Sample size and subject to item ratio in principal components analysis.

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

Statisticians have wrestled with the question of sample size in exploratory factor analysis and principal component analysis for decades, some looking at total N, some at the ratio of subjects to items. Although many articles attempt to examine this issue, few examine both possibilities comprehensively enough to be definitive. This study examines a previously published data set to examine whether N or subject to item ratio is more important in predicting important outcomes in PCA. The results indicate an interaction between the two, where the best outcomes occur in analyses where large Ns and high ratios are present.

Exploratory factor analysis (EFA) and Principal Components analysis (PCA) have both been important tools for researchers for the better part of a century now, and have become increasingly common with ubiquitous access to computing. Yet EFA and PCA remain oddities in quantitative analysis, as there are no inferential statistical tests, and no way to calculate or control the probability of making an error of inference. While the field as a whole aspires to maintain error rates of 5% or less, we have no way of knowing what proportion of EFAs or PCAs result in errors of inference.

It is crucial that statisticians use sound methodology when conducting studies involving EFA or PCA to minimize error rates and maximize the generalizability to the population of interest. The goal of this paper is to examine how and whether sample size affects the goodness of several important outcomes relating to principal components analysis. PCA is one of the most commonly used exploratory data reduction procedures used in the social sciences, and is conceptually and mathematically distinct from exploratory factor analysis, although the conclusions reached in this paper should generalize to EFA.

Why size matters

Larger samples are better than smaller samples (all other things being equal) because larger samples tend to minimize the probability of errors, maximize the accuracy of population estimates, and increase the generalizability of the results. Unfortunately, there are few sample size guidelines for researchers using EFA or PCA, and many of these have minimal empirical evidence (e.g., Guadagnoli & Velicer, 1988).

This is problematic because statistical procedures that create optimized linear combinations of variables (such as multiple regression, canonical correlation, and EFA\PCA) tend to "overfit" the data. This means that these procedures optimize the fit of the model the given data; yet no sample is perfectly reflective of the population. Thus, this overfitting can result in erroneous conclusions if models fit to one data set are applied to others. In multiple regression this manifests itself as inflated R^2 (shrinkage) and mis-estimated variable regression coefficients (Cohen & Cohen, 1983, p. 106). In EFA or PCA this "overfitting" can result in erroneous conclusions in several ways, including the extraction of erroneous factors or mis-assignment of items to factors (e.g., Tabachnick & Fidell, 2001, p. 588)

The ultimate concern is error. At the end of the analysis, if one has too small a sample, errors of inference can easily occur, particularly with techniques such as EFA or PCA.

Published sample size guidelines

In multiple regression texts some authors (e.g., Pedhazur, 1997, p. 207) suggest subject to variable ratios of 15:1 or 30:1 when generalization is critical. But there are few explicit guidelines such as this for EFA or PCA (Baggaley, 1983). Two different approaches have been taken: suggesting a minimum total sample size, or examining the ratio of subjects to variables, as in multiple regression.

Comfrey and Lee (1992) suggest that "the adequacy of sample size might be evaluated very roughly on the following scale: 50 – very poor; 100 – poor; 200 – fair; 300 – good; 500 – very good; 1000 or more – excellent" (p. 217). Guadagnoli and Velicer (1988) review several studies that conclude that absolute minimum sample sizes, rather than subject to item ratios, are more relevant. These studies range in their recommendations from an N of 50 (Barrett & Kline, 1981) to 400

(Aleamoni, 1976).

The case for ratios. There are few in the multiple regression camp who would argue that total *N* is a superior guideline than the ratio of subjects to variables, yet individuals focusing on the PCA and/or EFA methodologies occasionally vehemently defend this position. It is interesting precisely because the general goal for both analyses are the same: to take individual variables and create optimally weighted linear composites. While the mathematics and procedures differ in the details, the essence and the pitfalls are the same. Both EFA/PCA and multiple regression experience shrinkage, the over-fitting of the estimates to the data (Bobko & Schemmer, 1984), both suffer from lack of generalizability and inflated error rates when sample size is too small.

We find absolute sample sizes simplistic given the variance in the types of scales researchers examine. Each scale differs in the number of factors or components, the number of items on each factor, the magnitude of the item-factor correlations, and the correlation between factors, for example. This discomfort has led some authors to focus on the ratio of subjects to items, or more recently, the ratio of subjects to parameters (as each item will have a loading for each factor or component extracted), as authors do with regression, rather than absolute sample size when discussing guidelines concerning EFA and PCA.

