Author Guidelines for Reporting Scale Development and Validation Results in the Journal of the Society for Social Work and Research

Peter Cabrera-Nguyen, Washington University in St. Louis

Available at: https://works.bepress.com/elian_cabrera/4/
Author Guidelines for Reporting Scale Development and Validation Results in the *Journal of the Society for Social Work and Research*

Peter Cabrera-Nguyen
Washington University in St. Louis

Reliable and valid measurement is critical for advancing social work research and evidence-based social work practice. However, the quality of available evidence is largely determined by the study designs used in intervention research. A well-designed intervention study includes optimal sampling techniques, data analysis using appropriate statistical methods, steps to enhance internal and external validity, and psychometrically sound assessment instruments. Ensuring intervention studies include assessment instruments with strong psychometric properties may, ultimately, enable practitioners to assess client target problems with greater precision. Further, increasing the availability of instruments with demonstrated reliability and validity may also help practitioners select evidence-based interventions that best match the needs of their clients. Therefore, it is important for social work researchers engaged in scale development and validation to conduct their research and report their findings in a reliable manner that allows other researchers to replicate those findings. The guidelines presented in this article are intended to assist *JSSWR* authors in this endeavor.

These guidelines are limited to the latent variable approach (i.e., exploratory factor analysis [EFA] and confirmatory factor analysis [CFA]) to scale development and validation, and these suggestions do not cover item-response theory approaches. There is little consensus in the literature regarding these guidelines. Indeed, some aspects are still intensely debated. Therefore, these guidelines should be interpreted as a framework for reporting findings—from early-stage scale development to later-stage scale validation research.

**Guidelines**

1. Precisely define the target construct.
2. Justify the need for your new measure. For example, if measures of the construct exist in the literature, explain the value added by your new scale. How might the new measure enhance the substantive knowledge base or social work practice?
3. Indicate that you have submitted your initial pool of items to expert review (Worthington & Whittaker, 2006). Report (a) the number of items in the preliminary pool; (b) the number of expert reviewers and their qualifications; and (c) any major changes to your initial item pool following the review (e.g., a substantial decrease in the number of items, changes to the original item response format, overhaul of item pool due to experts’ assessment regarding content validity).
4. Report the name and version of the statistical software package used for all analyses.
5. Identify and justify the sampling strategy (e.g., convenience, snowball) and sampling frame. Report standard sample demographic characteristics as well as other salient sample characteristics (e.g., “participants were advanced-standing MSW students at a large public Midwestern university concentrating in social service administration”).
6. Discuss relevant data preparation and screening procedures. For instance, do the data meet the appropriate assumptions for factor analysis? If not, what actions were taken? Report tests of factorability if appropriate (e.g., report Bartlett’s test of sphericity).
7. Provide all dates of data collection.
8. Avoid use of principal components analysis (PCA) as a precursor to CFA (Costello & Osborne, 2005; Worthington & Whittaker, 2006). Instead, start with EFA to assess the underlying factor structure and refine the item pool. EFA should be
followed by CFA using a different sample (or samples) to evaluate the EFA-informed a priori theory about the measure’s factor-structure and psychometric properties. (Costello & Osborne, 2005; Henson & Roberts, 2006; Worthington & Whittaker, 2006). For CFA, authors should specify an a priori hypothesized model and a priori competing models (Jackson, Gillaspy, & Purc-Stephenson, 2009).

9. Guidelines for reporting EFA results are presented in Table 1.
10. Guidelines for reporting CFA results are presented in Table 2.
11. Describe the matrix (or matrices) you analyzed (e.g., covariance, correlation). Include matrices in the manuscript if feasible; otherwise, indicate these data are available upon request.
12. Report the amount of missing data and describe how missing data were handled. For a review of practices for handling missing data, see Sterne and colleagues (2009), Rose and Fraser (2008), and Horton and Kleinman (2007). Provide a rationale for your approach to handling missing data. Authors are encouraged to consider using multiple imputation or model estimation with full-information maximum likelihood (FIML).
13. Compare your CFA model with the alternative or competing models. Do competing models fit the data better or worse than your model (e.g., does your four-factor model of acculturation fit the data better than a two-factor model or a one-factor model)? Identify the preferable model based on appropriate fit statistics (e.g., chi-square difference test for nested models, Akaike information criterion for non-nested models), parsimony, and relevant theory.
15. Report how methodological limitations may have impacted findings regarding your measure’s psychometric properties (e.g., note potential repercussions of suboptimal sampling techniques, discuss implications of using listwise deletion to handle missing data instead of multiple imputation or FIML).
16. Discuss directions for future research (e.g., if appropriate, testing your scale for measurement invariance by conducting CFA on different populations).

