Principal Components Analysis, Exploratory Factor Analysis, and Confirmatory Factor Analysis

by Frances Chumney

Principal components analysis and factor analysis are common methods used to analyze groups of variables for the purpose of reducing them into subsets represented by latent constructs (Bartholomew, 1984; Grimm & Yarnold, 1995). Even though PCA shares some important characteristics with factor analytic methods such as exploratory factor analysis (EFA) and confirmatory factor analysis (CFA), the similarities between the two types of methods are superficial. The most important distinction to make is that PCA is a descriptive method, whereas EFA and CFA are modeling techniques (Unkel & Trendafilov, 2010). Together, PCA, EFA, and CFA are used to analyze multiple variables for the purposes of data reduction, scale construction and improvement, and evaluation of validity and psychometric utility (Brown, 2006; Brown, Chorpita, & Barlow, 1998). This paper provides a brief review of PCA, EFA, and CFA, describes the appropriate problems to which each might be correctly applied, and discusses the similarities and differences between these three methods.

PRINCIPAL COMPONENTS ANALYSIS

Principal components analysis (PCA; Goodall, 1954) is a method for explaining the maximum amount of variance among a set of items by creating linear functions of those items for the purpose of identifying the smallest number of linear functions necessary to explain the total variance observed for the item set in the correlation matrix (Grimm & Yarnold, 1995). Put another way, PCA identifies the smallest number of factors or components necessary to explain as much (or all) of the variance as possible. In this context, a factor or component is a set of variables that, when combined in a linear fashion, explains some portion of the observed variance (Mulaik, 1990). The foundational assumption of PCA is that the total variance of each variable or item can be explained by summing the true variance and error variance which characterizes that item (Hotelling, 1933).

When identifying factors or components, PCA first identifies the linear combination of variables that explains the largest proportion of total variance, and that factor is known as the first component. For a component with 6 indicators, the linear function takes the form:

$$C = l_1 i_1 + l_2 i_2 + l_3 i_3 + l_4 i_4 + l_5 i_5 + l_6 i_6$$

where *C* is a component or outcome of the linear function, *l* is an item loading, and *i* is an item. The second component is the linear combination of variables that explains the next largest proportion of variance that is not explained by the first component, and so on (Grimm & Yarnold, 1995). In PCA, each component is called an eigenvector; the portion of the total variance explained by each eigenvector is its eigenvalue (Wold, 1987). However, because it is possible to identify a large number of components, criteria have been developed for evaluating whether an item should be considered part of a component and whether a component should be retained as important for explaining variance.

Determination of the number of components to extract from a PCA may be completed a priori or by following a decision rule after the results of the analysis are produced. A priori decision rules that might be established for this task are the percentage of variance criterion and the a priori criterion. The percentage of variance criterion dictates that extracted factors are retained from an analysis until those factors combine to account for a specific percentage of the total variance (Grimm & Yarnold, 1995). The a priori criterion (Hair, Anderson, Tatham, & Black, 1992) is the specification before an analysis is conducted of a certain number of components that should be extracted. Ideally, the a priori criterion process is guided by theory or previous research. Decision rules that can be applied after an analysis is complete are Kaiser's stopping rule (Kaiser, 1960) and the scree test. Kaiser's rule simply means that all eigenvectors with an eigenvalue of 1.0 or greater are extracted from the data and retained as part of the solution. The value of 1.0 is not arbitrary in this context; 1.0 is the variance of a standardized variable. The scree test (Cattell, 1966) is dependent on a scree plot, or graphical display of the eigenvalues for successive eigenvectors. In most cases, the scree plot will visually depict a quick decline in eigenvalues followed by a series of less dramatic decreases. The eigenvector which represents the transition between the two trends (aka, the "elbow") and all successive eigenvectors are dropped (Gorsuch, 1983).

