How to Factor-Analyze Your Data Right: Do's, Don'ts, and How-To's.

Cómo analizar factorialmente tus datos: qué significa, qué no significa, y cómo hacerlo.

Masaki Matsunaga
Rikkyo University

ABSTRACT

The current article provides a guideline for conducting factor analysis, a technique used to estimate the population-level factor structure underlying the given sample data. First, the distinction between exploratory and confirmatory factor analyses (EFA and CFA) is briefly discussed; along with this discussion, the notion of principal component analysis and why it does not provide a valid substitute of factor analysis is noted. Second, a step-by-step walk-through of conducting factor analysis is illustrated; through these walk-through instructions, various decisions that need to be made in factor analysis are discussed and recommendations provided. Specifically, suggestions for how to carry out preliminary procedures, EFA, and CFA are provided with SPSS and LISREL syntax examples. Finally, some critical issues concerning the appropriate (and not-so-appropriate) use of factor analysis are discussed along with the discussion of recommended practices.

Key words: Confirmatory and Exploratory Factor Analysis ó LISREL - Parallel Analysis - Principal Component Analysis - SPSS

RESUMEN

El presente artículo provee una guía para conducir análisis factorial, una técnica usada para estimar la estructura de las variables a nivel de la población que subyacen a los datos de la muestra. Primero, se hace una distinción entre análisis factorial exploratorio (AFE) y análisis factorial confirmatorio (AFC) junto con una discusión de la noción de análisis de componentes principales y por qué este no reemplaza el análisis de factores. Luego, se presenta una guía acerca de cómo hacer análisis factorial y que incluye decisiones que deben tomarse durante el análisis factorial. En especial, se presentan ejemplos en SPSS y LISREL acerca de cómo llevar a cabo procedimientos preliminares, AFE y AFC. Finalmente, se discuten asuntos clave en relación con el uso apropiado del análisis de factores y practicas recomendables.

Palabras clave: Análisis factorial confirmatorio y exploratorio, LISREL, análisis paralelo, análisis de componentes principales, SPSS.

Factor analysis is a broad term representing a variety of statistical techniques that allow for estimating the population-level (i.e., unobserved) structure underlying the variations of observed variables and their interrelationships (Gorsuch, 1983; Kim & Mueller, 1978). As such, it is "intimately involved with questions of validity . . . [and] is at the heart of the measurement of psychological constructs" (Nunnally, 1978, pp. 112-113). In other words, factor analysis provides a diagnostic tool to evaluate whether the collected data are in line with the theoretically expected pattern, or structure, of the target construct and thereby to determine if the measures used have indeed measured what they are purported to measure.

Despite its importance, factor analysis is one of the most misunderstood and misused techniques. Although an increasing number of scholars have come to realize and take this issue seriously, evidence suggests that an overwhelming proportion of the research presented and published across fields still harbors ill-informed practices (Fabrigar, Wegener, MacCallum, & Strahan, 1999; Henson & Roberts, 2006; Park, Dailey, & Lemus, 2002; see Preacher & MacCallum, 2003, for a review). In hopes of making a difference in this trend, the current article provides a guideline for the "best practice" of factor analysis, along with illustrations of the use of widely available statistical software packages.

More specifically, this article first discusses the broad distinction among techniques called factor analysis—exploratory and confirmatory—and also notes the importance to distinguish factor analysis from principal component analysis, which serves a fundamentally different function from that of factor analysis. The next section illustrates a "hybrid" approach to factor analysis, where researchers initially run an exploratory factor analysis and follow-up on its results using a confirmatory factor analysis with separate data. This part provides detailed instructions of software usage (e.g., which pull-down menu in SPSS should be used to run an exploratory factor analysis or how to command LISREL to examine one's CFA model) in a step-by-step walk-through for factor analysis. Finally, a brief discussion on recommended "do's" and "don'ts" of factor analysis is presented.

IDENTIFYING TWO SPECIES OF FACTOR ANALYSIS

There are two methods for "factor analysis": Exploratory and confirmatory factor analyses (Thompson, 2004). Whereas both methods are used to examine the underlying factor structure of the data, they play quite different roles in terms of the purpose of given research: One is used for theory-building, whereas the other is used primarily for theory-testing.

Exploratory factor analysis (EFA) is used when researchers have little ideas about the underlying mechanisms of the target phenomena, and therefore, are unsure of how variables would operate vis-à-vis one another. As such, researchers utilize EFA to identify a set of unobserved (i.e., latent) factors that reconstruct the complexity of the observed (i.e., manifest) data in an essential form. By "essential form," it means that the factor solution extracted from an EFA should retain all important information available from the original data (e.g., between-individual variability and the covariance between the construct under study and other related constructs) while unnecessary and/or redundant information, as well as noises induced by sampling/measurement errors, are removed. Stated differently, EFA is a tool intended to help *generate a new theory* by exploring latent factors that best accounts for the variations and interrelationships of the manifest variables (Henson & Roberts, 2006).

Note that this type of factor analysis is used to estimate the unknown structure of the data. This is a critical point that distinguishes EFA from principal component analysis (PCA), which is often confused with EFA and therefore misused as its substitute or variant (Henson & Roberts, 2006). PCA, however, is fundamentally different from EFA because unlike factor analysis, PCA is used to summarize the information available from the given set of variables and reduce it into a fewer number of components (Fabrigar et al., 1999); see Figure 1 for a visual image of this distinction.

An important implication is that, in PCA, the observed items are assumed to have been assessed without measurement error. As a result, whereas both PCA and EFA are computed based on correlation matrices, the former assumes the value of 1.00 (i.e., perfect reliability) in the diagonal elements while the latter utilizes reliability estimates. Thus, PCA does not provide a substitute of EFA in either theoretical or statistical sense.

On the other hand, confirmatory factor analysis (CFA) is used to *test an existing theory*. It hypothesizes an a priori model of the underlying structure of the target construct and examines if this model fits the data adequately (Bandalos, 1996). The match between the hypothesized CFA model and the observed data is evaluated in the light of various fit statistics. Using those indices, researchers determine if their model represents the data well enough by consulting accepted standards (for discussions on the standards for CFA model evaluation, see Hu & Bentler, 1999; Kline, 2005, pp. 133-145; Marsh, Hau, & Wen, 2004).

