

# Practical Assessment, Research & Evaluation

A peer-reviewed electronic journal.

Copyright is retained by the first or sole author, who grants right of first publication to the *Practical Assessment, Research & Evaluation*. Permission is granted to distribute this article for nonprofit, educational purposes if it is copied in its entirety and the journal is credited.

Volume 10 Number 7, July 2005

ISSN 1531-7714

## Best Practices in Exploratory Factor Analysis: Four Recommendations for Getting the Most From Your Analysis

Anna B. Costello and Jason W. Osborne  
North Carolina State University

Exploratory factor analysis (EFA) is a complex, multi-step process. The goal of this paper is to collect, in one article, information that will allow researchers and practitioners to understand the various choices available through popular software packages, and to make decisions about “best practices” in exploratory factor analysis. In particular, this paper provides practical information on making decisions regarding (a) extraction, (b) rotation, (c) the number of factors to interpret, and (d) sample size.

Exploratory factor analysis (EFA) is a widely utilized and broadly applied statistical technique in the social sciences. In recently published studies, EFA was used for a variety of applications, including developing an instrument for the evaluation of school principals (Lovett, Zeiss, & Heinemann, 2002), assessing the motivation of Puerto Rican high school students (Morris, 2001), and determining what types of services should be offered to college students (Majors & Sedlacek, 2001).

A survey of a recent two-year period in PsycINFO yielded over 1700 studies that used some form of EFA. Well over half listed principal components analysis with varimax rotation as the method used for data analysis, and of those researchers who report their criteria for deciding the number of factors to be retained for rotation, a majority use the Kaiser criterion (all factors with eigenvalues greater than one). While this represents the norm in the literature (and often the defaults in popular statistical software packages), it will not always yield the best results for a particular data set.

EFA is a complex procedure with few absolute guidelines and many options. In some cases, options vary in terminology across software packages, and in many cases particular options are not well defined. Furthermore, study design, data properties, and the questions to be answered all have a bearing on which procedures will yield the maximum benefit.

The goal of this paper is to discuss common practice in studies using exploratory factor analysis, and provide practical information on best practices in the use of EFA. In particular we discuss four issues: 1) component vs. factor extraction, 2) number of factors to retain for rotation, 3) orthogonal vs. oblique rotation, and 4) adequate sample size.

### BEST PRACTICE

#### *Extraction: Principal Components vs. Factor Analysis*

PCA (principal components analysis) is the default method of extraction in many popular statistical software packages, including SPSS and SAS, which likely contributes to its popularity. However, PCA is

not a true method of factor analysis and there is disagreement among statistical theorists about when it should be used, if at all. Some argue for severely restricted use of components analysis in favor of a true factor analysis method (Bentler & Kano, 1990; Floyd & Widaman, 1995; Ford, MacCallum & Tait, 1986; Gorsuch, 1990; Loehlin, 1990; MacCallum & Tucker, 1991; Mulaik, 1990; Snook & Gorsuch, 1989; Widaman, 1990, 1993). Others disagree, and point out either that there is almost no difference between principal components and factor analysis, or that PCA is preferable (Arrindell & van der Ende, 1985; Guadagnoli and Velicer, 1988; Schoenmann, 1990; Steiger, 1990; Velicer & Jackson, 1990).

We suggest that factor analysis is preferable to principal components analysis. Components analysis is only a data reduction method. It became common decades ago when computers were slow and expensive to use; it was a quicker, cheaper alternative to factor analysis (Gorsuch, 1990). It is computed without regard to any underlying structure caused by latent variables; components are calculated using all of the variance of the manifest variables, and all of that variance appears in the solution (Ford et al., 1986). However, researchers rarely collect and analyze data without an *a priori* idea about how the variables are related (Floyd & Widaman, 1995). The aim of factor analysis is to reveal any latent variables that cause the manifest variables to covary. During factor extraction the shared variance of a variable is partitioned from its unique variance and error variance to reveal the underlying factor structure; only shared variance appears in the solution. Principal components analysis does not discriminate between shared and unique variance. When the factors are uncorrelated and communalities are moderate it can produce inflated values of variance accounted for by the components (Gorsuch, 1997; McArdle, 1990). Since factor analysis only analyzes shared variance, factor analysis should yield the same solution (all other things being equal) while also avoiding the inflation of estimates of variance accounted for.