In addition to deciding how many eigenvectors to extract, the researcher must decide how many/which items to include on each eigenvector or component. This decision is generally determined by the factor loading coefficient which describes the relationship (correlation) between each item and the eigenvector (Wold, 1987). Loadings may be positive or negative, and can have absolute values that range from 0.00 to 1.00. The factor loading coefficient is interpreted in the same manner as a correlation coefficient; the coefficient can also be used to calculate r^2 for the item to determine the portion of variance that is shared by the item and the component. When evaluating factor loading coefficients, researchers often rely on a general rule of thumb that the absolute value of a factor loading should be $\geq .30$ (Grimm & Yarnold, 1995) in order to be retained as an item on the component and included in interpretation of the latent variable represented by that component. However, this rule of thumb may not be a good practice in all instances. Because it is known that the level of significance associated with a correlation coefficient is influenced by the sample size, and because the factor loading coefficient is a correlation, it follows that the point at which a factor loading implies a significant relationship between the item and component is dependent upon sample size. Thus, some researchers have argued that the absolute value used for determining a relationship between an item and component should be at least as large as the value of the correlation necessary for $p < .05$ Type I error rate, given the sample size (Stevens, 1986). Compared to liberal criterions (close to 0.0), a more conservative criterion is important because it helps control the overall Type I error rate associated with the analysis.

In addition to evaluating the strength of the relationship between an item and its component, it is necessary to consider the need to interpret the latent variables represented by the components and eigenvectors. The relationship of items that allows for the most straightforward and interpretable solution is known as simple structure (Thurstone, 1947), and occurs when each variable has a loading close to zero on all but one eigenvector, all variables have loadings close to zero on most eigenvectors, and all items that load on a eigenvector have loadings closer to 1 than 0. If a simple structure is not identified, it is possible to rotate the eigenvectors to achieve simple structure and increase the interpretability of the factors. Eigenvectors can be rotated using orthogonal (uncorrelated, e.g., varimax, quartimax) or oblique (correlated, e.g., promax) methods (Osborne & Costello, 2009). In rotation, orthogonal and oblique refer to the relationships between scores on the various eigenvectors. Rotation is a mathematical manipulation meant to minimize the factor loadings close to 0 and maximize the loadings that are close to 1.0 for the purpose of simplifying interpretability of factors without changing the solution (Brown, 2006). Though useful, the concept of rotation raises the question of factor indeterminacy, a common criticism of PCA. Factor indeterminacy (Brown, 2006; Maraun, 1996; Steiger, 1979) means there is not a single correct solution to solving the puzzle of the relationships between the items.

Applications of PCA

The purpose of PCA is to identify underlying dimensions that explain response patterns (Wold, 1987). Thus, PCA is appropriate for use when applied to a set of responses such as those obtained from a questionnaire. In addition to simply identifying factors, PCA is useful for examining the number of components, identifying which items make up each component and how strongly they relate to the component, and investigating the strength of the relationship between components (Grimm & Yarnold, 1995). In these ways, PCA is most useful when used as a descriptive tool for the process of measure development by contributing to the researcher's understanding of the strengths and weaknesses of the measure in terms of content validity and structural/factorial validity (Aladwani & Palvia, 2002).

EXPLORATORY FACTOR ANALYSIS

Exploratory factor analysis (EFA; Bartholomew, 1984) is a data-driven, exploratory method for determining the number of common factors underlying a response set as well as the relationship between individual items and those common factors (Fabrigar, Wegener, MacCallum, & Strahan, 1999; Kline, 2011). The purpose of EFA is to evaluate the dimensionality of a response set by identifying interpretable factors necessary to explain the relationships between responses. In EFA, the observed variables are called *indicators* and the extracted factors are assumed to be the cause for the observed responses (Brown, 2006). EFA is based on the common factor model (Jöreskog, 1969; Thurstone, 1947) and is represented by the equation:

$$y_j = \lambda_{j1}\eta_1 + \lambda_{j2}\eta_2 + \cdots + \lambda_{jm}\eta_m + \varepsilon_j,$$

where y is an indicator, λ is a factor loading, η is a factor, and ε is the unique variance of y (Brown, 2006). Thus, the foundational assumption of EFA is that the total variance of each variable or item can be explained by summing the common variance, the specific variance, and the error variance associated with that item (Grimm & Yarnold, 1995). In this context, common variance is the variance of an item that is shared with other items, specific variance is the variance of an item that does not correlate with other items, and error variance is the portion of the total variance attributed to random variation.