Figure 1. *Conceptual distinction between factor analysis and principal component analysis. Note. An oval represents a latent (i.e., unobserved) factor at the population level, whereas a rectangle represents an observed variable at the sample level. An arrow represents a causal path. Note that the observed items in factor analysis are assumed to have been measured with measurement errors (i.e., s), whereas those in principal component analysis are not.*
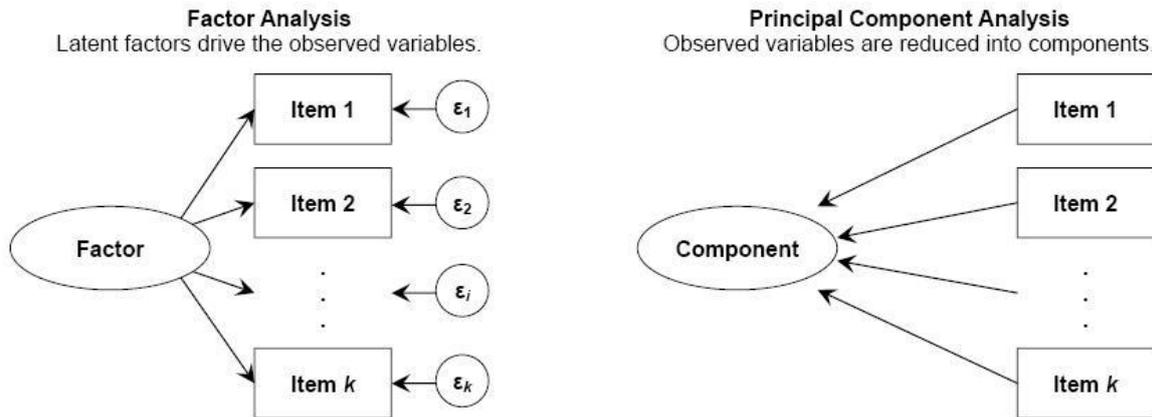
WALKING THROUGH FACTOR ANALYSIS: FROM PRE-EFA TO CFA

The EFA-CFA distinction discussed just above suggests a certain approach to identifying the target construct's factor structure. This approach utilizes both EFA and CFA, as well as PCA, at different steps such that: (a) an initial set of items are first screened by PCA, (b) the remaining items are subjected to EFA, and (c) the extracted factor solution is finally examined via CFA (Thompson, 2004). This "hybrid" approach is illustrated below. In so doing, the current article features computer software packages of SPSS 17.0 and LISREL 8.55; the former is used for PCA and EFA, where as the latter is for CFA.

Step 1: Generating and Screening Items

Item-generating procedures. First, researchers need to generate a pool of items that are purported to tap the target construct. This initial pool should be expansive; that is, it should include as many items as possible to avoid missing important aspects of the target construct and thereby maximize the face validity of the scale under study (for discussions of face validity, see Bornstein, 1996; Nevo, 1985). Having redundant items that appear to tap the same factor at this point poses little problem, while missing anything related to the target construct can cause a serious concern. Items may be generated by having a sample of the relevant population provide examples of manifest behaviors or perceptions caused by the target construct, or by the researchers based on the construct's conceptualization.

Next, researchers use the generated pool of items to collect quantitative data (i.e., interval data or ordinal data with a sufficiently large number of scale points; see Gaito, 1980; Svensson, 2000). Guadagnoli and Velicer (1988) suggest that the stability of component patterns (i.e., the likelihood for the obtained solution to hold across independent samples) is largely determined by the absolute sample size; although the recommended "cutoff value" varies widely, scholars appear to agree that a sample size of 200 or less is perhaps not large enough (and an N of 100 or less is certainly too small) in most situations (for reviews, see Bobko & Schemmer, 1984; Guadagnoli & Velicer, 1988).

Screening items. Once data are collected with a sufficiently large sample size, the next step is item-screening— to reduce the initial pool to a more manageable size by trimming items that did not emerge as expected. For this item-reduction purpose, PCA provides an effective tool, for it is a technique designed exactly to that end (i.e., reduce a pool of items into a smaller number of components with as little a loss of information as possible). To run this procedure, load the data into SPSS; then select "Analyze" from the top menu-bar and go "Dimension Reduction" "Factor"; because the SPSS default is set to run PCA, no

change is needed here except re-specifying the rotation method to be õPromaxö (see Figure 2).

Figure 2. *SPSS screen-shot for specifying PCA/EFA with the promax rotation method*
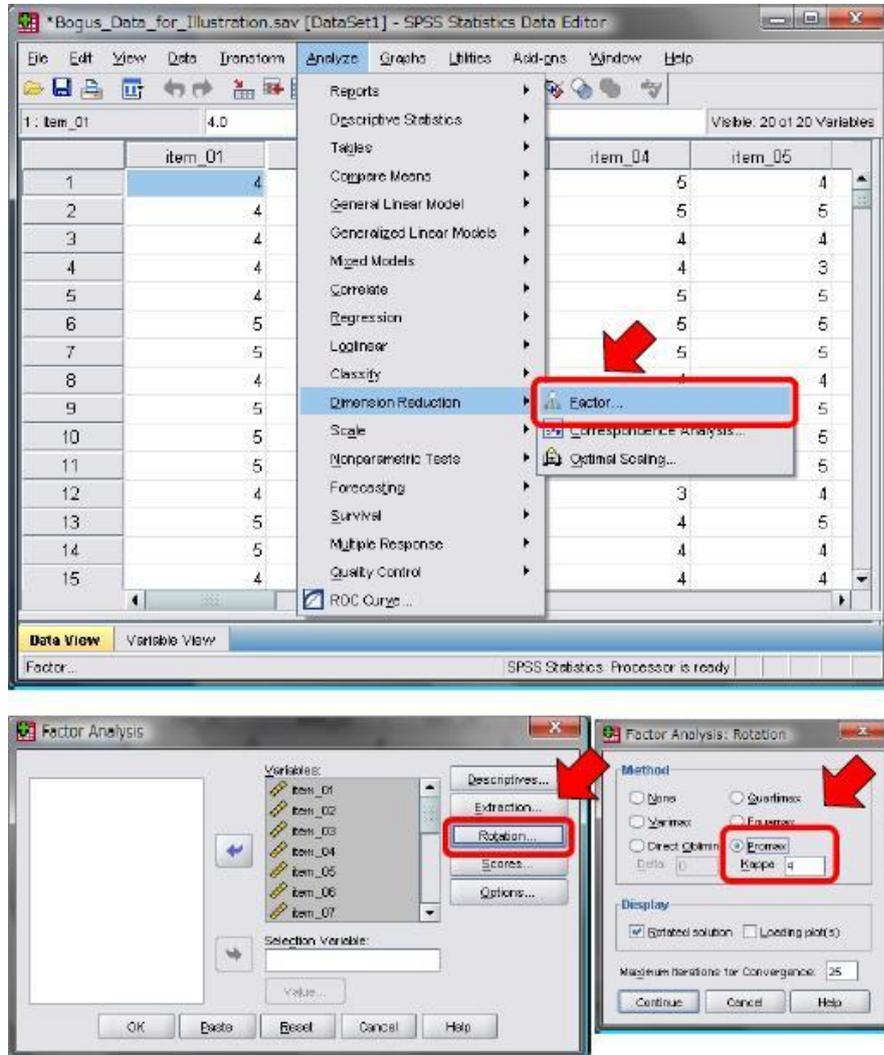
Promaxö is one of the rotation methods that provide solutions with correlated components/factors (called oblique solutions).1 The SPSS default is set to õVarimaxö rotation method, which classifies items into components in such a way that the resultant components are orthogonal to each other (i.e., no correlations among components) (Pett, Lackey, & Sullivan, 2003). This option, however, has at least three problems: (a) in almost all fields of social science, any factor/construct is to some extent related to other factors, and thus, arbitrarily forcing the components to be orthogonal may distort the findings; (b) even if the dimensions or sub-factors of the construct under study are indeed uncorrelated, such patterns should emerge naturally (not as an artifact of the researcher's choice) out of the promax rotation anyhow; and (c) although orthogonally rotated solutions are considered less susceptible to sampling error and hence more replicable, utilizing a large sample should address the concern of replicability (Hetzel, 1996; Pett et al., 2003). Therefore, it is suggested that the rotation method be specified to õpromax,ö which begins with a varimax solution and then raise the factor loadings to a stated power called kappa ( ), typically 2, 4, or 6; as a result, loadings of small magnitudeô .20, sayô will become close to zero (if is set

at 4, the loading of .20 will be .0016), whereas large loadings, although reduced, remain substantial; the higher the , the higher the correlations among factors/components. In computation, the promax method operates to obtain the solution with the lowest possible kappa so that the resultant factors/components are maximally distinguishable (Comrey & Lee, 1992).