### ***Choosing a Factor Extraction Method***

There are several factor analysis extraction methods to choose from. SPSS has six (in addition to PCA; SAS and other packages have similar options): unweighted least squares, generalized least squares, maximum likelihood, principal axis factoring, alpha factoring, and image factoring. Information on the

relative strengths and weaknesses of these techniques is scarce, often only available in obscure references. To complicate matters further, there does not even seem to be an exact name for several of the methods; it is often hard to figure out which method a textbook or journal article author is describing, and whether or not it is actually available in the software package the researcher is using. This probably explains the popularity of principal components analysis – not only is it the default, but choosing from the factor analysis extraction methods can be completely confusing.

A recent article by Fabrigar, Wegener, MacCallum and Strahan (1999) argued that if data are relatively normally distributed, maximum likelihood is the best choice because “it allows for the computation of a wide range of indexes of the goodness of fit of the model [and] permits statistical significance testing of factor loadings and correlations among factors and the computation of confidence intervals.” (p. 277). If the assumption of multivariate normality is “severely violated” they recommend one of the principal factor methods; in SPSS this procedure is called “principal axis factors” (Fabrigar et al., 1999). Other authors have argued that in specialized cases, or for particular applications, other extraction techniques (e.g., alpha extraction) are most appropriate, but the evidence of advantage is slim. In general, ML or PAF will give you the best results, depending on whether your data are generally normally-distributed or significantly non-normal, respectively.

### ***Number of Factors Retained***

After extraction the researcher must decide how many factors to retain for rotation. Both overextraction and underextraction of factors retained for rotation can have deleterious effects on the results. The default in most statistical software packages is to retain all factors with eigenvalues greater than 1.0. There is broad consensus in the literature that this is *among the least accurate methods* for selecting the number of factors to retain (Velicer & Jackson, 1990). In monte carlo analyses we performed to test this assertion, 36% of our samples retained too many factors using this criterion. Alternate tests for factor retention include the scree test, Velicer’s MAP criteria, and parallel analysis (Velicer & Jackson, 1990). Unfortunately the latter two methods, although accurate and easy to use, are not available in the most frequently used statistical software and must be

calculated by hand. Therefore the best choice for researchers is the scree test. This method is described and pictured in every textbook discussion of factor analysis, and can also be found in any statistical reference on the internet, such as StatSoft's electronic textbook at <http://www.statsoft.com/textbook/stathome.html>.

The scree test involves examining the graph of the eigenvalues (available via every software package) and looking for the natural bend or break point in the data where the curve flattens out. The number of datapoints *above* the "break" (i.e., not including the point at which the break occurs) is usually the number of factors to retain, although it can be unclear if there are data points clustered together near the bend. This can be tested simply by running multiple factor analyses and setting the number of factors to retain manually – once at the projected number based on the a priori factor structure, again at the number of factors suggested by the scree test if it is different from the predicted number, and then at numbers above and below those numbers. For example, if the predicted number of factors is six and the scree test suggests five then run the data four times, setting the number of factors extracted at four, five, six, and seven. After rotation (see below for rotation criteria) compare the item loading tables; the one with the "cleanest" factor structure – item loadings above .30, no or few item crossloadings, no factors with fewer than three items – has the best fit to the data. If all loading tables look messy or uninterpretable then there is a problem with the data that cannot be resolved by manipulating the number of factors retained. Sometimes dropping problematic items (ones that are low-loading, crossloading or freestanding) and rerunning the analysis can solve the problem, but the researcher has to consider if doing so compromises the integrity of the data. If the factor structure still fails to clarify after multiple test runs, there is a problem with item construction, scale design, or the hypothesis itself, and the researcher may need to throw out the data as unusable and start from scratch. One other possibility is that the sample size was too small and more data needs to be collected before running the analyses; this issue is addressed later in this paper.