EFA is an iterative estimation process during which the correlation matrix of the observed data is analyzed using PCA, communalities are estimated for the factors extracted during the PCA, the communalities replace the item variance values on the diagonal of the correlation matrix, and the process repeats by analyzing the adjusted matrix using PCA. This process continues to repeat until the changes in the communalities derived from each PCA is minimal (Grimm & Yarnold, 1995). The communality of a variable is the proportion of the variable that is not attributed to the variable's uniqueness, which is conceptualized as the sum of the specific variance and error variance (Russell, 2002). Alternatively, the communality of a variable can be understood as the sum of the variable's squared factor loadings across all factors (Grimm & Yarnold, 1995). Thus, communality is defined as [1 – (specific variance + error variance)], and represents the portion of the variable's variance account for by the extracted factors. Several methods of estimation are available for extracting factors through EFA, including maximum likelihood (ML), principal factors (PF), and least squares (Unkel & Trendafilov, 2010). ML implies an assumption of multivariate normality, which means that item responses are normally distributed, as are their linear combinations (Brown, 2006). When this assumption is violated, ML frequently leads to unstable estimates.

Applications of EFA

EFA is particularly suited to exploratory analyses aimed to determine the number of dimensions underlying a response set, the subjective meaning of each dimension, how the items relate to the dimensions, and how the dimensions relate to each other (Mulaik, 1990). EFA is also an appropriate tool for the evaluation of a measure's content validity, as the extracted factors represent the dimensions that they measure (Floyd & Widaman, 1995; Grimm & Yarnold, 1995).

Exploratory Factor Analysis vs. Principal Components Analysis

In many ways, EFA and PCA are very similar. For instance, both EFA and PCA attempt to reduce a set of observed data into components, rotation can be used with both methods to achieve simple structure and increase the interpretability of the results, and both methods lack strong empirical evidence for cut-off rules to use in the determination of how large an item's loading should be on a factor (or component, as is the case with PCA; Grimm & Yarnold, 1995).

Despite these similarities, EFA differs from PCA in important ways. EFA differs from PCA in its approach to factor extraction, in that EFA analyzes the covariance among variables to produce either correlated or uncorrelated factors and PCA analyzes the variance of the variables to produce only uncorrelated components (Brown, 2006). Thus, in PCA, extracted components are linear combinations of the raw data, and in EFA extracted factors are an explanation of the raw data and the relationships within them. EFA is a more elegant analysis, but PCA is often viewed favorably because it is simpler, does not converge to improper solutions, allows for a direct calculation of participant scores on components, and often leads to the same interpretations as EFA (Brown, 2006). EFA also differs from PCA in that PCA is more appropriate when the goal is to refine a measure (e.g., identify dimensions that need more items, reduce the total number of items for future administrations of a measure), and EFA is more appropriate for reducing the data from multiple variables into fewer variables that are representative of the larger variable group (Brown, 2006). Further, because EFA is grounded in common factor theory (and includes an error term), it is a more appropriate tool (compared to PCA) when measurement error is expected in the data set (Brown, 2006).

CONFIRMATORY FACTOR ANALYSIS

Confirmatory factor analysis (CFA) is a statistical method appropriate for testing whether a theoretical model of relationships is consistent with a given set of data (Brown, 2006). Like EFA, CFA is grounded in the common factor model (Thurstone, 1947) and assumes the relationships observed between variables exist because they are influenced by the same underlying construct, which is represented by the latent construct (Brown, 2006). In order to estimate a CFA model, the researcher must specify the hypothesized characteristics of factor loadings, the relationships between factors, and measurement error (Grimm & Yarnold, 1995). CFA is applicable to instances of single samples as well as multiple samples, and enables the testing of hypotheses related to the consistency of relationships across groups (Mellenbergh, 1989; Meredith, 1993; Sörbom, 1974).