With thus generated PCA output, researchers examine factor loadings; the purpose at this step is to identify items that have not loaded on to any of the major components with sufficiently large factor loadings and remove those items. Obliquely rotated solutions come along with two matrices, in addition to the unrotated component matrix: A pattern matrix and a structure matrix. The pattern matrix represents the variance in an observed item accounted for by the components identified via PCA, whereas the structure matrix contains coefficients made by both such õcommonö components/factors and an idiosyncratic factor that uniquely affects the given item; for this reason, although it often appears easier to interpret the results of PCA (and EFA) based on the pattern matrix alone, it is essential to examine both matrices to derive an appropriate interpretation (Henson & Roberts, 2006).

The next question to be answered in the item-screening procedure is a controversial one; that is, how large should an itemøs factor loading be to retain that item in the pool? Unfortunately, there seems no one set answer in the literature (Comrey & Lee, 1992; Gorsuch, 1983). Ideally, researchers should retain items that load clearly and strongly onto one component/factor while showing small to nil loadings onto other components/factors. More often than not, however, researchers find themselves in a situation to make some delicate, and in part subjective, decision. For example, an item may cross-load (i.e., having large factor loadings onto multiple components/factors), or its primary loading is not as large to call it õclearly loadedö; thus, there is a certain degree of judgment-call involved in this procedure.

One widely utilized approach is to focus on the highest loading with a cutoff. If an itemøs highest factor loading is greater than an a priori determined cutoff value, then researchers retain that item in the pool. On a conventional liberal-to-conservative continuum, setting the cutoff at .40 (i.e., items with a factor loading of .40 or greater is retained) is perhaps the lowest acceptable threshold, whereas .60 or.70 would be the limit of the conservative end. Another approach is to examine both the highest and second highest factor loadings. For example, in many social scientific studies, the .5/.2 or .6/.3 rule seems to constitute a norm, though studies employing a .6/.4 criterion are not uncommon (Henson & Roberts, 2006; Park et al., 2002). That is, an item is retained if its primary

loading is greater than .5-.6 and also if its second highest factor loading is smaller than .2-.3. This approach is more sophisticated than the first one in that it considers the larger pattern of factor loadings and partly addresses the issue of cross-loadings. The last approach is to focus on the discrepancy between the primary and secondary factor loadings, and retain items if their primary-secondary discrepancy is sufficiently large (usually .3-.4). Whichever approach researchers decide to take, it is important that they clearly state the approach they have taken in reporting the findings for both research fidelity and replicabilityøs sake.

At this point, the initial pool of items should be reduced in size. Before moving on to EFA, however, it is strongly recommended that researchers take a moment to review the remaining items and examine if it indeed makes theoretical sense to retain each item. By õmaking sense theoretically,ö it is intended that for each item, researchers must be able to provide a conceptual explanation for how the target construct drives the variation in the behavior/perception described by the respective item (Bornstein, 1996).

Step 2: Exploring the Underlying Structure of the Data

Having completed the item-generation and item-screening procedures illustrated above, the next major step for researchers to take is to conduct an exploratory factor analysis (EFA). Main purposes of this step include (a) determining the number of factors underlying the variation in and correlations among the items, (b) identifying the items that load onto particular factors, and (c) possibly removing items that do not load onto any of the extracted factors (Thompson, 2004). Before discussing how to achieve these goals in EFA, however, it is due to illustrate some preliminary considerations and procedures.

Collecting data for EFA. First, an entirely new set of empirical data should be collected for EFA. Although it for sure is enticing to subject a data set to PCA and then use the same data (with a reduced number of items) to EFA, such a õrecyclingö approach should be avoided because it capitalizes on chance. On the other hand, if similar component/factor structure patterns are obtained across multiple samples, it provides strong evidence that supports thus obtained solution (MacCallum, Widaman, Zhang, & Hong, 1999).

The question of sample size is to be revisited here. Gorsuch (1983) maintains that the sample size for an EFA should be at least 100. Comrey and Lee (1992) suggest that an N of 100 is õpoor,ö 200 is õfair,ö 300 is õgood,ö 500 is õvery good,ö and 1,000 or more is õexcellent.ö In a related vein, Everitt (1975) argued that the ratio of the sample size to the number of items (p) should be at least 10. These

recommendations are, however, ill-directed ones, according to MacCallum et al. (1999). Those authors suggest that the appropriate sample size, or N:p ratio, for a given measurement analysis is actually a function of several aspects of the data, such as how closely items are related to the target construct; if the items squarely tap the construct, the required sample size would be small, whereas a greater N would be needed if the correlations between items and the construct were small. According to MacCallum et al., if there are a good number of items per latent factor (i.e., preferably five or more items tapping the same factor) and each of those items are closely related to the factor in question, a sample size of 100-200 may be sufficient; as these conditions become compromised, larger samples would be necessary to obtain a generalizable factor solution.

Determining the number of factors. After collecting a new sample, researchers run an EFA. To do so, the õextractionö method needs to be changed from õprincipal componentö to õprincipal axisö; in addition, the rotation method should be specified to the õpromax.ö Unlike PCA, principal-axis factoring zeros in on the common variance among the items and delineates the latent factors underlying the data (Fabrigar et al., 1999; see also Figure 1).

Once the EFA output is generated, researchers need to determine how many factors should be retained. In any factor analysis, the total number of possible factors is equal to that of the items subjected to the analysis (e.g., if 10 items are factor-analyzed, the total number of factors is 10). Most of those factors, however, would not contribute appreciably to account for the data's variance or not be readily interpretable; generally, those insubstantial factors represent spurious noise or measurement error. Thus, it is critical to sieve those non-essential noises/errors out while extracting factors that are indeed theoretically meaningful.

There are a number of rules/criteria introduced in the literature that help determine the number of factors to retain (see Zwick & Velicer, 1986, for a review). The most frequently used strategy (and SPSS default) is to retain all factors whose computed eigenvalue is greater than 1.0. This ruleô also known as Kaiser-Guttman criterionô is, however, not an optimal strategy to identify the true factor structure of data, because it is known to overestimate the number of latent factors (Hayton, Allen, & Scarpello, 2004). Other factor-retention strategies include scree test (Cattell, 1966), minimum-average partial correlation (Velicer, 1976), Bartlett's 2 test (Bartlett, 1950, 1951), RMSEA-based maximum-likelihood method (Park et al., 2002), and parallel analysis (Horn, 1965; Turner, 1998).