Conclusion

Scale development and validation using EFA and CFA is a complex process involving many choices regarding (a) data screening procedures; (b) model fit statistics; (c) statistical tests for comparing competing models; and (d) the next appropriate step in the scale development and validation process. It is difficult to develop a decision tree that adequately specifies the entire universe of choices. Authors should use these guidelines as a roadmap, but they should also be familiar with ongoing developments and debates in the psychometric literature.

The following resources may be useful for researchers involved in scale development and validation:

- SEMNET - an active online discussion group about all things related to structural equation modeling (SEM; including CFA) for novices and experts alike. Available at http://alabamamaps.ua.edu/archives/semnet.html
- UCLA Academic Technology Services – an invaluable resource for an array of methodologies including EFA/CFA/SEM. This site offers a wide range of online seminars on topics such as conducting EFA/CFA using various versions of Mplus, and carrying out multiple imputation of missing data using a range of software packages. Available at http://www.ats.ucla.edu/stat/seminars/default.htm
- StatNotes – a good, general resource for quantitative data analysis. An overview of factor analysis is available at http://faculty.chass.ncsu.edu/garson/PA765/factor.htm

Finally, authors interested in cross-cultural scale development and validation are encouraged to refer to classic articles by Alegría and colleagues (2004) and Canino and Bravo (1994).
Table 1
Reporting Guidelines for Exploratory Factor Analysis

1. **How large a sample?** One common rule of thumb is to ensure a person-to-item ratio of 10:1. Another rule of thumb is that \( N = 300 \) is usually acceptable (Worthington & Whittaker, 2006). However, some researchers have criticized these sample size rules of thumb, noting the appropriate sample size is dependent on the features of the gathered data. These researchers recommend obtaining the largest possible sample because the adequacy of the sample size cannot be determined until *after* the data have been analyzed (Henson & Roberts, 2006).

2. **Run EFA . . . or not.** Run a preliminary EFA to determine if further data collection is required based on the following criteria: (a) If communalities are greater than .50 or there are 10:1 items per factor with factor loadings of roughly \(|.4|\), then a sample size of 150 to 200 is likely to be adequate; (b) If communalities are all at least .60 or there are a minimum of 4:1 items per factor with factor loadings above \(|.6|\), then even smaller sample sizes may suffice; (Worthington & Whittaker, 2006). Report if additional data collection was necessary due to inadequate sample size. If so, report the new participants’ sociodemographic characteristics and test for differences between groups using standard statistical procedures (e.g., \( t \) tests).

3. **Give EFA details.** Report the specific rotation strategy used (e.g. varimax, geomin). Justify the decision to use an orthogonal or oblique solution. One recommendation is to always begin with an oblique rotation, empirically assess factor intercorrelations, and report them before deciding upon a final rotation solution (Henson & Roberts, 2006; Worthington & Whittaker, 2006). Some researchers argue oblique rotation is always the best approach because (a) factor intercorrelations are the norm in social sciences and (b) both approaches yield the same result if the factors happen to be uncorrelated (Costello & Osborne, 2005). Conversely, other researchers contend that orthogonal rotation is preferable because fewer parameters are estimated—orthogonal rotation is more parsimonious and amenable to replication (Henson & Roberts, 2006). Similarly, some researchers warn against relying on a statistical software package’s default settings to determine the appropriate type of oblique rotation (Henson & Roberts, 2006; Worthington & Whittaker, 2006). Others state that doing so is “fine” (Costello & Osborne, 2005, p.3). Given the lack of consensus, it is probably best to describe what you do and defend your approach on substantive grounds, if possible..

4. **Report the whole factor pattern/structure.** Always report the whole factor pattern/structure matrix, including all of the items in the analysis. It is recommended that authors report this information in a chart following the example provided by Henson and Roberts (2006) on page 411.

5. **Criteria for deleting (crossloaded) items.** Report any deleted items and the criteria used for deletion. Crossloading items with values \( \geq .32 \) on at least two factors should generally be candidates for deletion, especially if there are other items with factor loadings of .50 or greater (Costello & Osborne, 2005). Rerun the EFA each time an item is deleted.

6. **Criteria for number of factors.** Report the number of factors retained and justify this decision using multiple criteria (eigenvalue > 1, scree test, parallel analysis, rejection of a factor with fewer than 3 items, etc). Reporting the eigenvalue > 1 rule alone is inadequate because it is among the least accurate criteria for assessing factor retention (Costello & Osborne, 2005; Henson & Roberts, 2006).

