Journal of Modern Applied Statistical Methods

Volume 8 | Issue 2 Article 10

11-1-2009

Relationship between Internal Consistency and Goodness of Fit Maximum Likelihood Factor Analysis with Varimax Rotation

Gibbs Y. Kanyongo
Duquesne University, kanyongog@duq.edu

James B. Schreiber
Duquesne University, schreiberj@duq.edu

Journal of Modern Applied Statistical Methods
November 2009, Vol. 8, No. 2, 463-468

1538 – 9472/09/$95.00

Relationship between Internal Consistency and Goodness of Fit Maximum Likelihood Factor Analysis with Varimax Rotation

Gibbs Y. Kanyongo James B. Schreiber
Duquesne University

This study investigates how reliability (internal consistency) affects model-fitting in maximum likelihood exploratory factor analysis (EFA). This was accomplished through an examination of goodness of fit indices between the population and the sample matrices. Monte Carlo simulations were performed to create pseudo-populations with known parameters. Results indicated that the higher the internal consistency the worse the fit. It is postulated that the observations are similar to those from structural equation modeling where a good fit with low correlations can be observed and also the reverse with higher item correlations.

Key words: Factor structure; Matrices; Scree plot; Parallel Analysis.

Introduction

The purpose of the study is to investigate how reliability (internal consistency) affects model-fitting in maximum likelihood exploratory factor analysis (EFA). The study seeks to accomplish this through integrating and extending the work of Kanyongo (2006) on reliability and number of factors extracted and Fabringer, Wegener, MacCallum, Strahan's (1999) work on model-fitting and the number of factors extracted in exploratory factor analysis.

Internal Consistency

Henson (2001) noted that reliability is often a misunderstood measurement concept. There are several forms of reliability coefficients, but some of the most commonly used are internal consistency estimates. Internal consistency estimates relate to item homogeneity, or the degree to which the items on a test jointly measure the same construct (Henson, 2001). Thus, in the classical test theory, reliability is concerned with score consistency. The classical conceptualization of score reliability relates the concept of score consistency to true scores. Allen and Yen (1979) defined a person's true score as the theoretical average obtained from an infinite number of independent testings of the same person with the same test.

Many authors conceptualize three sources of measurement error within the classical framework: content sampling of items, stability across time, and interrater error (Henson, 2001). Content sampling refers to the theoretical idea that the test is made up of a random sampling of all possible items that could be on the test. If that is the case, the items should be highly interrelated because they assess the same construct of interest. This item interrelationship is typically called internal consistency, which suggests that the items on a measure should correlate highly with each other if they truly represent appropriate content sampling (Henson, 2001). If the items are highly correlated, it is theoretically assumed that the construct of interest has been measured to some degree of consistency, that is, the scores are reliable.

Internal consistency estimates are intended to apply to test items assumed to represent a single underlying construct, thus, the

Gibbs Y. Kanyongo is an Assistant Professor in the Department of Educational Foundations and Leadership in the School of Education. Email: kanyongog@duq.edu. James B. Schreiber is an Associate Professor in the Department of Foundations and Leadership in the School of Education. Email: schreiberj@duq.edu.

use of these estimates with speeded tests is inappropriate due to the confounding of construct measurement with testing speed. Furthermore, for tests that consist of scales measuring different constructs, internal consistency should be assessed separately for each scale (Henson, 2001).

Exploratory Factor Analysis (EFA)

The primary purpose of EFA is to arrive at a more parsimonious conceptual understanding of a set of measured variables by determining the number and nature of common factors needed to account for the pattern of correlations among the measured variables (Fabringer, et al., 1999). EFA is based on the common factor model (Thurstone, 1947). The model postulates that each measured variable in a battery of measured variables is a linear function of one or more common factors and one unique factor.

Fabringer, et al. (1999) defined common factors as unobservable latent variables that influence more than one measured variable in the battery and accounts for the correlations among the measured variables, and unique factors as latent variables that influence only one measured variable in a battery and do not account for correlations among measured variables. Unique factors are assumed to have two components; a specific component and an error of measurement component, the unreliability in the measured variable. The common factor model seeks to understand the structure of correlations among measured variables by estimating the pattern of relations between the common factor(s) and each of the measured variables (Fabringer, et al., 1999).

Previous Work

Kanyongo investigated the influence of internal consistency on the number of components extracted by various procedures in principal components analysis. Internal consistency reliability coefficients are not direct measures of reliability, but are theoretical estimates based on classical test theory. IC addresses reliability in terms of consistency of scores across a given set of items. In other words, it is a measure of the correlation between subsets of items within an instrument.

