Using Principal Components Analysis in Program Evaluation: Some Practical Considerations

J. Thomas Kellow

Assistant Professor of Research and Statistics

Mercer University

145 Peachtree Park #2602

Atlanta, GA 30309

832-277-1727

kellow_jt@mercer.edu

Abstract

Principal Components Analysis (PCA) is widely used by behavioral science researchers to assess the dimensional structure of data and for data reduction purposes. Despite the wide array of analytic choices available, many who employ this method continue to rely exclusively on the default options recommended in dominant statistical packages. This paper examines alternative analytic strategies to guide interpretation of PCA results that expand on these default options, including (a) rules for retaining factors or components and (b) rotation strategies. Conventional wisdom related to the interpretation of pattern/structure coefficients also is challenged. Finally, the use of principal component scores in subsequent

analyses is explored. A small set of actual data is used to facilitate illustrations and discussion.

Despite the increasing popularity of confirmatory factor analytic (CFA) techniques, principal components analysis (PCA) continues to enjoy widespread use (Kellow, 2004; Thompson, 2004). Researchers who employ PCA are typically interested in (a) assessing the dimensional structure of a dataset (Dunteman, 1989) or (b) reducing a large number of variables into a smaller set of linear combinations (components) for subsequent analyses (e.g., multiple regression). For instance, an evaluator may have occasion to develop a new instrument and wish to ascertain the number and features of the underlying dimensions represented in the data. At other times an existing measure is modified or shortened and the sample data are used to explore the extent to which the structure of the original version has or has not been substantively altered (although CFA is a stronger method for this purpose). The PCA approach also is useful for creating new variables that are linear combinations of a set of highly correlated original variables. These new composite variables may then be used in subsequent analyses. As Stephens (1992) notes, "… if there are 30 variables (whether predictors or items), we are undoubtedly not measuring 30 different constructs, hence, it makes sense to find some variable reduction scheme that will indicate how the variables cluster or hang together" (p. 374). Use of PCA helps to solve at least two problems. First, the presence of multicollinearity (high inter-item or variable correlations) leads to inflated standard errors for the measured variables when conducting statistical analyses, which increases the probability of Type II errors (non-significance when a significant difference exists in the population). Second, when one is using a large set of variables to predict or explain another variable (or set of variables) as opposed to a smaller set of composites, one pays a price in terms of the degrees of freedom used in the analysis. All other things being equal, the more degrees of freedom expended the smaller the value of the omnibus test statistic (e.g., *F*) that results from the analysis (Stephens, 1992).

There are a number of important issues related to the data in hand that need to be addressed (e.g., linearity; absence of outliers) before invoking PCA, and readers are referred to Tabachnick and Fidell (2001) for an excellent overview of these considerations. Once PCA is determined to be appropriate, the analysis proceeds in a series of sequential steps—several options are available to researchers at each step. Too often researchers rely on the default options provided in the major statistical packages and fail to examine other options that may allow for fuller exploitation of the data. The purpose of the present paper is to briefly explore the options available to analysts with respect to (a) rules for retaining principal components and (c) rotation strategies. In addition, conventional wisdom related to the interpretation of pattern/structure coefficients is challenged on substantive grounds. Finally, we briefly explore how PCA may be used to derive component scores for further data analysis.

Heuristic Data

For heuristic purposes, real data from a recent evaluation of a large alternative education campus are used in the analyses. These data consist of responses from a sample of 36 teachers and administrators on six items taken from the School Culture Quality Survey (SCQS) developed by Katzenmeyer (1994). The items represent two of the four subscales on the instrument: Shared Vision (SV) and Facilitative Leadership (FL). We want to emphasize that one important consideration in invoking PCA is the subject to variable ratio. Since PCA capitalizes on chance associations—as do all members of the general linear model family—this ratio is ideally at least 10:1 (see Thompson, 2004, for a review of other factors that also are relevant). The present ratio of 6:1 is far from ideal, but will suffice for the present discussion.