### **Rotation**

The next decision is rotation method. The goal of rotation is to simplify and clarify the data structure.

Rotation *cannot* improve the basic aspects of the analysis, such as the amount of variance extracted from the items. As with extraction method, there are a variety of choices. Varimax rotation is by far the most common choice. Varimax, quartimax, and equamax are commonly available orthogonal methods of rotation; direct oblimin, quartimin, and promax are oblique. Orthogonal rotations produce factors that are uncorrelated; oblique methods allow the factors to correlate. Conventional wisdom advises researchers to use orthogonal rotation because it produces more easily interpretable results, but this is a flawed argument. In the social sciences we generally expect some correlation among factors, since behavior is rarely partitioned into neatly packaged units that function independently of one another. Therefore using orthogonal rotation results in a loss of valuable information if the factors are correlated, and oblique rotation should theoretically render a more accurate, and perhaps more reproducible, solution. If the factors are truly uncorrelated, orthogonal and oblique rotation produce nearly identical results.

Oblique rotation output is only slightly more complex than orthogonal rotation output. In SPSS output the *rotated factor matrix* is interpreted after orthogonal rotation; when using oblique rotation the *pattern matrix* is examined for factor/item loadings and the *factor correlation matrix* reveals any correlation between the factors. The substantive interpretations are essentially the same.

There is no widely preferred method of oblique rotation; all tend to produce similar results (Fabrigar et al., 1999), and it is fine to use the default delta (0) or kappa (4) values in the software packages. Manipulating delta or kappa changes the amount the rotation procedure "allows" the factors to correlate, and this appears to introduce unnecessary complexity for interpretation of results. In fact, in our research we could not even find any explanation of when, why, or to what one should change the kappa or delta settings.

### **Sample Size**

To summarize practices in sample size in EFA in the literature, we surveyed two years' worth of PsychINFO articles that both reported some form of principal components or exploratory factor analysis and listed both the number of subjects and the number of items analyzed ( $N = 303$ ). We decided the

best method for standardizing our sample size data was subject to item ratio, since we needed a criterion for a reasonably direct comparison to our own data analysis. In the studies reporting scale construction, the number of items in the initial item pool were recorded rather than the number of items kept for the final version of the scale, since the subject to item ratio is determined by how many items each subject answered or was measured on, not how many were kept after analysis. The results of this survey and are summarized in Table 1. A large percentage of

researchers report factor analyses using relatively small samples. In a majority of the studies in our survey (62.9%) researchers performed analyses with subject to item ratios of 10:1 or less, which is an early and still-prevalent rule-of-thumb many researchers use for determining *a priori* sample size. A surprisingly high proportion (almost one-sixth) reported factor analyses based on subject to item ratios of only 2:1 or less. The effects of small samples on EFA analyses are discussed later in this paper.

Table 1: *Current Practice in Factor Analysis*

Subject to item ratio	% of studies	Cumulative %
2:1 or less	14.7%	14.7%
> 2:1, ≤ 5:1	25.8%	40.5%
> 5:1, ≤ 10:1	22.7%	63.2%
> 10:1, ≤ 20:1	15.4%	78.6%
> 20:1, ≤ 100:1	18.4%	97.0%
> 100:1	3.0%	100.0%

Strict rules regarding sample size for exploratory factor analysis have mostly disappeared. Studies have revealed that adequate sample size is partly determined by the nature of the data (Fabrigar et al., 1999; MacCallum, Widaman, Zhang, & Hong, 1999). In general, the stronger the data, the smaller the sample can be for an accurate analysis. “Strong data” in factor analysis means uniformly high communalities without cross loadings, plus several variables loading strongly on each factor. In practice these conditions can be rare (Mulaik, 1990; Widaman, 1993). If the following problems emerge in the data, a larger sample can help determine whether or not the factor structure and individual items are valid:

1) Item communalities are considered “high” if they are all .8 or greater (Velicer and Fava, 1998) – but this is unlikely to occur in real data. More common magnitudes in the social sciences are low to moderate communalities of .40 to .70. If an item has a

communality of less than .40, it may either a) not be related to the other items, or b) suggest an additional factor that should be explored. The researcher should consider why that item was included in the data and decide whether to drop it or add similar items for future research. (Note that these numbers are essentially correlation coefficients, and therefore the magnitude of the loadings can be understood similarly).