The study employed the use of Monte Carlo simulations to generate scores at different levels of reliability. The number of components extracted by each of the four procedures, scree plot, Kaiser Rule, Horn's parallel analysis procedure and modified Horn's parallel analysis procedure was determined at each reliability level. In his study, Kanyongo (2006) found mixed results on the influence of reliability on the number of components extracted. However, generally, when component loading was high, an improvement in reliability resulted in improvement of the accuracy of the procedures especially for variable-to-component ratio of 4:1.

The Kaiser procedure showed the greatest improvement in performance although it still had the worst performance at any given reliability level. When the variable-to-component ratio was 8:1, reliability did not impact the performance of the scree plot, Horn's parallel analysis (HPA) or modified Horn's parallel analysis (MHPA) since they were 100% accurate at all reliability levels. When component loading was low, it was not clear what the impact of reliability was on the performance of the procedures.

The work of Fabringer, et al. (1999) involved an examination of the use of exploratory factor analysis (EFA) in psychological research. They noted that a clear conceptual distinction exists between principal factor analysis (PCA) and EFA. When the goal of the analysis is to identify latent constructs underlying measured variables, it is more sensible to use EFA than PCA. Also, in situations in which a researcher has relatively little theoretical or empirical basis to make strong assumptions about how many common factors exist or what specific measured variables these common factors are likely to influence, EFA is probably a more sensible approach than confirmatory factor analysis (CFA**).**

Fabringer, et al. (1999) pointed that in EFA; sound selection of measured variables requires consideration of psychometric properties of measures. When EFA is conducted on measured variables with low communalities, substantial distortion in results occurs. One of the reasons why variables may have low communalities is low reliability. Variance due to

random error cannot be explained by common factors; and because of this, variables with low reliability will have low communality and should be avoided.

Fabringer, et al. (1999) also noted that although there are several procedures that are available for model-fitting in EFA, the maximum likelihood (ML) method of factor extraction is becoming increasingly popular. ML is a procedure used to fit the common factor model to the data in EFA. ML allows for the computation of a wide range of indices of the goodness of fit of the model. ML also permits statistical significance testing of factor loadings and correlations among and the computation of confidence intervals for these parameters (Cudeck & O'Dell, 1994). Fabringer, et al. (1999) pointed out that the ML method has a more formal statistical foundation than the principal factors methods and thus provides more capabilities for statistical inference, such as significance testing and determination of confidence intervals.

In their work, Frabinger, et al. (1999) further stated that, ideally, the preferred model should not just fit the data substantially better than simple models and as well as more complex models. The preferred model should fit the data reasonably well in an absolute sense. A statistic used for assessing the fit of a model in ML factor analysis solutions is called a goodness of fit index.

There are several fit indices used in ML factor analysis and one of them is the likelihood ratio statistic (Lawley, 1940). If sample size is sufficiently large and the distributional assumptions underlying ML estimation are adequately satisfied, the likelihood ratio statistic approximately follows a Chi-square distribution if the specified number of factors is correct in the population (Fabringer, et al., 1999). They noted that, if this is not the case, a researcher should exercise caution in interpreting the results because a preferred model that fits the data poorly might do so and because the data do not correspond to assumptions of the common factor model.Alternatively, it might suggest the existence of numerous minor common factors. Fabringer, et al. also suggested that "with respect to selecting one of the major methods of fitting the common factor model in EFA (i.e., principal factors, iterated principal factors, maximum likelihood), all three are reasonable approaches with certain advantages and disadvantages. Nonetheless, the wide range of fit indexes available for ML EFA provides some basis for preferring this method" (p.283). Since ML EFA has potential to provide misleading results when assumptions of multivariate normality are severely violated, the recommendation is that the distribution of the measured variables should be examined prior to using the procedure. If non-normality is severe (skew>2; kurtosis>7), measured variables should be transformed to normalize their distributions (Curran, West & Finch, 1996).

Fabringer, et al. (1999) noted that the root mean square error of approximation (RMSEA) fit index and the expected cross-validation index (ECVI) provide a promising approach for assessing fit of a model in determining the number of factors in EFA. They recommended that "In ML factor analysis; we encourage the use of descriptive fit indices such as RMSEA and ECVI along with more traditional approaches such as the scree plot and parallel analysis" (p.283). Based on this recommendation, this study uses these fit indices along with the scree plot and parallel analysis to assess the accuracy of determining the number of factor at a given level of reliability.

Research Question

The main research question that this study intends to answer is: As the internal consistency of a set of items increases, does the fit of the data to the exploratory factor analysis improve? To answer this question, a Monte Carlo simulation study was conducted which involved the manipulation of component reliability ($\rho_{xx'}$) loading (a_{ij}), variable-to-component ratio ($p:m$). The number of variables (p) was made constant at 24 to represent a moderately large data set.