Rules for Retaining Components

In the initial extraction process, PCA will derive as many components as the number of measured variables. After the initial components are extracted, the analyst must decide on how many components should be retained to meaningfully represent the original correlation matrix. The initial component eigenvalues, percent of variance accounted for, and cumulative variance accounted for are provided in Table 1. According to Stevens, "probably the most widely used criterion is that of Kaiser (1960): Retain only those components whose eigenvalues are greater than 1" (1992, p. 378). This is the default option in many statistical packages (e.g., SPSS). Other methods for retaining factors, however, may be more defensible and perhaps meaningful in interpreting the data. Indeed, after reviewing empirical findings on its utility, Preacher and McCallum (2003) report that "the general conclusion is that there is little justification for using the Kaiser criterion to decide how many factors to retain" (p. 23).

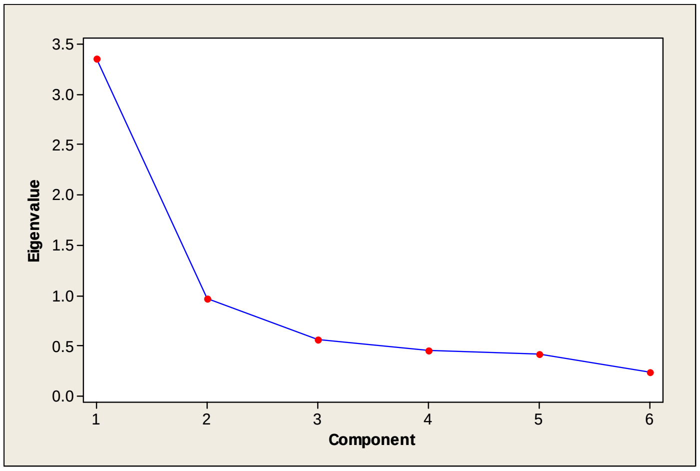
Table 1.

Initial component extraction statistics

Initial Eigenvalues	% of Variance	Cumulative %
3.363	56.048	56.048
.973	16.211	72.259
.558	9.296	81.555
.449	7.477	89.033
.419	6.979	96.012
.239	3.988	100.000

One reasonable alternative to Kaiser's rule is Cattell's (1966) scree test, which provides a graphical representation of the eigenvalues relative to their magnitude (this option is available in most major statistical packages). The basic idea is to plot eigenvalues on the ordinate (y axis) of a bivariate scatter with order of magnitude represented on the abscissa (x axis). Then, a visual inspection of the scree plot is undertaken to identify a point at which an inflection occurs that signifies a flattening of the line of best fit. Eigenvalues that occur *before* the first value that signifies a flattening are then retained (Stevens, 1992). An example is provided in Figure 1. For the present data, the Kaiser criterion would suggest retaining a single component, while the scree plot suggests retaining two components.

Figure 1. Scree plot for the SCQS data



A fairly common technique noted in the literature (Kellow, 2004) combines the two approaches. Eigenvalues greater than one are initially retained, and the scree test is used subsequently to assess the tenability of the model. Because eigenvalues represent reproduced variance, this is equivalent to setting a minimum level of acceptable variance reproduced by a component. The second stage evaluates the parsimony of the solution relative to the contribution of each component to reproducing the original variance in the data. A potential disadvantage of this approach is the arbitrary criterion of retaining eigenvalues greater than one in the first stage. Because PCA studies typically rely on sample data, eigenvalues (reproduced variance) should be expected to change (even with large samples) slightly from sample to sample. In addition, the interpretation of what constitutes a "meaningful" amount of variance accounted for (which eigenvalues represent) is inherently subjective (Thompson, 2002). Indeed, in the present data the second eigenvalue is .973, which is virtually indistinguishable from one. Moreover, since the measured variables conceptually are indicators of two different components, it seems reasonable (and supported by the scree interpretation) to retain both.

Rotation Strategies

Once an appropriate number of components have been determined, the analyst is charged with the task of interpreting the components. This process often is facilitated by geometrically rotating the factors to obtain a sharper conceptual solution. Because the starting point for locating factors in geometric space is arbitrary, rotating the factors does not change the overall variance explained by the components, although the eigenvalues associated with the respective components are not necessarily the same as the unrotated solution (Thompson, 1996). Two methods of rotation are available: (a) orthogonal and (b) oblique.