2) Tabachnick and Fidell (2001) cite .32 as a good rule of thumb for the minimum loading of an item, which equates to approximately 10% overlapping variance with the other items in that factor. A “crossloading” item is an item that loads at .32 or higher on two or more factors. The researcher needs to decide whether a crossloading item should be dropped from the analysis, which may be a good choice if there are several adequate to strong loaders (.50 or better) on each factor. If there are several crossloaders, the items

may be poorly written or the *a priori* factor structure could be flawed.

3) A factor with fewer than three items is generally weak and unstable; 5 or more strongly loading items (.50 or better) are desirable and indicate a solid factor. With further research and analysis it may be possible to reduce the item number and maintain a strong factor; if there is a very large data set.

In general, we caution researchers to remember that EFA is a “large-sample” procedure; generalizable or replicable results are unlikely if the sample is too small. In other words, more is better.

## **METHODOLOGY**

The purpose of this section is to empirically demonstrate the effects of various recommendations we made above.

For this study we used real data to examine the effects of 1) principal components analysis vs. maximum likelihood (ML) extraction, 2) orthogonal vs. oblique rotation, and 3) various sample sizes. Our goal was to utilize methods that most closely simulate real practice, and real data, so that our results will shed light on the effects of current practice in research. For our purposes it was important to have a known population. We operationally defined a very large sample as our “population” to provide population parameters for comparison with EFA results.

**Data Source.** Data for this study were drawn from the first follow-up data set of the National Education Longitudinal Study (NELS88; for an overview of this sample, see NCES, 1992), using a sample of 24,599 students who completed Marsh's Self-Description Questionnaire (SDQ II). Marsh's self-concept scale is constructed from a hierarchical facet model of a dimensionalized self; it draws on both generalized and domain-specific self-concepts and has well-known and desirable psychometric properties (e.g., Marsh, 1990). We chose this data set because it is very large, accessible to everyone, and easily adapted for this study; the results of our analyses are applicable across all fields of social science research.

The full data set of 24,599 students was operationally defined as the population for the purposes of this study. In the NELS88, five of Marsh's subscales were used (relations with parents, language self-concept, math self-concept, opposite sex

relationships, same sex relationships). We excluded the last two subscales from our analyses because there are different subscales for males and females.. The remaining subscales (parents, language, math) show good reliability, both in other studies (Marsh, 1990) and in this particular data set, with Cronbach's alphas of .84 to .89. This measure also shows a very clear factor structure. When analyzed using maximum likelihood extraction with direct oblimin rotation these three subscales form three strong factors (eigenvalues of 4.08, 2.56, and 2.21). Factor loadings for this scale are also clear, with high factor loadings (ranging from .72 to .87, .72 to .91, and .69 to .83 on the three factors) and minimal cross-factor loadings (none greater than .17).

**Extraction and rotation.** The entire “population” of 24,599 subjects was analyzed via principal components analysis (PCA) and maximum likelihood (ML) extraction methods, followed by both orthogonal (varimax) and oblique (direct oblimin) rotations.

**Sample size.** Samples were drawn from the entire population via random sampling with replacement. We extracted twenty samples at the 2:1, 5:1, 10:1, and 20:1 subject to item ratios, creating sample sizes of  $N = 26, 65, 130,$  and  $260$ . The samples drawn from the population data were analyzed using maximum likelihood extraction with direct oblimin rotation, as suggested above. For each sample, the magnitude of the eigenvalues, the number of eigenvalues greater than 1.0, the factor loadings of the individual items, and the number of items incorrectly loading on a factor were recorded. In order to assess accuracy as a function of sample size, we computed average error in eigenvalues and average error in factor loadings. We also recorded aberrations such as Heywood cases (occasions when a loading exceeds 1.0) and instances of failure for ML to converge on a solution after 250 iterations.