Unfortunately, these rules and criteria often lead to different conclusions regarding the number of factors to retain, and many of them have been criticized for one reason or another. For example, scree test relies on the researcher's subjective decision and eyeball interpretation of the scree plot; in addition, similar to the Kaiser-Guttman criterion, it often results in overextracting factors. Bartlett's 2 test is based on the chi-square significance test, which is highly susceptible to the sample size; if the N is large enough, almost any analysis produces statistically significant results and therefore provides little information (see Zwick & Velicer, 1986). RMSEA-based maximum-likelihood method, although relatively accurate, is available only if the maximum-likelihood estimator is used.

Fortunately, research suggests that, among the rules/criteria mentioned above, parallel analysis (PA) provides the most accurate approach (Henson & Roberts, 2006). Basic procedures of PA are as follows: First, researchers run an EFA on their original data and record the eigenvalues of thus extracted factors; next, õparallel dataö are generatedô this is an artificial data set which contains the same number of variables with the same number of observations as the researchers' original data, but all variables included in this õparallel dataö set are random variables; the õparallel dataö are factor-analyzed and eigenvalues for factors are computed; this procedure of generating õparallel dataö and factor-analyzing them is repeated, usually 500-1000 times, and the eigenvalues of each trial are recorded; then, the average of those eigenvalues are compared to those for the factors extracted from the original data; if the eigenvalue of the original data's factor is greater than the average of the eigenvalues of the õparallel factorö (i.e., factor of the same rank extracted from the õparallel dataö), that factor is retained; if the eigenvalue of the original data's factor is equal to or smaller than the average, that factor is considered no more substantial than a random factor and therefore discarded (see Hayton et al., 2004, for a more detailed discussion).

Reviews of previous studies suggest that PA is one of theô if not *the*ô most accurate factor-retention method (e.g., Hayton et al., 2004; Henson & Roberts, 2006; Fabrigar et al., 1999). On this ground, it is recommended that researchers running an EFA should utilize PA as the primary method to determine the number of factors underlying the variance in given data, along with qualitative scrutiny such as examination of the interpretability of factors and theoretical expectations regarding the construct under study. In performing PA, researchers should consult the Hayton et al. (2004) article, which provides a template of SPSS syntax to run PA. As an alternative, there is a free software program developed by Watkins (2008). After downloading and starting the program, users enter the number of variables and subjects of their original data,

specify the number of replications of the "parallel data" generation, and click the "calculate" button; the program output shows the average of the eigenvalues of "parallel data" as specified in the previous step (see Figure 3). Researchers then compare those computed averages of

eigenvalues and those for their original data to determine which factors are to be retained and which are to be discarded.

Figure 3. A screen-shot of an outcome of Watkins' (2008) Monte-Carlo-based program for parallel analysis
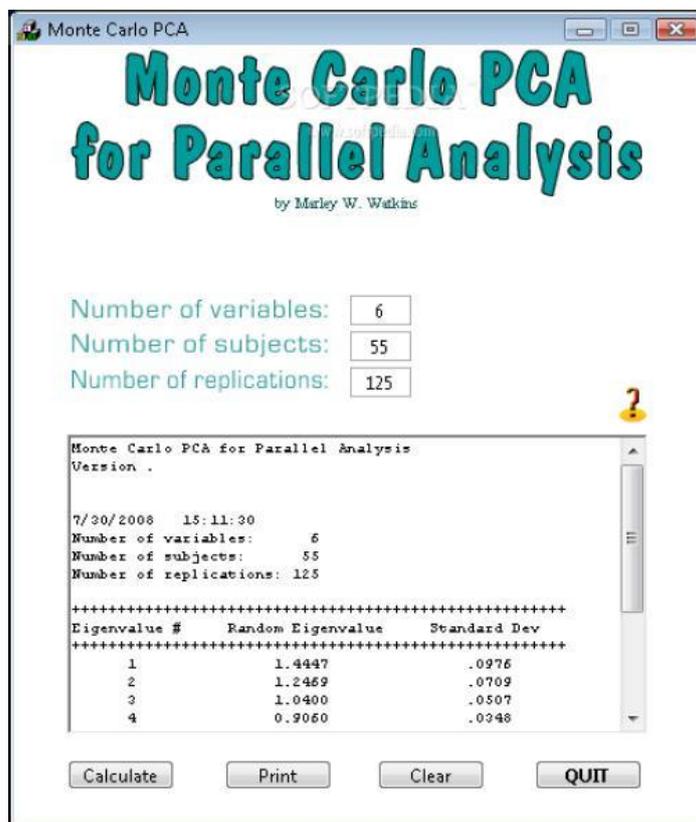
Determining the factor-loading patterns and trimming items (if necessary). Now that researchers have identified the number of factors to retain, they are ready to take the final step for an EFA, which is to determine which items load onto those factors. Similar to the item-trimming procedure in PCA, researchers examine both pattern and structure matrices of factor loadings. Then, the items to retain are determined based on some criteria (see the earlier discussion on various criteria to evaluate an item's factor loading). The resultant pool should contain only items that tap theoretically meaningful and interpretable factors, but not those that reflect insubstantial noises or measurement/sampling errors.

Step 3: Confirming the Factor Structure of the Data

After the PCA and EFA steps discussed just above, the third and final step is a confirmatory factor analysis

(CFA), by which researchers construct an explicit model of the factor structure underlying the given data and statistically test its fit (Russell, 2002). To run a CFA, researchers need to choose a software program, because there are several commercially-available programs specialized for CFA (or more generally, SEM). In this article, explanations and illustrations, as well as syntax templates, are provided based on LISREL, for it is: (a) one of the most widely utilized CFA/SEM software programs; (b) less expensive than other commercially-available programs (e.g., AMOS costs nearly $1,000 for the 1-yesr license of academic package, whereas the regular version of LISREL is available for $495); and (c) evaluated as superior to some of its rival software in terms of standard error calculation and parameter estimation (see, e.g., Byrne, 1998; von Eye & Fuller, 2003).

Specifying the model. Unlike EFA (which works best when researchers have little ideas about how the items are structured), in CFA, researchers must have an a priori theory on the factor structure underlying the given data (see Levine, 2005, for a more detailed discussion regarding the distinction between EFA and CFA). As such, researchers running a CFA need to specify a number of parameters to be estimated in the model.

First, the number of latent factors needs to be determined. This decision should be informed both by the results of the EFA, particularly that of the parallel analysis (Hayton et al., 2004), at the prior step of the analysis and the theory or conceptualization regarding the construct under study (i.e., whether it is conceptualized as a monolithic, unidimensional construct or as a multifaceted construct that embraces multiple interrelated sub-factors).