Gorsuch (1983, p.332) and Hatcher (1994, p. 73) recommend a minimum subject to item ratio of at least 5:1 in EFA, but they also have stringent guidelines for when this ratio is acceptable, and they both note that higher ratios are generally better. There is a widely-cited rule of thumb from Nunnally (1978, p. 421) that the subject to item ratio for exploratory factor analysis should be at least 10:1, but that recommendation was not supported by published research. There is no one ratio that will work in all cases; the number of items per factor and communalities and item loading magnitudes can make any particular ratio overkill or hopelessly insufficient (MacCallum, Widaman, Preacher, & Hong, 2001).

Previous research on ratios. Unfortunately, much of the literature that has attempted to address this issue, particularly the studies attempting to dismiss subject:parameter ratios, use flawed data. We will purposely not cite studies here, but consider it sufficient to say that many of these studies either tend to use highly restricted ranges of subject:item or subject:parameter ratios or fail to adequately control for or vary other confounding variables (e.g., factor loadings, number of items per scale or per factor/component) or restricted range of *N*. Some of these studies purporting to address subject to item ratio fail to actually test subject to item ratio in their analyses.

Thus researchers seeking guidance concerning sufficient sample size in EFA or PCA are left between two entrenched camps-- those arguing for looking at total sample size and those looking at ratios. This is unfortunate, because both probably matter in some sense, and ignoring either one can have the same result: errors of inference. Failure to have a representative sample of sufficient size results in unstable loadings (Cliff, 1970), random, non-replicable factors (Aleamoni, 1976; Humphreys, Ilgen, McGrath, & Montanelli, 1969), and lack of generalizability to the population (MacCallum, Widaman, Zhang, & Hong, 1999).

EFA and PCA in practice

If one were to take either set of guidelines (e.g, 10:1 ratio or a minimum N of 400 - 500) as reasonable a casual perusal of the published literature shows that a large portion of studies come up short. One can easily find articles utilizing EFA or (more commonly) PCA based on samples with fewer subjects than items or parameters estimated that nevertheless draw substantive conclusions based on these questionable analyses. Many more have hopelessly insufficient samples by either guideline.

For example, Ford, MacCallum, and Tait (1986) examined common practice in factor analysis in industrial and organizational psychology during the ten year period of 1974 - 1984. They found that out of 152 studies utilizing EFA or PCA, 27.3% had a subject to item ratio of less than 5:1; 56% had a ratio of less than 10:1. A similar, more recent survey of 1076 journal articles utilizing PCA or EFA in psychology revealed that 40.5% of peer-reviewed, published studies utilized less than a 5:1 subject to item ratio, and 63.2% utilized 10:1 or under (Costello & Osborne, 2003). Given the stakes and the empirical evidence on the consequences of insufficient sample size, this is not exactly a desirable state of affairs.

The Present Study

This study focuses on one particularly interesting and well-executed study on this issue—that of Guadagnoli and Velicer (1988). In this study, the authors used monte carlo methods to examine the effects of number of components (3, 6, 9, 18), the number of variables (36, 72, 108, and 144), average item-component correlation (.40, .60, or .80), and number of subjects (*Ns* of 50, 100, 150, 200, 300, 500, and 1000) on the stability of component patterns in principal components analysis. In these analyses each item loaded on only one component, all items loaded equally on every component, and each component contained an equal number of variables. This study represents one of the few studies to manipulate all of these important aspects across the range of variation seen in the literature (with the two possible exceptions: first, people often have less than 36 items in an EFA or PCA analysis, and second, the factor loading patterns are rarely as clear and homogenous as in these data).

Guadagnoli and Velicer's (1988) study was also interesting in that they used several different high-quality fit/agreement

indices. Equally interesting is the authors' strong assertion that total sample size is critical, although they never actually operationalize subject to item ratio, nor test whether total N is a better predictor of important outcomes than subject to item ratio, although given their data it was possible to do so. Finally, the authors' complete data tables were published, allowing for reanalyses of the data.

The goal of this study is to directly examine competing claims regarding the importance of sample size to PCA -- to determine whether either overall sample size or subject to item ratio uniquely contribute to the "goodness" of outcomes in PCA, beyond the contributions of other important variables, such as number of variables or components and average item loading that have been identified as important in the literature.