Methodology

The underlying population correlation matrix was generated for each possible p, $p:m$ and a_{ij} combination, and the factors upon which this population correlation matrix was based were independent of each other. RANCORR program

by Hong (1999) was used to generate the population matrix as follows.

The component pattern matrix was specified with component loading of .80 and variable-to-component ratio of 8:1. After specifying the component pattern matrix and the program was executed, a population correlation matrix is produced. After the population correlation matrix was generated as described in the above section, the MNDG program (Brooks, 2002) was then used to generate samples from the population correlation matrix. Three data sets for reliability of .60, .80, and 1.00, each consisting of 24 variables and 300 cases were generated. Based on the variable-to-component ratio of 8:1, each dataset had 3 factors built in.

Analysis

An exploratory factor analysis, maximum likelihood, varimax rotation and a three factor specification, was used for each of the three data sets; coefficient alpha = .6, .8, and 1.0. Two goodness-of-fit indices were chosen for this analysis RMSEA and ECVI. RMSEA was chosen because it is based on the predicted versus observed covariances which is appropriate given that nested models are not being compared.

Hu and Bentler (1999) suggested RMSEA <= .06 as the cutoff for a good model fit. RMSEA is a commonly utilized measure of fit, partly because it does not require comparison with a null model. ECVI was chosen because it is based on information theory; the discrepancy between models implied and observed covariance matrices: the lower the ECVI, the better the fit. Finally, the Chi-square and degrees of freedom are provided for each analysis. The data were also submitted to a PCA using the scree plot and parallel analysis to assess the accuracy of determining the number of common factors underlying the data sets.

Results

The results in Table 1 show that the two measures of goodness-of-fit used in this study (RMSEA) and (ECVI) both display the same pattern; the smaller the alpha, the better the model fit. The best fit was obtained for alpha of 0.6, RMSEA (0.025) and ECVI (1.44). As alpha increased from 0.6 to 1.0, both indices increased; an indication that the model fit became poor. Based on Hu and Bentler's (1999) recommendation that the cutoff for a good fit be RMSEA <= 0.06, results here show that only alpha of 0.6 had a good fit. The goodness-of-fit indices therefore suggest that the three-factor model is acceptable at alpha value of 0.6.

Table 1: Goodness-of-Fit Indices

Alpha	Chi-Square (df)	RMSEA	ECVI
.6	247.153 (207)	.025	1.44
.8	436.535 (207)	.061	2.07
1.0	736.385 (207)	.092	3.07

Along with goodness-of-fit-indices, the dataset with the best fit was submitted to principal components analysis through the scree plot and parallel analysis. Results of the scree plot analysis are displayed in Figure 1 while parallel analysis results are shown in Table 2. The scree plot shows a sharp drop between the third and fourth eigenvalues; an indication that there were three distinct factors in the data sets. These results confirm the three-factor model as the best model for these data.

To interpret results of parallel analysis, real data eigenvalues must be larger than random data eigenvalues for them to be considered meaningful eigenvalues. Table 2 shows that the first three eigenvalues expected for random data (1.55, 1.46 and 1.39) fall below the observed eigenvalues for all the three values of alpha. However, the forth eigenvalue of the random data (1.34) is greater than the observed eigenvalues of all the three alpha values. Again, the results further confirm the three-factor model as the best model.

Conclusion

Results in this study were inconsistent with our original ideas of the pattern of goodness of fit and internal consistency. It was anticipated that high internal consistency would yield a better fit.