Second, researchers need to specify the patterns in which each item loads onto a particular factor. In CFA, there usually is an a priori expectation about how each item loads onto the hypothesized factor(s) such that each item has its unique pattern of non-zero factor loadings and zero loadings; this is a critical difference between CFA and EFA, because in the latter, all items load onto all factors (see Figure 4). Note in the CFA model in Figure 4 (left panel), no arrows are drawn from Factor A to Items 5, 6, and 7; likewise, Factor B is specified to underlie only Items 5-7 and not Items 1-4. This specification reflects researchers' expectation that: (a) there are two interrelated factors underlying the data under study; (b) the first four items included in the data tap Factor A and the remaining ones tap Factor B; and (c) any correlations between Items 1-4 with Factor B, as well as those between Items 5-7 with Factor A, are null at the population level (hence, the zero-loadings) and therefore should be ignored. Stated differently, in CFA, not drawing the factor-loading path between an item and a factor is as important and meaningful in terms of theoretical weight as specifying which items should load onto particular latent factors (Bandalos, 1996; Kline, 2005).

Figure 4. Conceptual distinction between confirmatory factor analysis (left) and exploratory factor analysis with an oblique rotation (right). Note. Ovals represent unobservable latent factors, whereas the rectangles represent observed items. Double-headed arcs indicate correlations. Arrows represent factor loadings
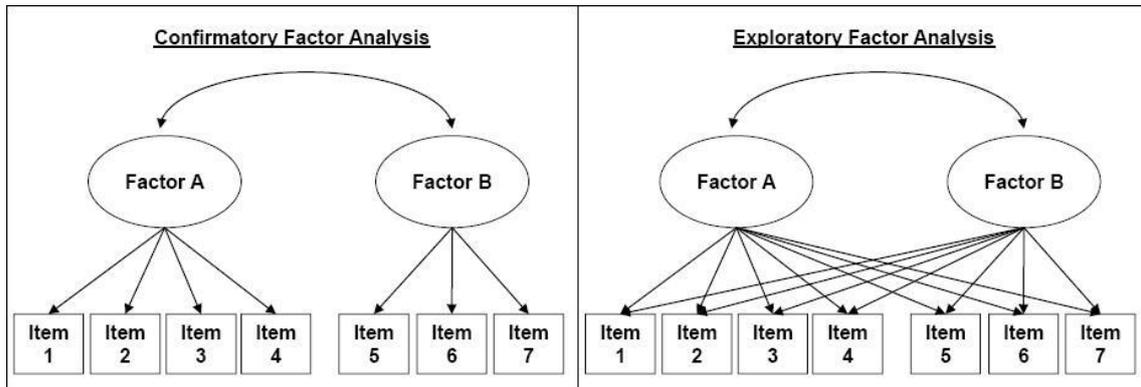
A third step to follow after researchers have specified the number of latent factors and factor-loading patterns (including the specification of zero loadings) is to execute the analysis. In so doing, researchers use the syntax. The current article provides a simple walk-through of a LISREL syntax based on the model shown in Figure 4, which features seven items and hypothesizes four of them (Items 1-4) tap a latent factor (Factor A) while the other three items (Items 5-7) load onto another factor (Factor B); additionally, those two latent factors are specified to be related to each other and no cross-loadings are allowed (readers interested in learning LISREL syntax more fully than what this article covers should be directed to Byrne, 1998). Figure 5 provides a LISREL syntax that specifies such a model; several key points of this syntax are illustrated below.

Walk-through of LISREL syntax. The command in the first line of the syntax ("TI") specifies the title of the analysis. Although this command is optional, it often helps to put an easily identifiable title to each analysis in back-referencing the results of multiple analyses. The command in the second line ("DA") is more important, as it feeds key features of the data into LISREL; "NI" specifies the number of items included in the data, while "NO" refers to the number of observations. "NG" refers to the number of groups to be analyzed; unless researchers are interested in analyzing multiple discrete groups, this should be the default (i.e., "NG=1"). "MA" specifies the type of matrix to

be analyzed; typically, a covariance matrix ("CM") is analyzed in CFA/SEM, so this part should be specified as "MA=CM."

The command in the third line, "LA," specifies the label of the items read into LISREL. It should be noted that, depending on the version of LISREL, there is a restriction to the number of letters that can be used for an item (e.g., only up to eight letters can be used in LISREL version 8.55). It is generally a desirable practice to use simple and systematic labels.

The command in the next lines ("CM") specifies the location of the data file to be used in the analysis. Note that the extension of the file (".COV") indicates that it is a covariance-matrix file, which can be extracted from an SPSS data file using a function of LISREL called PRELIS (see Byrne, 1998, for more detailed instructions regarding the use of PRELIS). The next command ("SE") specifies the sequence for the items to be read in the analysis; because model parameters are specified by matrix numbers (see below), it is advisable that researchers organize the items in a sensible manner at this point.
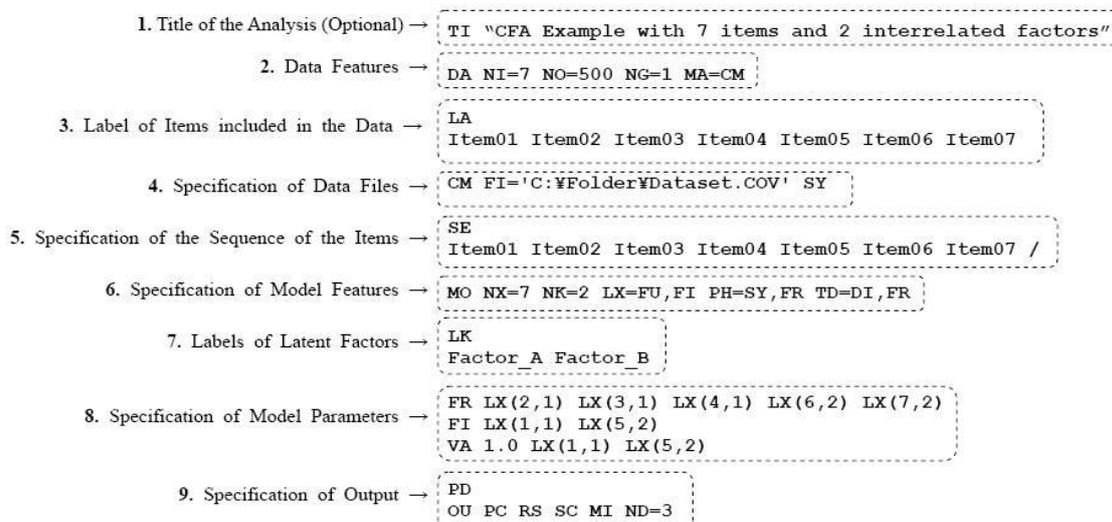
The "MO" command specifies the key features of the model to be analyzed. "NX" indicates the number of observed variables used in the analysis (which is specified as "NX=7" as there are seven items in the hypothetical model used as an example in this article), whereas "NK" command refers to the number of latent factors to be modeled. The next three commands ("LX," "PH," and "TD") specify how various parameters of CFA/SEM should be estimated; "LX" indicates the matrix of lambda parameters, or factor loadings; "PH" indicates the phi matrix, or the matrix of the latent factors' variances and covariances; and "TD" indicates the matrix of error

variances and covariances (for more detailed discussions of these different parameter matrices, see Byrne, 1998; Kline, 2005).