<div align="center">

METHODS

</div>

The data for this study were taken directly from published data in Guadagnoli and Velicer (1988). These data were generated via monte carlo methods outlined in their article in detail. In general, the authors generated multiple data sets, each set representing a specific combination of conditions, discussed below. These data sets were then subjected to PCA, and the outcomes were recorded for analysis.

Variables included by the authors

Number of factors (m). The number of components examined included 3, 6, 9, and 18.

Loadings. The authors used loadings of .40, .60, and .80. It should be noted that in these data sets, items not intended to load on a component were assigned a loading of 0.00, making these pattern matrices artificially clear.

Number of items (p). The number of items in the analyses included 36, 72, 108, and 144.

Number of subjects (N). The number of subjects in the analyses included 50, 100, 150, 200, 300, 500, and 1000. Note however, that certain cases were omitted or altered by the authors, such as when N was less than the number of items in the analysis.

Pattern comparison (g^2). In order to compare sample component patterns with population component patterns, the average of the squared differences between the two matrices was computed. Furthermore, the authors identified $g^2 = .01$ as the maximum value that indicates acceptable fit.

Pattern agreement (kappa). Salient variables (loadings > .40) and non-salient variables (loadings < .40) were identified and noted in decision tables. These decision tables were then compared to the population decision table via the kappa statistic. As kappa approaches 1.0 the two matrices become more in agreement with each other. A 0 indicates random chance level of agreement, and negative kappas indicate poorer than chance agreement.

Type I errors. The authors calculated the percent of variables that should not have been considered salient but were in a particular data set, indicating Type I error classifications.

Type II errors. The authors also calculated the percent of variables that should have been considered salient but were not found to be so, indicating Type II error classifications.

Variables calculated for this analysis

For our purposes we calculated the following variables based on the information obtained from the data set:

Subject-to-item ratio. The ratio of the number of subjects per item in a particular analysis was calculated from the information given.

Variable-to-component ratio. As some authors have argued that the number of variables per component or factor is important (see Guadagnoli & Velicer, 1988) we included this variable in analyses.

Extra matrices. In describing their data generation procedures Guadagnoli and Velicer (1988) indicated that under certain conditions "the component patterns did not possess a structure defined well enough for a one-to-one component match with the population component structure to be attained." (p. 267). In other words, certain data sets produced errors of inference regarding the number of factors extracted from the data. The authors discarded these data matrices and replaced them until 5 good matrices for a particular set of criteria were obtained. They noted cases where up to 10 additional matrices were required before 5 good matrices were obtained, and cases where 10 or more matrices were required (a phenomenally high error rate). From an applied research point of view, this could be viewed as an important outcome, where a researcher would find results that differ radically from the population, and thus should be examined as a variable of interest. Thus, this variable was coded into the data set for the current analyses as *0* (no extra matrices), *1* (up to 10 extra matrices required), or *2* (more than 10 extra matrices required) as that is how this information was

reported.

Correct factor structure. What many people are looking for when they do a EFA or PCA analysis is the pattern—what variables "load" on what components. Researchers are generally less interested in the absolute magnitude of the loading (above a certain "salient" level-- that is a source of debate in and of itself) than which variable goes with which factor. Thus, we included information in our data set that indicated when this had or had not occurred, based on the number of Type I and Type II errors. Matrices that had no errors were considered "correct," while matrices with errors were considered not correct. Some might think this a strict criterion, and they are correct. However, the presence of these errors can significantly alter the interpretation of an exploratory analysis (either EFA or PCA). In this study, 34.3% of the cases failed to faithfully replicate the pattern found in the population.

Data

The authors generated five samples for each of the 205 valid conditions described. The average g^2, kappa, type I error, type II error for the five samples in each condition were reported in tables. Thus, these results represent analyses of the data aggregated across five samples in each condition.

RESULTS

Main effects

With the exception of the newly-calculated variables described above, we attempted to faithfully reproduce the authors' analyses. We performed multiple regression analyses on the dependent variables (g^2, kappa, Type I error, and Type II error), a binomial logistic multiple regression predicting correct factor structure, and a multinomial logistic regression predicting the presence of extra matrices. As in the original article, we examined all possible two-way interactions. The difference is that we now simultaneously can examine total N and subject to item ratio for their unique and joint contributions to the goodness of PCA outcomes.