Finally, a global assessment of the correctness or incorrectness of the factor structure was made. If a factor analysis for a particular sample produced three factors, and the items loaded on the correct factors (all five parent items loaded together on a single factor, all language items loaded together on a single factor, all math items loaded together on a single factor), that analysis was considered to have produced the correct factor structure (i.e., a researcher drawing that sample, and performing that analysis, would draw

the correct conclusions regarding the underlying factor structure for those items). If a factor analysis produced an incorrect number of factors with eigenvalues greater than 1.0 (some produced up to 5), or if one or more items failed to load on the appropriate factor, that analysis was considered to have produced an incorrect factor structure (i.e., a researcher drawing that sample, and performing that analysis, would not draw the correct conclusions regarding the underlying factor structure).

## RESULTS AND DISCUSSION

**Extraction Method.** The results of these analyses are

presented in Table 2. Both extraction methods produced identical eigenvalues after extraction, as expected. However, PCA resulted in a significantly higher total variance accounted for. Item loadings were also higher for PCA with both orthogonal (varimax) and oblique (direct oblimin) rotations. This happens because PCA does not partition unique variance from shared variance so it sets all item communalities at 1.0, whereas ML estimates the level of shared variance (communalities) for the items, which ranged from .39 to .70 – all much

Table 2: Comparison of Extraction and Rotation Methods (N = 24,599)

Rotation Method		Principal Components		Maximum Likelihood	
		Orthogonal	Oblique	Orthogonal	Oblique
Variance Accounted for after Rotation		68.0%	*	59.4%	*
Item Loadings					
Factor 1	Item 1	0.84	0.84	0.77	0.78
	Item 2	0.86	0.88	0.82	0.84
	Item 3	0.87	0.87	0.84	0.85
	Item 4	0.72	0.72	0.61	0.61
Factor 2	Item 1	0.91	0.92	0.89	0.90
	Item 2	0.89	0.89	0.86	0.86
	Item 3	0.90	0.91	0.88	0.88
	Item 4	0.72	0.72	0.60	0.60
Factor 3	Item 1	0.77	0.78	0.71	0.72
	Item 2	0.76	0.78	0.66	0.68
	Item 3	0.83	0.84	0.82	0.83
	Item 4	0.69	0.68	0.59	0.58
	Item 5	0.79	0.80	0.73	0.75

\* cannot compute total variance after oblique rotation due to correlation of factors

fewer than 1. For this data set the difference between PCA and ML in computing item loadings is minimal, but this is due to the sheer size of the sample (N = 24,599). With smaller data sets the range – and room for error – is much greater, as discussed below. In these analyses, factor analysis produced an average variance accounted for of 59.8%, while principal components analysis showed a variance accounted for

of 69.6% -- an over-estimation of 16.4%.<sup>1</sup> PCA also produced inflated item loadings in many cases. In sum, ML will produce more generalizable and reproducible results, as it does not inflate the variance estimates.

**Rotation Method.** We chose a scale designed to have orthogonal factors. Oblique rotation produced results



nearly identical to the orthogonal rotation when using the same extraction method, as evident in Table 2. Since oblique rotation will reproduce an orthogonal solution but not vice versa, we recommend oblique rotation.

**Sample size.** In order to examine how sample size affected the likelihood of errors of inference regarding factor structure of this scale, an analysis of variance was performed, examining the number of samples producing correct factor structures as a function of the sample size. The results of this analysis are presented in Table 3. As expected, larger samples tended to produce solutions that were more accurate. Only 10% of samples in the smallest (2:1) sample produced correct solutions (identical to the

population parameters), while 70% in the largest (20:1) produced correct solutions. Further, the number of misclassified items was also significantly affected by sample size. Almost two of thirteen items on average were misclassified on the wrong factor in the smallest samples, whereas just over one item in every two analyses were misclassified in the largest samples. Finally, two indicators of extreme trouble—the presence of Heywood cases (factor loadings greater than 1.0, an impossible outcome) or failure to converge, were both exclusively observed in the smaller samples, with almost one-third of analyses in the smallest sample size category failing to produce a solution.