Orthogonal rotation constrains the obtained solution such that the obtained factors are *uncorrelated*. The overwhelming choice of analysts who opt for an orthogonal solution is the varimax procedure, which is the default option in most popular statistical packages (Kellow, 2004; Russell, 2002; Thompson, 2004). For various reasons (Tabachnick & Fidell, 2001; Stevens, 1992) varimax is generally an excellent choice if one prefers an orthogonal solution, although other options are available.

In contrast to orthogonal solutions, oblique rotation solutions allow for factors to be *correlated*. At times, the quest for simple structure is inhibited by the assumption of uncorrelated factors. "Typically this is indicated by variables having…coefficients that are large in absolute value on two or more factors (which is sometimes called multivocal vs. univocal)" (Thompson, 2004, p. 42). The use of an oblique solution, such as oblimin or promax (see Tabachnick & Fidell, 2001, for an overview) often best captures the reality of the construct(s) being investigated. Rarely does one assume that multidimensional constructs, such as school climate, are composed of dimensions that are completely independent of one another. Most statistical programs will provide an estimate of the correlation between components when an oblique rotation is requested. Tabachnick and Fidell recommend performing such an analysis and examining the correlations for values of .32 and above, indicating at least 10% of overlap between the components (although, again, this is an arbitrary value).

In order to interpret the principal components one must consult the correlations between variables and components, often referred to as "loadings." As noted by Thompson (1996), these coefficients are merely "weights" assigned to variables to indicate their importance. However, this obscures a very important difference between these values when oblique as opposed to orthogonal rotational strategies are used.

If an orthogonal rotation is used, the correlation between a variable and a component represents the *total* contribution of the variable to the respective component (called a structure coefficient). In the case of orthogonal rotation, the components will be uncorrelated and the structure coefficients and pattern coefficients will be identical**.** In contrast, when an oblique rotation is employed, the correlation coefficient associated with a particular variable and a component indicates the *unique* contribution of that variable to the component after partialling out the variance attributable to the variable's covariance with other components (called a pattern coefficient) (Tabachnick & Fidell, 2001). This is analogous to regression analysis, where the beta (ß) weights indicate the contribution of individual predictors in "explaining" the criterion variable. If the predictors are perfectly uncorrelated, these weights indicate both the total and unique contribution of a predictor variable. However, when the individual predictor variables are correlated with one another—which is usually the case—the weights indicate the unique contribution of the variable to explaining the criterion in the presence of other predictors.

Table 2 provides the unrotated component pattern and structure matrix and rotated matrices for the present data using both orthogonal (varimax) and oblique (oblimin and promax) rotations. Inspection of the unrotated matrices (which are identical) indicates that the items tend to correlate highly with both components, despite the postulated existence of two separate components. The varimax rotation provides a fairly clear differentiation between the components with the exception of the sixth indicator, which is fairly highly correlated with both components. Note that the pattern and structure coefficients are identical, which confirms our earlier statement about the identity relationship between the two matrices when components are rotated to be orthogonal (uncorrelated). Relaxing the assumption of uncorrelated factors by invoking the oblimin procedure results in a slightly better fit (*r* between factors = .42). Inspection of the pattern and structure coefficients for the oblique rotation reveals, indeed, that these coefficients are *not* the same because of the correlation between the two components, as mentioned earlier. The promax solution provides an even more parsimonious fit of the data (*r* between factors = .53). It should be noted that in both oblique methods, one may alter the degree of correlation allowed between the components by manipulating a parameter called *delta* in oblimin and the *pivot power* in promax. We (shamefully) have provided examples using the default values for the sake of brevity; however, the interested reader is referred to Kim and Mueller (1978) and Tabachnick and Fidell (2001) for further explication.

Table 2.