The results of these analyses are presented in Tables 1-3. As Table 1 indicates, the number of components was not significant predictor of any dependent variable once other variables were controlled for. As previous research has reported, item loading magnitude accounted for significant unique variance in the expected direction in all but one case, and in most cases was the strongest unique predictor of congruence between sample and population. Specifically, as item loadings increased, average squared discrepancy between population and sample results (g^2) decreased, agreement (kappa) increased, Type II errors decreased, and the odds of getting the correct component pattern increased dramatically.

Table 1: Predictors of component pattern stability—main effects

Dependent variable:	Number of components	Loadings	Number of variables	Number of subjects	Subject: item ratio	Variable: component ratio
g^2	-.08	-.41***	-.26	-.11	-.37***	-.25
	(.15**)	(-.45***)	(-.30***)	(-.41***)		
Kappa	-.09	.62***	.16	.20	.14	-.08
	(-.01)	(.62***)	(.03)	(.31***)		
Type I error	-.12	-.22	-.23	.12	-.36***	-.20
	(-.12)	(-.22***)	(-.23***)	(-.17***)		
Type II error	.09	-.67***	-.03	-.29***	-.03	.10
	(-.01)	(-.67***)	(.07)	(-.31***)		
Correct pattern[1]	1.14	.58***	1.01	1.00	1.44**	1.15
	(1.10)	(.99)	(.99)	(1.01***)		
		.76***				
		(.99)				

Note: Statistics reported represent betas (standardized regression coefficients) when all predictors are in the equation. Betas in parentheses are from regression equations with two ratio variables removed. * $p < .05$, ** $p < .01$, *** $p < .001$.

1. Odds ratio reported. For loadings, as there was only .40, .60, and .80 for values, this was considered a categorical variable. Thus, the first odds ratio represents the relative odds of getting correct pattern structures with a .40 vs. a .80 average loading, while the second odds ratio represents the relative odds of getting correct pattern structures with a .60 vs. an .80 average loading.

Contrary to other studies, neither the number of variables nor N had a significant unique effect when all other variables were held constant (except for the relationship between N and the odds of a Type II error). The lack of findings for these

two variables might be directly attributable to the presence of the ratios of subject to item and variable to component, which would likely share variance. To test this hypothesis, a blockwise multiple regression was performed entering number of components, loadings, number of variables, and N in block 1, and subject to item ratio and variable to factor ratio in block 2. Number of variables was a significant predictor in two of the five analyses where the two ratio variables were not in the equation. Total N was significant in all five analyses where the two ratio variables were not in the equation.

The ratio of subjects to items had a significant and substantial influence on three outcomes in the expected direction. As subject to item ratio increased, the squared discrepancy between population and sample matrices decreased, the odds of a Type I error decreased, and the odds of getting a correct component pattern matrix increased. Finally, the ratio of variables to components had no unique effect.

The multinomial logistic regression predicting the need for extra matrices did not identify any significant predictors, nor did a binomial logistic regression analysis predicting the need for any extra matrices or none. Whatever the reason for this lack of results, it is clear that this outcome is related to subject to item ratio, number of subjects, and the ratio of variables to components (all $p < .001$ when analyzed individually) , as Table 2 shows. Note that this event only occurred when loadings were relatively weak (.40), and thus loadings was held constant.

Table 2 The relationship between the number of extra matrices drawn and subject:item ratio.

	10 + extra	1-9 extra	No extra
S:I ratio			
<5:1	8	7	30
5:1 -10:1		1	14
>10:1			9
# subjects			
50	2		1
100	2	1	3
150	2	4	6
200	2		10
300		2	10
500		1	11
1000			12
V:F ratio			

<10:1	8	7	15
10:1-19:1		1	27
>19:1			11

Interactions

To test for interaction effects, a blockwise multiple regression analysis was performed entering all main effects in block 1 and all interactions in block 2. In all cases, when block 2 was entered there was a significant change in R and R^2 (all $p <$ .0001).

As the results in Table 3 show, there were several interesting interactions present in these data. There were no significant interactions including the ratio of variables to components.

Table 3: Predictors of component pattern stability—interactions				
Interaction	g^2	**Kappa**	**Type I error**	**Type II error**
# components x loadings	.003	---	---	---
# components x # variables	---	---	---	---
# components x # subjects	---	---	---	---
# components x S:I ratio	---	---	---	---
Loading x # variables	.0001	---	.0001	---
Loading x # subjects	---	.004	---	.0001
Loading x S:I ratio	.007	---	.003	---
# variables x # subjects	.0001	.03	.003	---
# variables x S:I ratio	---	---	---	---
# subjects x S:I ratio	.0001	.0001	.0001	.0001

Number of components and component loadings. There was a significant interaction between the number of components extracted and the magnitude of component loadings. The nature of the interaction indicated that more components tended to inflate g^2 when loadings were relatively weak, but had less of an effect when the loadings were very strong.