7. **Explained variance.** Report the variance explained by the factors.

8. **In general, describe your decisions.** EFA is a complex, iterative, and subjective process. Therefore, “it is very important that researchers [and reviewers] be able to independently evaluate the results obtained in an EFA study. This can, and should, occur on two levels. Given the myriad subjective decisions necessary in EFA, independent researchers should be able to evaluate the analytic choices of authors in the reported study. Second, independent researchers should be able to accurately replicate the study on new data, or even employ a CFA” (Hensen & Roberts, 2006, p. 400).

Every decision should be thoroughly reported and justified. When in doubt, err on the side of over reporting.
### Reporting Guidelines for Confirmatory Factor Analysis

1. **Describe and justify the theoretical model.** Report hypothesized factor structure. Provide theoretical and empirical justification (e.g., results of preliminary EFAs) for your hypothesis. In addition, report a priori competing models.

2. **Describe the parameterization.** Provide a comprehensive description of the a priori parameter specification. Identify fixed parameters, free parameters, and constrained parameters. For example, indicate if you freed the errors of any items to correlate.

3. **Include a figure.** Include a figure of each CFA model being tested using Kline’s (2005) graphical conventions if feasible.

4. **Identification.** Demonstrate model identification (e.g., $df > 0$; scaling of factors; assess and report the “t-rule”; the two-indicator rule). Necessary and sufficient conditions for model identification may vary for certain types of CFA models. When in doubt, authors should consult Brown’s (2006) CFA text or Kline’s (2005) SEM text for guidance.

5. **Select an estimator based on distributional patterns and assumptions.** Report the estimator used (e.g., ML, WLSMV) and justify your choice based on distributional assumptions. It is not appropriate to report that you relied on your statistical software’s default setting.

6. **Use multiple fit indices.** After estimating a model, always report multiple fit indices (e.g., model $X^2$, $df$, $p$, CFI/TLI, RMSEA, SRMR). Report all appropriate fit indices, not just those favorable to your hypotheses (Jackson et al., 2009). For example, do not report acceptable CFI and TLI scores while omitting a relevant fit index with a suboptimal value.

7. **What is acceptable fit?** For model fit indices, authors should generally use the cut-off values recommended by Hu and Bentler (1999) and endorsed by Brown (2006), assuming ML estimation:
   a. CFI/TLI $\geq .95$
   b. RMSEA $\leq .06$
   c. SRMR $\leq .08$

8. **Localized strain?** When reporting model fit, include an assessment for localized areas of strain by examining standardized residuals. Standardized residuals greater than 1.96 (for $p < .05$) indicate areas of strain (Harrington, 2009). Report the absence of localized strain, if appropriate; otherwise, note localized areas of strain by reporting the relevant standardized residuals.

9. **Parameter estimates and SEs.** When reporting factor loadings and other parameter estimates, always report the unstandardized estimates, their $p$ values, and the standard errors. In addition, include the standardized estimates when appropriate. Be sure to report all parameter estimates, even those that are nonsignificant (Brown, 2006; Jackson et al., 2009).

10. **Assessing the validity of the factor solution.** Comment on the new measure’s convergent and discriminant validity based on parameter estimates. For instance, factor correlations $\geq .80$ may indicate poor discriminant validity (Brown, 2006). In addition, strong factor loadings that do not crossload may indicate good convergent validity. One rule of thumb is that factor loadings $< .40$ are weak and factor loadings $\geq .60$ are strong (Garson, 2010).

11. **Other measures.** Report squared multiple correlations and comment on the measure’s reliability (e.g., report Raykov’s Rho if appropriate).

12. **Respecification: Caution!** Report any post-hoc respecifications to improve model fit based on modification indices. Justify the respecifications on theoretical or conceptual grounds (Jackson et al., 2009). Respecification to allow for correlated errors is not supportable without strong pragmatic justification (e.g., items contain similar words or phrases). Note that respecification precludes comparing the model with your a priori specified competing models. Report improvements in appropriate model fit indices for respecified models (e.g., chi-square difference test).

---

*a Authors should consult the current SEM literature to stay abreast of the ongoing debate over appropriate cut-off values. Authors who choose to use lower cut-offs should provide a thorough rationale supported by ample citations from the recent SEM literature (e.g., Marsh, Hau, & Wen, 2004; Fan & Sivo, 2007). For recommended cut-off scores across a variety of circumstances, authors are referred to Yu (2002).*
References