Rotated component pattern and structure matrices using orthogonal and oblique methods

Pattern Matrices

	Unrotated		Varimax		Oblimin		Promax	
Variable	I	II	I	II	I	II	I	II
SV 1	.778	-.410	.863	.172	.894	-.035	.922	-.085
SV 2	.721	-.430	.831	.121	.870	-.081	.902	-.132
SV 3	.857	-.169	.771	.409	.743	.243	.742	.213
FL 1	.570	.665	.023	.876	-.167	.933	-.255	.983
FL 2	.734	.342	.355	.728	.223	.690	.168	.709
FL 3	.800	.179	.508	.643	.407	.561	.368	.563

Structure Matrices

SV 1	.778	-.410	.863	.172	.879	.340	.877	.401
SV 2	.721	-.430	.831	.121	.836	.284	.832	.343
SV 3	.857	-.169	.771	.409	.845	.554	.854	.604
FL 1	.570	.665	.023	.876	.224	.863	.263	.849
FL 2	.734	.342	.355	.728	.513	.784	.542	.797
FL 3	.800	.179	.508	.643	.642	.731	.665	.757

SV = Shared Vision FL = Facilitative Leadership

As noted by Thompson (2004), “Persons first learning of rotation are often squeamish about the ethics of this procedure” (p. 40). It should be stressed however, that component rotation simply expresses the data in a different dimensional space. The wise analyst would do well to go beyond default settings by exploring both orthogonal *and* oblique rotation strategies.

Interpretation of Factor Pattern/Structure Coefficients

Although the misuse of language associated with these coefficients is problematic, a second issue, related to their substantive interpretation, is more troublesome. It may be argued that the most artful aspect of PCA is the determination of the salience of variables as they relate to derived factors. Analysts often use an *absolute criterion* for deciding to retain variables that is *absolutely arbitrary* (Hogarty, Kromrey, Ferron, & Hines, 2004). Two of the most popular criteria are pattern/structure coefficients greater than $|.3|$ or coefficients greater than $|.4|$. The former rule appears to be attributable to Nunnally (1982), who claimed that “It is doubtful that loadings (sic) of any smaller size (.30) should be taken seriously, because they represent less than 10 percent of the variance” (p. 423). The latter criterion can be traced to Stevens (1992), who stated:

> It would seem that one would want in general a variable to share *at least* 15% of its variance with the construct (factor) it is going to be used to help name. This means only using loadings (sic) which are about .4 or greater for interpretation purposes. (p. 384)

A basic problem with this approach is the dichotomous decision making process that it encourages: $|.31|$ is good and $|.29|$ is bad. In the present example, the factor pattern/structure coefficients clearly meet the previous criteria; however, this is not always the case. Happily, a recent review of practice (Kellow, 2004) indicates that at least some researchers employing PCA (about 25%) refuse to be bound by such restraints and, instead, rely on logical interpretation within the context of the phenomena being investigated to guide their interpretation of the salience of component pattern/structure coefficients. Adhering to conventional “rules of thumb” such as interpreting coefficients based only on their magnitude in comparison with a strict criterion seems to belie the “exploratory” spirit of PCA.

Using Component Scores in Subsequent Analyses

Once the dimensional structure of the data has been determined, there are several methods for determining individual scores on these dimensions for further analyses. For the sake of brevity, all analyses will be conducted using multiple regression, however, the results easily generalize to all GLM methods (e.g., ANOVA; discriminant analysis) provided that statistical assumptions for each method are satisfied. We will use the data for the SCQS to predict the responses on a single item measuring ov*erall effectiveness* of the school principal using an 8-point scale ranging from 0 (*not at all effective*) to 7 (*highly effective*).

Before pursuing the notion of component scores, imagine a statistically naïve evaluator who might decide to use all six variables to predict the criterion of overall effectiveness. These results are provided in the first section of Table 3. Using all six items results in a large R^2 (.54) and statistically significant omnibus F value (5.59, $p < .001$). Notice, however, that *none* of the individual ß weights for the items would be considered statistically significant at conventional levels. This rather perplexing event happens because of the high inter-correlations between the items, particularly with such a small sample size. This phenomenon, known as multicollinearity, results in a very unstable regression solution.