Component loadings and the number of variables. This interaction indicated that, while stronger component loadings are related to lower g^2 and Type I error rates, loadings had less of an effect as the number of variables increased.

Component loadings and the number of subjects. This interaction indicated that, while stronger component loadings are related to higher kappas and lower Type II error rates, loadings had less of an effect as the number of subjects increased.

Component loadings and the ratio of subjects to variables. This interaction indicated that, while stronger component loadings were generally related to lower g^2 and Type I error rates, loadings had less of an effect as the ratio of subjects to variables increased.

Number of variables and the number of subjects. While the number of variables was generally related to more favorable outcomes (lower g^2, higher kappa, and lower Type I error rates), as the number of subjects increased the effect of the number of variables decreased.

Number of subjects and the ratio of subjects to variables. While increasing ratios of subjects to variables was generally related to more favorable outcomes (lower g^2, higher kappa, and lower Type I and Type II error rates), as N increased, this effect became less important.

DISCUSSION

While the original authors of this study concluded that the only two important factors in determining the correspondence between a PCA and the population were the raw number of subjects and the magnitude of the component loadings, the examination of the ratio of subjects to variables and variables to components and their various interactions tell a slightly more subtle story.

First, while the magnitude of component loadings has a large influence on goodness of the analyses, the raw number of subjects had a significant influence on the average percent of Type II errors. The ratio of subjects to variables had a significant unique effect on g^2, Type I error rates, and obtaining the correct loading pattern. In looking at Table 1 it is difficult to dismiss component loadings and the ratio of subjects to variables as the most consistent predictors of these variables. Equally notable was the relative lack of unique impact of N once the ratio of subject to variables was accounted for.

These main effects were in some ways qualified by several interactions. For example, the ratio of subjects to variables appeared to have a larger effect when the raw number of subjects was lower, the number of subjects appeared to have less of an effect when there were fewer variables in the analysis, and the number of variables, the number of subjects, and the subject to variable ratio had a larger effects when the component loadings were smaller.

The interaction of N and subject to variable ratio was particularly interesting. Although the ratio of subject to variable is an important predictor of the goodness of a PCA or EFA, it appears that as total N increases, this ratio becomes less important (the converse is also true-- as the subject to item ratio increases, total N becomes less important). In some sense, then, authors from both sides of this debate are correct-- total N matters (but more so when subject:item ratio is low), and the ratio of subjects to items matters (but more so when N is relatively low), and if you have a large N or large ratio, your results will be more reliable. It should be clear from both the main effect and interaction analyses that it is difficult to dismiss any of these factors in discussing the reproducibility of population values and patterns in sample analyses.

Caveats

It should also be noted that, although these data were superior than the data used in many other articles on the topic, they are not necessarily ideal. For example, the original authors chose to analyze large number of items (36 minimum, 144 maximum), which may not be representative of what researchers generally investigate. The number of factors was generally large, ranging from 3 to 18, ignoring single-or two-factor analyses. Finally, the median subject to item ratio was 3.5 (with a range of 1.04:1 to 27.78:1), which is far from ideal. This general restriction of range and the highly skewed nature of the ratio variable may lead to an underestimation of the effect of this ratio, as most guidelines call for at least 10:1 or more.

It is also important to note that the results reported here reflect principal component analyses. Many statisticians and methodologists will point out that EFA and PCA are distinct procedures. However, from a practitioner point of view, the mathematics and processes behind each are related and similar, in practice the outcome of a PCA and EFA is often identical, and these results relating to PCA *should* generalize to EFA handily. However, the caveat is that the effect of these variables on EFA has not been rigorously demonstrated yet.

