exploratory factor analysis: CreateSpace Independent Publishing Platform; 2014.
8. Tabachnick BG, Fidell LS. Using Multivariate Statistics: Pearson; 2013.

9. Plonsky L. Advancing Quantitative Methods in Second Language Research: Taylor & Francis; 2015.
10. Osborne JW, Fitzpatrick DC. Replication analysis in exploratory factor analysis: What it is and: Why it makes your analysis better. Practical Assessment, Research and Evaluation. 2012;17(15):1-8.
11. Pahlevan Sharif S, Naghavi N, Sharif-Nia H. Concerns regarding the validity of nutrition self-efficacy questionnaire among Iranian elderly population. Journal of Health, Population and Nutrition. 2022 Mar 25;41(1):11.
12. Koran J. Indicators per factor in confirmatory factor analysis: More is not always better. Structural Equation Modeling: A Multidisciplinary Journal. 2020;27(5):765-72.
13. Shi D, Maydeu-Olivares A, Rosseel Y. Assessing fit in ordinal factor analysis models: SRMR vs. RMSEA. Structural Equation Modeling: A Multidisciplinary Journal. 2020;27(1):1-15.
14. Lose G, Klarskov N. Why published research is untrustworthy. International Urogynecology Journal. 2017;28(9):1271-4.