Figure 1: Results of the Scree Plot

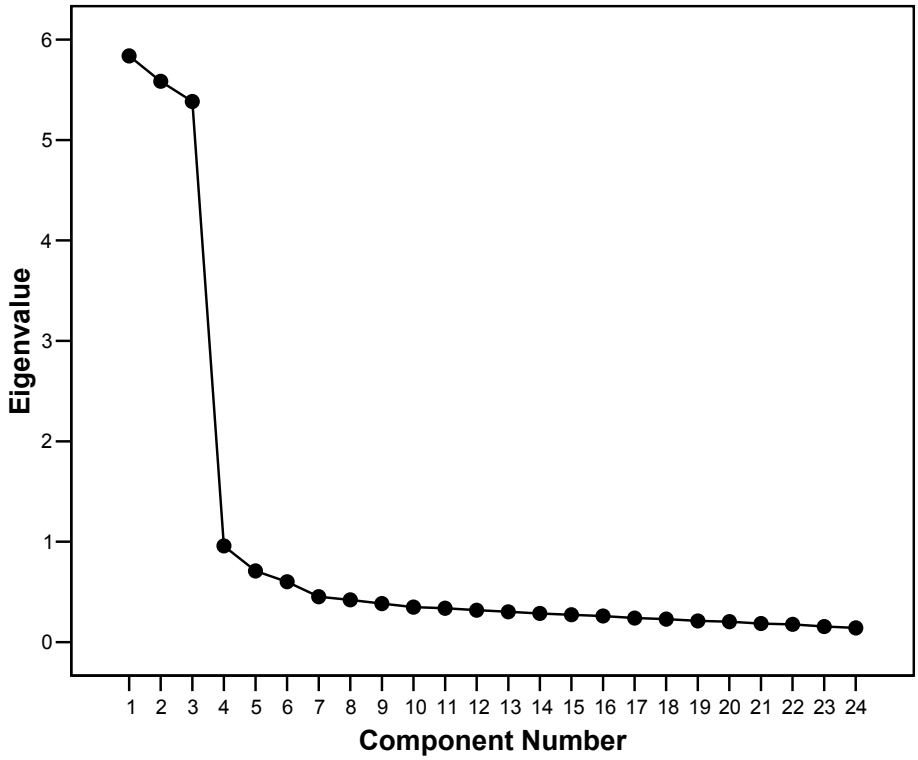

Table 2: Parallel Analysis Results

		Eigenvalues			
		1	2	3	4
Random Data		1.55	1.46	1.39	1.34
Real Data	.6	3.87	3.63	3.31	1.07
	.8	4.66	4.59	4.12	1.32
	1.0	5.84	5.58	5.38	.96

However, the findings seem logical because based on one author's experience with structural equation modeling, a perfect fit can exist when all variables in the model were completely uncorrelated if the variances are not constrained. Also, the lower the correlations stipulated in the model, the easier it is to find good fit. The stronger the correlations, the more power there is within structural equation modeling to detect an incorrect model. In essence, the higher the correlations, the more likely it is to incorrectly specify the model and observe a poor fit based on the indices. Second, if correlations are low, the researcher may lack the power to reject the model at hand.

Also, results seem to confirm what other researchers have argued in the literature. For example, Cudeck and Hulen (2001) noted that if a group of items has been identified as one-dimensional, the internal consistency of the

collection of items need not be high for factor analysis to be able to identify homogenous sets of items in a measuring scale. Test reliability is a function of items. Therefore, if only a few items have been identified as homogeneous by factor analysis, their reliability may not be high.

If ML with exploratory factor analysis including goodness-of-fit analyses are to be used more extensively in the future, a great deal of work must to be done to help researchers make good decisions. This assertion is supported by Fabringer, et al. (1999) who noted that, "although these guidelines for RMSEA are generally accepted, it is of course possible that subsequent research might suggest modifications" (p.280).

Limitations of Current Research

Since the study involved simulations, the major limitation of the study, like any other simulation study is that the results might not be generalizable to other situations. This is especially true considering the fact that the manipulation of the parameters for this study yielded strong internal validity thereby compromising external validity. However, despite this limitation, the importance of the findings cannot be neglected because they help inform researchers on the need to move away from relying entirely on internal consistency as a measure of dimensionality of data to an approach where other analyses are considered as well. This point was reiterated by Cudeck and Hulin (2001) who stated that a reliable test need not conform to a one-factor model and conversely items that fit a single common factor may have low reliability.

References

Allen, M. J., & Yen, W. M. (1979). *Introduction to measurement theory*. Monterey, CA: Brooks/Cole.

Brooks, G. P. (2002). *MNDG*. [http://oak.cats.ohiou.edu/~brooksg/mndg.htm]. (Computer Software and Manual).

Cudek, R., & Hulen, C. (2001). Measurement. *Journal of Consumer Psychology*, *10*, 55–69.

Cudek, R., & O'Dell, L. L. (1994). Applications of standard error estimates in unrestricted factor analysis: Significance tests for factor loadings and correlations. *Psychological Bulletin*, *115*, 475-487.

Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods*, *1*, 16-29.

Fabringer, R. L., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4*, 272-299.

Henson, R. K. (2001). Understanding internal consistency reliability estimates: A conceptual primer on coefficient alpha. *Measurement and Evaluation in Counseling and Development*, *34*, 177-189.

Hong, S. (1999). Generating correlation matrices with model error for simulation studies in factor analysis: A combination of the Tucker-Koopman-Linn model and Wijsman's algorithm. *Behavior Research Methods, Instruments & Computers*, *31*, 727-730.

Hu, L., & Bentler, M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1-55.

Kanyongo, G. Y. (2006). The influence of reliability of four rules for determining the number of components to retain. *Journal of Modern Applied Statistical Methods*, *2*, 332-343.

Lawrey, D. N. (1940). The estimation of factor loadings by the method of maximum likelihood. *Proceedings of the Royal Society of Edinborough*, *60A*, 64-72.

Thurstone, L. L. (1947). *Multiple factor analysis*. Chicago, IL: University of Chicago Press.