The next command, "LK," specifies the labels put on the latent factors. The same restriction of the number of letters as that for the "LA" command applies (see above), and again, it is recommendable to put easily recognizable and meaningful labels.

The next series of commands ("FR," "FI," and "VA") provide the heart of LISREL syntax, as they specify how the parameters should be estimated and thereby determine how the analysis should be executed. "FR" command specifies the parameters to be freely estimated; in a CFA context, these parameters represent the non-zero factor-loadings, or arrows drawn from a latent factor to observed items. Factor-loading pattern is specified in LISREL by matrix. As noted above, "LX" refers to the matrix of factor loadings (i.e., lambda parameters); under the "FR" command, each "LX" command indicates a particular element in the lambda matrix. For example, **LX(2,1)** indicates the factor loading of the second item in the given data (Item 2, in this article's example) onto the first factor hypothesized in the model (Factor A). Note in Figure 5, only Items 1-4 are specified to load onto Factor A, while Items 5-7 load onto Factor B alone. Recall it is as meaningful *not* to draw a factor-loading arrow from one factor to an item as drawing an arrow, because it equally impacts the model results. In other words, in the CFA context, zero loadings (which are not specified explicitly by researchers in the given syntax) are as essential as the non-zero loadings.

Figure 5. *An example LISREL syntax.*

1. Title of the Analysis (Optional) →	`TI "CFA Example with 7 items and 2 interrelated factors"`
2. Data Features →	`DA NI=7 NO=500 NG=1 MA=CM`
3. Label of Items included in the Data →	`LA` `Item01 Item02 Item03 Item04 Item05 Item06 Item07`
4. Specification of Data Files →	`CM FI='C:¥Folder¥Dataset.COV' SY`
5. Specification of the Sequence of the Items →	`SE` `Item01 Item02 Item03 Item04 Item05 Item06 Item07 /`
6. Specification of Model Features →	`MO NX=7 NK=2 LX=FU,FI PH=SY,FR TD=DI,FR`
7. Labels of Latent Factors →	`LK` `Factor_A Factor_B`

8. Specification of Model Parameters →	`FR LX(2,1) LX(3,1) LX(4,1) LX(6,2) LX(7,2)`
	`FI LX(1,1) LX(5,2)`
	`VA 1.0 LX(1,1) LX(5,2)`

9. Specification of Output →	`PD` `OU PC RS SC MI ND=3`

In contrast, "FI" command specifies the parameters that are fixed to a certain value and not estimated in the analysis; the value to be fixed at is specified by the "VA" command. Note that in the example syntax (Figure 5), two of the factor loadings—factor loading of Item 1 to Factor A ("LX(1,1)") and that of Item 5 to Factor B ("LX(5,2)")—are fixed at the value of 1.0. In CFA, one factor loading per factor needs to be fixed at a certain value to determine the scale of the respective factor and thereby identify it (see Kline, 2005, for a more detailed explanation of identification of latent factors).[3]

The last section in the LISREL syntax specifies the information to be included in the output. If the "PD" command is included in the syntax, a path diagram will be produced. The "OU" command specifies the type of information that will be contained in the output: "PC" is the request for the parameter correlations; the "RS" command invokes the fit-related information such as the residuals, QQ-plot, and fitted covariance matrix; the "SC" command is the request for completely standardized solutions of model parameters; the "MI" command adds to the output the information of modification indices regarding all estimated parameters; finally, the "ND" command specifies the number of decimal places displayed in the output.

Evaluating the model and parameter estimates. Having made these specifications, researchers are ready to run a CFA using LISREL (for instructions regarding the use of other major software programs, see Byrne, 1994, 2001). Next, they need to examine the analysis output; first the fit of the overall model to the data, and then the direction, magnitude, and statistical significance of parameter estimates. The initial step in this model-evaluation procedure is to examine if the model under the test fits the data adequately (Kline, 2005).

There are a number of fit indices and evaluation criteria proposed in the literature, and most of the software programs, such as AMOS, LISREL, or EQS, produce more than a dozen of those indices; and yet evaluation of the fit of a given CFA model marks one of the most hotly debated issues among scholars (e.g., Fan, Thompson, & Wang, 1999; Hu & Bentler, 1998, 1999; Hu, Bentler, & Kano, 1992; Marsh et al., 2004). Therefore, it is critical that researchers not only know what each of the major fit indices represents but also have the ability to interpret them properly and utilize them in an organic fashion to evaluate the model in question (for reviews of CFA/SEM fit indices, see Hu & Bentler, 1998; Kline, 2005).

The chi-square statistic provides the most conventional fit index; it is called "exact fit index" and indicates the degree of discrepancy between the data's variance/covariance pattern and that of the model being tested. Although conceptually simple and easy to interpret (i.e., if the computed χ^2 value is statistically significant, the model is considered discrepant from the population's true covariance structure), the practice of drawing solely on this index has been criticized because the χ^2 test is highly susceptible to the impact of sample size: The larger the sample size, the more likely the results of the test become significant, regardless of the model's specific features (Russell, 2002). Additionally, research has shown that the results of a χ^2 test are vulnerable to the violation of some assumptions (e.g., non-normality).

To compensate this problem associated with the exact fit index, Hu and Bentler (1999) suggest that researchers using a CFA/SEM should employ what they call "two criteria" strategy. That is, in addition to the information regarding the exact fit of a model (i.e., χ^2 value), researchers should examine at least two different types of fit indices and thereby evaluate the fit of the model. There are several "clusters" of fit indices such that all indices included in a cluster reflect some unique aspect of the model, while different clusters help examine the model from different angles (Kline, 2005). *Root mean square error of approximation* (RMSEA; Steiger, 1990) "estimates the amount of error of approximation per model degree of freedom and takes sample size into account" (Kline, 2005, p. 139) and represents the cluster called "approximate fit index." Unlike the exact fit index, this type of fit index evaluates the model in terms of how close it fits to the data. Hu and Bentler recommend that RMSEA should be .06 or lower, though Marsh et al. (2004) suggest that.08 should be acceptable in most circumstances (see also Thompson, 2004).

The second cluster of fit index is called "incremental fit index," which represents the degree to which the tested model accounts for the variance in the data vis-à-vis a baseline model (i.e., a hypothetical model that features no structural path, factor loading, or inter-factor correlations at all). Major incremental fit indices include *comparative fit index* (CFI; Bentler, 1990), *Tucker-Lewis index* (TLI; Tucker & Lewis, 1973), and *relative noncentrality index* (RNI; McDonald & Marsh, 1990). Hu and Bentler (1999) suggest that, for a model to be considered adequate, it should have an incremental fit index value of .95 or higher, although the conventional cutoff seen in the literature is about .90 (Russell, 2002).

A third cluster of model fit index, called residual-based index, focuses on covariance residuals, or discrepancies between observed covariances in the data and the covariances predicted by the model under study. The most widely utilized residual-based index is the *standardized root mean square residual* (SRMR), which indicates the average value of the standardized residuals

between observed and predicted covariances. Hu and Bentler (1999) and Kline (2005) both suggest that SRMR should be less than .10.