Finally, it should be noted that even the original authors noted the artificially "clean" nature of the patterns being replicated. No researcher working with empirical data will see patterns of 0s and .60s, for example. In fact, many researchers will not see average loadings of .80 at all; anything over .50 is generally classified as a "strong" item loading. Moderate and weak loadings, which are frequently the rule in behavioral research, range from .32 (which equates to approximately 10% of the variance accounted for) up to .50. Guadagnoli and Velicer (1988) also completely ignore crossloaders, which are items that load above the .32 level on more than one factors. These are particularly prone to occur during the initial research phase of measure construction, when items are being tested for inclusion. The incidence of crossloadings can be dramatically effected by both sample size and subject:item ratio, particularly when item loadings are in the low to moderate range (Costello & Osborne, 2003). Frequently crossloaders disappear when the sample size is adequate, but when the sample size or subject:item ratio is small there is no way to determine whether a crossloading item is the result of a sampling error or the indication of a poor item.

Past studies have found that variables such as the number of items per component/factor, and the magnitude of the item loadings tend to reduce the sample size needed for valid inference. We do not take issue with these previous findings. While it is possible for researchers to control the number of items if the researchers is also a scale designer, it seems to be

even easier to control sample size in many cases.

CONCLUSIONS

Although some authors eschew the concept of subject to variable ratio as an important influence in the "goodness" of exploratory factor analysis or principal components analysis, that seems short-sighted and simplistic. Researchers need to remember that EFA and PCA (and other techniques like structural equation modeling) are large-sample techniques, not well-suited to the small sample sizes some researchers use them on. These analyses demonstrate that, at least for this data set (which we believe is one of the better ones out there), holding all other variables constant, subject to variable ratio makes a significant contribution beyond that of mere sample size, particularly when overall sample size is not overwhelmingly large.

But these effects do not show evidence of a "critical mass" or "critical ratio"-- they do not plateau, the lines are not asymptotic. There are diminishing returns, but even at large subject to item ratios and *N*s (such as 20:1 ratio or N > 1000) and with unrealistically strong factor loadings and clear factor structures, EFA and PCA can produce error rates up to 30% (Costello & Osborne, 2003), leaving room for improvement via larger samples.

Thus, the most valid conclusion regarding sample size is that more is always better. Period. If subject to item ratios appeal intuitively to some researchers, and if it leads researchers to utilize samples of a more appropriate size, it is useful. Why not encourage this way of thinking?

References

Aleamoni, L. M. (1976). The relation of sample size to the number of variables in using factor analysis techniques. *Educational and Psychological Measurement, 36*, 879-883.

Baggaley, A. R. (1983). Deciding on the ratio of number of subjects to number of variables in factor analysis. *Multivariate Experimental Clinical Research, 6*(2), 81-85.

Barrett, P. T., & Kline, P. (1981). The observation to variable ratio in factor analysis. *Personality study and group behavior, 1*, 23-33.

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

Cliff, N. (1970). The relation between sample and population characteristic vectors. *Psychometrika, 35*, 163-178.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Comfrey, A. L., & Lee, H. B. (1992). *A First Course in Factor Analysis*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Costello, A. B., & Osborne, J. W. (2003). *Exploring best practices in Factor Analysis: Four mistakes applied researchers make*. Paper presented at the Paper presented at the annual meeting of the American Educational Research Association, Chicago, Ill, April.

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*, 291-314.

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

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

Hatcher, L. (1994). *A Step-by-Step Approach to Using the SAS® System for Factor Analysis and Structural Equation Modeling*. Cary, N.C.: SAS Institutte, Inc.

Humphreys, L. G., Ilgen, D., McGrath, D., & Montanelli, R. (1969). Capitalization on chance in rotation of factors. *Educational and Psychological Measurement, 29*(2), 259-271.

MacCallum, R. C., Widaman, K. F., Preacher, K. J., & Hong, S. (2001). Sample size in factor analysis: The role of model error. *Multivariate Behavioral Research, 36*, 611-637.

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

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

Pedhazur, E. J. (1997). *Multiple Regression in Behavioral Research: Explanation and Prediction*. Fort Worth, TX:

Harcourt Brace College Publishers.

Tabachnick, B. G., & Fidell, L. S. (2001). *Using Multivariate Statistics* (4th ed.). New York:: Harper Collins.

Author Contact information:

Jason W. Osborne,
Dept of Curriculum and Instruction
North Carolina State University
Poe Hall 602, Campus Box 7801
Raleigh NC 27695-7801

919-515-1714
jason_osborne@ncsu.edu

Author notes:

As often happens in science, the impetus for this paper was a methodological debate arising out of the second author's Master's thesis. We decided to take an empirical and scholarly approach to informing the debate. Communication regarding this paper should be directed via email to jason_osborne@ncsu.edu or blandy_costello@ncsu.edu.