Put simply, CFA works by evaluating how well the researcher-specified relationships recreate the covariance matrix of the observed data (Brown, 2006). Though several methods are available for the purpose of estimating CFA models, the most common estimation method applied with CFA is ML (Babakus, Ferguson, & Jöreskog, 1987; Brown, 2006), due in part to the fact that ML allows for an empirical, statistical evaluation of model fit to the data. The overall model fit index for a CFA estimated with ML is χ^2, for which a *p* value can be obtained. The hypothesis being tested when a CFA model is fit to data is that there is no difference between the model and the data. Thus, a significant *p* value indicates that the specified model does not reproduce the observed data (Mulaik, James, Van Alstine, Bennet, Lind, & Stilwell, 1989). A disadvantage associated with the χ^2 statistic is that it is sensitive to sample size and will sometimes yield with a significant *p* value due to sample size (e.g., Alwin & Jackson, 1980; Bentler, 1990). Fortunately, the χ^2 is not the only statistic available for the evaluation of CFA model fit. Other model fit indices appropriate for CFA include RMSEA (Steiger & Lind, 1980), the Tucker-Lewis coefficient (TLC; Tucker & Lewis, 1973), and the normed comparative fit index (CFI; Bentler, 1990). The specifics of these fit statistics is beyond the scope of this paper; they are mentioned here only as examples. Fit indices such as these are typically referred to as measures of comparative fit because they quantify the degree to which the specific model fits the data better than the typical null model (no common factors, item covariance explained by sampling error alone) fits the data (Tanaka, 1993). Two other important topics relevant to CFA but beyond the scope of this paper include model identification and model comparison.

Applications of CFA

CFA is particularly suited to testing hypotheses about the relationships between indicators (observed data) and factors, and the relationships between factors (Brown, 2006). In addition to evaluating whether a specific model fits the data, CFA can be used to compare the fit of multiple models to determine which provides the best fit to the data (Schreiber, Nora, Stage, Barlow, & King, 2006), or to compare the fit of a model to data from multiple groups (Jöreskog, 1971). With regard to instrument development, CFA is often used as part of the process for evaluating content, convergent, and discriminant validity (Cole, 1987). During the past decade, the applications of CFA have expanded to include analysis of multi-trait multi-method data (MTMM; Bagozi & Yi, 1990; Wothke, 1996), which consists of using different models for different factors.

Confirmatory Factor Analysis vs. Exploratory Factor Analysis & Principal Components

Like PCA and EFA, CFA is often used for instrument development (e.g., construct validation; Judd, Jessor, & Donovan, 1986). Because CFA and EFA are based on the same model, they are more similar than are CFA and PCA, and it is likely that the results of EFAs are more likely to generalize to the CFA framework than are results of PCAs (Floyd & Widaman, 1995). The important difference between CFA and EFA is that EFA is a tool for developing theories and CFA is a tool for theory testing (Bollen, 1989). A way in ≠EFA assumes they are independent (Grimm & Yarnold, 1995).

CFA is distinct from PCA and EFA in that it constitutes a method of hypothesis testing applicable when the focus of the hypothesis is the structural relationship among variables. More specifically, CFA allows the researcher to specify how items relate to factors and how factors relate to each other (Bollen, 1989; Grimm & Yarnold, 1995). This is very different from exploratory methods like PCA and EFA, with which a priori hypotheses cannot be tested. Thus, PCA and EFA are methods for theory development and CFA is better suited to theory testing.

CONCLUSION

On the surface, PCA, EFA, and CFA appear to be similar methods for reaching the same goal. This is particularly true of PCA and EFA given that both are exploratory procedures which often yield the same results. However, Mulaik (1990) warns against being satisfied with methods that merely approximate outcomes, as PCA does. The primary differences between these three methods have been outlined here and should not be overlooked. Though it is true that these methods do have some characteristics in common, they constitute three distinct analyses that can be used together to strengthen research instruments and outcomes.

References

Aladwani, A. M., & Palvia, P. C. (2002). Developing and validating an instrument for measuring user-perceived web quality. *Information & Management*, *39*, 467-476.

Alwin, D. F., & Jackson, D. J. (1990). Measurement models for response errors in surveys: Issues and applications. In K. F. Schuessler (Ed.), *Sociological methodology, 1980* (pp.68-119). San Francisco: Jossey-Bass.

Babakus, E., Ferguson, Jr., C. E., & Jöreskog, K. G. (1987). The sensitivity of confirmatory maximum likelihood factor analysis to violations of measurement scale and distributional assumptions. *Journal of Marketing Research, 24*, 222-228.