DISCUSSION

The current article has discussed the notion of factor analysis, with a note on the distinction between exploratory and confirmatory factor analyses (as well as principal component analysis). A step-by-step walk-through that leads interested readers from the preliminary procedures of item-generation and item-screening via PCA to EFA and CFA was provided. To complement these discussions, this final section provides general recommendations concerning the use of exploratory and confirmatory factor analyses, followed by some concluding remarks on the importance of conducting EFA/CFA properly.

General Recommendations

There are a number of issues concerning the use of EFA/CFA. Some of them already have been discussed earlier in this article; but all recommendations mapped out below should be noted nonetheless, should researchers wish to factor-analyze their data properly.

Sample size. As repeated several times in the current article, researchers using factor analysis, either EFA or CFA or both, should strive to gather as large data as possible. Based on his review of the back numbers of *Personality and Social Psychology Bulletin*, Russell (2002) points to õa general need to increase the size of the samplesö (p. 1637). Similarly, MacCallum and his colleagues caution that the sample size of the factor analyses reported in the psychology literature is often too small and it suggests a considerable risk of misspecification of the model and bias in the existing measurement scales (MacCallum et al., 1999). In view of the results of these and other quantitative reviews of the literature (e.g., Fabrigar et al., 1999), it is strongly suggested that researchers using factor analysis should make every effort to increase their studyøs sample size.

Factor extraction method. In running an EFA, it is still not uncommon to see PCA used as a substitute. PCA is, however, fundamentally different from factor analysis, as it has been designed to summarize the observed data with as little a loss of information as possible, not to identify unobservable latent factors underlying the variations and covariances of the variables (Kim & Mueller, 1978; Park et al., 2002). PCA should only be used in reducing the number of items included in the given scale (i.e., item-screening).

Determination of the number of factors. For both EFA and CFA, the number of latent factors should be determined primarily on the ground of theoretical expectations and conceptualization of the target construct. At the same time, in cases where the constructøs true nature is uncertain or debated across scholars, empirical exploration via EFA is duly called for. In such cases, researchers should *not* rely on the Keiser-Guttman (i.e., eigenvalue $\times$ 1.0) rule; although this is set as the default criterion of factor extraction in most statistical software packages, including SPSS and SAS, it is known to produce inaccurate results by overextracting factors (Fabrigar et al., 1999; Gorsuch, 1983). Instead, this article recommends that researchers running an EFA should utilize parallel analysis (PA; Turner, 1998). Research shows that PA often provides one of the most accurate factor-extraction methods; although it has long been underutilized partly due to the complexity of computation, recent development of freely available computer programs and SPSS tutorials greatly have reduced the burden to conduct a PA (Hayton et al., 2004; Watkins, 2008).

Rotation methods. The õorthogonal-versus-obliqueö debate marks one of the most hotly debated issues among scholars (see Comrey & Lee, 1992; Fabrigar et al., 1999; Henson & Roberts, 2006; Park et al., 2002; Preacher & MacCallum, 2003). This article suggests that any EFA should employ an oblique-rotation method, even if the conceptualization of the target construct suggests that factors should be unrelated (i.e., orthogonal to one another) for three reasons. First, as noted earlier, almost all phenomena that are studied in social sciences are more or less interrelated to one another and completely orthogonal relationships are rare; therefore, imposing an orthogonal factor solution is likely to result in biasing the reality. Second, if the construct under study indeed features unrelated factors, this orthogonality should be empirically verified (i.e., if factors are indeed unrelated, it should be revealed via EFA employing an oblique-rotation method). Third, because in most CFAs latent factors are specified to be interrelated (cf. Figure 4), employing an oblique-rotation method helps maintain conceptual consistency across EFA and CFA within the hybrid approach introduced in the current article (i.e., exploring the data via EFA first, followed by CFA).

Estimation methods. Similar to the factor-extraction method in EFA, what estimator should be used in CFA marks a controversial topic (Kline, 2005). Nevertheless, the literature appears to suggest that the use of maximum-likelihood (ML) estimation method provides the de facto standard (Bandalos, 1996; Russell, 2002). Although there are a number of estimators proposed and utilized in the literature, the vast majority of the published CFA studies employ ML and research suggests that it produces accurate results in most situations (Fan et al., 1999; Levine, 2005; Thompson, 2004; see Kline, 2005, for a review of various estimation methods).

Model fit evaluations. Whereas the specific standards employed to evaluate a given CFA model differs across fields and influenced by numerous factors (e.g., development of the theory/construct in question), researchers seem to agree at least that multiple criteria should be used to evaluate a model (Fan et al., 1999; Hu & Bentler, 1999). The current article suggests that a CFA model should be evaluated in the light of its exact fit (i.e., χ^2 value), RMSEA, one of the incremental fit indices (CFI, TLI, or RNI), and SRMR (see also Kline, 2005). If the model exhibits an adequate fit with regard to all of those indicesô that is, the computed χ^2 value is not statistically significant, RMSEA is smaller than .06, CFI/TLI/RNI is greater than .95, and SRMR is smaller than .10ô then, researchers can confidently claim that it represents the latent factor structure underlying the data well. Perhaps some criteria may be loosened without causing overly drastic consequences; for example, RMSEA smaller than .08 should be considered acceptable under most circumstances and so is CFI/TLI/RNI greater than .90 (see Fan et al., 1999; Marsh et al., 2004). In a related vein, it seems noteworthy that the number of items being analyzed in a given CFA is negatively associated with the model's goodness of fit. In other words, generally speaking, the more the items, the worse the model fit (Kenny & McCoach, 2003). This finding points to the importance of the item-generating and item-screening procedures, because it illuminates that not only selecting quality items helps the model to fit well, but also failing to sieve unnecessary items out eventually results in harming the model and therefore impedes the analysis.

Concluding Remarks

Invented in the first decade of the 20th century (Spearman, 1904), factor analysis has been one of the most widely utilized methodological tools for quantitatively oriented researchers over the last 100 years (see Gould, 1996, for a review of the history of the development of factor analysis). This article illuminated the distinction between two species of factor analysis: exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). As a general rule, it is advisable to use CFA whenever possible (i.e., researchers have some expectations or theories to draw on in specifying the factor structure underlying the data) and, as Levine (2005) maintains, the conditions required for the use of CFA are readily satisfied in most cases where factor analysis is used. After all, researchers should have at least some idea about how their data are structured and what factors underlie their observed patterns. Thus, researchers interested in identifying the underlying structure of data and/or developing a valid measurement scale should consider using CFA as the primary option (Thompson, 2004).

At the same time, readers should keep in mind that there are times and places wherein the use of EFA is most appropriate. Unfortunately, however, reviews of the literature suggest that in many cases where EFA is used, researchers tend to make inappropriate decisions; they may use PCA instead of EFA, enforce an orthogonally-rotated solution, draw on the Keiser-Guttman rule by uncritically using the default setting of the software package, or simply collecting too few samples (see, e.g., Fabrigar et al., 1999; Gorsuch, 1983; Hetzel, 1996; Park et al., 2002; Preacher & MacCallum, 2003). These ill-informed practices (and other mistakes discussed in this article and also in its cited references) can severely harm the analysis by inducing bias and distorting results. It is hoped that the walk-through illustrated in the current article and the discussions provided therein will give young scholars an idea about how to factor-analyze their data properly and break the cycle of reproducing malpractices that have been repeatedly made and critiqued in the literature of factor analysis.