Table 3: *The Effects of Subject to Item Ratio on Exploratory Factor Analysis*

Variable:	2:1	5:1	10:1	20:1	$F_{(3,76)} =$	$\eta^2 =$
% samples with correct factor structure	10%	40%	60%	70%	13.64***	.21
Average number of items misclassified on wrong factor	1.93	1.20	0.70	0.60	9.25***	.16
Average error in eigenvalues	0.41	0.33	.20	.16	25.36***	.33
Average error in factor loadings	.15	.12	.09	.07	36.38***	.43
% analyses failing to converge after 250 iterations	30%	0%	0%	0%	8.14***	.24
% with Heywood cases	15%	20%	0%	0%	2.81	.10

\*\*\*  $p < .0001$

## CONCLUSION

Conventional wisdom states that even though there are many options for executing the steps of EFA, the actual differences between them are small, so it doesn't really matter which methods the practitioner chooses. We disagree. Exploratory factor analysis is a complex procedure, exacerbated by the lack of inferential statistics and the imperfections of "real world" data.

While principal components with varimax rotation and the Kaiser criterion are the norm, they are not optimal, particularly when data do not meet assumptions, as is often the case in the social sciences.

We believe the data and literature supports the argument that optimal results (i.e., results that will generalize to other samples and that reflect the nature of the population) will be achieved by use of a true factor analysis extraction method (we prefer maximum likelihood), oblique rotation (such as direct oblimin), and use scree plots plus multiple test runs for information on how many meaningful factors might be in a data set. As for sample size, even at relatively large sample sizes EFA is an error-prone procedure. Our analyses demonstrate that at a 20:1 subject to item ratio there are error rates well above the field standard  $\alpha = .05$  level. The most replicable results are obtained by using large samples (unless you have

unusually strong data).

This raises another point that bears discussion: by nature and design EFA is *exploratory*. There are no inferential statistics. It was designed and is still most appropriate for use in exploring a data set. It is *not* designed to test hypotheses or theories. It is, as our analyses show, an error-prone procedure even with very large samples and optimal data. We have seen many cases where researchers used EFA when they should have used confirmatory factor analysis. Once an instrument has been developed using EFA and other techniques, it is time to move to confirmatory factor analysis to answer questions such as “does an instrument have the same structure across certain population subgroups?” We would strongly caution researchers against drawing substantive conclusions based on exploratory analyses. Confirmatory factor analysis, as well as other latent variable modeling techniques, can allow researchers to test hypotheses via inferential techniques, and can provide more informative analytic options.

In conclusion, researchers using large samples and making informed choices from the options available for data analysis are the ones most likely to accomplish their goal: to come to conclusions that will generalize beyond a particular sample to either another sample or to the population (or *a* population) of interest. To do less is to arrive at conclusions that are unlikely to be of any use or interest beyond that sample and that analysis.

## NOTES

<sup>1</sup> Total variance accounted for after rotation is only given for an orthogonal rotation. It is computed using sum of squares loadings, which cannot be added when factors are correlated, but with an oblique rotation the difference between principal components and factor analysis still appears in the magnitude of the item loadings.

## REFERENCES

Arrindell, W. A., & van der Ende, J. (1985). An Empirical-Test of the Utility of the Observations-to-Variables Ratio in Factor and Components-Analysis. *Applied Psychological Measurement*, *9*(2), 165-178.

Bentler, P. M., & Kano, Y. (1990). On the Equivalence of Factors and Components. *Multivariate Behavioral Research*, *25*(1), 67-74.

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*(3), 272-299.

Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment*, *7*(3), 286-299.

Ford, J. K., MacCallum, R. C., & Tait, M. (1986). The Application of Exploratory Factor-Analysis in Applied- Psychology - a Critical-Review and Analysis. *Personnel Psychology*, *39*(2), 291-314.

Gorsuch, R. L. (1990). Common Factor-Analysis Versus Component Analysis - Some Well and Little Known Facts. *Multivariate Behavioral Research*, *25*(1), 33-39.