Table 3.

Predicting "Overall Effectiveness" with SCQS scores using various score derivation methods

Method	Variables	ß weight	*p* value
All variables			
(R^2 = .54, F = 5.59, p = .00)	SV 1	.230	.131
	SV 2	.088	.678
	SV 3	.071	.717
	FL 1	.242	.190
	FL 2	.222	.190
	FL 3	.247	.145
Composite Scores			
(R^2 = .51, F = 17.19, p = .00)	SV	.589	.000
	FL	.189	.213
Orthogonal Component Scores			
(R^2 = .51, F = 17.2, p = .00)	Reg SV	.588	.000
	Reg FL	.407	.002
Oblique Component Scores			
	Reg SV	.533	.001
(R^2 = .51, F = 17.2, p = .00)	Reg FL	.272	.049

SV = Shared Vision FL = Facilitative Leadership

Suppose, instead, we use a simple method of constructing component scores by simply summing the individual responses of teachers to each of the three items on each of the two components. These results are presented in the second section of

Table 3. Note that the R^2 value is also substantial (.51) and the obtained F statistic is both statistically significant ($p < .00$) and much larger (17.19) than in the previous analysis using all six items as predictors. This is because each predictor we enter into the regression equation requires us to expend an additional degree of freedom, which in turn reduces the obtained F value (all other things being equal). From these results we would conclude, based on the ß weights and corresponding p values, that SV was both substantively and statistically significantly superior to FL in predicting overall effectiveness. Looks, however, can be deceiving. It so happens that the zero-order correlation between SV and the criterion is .697 (r^2 = .486), while the zero-order correlation between FL and the criterion is .527 (r^2 = .278). Both values are substantively and statistically significant ($p < .01$). Clearly, both composite variables have something to offer in predicting the criterion. But SV and FL also are highly correlated ($r = .574$). In standard multiple regression, the predictor with the highest correlation with the criterion (SV) always enters the solution first. The second predictor, if correlated with the first, is evaluated based on its *unique* contribution to predicting the criterion after controlling for variance accounted for by the first predictor.

The third section of Table 3 presents results using component scores derived from the varimax rotated (orthogonal) components. Most statistical packages offer several methods of obtaining empirically derived component scores, but in the case of PCA, all will yield identical results for a particular rotational strategy. The present scores were derived using a regression approach. This method uses the component pattern/structure coefficients to weigh the observed scores, forming an additive composite score for each component. In the case of orthogonal solutions, the scores will be uncorrelated across the respective components. Therefore, each component ß weight represents the total contribution of the variable in predicting the criterion *irrespective* of the other predictor. This may be confirmed by squaring the ß weights ($.588^2 = .346$ and $.407^2 = .166$), and obtaining the sum, which results in a value of .512—within rounding error of the obtained R^2 (.511). Orthogonal factor scores may be optimal in their advantage of being directly interpreted as the overall contribution of each component in predicting the criterion; however, as noted earlier, orthogonal solutions may not be optimal in fitting the original data. The choice of which component scores to use in subsequent analyses is determined by the initial selection of a rotation solution. Since we chose the oblique (promax) solution as the best fit of the data to the model, we are obliged to use the oblique component scores for subsequent analyses (B. Thompson, personal communication, December 15, 2004). These statistics are displayed in the final section of Table 3. Importantly, the obtained R^2 is *identical* to that obtained using the orthogonal scores, which reinforces the notion that, while different rotational strategies may distribute the variance accounted for by the components differently, the amount of variance explained by the original components remains constant.

Discussion

When used thoughtfully, PCA is a powerful tool for data analysts interested in exploring the dimensional structure of scale variables. Too often, we argue, persons using the technique rely blindly on the default options in popular statistics packages. The popular “Little Jiffy” combination advocated by Kaiser (1970), wherein components with eigenvalues greater than one are retained, and the varimax criterion, is atavistic in light of the complexity underlying components analysis. As we have demonstrated, the use of different criteria can impact the decisions made at various steps in the analytic sequence. While no interpretation or decision may be thought of as inherently superior, the use of multiple criteria acknowledges the subjective nature of interpreting PCA results. The same may be said of the tendency for analysts to set some *a priori* criterion for interpreting pattern/structure coefficients without considering the component structure as a whole.