REFERENCES

Bandalos, B. (1996). Confirmatory factor analysis. In J. Stevens (Ed.), *Applied multivariate statistics for the social sciences* (3rd ed., pp. 389-420). Mahwah, NJ: LEA.

Bartlett, M. S. (1950). Tests of significance in factor analysis. *British Journal of Psychology, 3*, 77-85.

Bartlett, M. S. (1951). A further note on tests of significance in factor analysis. *British Journal of Psychology, 4*, 1-2.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*, 238-246.

Bentler, P. M. (2000). Rites, wrongs, and gold in model testing. *Structural Equation Modeling, 7*, 82-91.

Bobko, P., & Schemmer, F. M. (1984). Eigen value shrinkage in principal component based factor analysis. *Applied Psychological Measurement, 8*, 439-451.

Boomsma, A. (2000). Reporting analyses of covariance structure. *Structural Equation Modeling, 7*, 461-483.

Bornstein, R. F. (1996). Face validity in psychological assessment: Implications for a unified model of validity. *American Psychologist, 51*, 983-984.

Byrne, B. M. (1994). *Structural equation modeling with EQS and EQS/Windows: Basic concepts, applications, and programming*. Thousand Oaks, CA: Sage.

Byrne, B. M. (1998). *Structural equation modeling with LISREL, PRELIS, and SIMPLIS: Basic concepts, applications, and programming*. Mahwah, NJ: LEA.

Byrne, B. M. (2001). *Structural equation modeling with AMOS: Basic concepts, applications, and programming*. New York: LEA.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1*, 245-276.

Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis*. Hillsdale, NJ: LEA. Do's, Don'ts, and How-To's of Factor Analysis 26.

Everitt, B. S. (1975). Multivariate analysis: The need for data and other problems. *British Journal of Psychiatry, 126*,237-240.

Fabrigar, L. R., 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.

Fan, X., Thompson, B., & Wang, L. (1999). Effects of sample size, estimation methods, and model specification on structural equation modeling fit indexes. *Structural Equation Modeling, 6*, 56-83.

Gaito, J. (1980). Measurement scales and statistics: Resurgence of an old misconception. *Psychological Bulletin, 87*, 564-567.

Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: LEA.

Gould, S. J. (1996). *The mismeasure of man: The definitive refutation of the argument of* The Bell Curve. New York: W. W. Norton.

Guadagnoli, E., & Velicer, W. F. (1988). Relation of sample size to the stability of component patterns. *Psychological Bulletin, 103*, 265-275.

Hayton, J. C., Allen, D. G., & Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational Research Methods, 7*, 191-205.

Henson, R. K., & Roberts, J. K. (2006). Use of exploratory factor analysis in published research: Common errors and some comment on improved practice. *Educational and Psychological Measurement, 66*, 393-416.

Hetzel, R. D. (1996). A primer on factor analysis with comments on patterns of practice and reporting. In B. Thompson (Ed.), *Advances in social science methodology: Vol. 4* (pp. 175-206). Greenwich, CT: JAI.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30*, 179-185.Do's, Don'ts, and How-To's of Factor Analysis 27.

Hu, L.-T., Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods, 3*, 424-453.

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

Hu, L.-T., Bentler, P. M., & Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted? *Psychological Bulletin, 112*, 351-362.

Kenny, D. A., & McCoach, D. B. (2003). Effects of the number of variables on measures of fit in structural equation modeling. *Structural Equation Modeling, 10*, 333-351.

Kim, J.-O., & Mueller, C. W. (1978). *Introduction to factor analysis: What it is and how to do it*. Newbury Park, CA: Sage.

Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York: Guilford.

Levine, T. R. (2005). Confirmatory factor analysis and scale validation in communication research. *Communication Research Reports, 22*, 335-338.

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods, 4*,84-99.

McDonald, R. P., & Marsh, H. W. (1990). Choosing a multivariate model: Noncentrality and goodness of fit. *Psychological Bulletin, 107*, 247-255.

Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In search of Golden Rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgenralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling, 11*, 320-341.

McDonald, R. P., & Ho, M.-H. R. (2002). Principles and practice in reporting structural equation analyses. *Psychological Methods, 7*, 64-82.Do's, Don'ts, and How-To's of Factor Analysis 28.

Mulaik, S. A. (2000). Objectivity and other metaphors of structural equation modeling. In R. Cudeck, S. Du Toit, & D. Sörbom (Eds.), *Structural equation modeling: Present and future* (pp. 59-78). Lincolnwood, IL: Scientific Software International.

Nevo, B. (1985). Face validity revisited. *Journal of Educational Measurement, 22*, 287-293.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

Park, H. S., Dailey, R., & Lemus, D. (2002). The use of exploratory factor analysis and principal components analysis in communication research. *Human Communication Research, 28*, 562-577.

Pett, M. A., Lackey, N. R., & Sullivan, J. J. (2003). *Making sense of factor analysis: The use of factor analysis for instrument development in health care research*. Thousand Oaks, CA: Sage.

Preacher, K. J., & MacCallum, R. C. (2003). Repairing Tom Swift's electric factor analysis machine. *Understanding Statistics, 2*, 13-43.

Russell, D. W. (2002). In search of underlying dimensions: The use (and abuse) of factor analysis in *Personality and Social Psychology Bulletin*.

Personality and Social Psychology Bulletin, *28*, 1629-1646.

Sörbom, D. (1989). Model modification. *Psychometrika, 54*, 371-384.

Spearman, C. (1904). General intelligence objectively determined and measured. *American Journal of Psychology, 15*, 201-293.

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research, 25*, 173-180.

Svensson, E. (2000). Comparison of the quality of assessment using continuous and discrete ordinal rating scales. *Biometrical Journal, 42*, 417-434.

Thompson, B. (2004). *Exploratory and confirmatory factor analysis*. Washington, DC: American Psychological Association.

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika, 38*, 1-10.

Turner, N. E. (1998).The effect of common variance and structure pattern on random data eigenvalues: Implications for the accuracy of parallel analysis. *Educational and Psychological Measurement, 58*, 541-568.

Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika, 41*, 321-327.

von Eye, A., & Fuller, B. E. (2003). A comparison of the SEM software packages LISREL, EQS, and Amos. In B. Pugesek, A. Tomer, & A. von Eye (Eds.), *Structural equations modeling: applications in ecological and evolutionary biology research* (pp. 355-391). Cambridge, UK: Cambridge University Press.

Watkins, M. (2008). Monte Carlo for PCA parallel analysis (Version 2.3) [Computer software]. Retrieved October 6, 2009 from www.softpedia.com/get/Others/Home-Education/Monte-Carlo-PCA-for-Parallel-Analysis.shtml.

Zwick, W. R., & Velicer, W. F. (1986). Factors influencing five rules for determining the number of components to retain. *Psychological Bulletin, 99*, 432-442