Gorsuch, R. L. (1997). Exploratory factor analysis: Its role in item analysis. *Journal of Personality Assessment*, *68*(3), 532-560.

Guadagnoli, E., & Velicer, W. F. (1988). Relation of Sample-Size to the Stability of Component Patterns. *Psychological Bulletin*, *103*(2), 265-275.

Loehlin, J. C. (1990). Component Analysis Versus Common Factor-Analysis - a Case of Disputed Authorship. *Multivariate Behavioral Research*, *25*(1), 29-31.

Lovett, S., Zeiss, A. M., & Heinemann, G. D. (2002). Assessment and development: Now and in the future, *Heinemann, Gloria D. (Ed); Zeiss, Antonette M. (Ed). (2002). Team performance in health care: Assessment and development. Issues in the practice of psychology. (pp. 385 400*

MacCallum, R. C., & Tucker, L. R. (1991). Representing Sources of Error in the Common-Factor Model - Implications for Theory and Practice. *Psychological Bulletin*, *109*(3), 502-511.

MacCallum, R. C., Widaman, K. F., Zhang, S. B., & Hong, S. H. (1999). Sample size in factor analysis. *Psychological Methods*, *4*(1), 84-99.

Majors, M. S., & Sedlacek, W. E. (2001). Using factor analysis to organize student services. *Journal of College Student Development*, *42*(3), 2272-2278.

Marsh, H. W. (1990). A multidimensional, hierarchical model of self-concept: Theoretical and empirical justification. *Educational Psychology Review*, *2*, 77-172.

McArdle, J. J. (1990). Principles Versus Principals of Structural Factor-Analyses. *Multivariate Behavioral Research*, *25*(1), 81-87.

Morris, S. B. (2001). Sample size required for adverse impact analysis. *Applied HRM Research*, *6*(1-2), 13-



32.

- Mulaik, S. A. (1990). Blurring the Distinctions between Component Analysis and Common Factor-Analysis. *Multivariate Behavioral Research*, 25(1), 53-59.
- National Center for Educational Statistics (1992) *National Education Longitudinal Study of 1988 First Follow-up: Student component data file user's manual*. U.S. Department of Education, Office of Educational Research and Improvement.
- Schonemann, P. H. (1990). Facts, Fictions, and Common-Sense About Factors and Components. *Multivariate Behavioral Research*, 25(1), 47-51.
- Snook, S. C., & Gorsuch, R. L. (1989). Component Analysis Versus Common Factor-Analysis - a Monte- Carlo Study. *Psychological Bulletin*, 106(1), 148-154.
- Steiger, J. H. (1990). Some Additional Thoughts on Components, Factors, and Factor- Indeterminacy. *Multivariate Behavioral Research*, 25(1), 41-45.
- Tabachnick, B. G., & Fidell, L. S. (2001). *Using Multivariate Statistics*. Boston: Allyn and Bacon.
- Velicer, W. F., & Fava, J. L. (1998). Effects of variable and subject sampling on factor pattern recovery. *Psychological Methods*, 3(2), 231-251.
- Velicer, W. F., & Jackson, D. N. (1990). Component Analysis Versus Common Factor-Analysis - Some Further Observations. *Multivariate Behavioral Research*, 25(1), 97-114.
- Widaman, K. F. (1990). Bias in Pattern Loadings Represented by Common Factor-Analysis and Component Analysis. *Multivariate Behavioral Research*, 25(1), 89-95.
- Widaman, K. F. (1993). Common Factor-Analysis Versus Principal Component Analysis - Differential Bias in Representing Model Parameters. *Multivariate Behavioral Research*, 28(3), 263-311.

## Citation

Costello, Anna B. & Jason Osborne (2005). Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis. *Practical Assessment Research & Evaluation*, 10(7). Available online: <http://pareonline.net/getvn.asp?v=10&n=7>

## Contact

Jason W. Osborne, Ph.D.  
602 Poe Hall  
Campus Box 7801  
North Carolina State University  
Raleigh, NC 27695-7801  
e-mail: [jason\\_osborne@ncsu.edu](mailto:jason_osborne@ncsu.edu)