Evaluation analysts are encouraged to explore a variety of options at each stage of the PCA process, and to allow informed judgment to guide the process rather than strict, arbitrary criteria. We have presented a few of these options in the present paper, but there are additional options that are infrequently used because they are not readily available in most packages. For instance, some have suggested a promising variant of the scree plot in which standard errors are computed to supplement interpretation of the number of components to retain (Nasser, Benson, & Wisenbaker, 2002).

The use of PCA to obtain composite scores is a valuable tool when dealing with correlated variables. As has been shown, the problem of multicollinearity can lead to some perplexing results, and the use of component scores can help to clarify these statistical dilemmas. In addition, the use of component scores rather than a large number of individual variables is better given the fact that, all other things being equal, using fewer predictors (in the regression case) makes for a more powerful analysis.

On a final note, we would like to reaffirm the thoughts of May (2004) on the presentation of statistical analyses to evaluation audiences. Increasingly, evaluators are becoming cognizant of the multivariate reality of evaluation contexts. To the extent possible, evaluators are obliged to honor this reality. It would, however, be a grave mistake to present component matrices and the like within the body of an evaluation report (unless one is afraid someone might actually read the report). While valuable, these data are best left to appendices, or as May suggests, a different evaluation report aimed at researchers rather than a non-technical audience. That being said, we hope this brief paper encourages the diligent evaluator to go beyond the "hegemony" of the default and explore the rich number of options available to analysts who choose to invoke PCA.

References

Cattell, R. B. (1966). The meaning and strategic use of factor analysis. In R.B. Cattell (Ed.), *Handbook of multivariate experimental psychology* (pp. 174-243). Chicago: Rand McNally.

Dunteman, G. H. (1989). *Principal components analysis*. Newbury Park, CA: Sage.

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.

Gorsuch, R. L. (1983). *Factor analysis*. Hillsdale, NJ: Earlbaum.

Hogarty, K. Y., Kromrey, J. D., Ferron, J. M., & Hines, C. V. (in press). Selection of variables in exploratory factor analysis: An empirical comparison of a stepwise and traditional approach. *Psychometrika.*

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*, 179-185.

Kaiser, H. F. (1970). A second generation Little Jiffy. *Psychometrika*, *35*, 401–415.

Katzenmeyer, W. (1994). *School culture quality survey.* Tampa, FL: Anchin Center.

Kim, J. O., & Mueller, C. W. (1978). Factor analysis: Statistical methods and practical issues. Newbury Park, CA: Sage.

Kellow, J. T. (2004, July). *Exploratory Factor Analysis in Two Prominent Journals: Hegemony by Default*. Paper presented at the annual meeting of the American Psychological Association, Honolulu, HI.

May, H. (2004). Making statistics more meaningful for policy research and program evaluation. *American Journal of Evaluation*, *25*, 525-540.

Nasser, F., Benson, J. & Wisenbaker, J. (2002). The performance of regression-based variations of the visual scree for determining the number of common factors. *Educational and Psychological Measurement, 62,* 397-419.

Nunnally, J. C. (1978). *Psychometric Theory* (2nd Ed.). New York: McGraw Hill.

Preacher, K. J., & MacCallum, R. C. (2003). Repairing Tom Swift's electronic factoring 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*, *28*, 1629-1646.

Stevens, J. (1992). *Applied multivariate statistics for the social sciences* (2^{nd}. Ed.). Hillsdale, NJ: Earlbaum.

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4^{th} Ed.). Needham Heights, MA: Allyn & Bacon.

Thompson, B., & Daniel, L. G. (1996). Factor analytic evidence for the construct validity of scores: A historical overview and some guidelines. *Educational and Psychological Measurement*, *56*, 197-208.

Thompson, B. (2002). What future quantitative social science research could look like: Confidence intervals for effect sizes. *Educational Researcher, 31*(3), 24-31.

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