Bagozzi, R. P., & Yi, Y. (1990). Assessing method variance in multitrait-multimethod matrices: The case of self-reported affect and perceptions at work. *Journal of Applied Psychology, 75*(5), 547-560.

Bartholomew, D. J. (1984). The foundation of factor analysis. *Biometrika, 71*(2), 221-232.

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

Bollen, K. A. (1989). *Structural equations with latent variables.* New York: Wiley.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research.* New York: Guilford Press.

Brown, T. A., Chorpita, B. F., & Barlow, D. H. (1998). Structural relationships among dimensions of the DSM-IV anxiety and mood disorders and dimensions of negative affect, positive affect, and autonomic arousal. *Journal of Abnormal Psychology*, *107*, 179-192.

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

Cole, D. A. (1987). Utility of confirmatory factor analysis in test validation research. *Journal of Consulting and Clinical Psychology, 55*, 584-594.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research, 4*(3), 272-299.*

Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment, 7*, 286-299.

Grimm, L. G., & Yarnold, P. R. (1995). *Reading and understanding multivariate statistics.* Washington, DC: American Psychological Association.

Goodall, D. W. (1954). Objective methods for the classification of vegetation: An essay in the use of factor analysis. *Australian Journal of Botany, 2*, 304-324.

Gorsuch, R. L. (1983). *Factor Analysis* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1992). *Multivariate data analysis with readings* (3rd ed.). New York: Macmillan.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology, 24*, 417-441, 498-520.

Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika, 34*, 183-202.

Jöreskog, K. G. (1971). Statistical analysis of sets of congeneric tests. *Psychometrika, 36*, 109-133.

Judd, C. M., Jessor, R., & Donovan, J. E. (1986). Structural equation models and personality research. *Journal of Personality, 54*, 49-198.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement, 20*, 141-151.

Kline, R. B. (2011). Principles and practice of structural equation modeling. New York: Guilford Press .

Maraun, M. D. (1996). Meaning and mythology in the factor analysis model. *Multivariate Behavioral Research, 31*(4), 603-616.

Mellenbergh, G. J. (1989). Item bias and item response theory. *International Journal of Educational Research, 13,* 121-143.

Meredith, W. (1993). Measurement invariance, factor analysis, and factorial invariance. *Psychometrika, 58*, 521-543.

Mulaik, S. A. (1990). Blurring the distinctions between component analysis and common factor analysis. *Multivariate Behavioral Research, 25*(1), 53-59.

Mulaik, S. A., James, L. R., Van Alstine, J., Bennett, N., Lind, S., & Stilwell, C. D. (1989), Evaluation of Goodness-of-Fit Indices for Structural Equation Models. *Psychological Bulletin*, 105, 430–445.

Osborne, J. W., & Costello, A. B. (2009). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Pan-Pacific Management Review, 12*(2), 131-146.

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.

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of Educational Research, 99,* 323-337.

Sörbom, D. (1974). A general method for studying differences in factor means and factor structure between groups. *British Journal of Mathematical and Statistical Psychology*, *27*, 221-239.

Steiger, J. H. (1979). Factor indeterminancy in the 1930's and the 1970's: Some interesting parallels. *Psychometrika, 44*, 157-167.

Steiger, J. H., & Lind, J. M. (1980). *Statistically based tests for the number of common factors.* Paper presented at the meeting of the Psychometric Society, Iowa City, IA.

Stevens, J. P. (1986). *Applied multivariate statistics for the social sciences.* Hillsdale, NJ: Erlbaum.

Tanaka, J. S. (1993). Multifaceted conception of fit in structural equation models. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 10-39). London: Sage.

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

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

Unkel, S., & Trendafilov, N. T. (2010). Simultaneous parameter estimation in exploratory factor analysis: An expository review. *International Statistical Review*, *78*(3), 363-382.

Wold, S., Esbensen, K., & Geladi, P. (1987). Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, *2*, 37-52.

Wothke, W. A. (1996). Models for multitrait-multimethod matrix analysis. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling: Issues and techniques* (ii. 7-56). Mahwah, NJ: Erlbaum.