9-2015

Performance of Number of Factors Procedures in Small Sample Sizes

Marc Thomas Porritt

LOMA LINDA UNIVERSITY
School of Behavioral Health
in conjunction with the
Faculty of Graduate Studies

Performance of Number of Factors Procedures in Small Sample Sizes

by

Marc Thomas Porritt

A Dissertation submitted in partial satisfaction of
the requirements for the degree
Doctor of Philosophy in Clinical Psychology

September 2015

Each person whose signature appears below certifies that this dissertation in his/her opinion is adequate, in scope and quality, as a dissertation for the degree Doctor of Philosophy.

____, Chairperson
Kendal C. Boyd, Associate Professor of Psychology

____
Brian Distelberg, Associate Professor of Counseling and Family Sciences

____
Sylvia M. Herbozo, Assistant Professor of Psychology

____
Holly E.R. Morrell, Assistant Professor of Psychology

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to Dr. Boyd for his thoughtful guidance and supervision of this project. I am also deeply appreciative to the members of my committee for their insightful feedback and support and above all their partience and understanding. The success of this project is also in large part due to aid of Dr. Robert MacCallum who has been more than generous in sharing his resources including data and syntax.

I also owe a debt of gratitude to my family. Especially to my wife, Ingrid, for her countless hours of revision and support and as well as her patience with this process; and my parents who have always believed in me and supported me, no matter what. I would also like to thank my children for being patient with their daddy and all the time he has had to spend away. Finally, I would like to acknowledge the guiding hand of my Lord and God without whom nothing would be possible.

CONTENT

FIGURES

Figures Page

TABLES

ABBREVIATIONS

EFA	Exploratory Factor Analysis
MAP	Minimum Average Partail
MAP4	Fourth Power Minimum Average Partail
PA	Parellel Anaylisis
PA95	95th Percentile Parellel Anaylisis
SL	Salient Loadings Criteria

ABSTRACT OF THE DISSERTATION

Performance of Number of Factors Procedures in Small Sample Sizes

by

Marc Thomas Porritt

Doctor of Philosophy, Graduate Program in Clinical Psychology
Loma Linda University, September 2015
Dr. Kendal C. Boyd, Chairperson

Recent studies have indicated that under the proper circumstances factor anaylisis may be accurately performed in samples as small as N = 9. However, all of these studies have extracted a pre-known number of factors, leaving an examination of determining the proper number of factors to future studies. The current study uses examines the following methods for determining the proper number of factors: Monte Carlo data to examine the performance of common versions of the Kaiser Rule, minimum average partial, parallel analysis and salient loading criteria under the conditions created by all possible combinations of method, model strength, overdetermination and sample size. Method performance was compared for overall accuracy (percent correct), and average discrepancy (mean difference from correct). ANOVA revealed that item level methods, including salient loading criteria and MAP procedures, maintain accuracy when model strength is at least moderate and overdetermiantion is high. Use of selected empirical methods for determining the number of factors is possible in small sample sizes only when overdetermination and model strength are adequately high, larger sample sizes should be preferred when possible.

CHAPTER ONE

INTRODUCTION

Factor analysis is a correlation method used to combine a number of variables into a limited number of factors that hypothetically represent real world constructs (e.g., personality traits, intellectual abilities). This technique has recently been called "arguably the most popular and useful method for identifying underlying dimensions that can account mathematically for behavior" (Widaman, 2012). Exploratory Factor Analysis (EFA) relies on direct observation of the data without placing prior constraints on the outcome. This freedom allows researchers to establish working hypotheses based on empirical observation, making EFA a preferable method for initial analyses where little is known about the constructs of interest.

Despite its popularity and effectiveness, applications of EFA have been limited as it has typically been considered a large sample method with recommended minimum sample sizes ranging from 100 – 1000. These guidelines seemingly preclude the use of EFA in areas where it is typically impractical or impossible to obtain large samples. However, recent research indicates that under appropriate conditions accurate EFA may be possible with much smaller samples. If this is the case, the use of EFA could clearly be expanded.

Sample Size and Error in EFA

Large sample sizes have traditionally been thought to increase the accuracy of estimates by creating a more representative sample and increasing statistical power. However, recent research indicates that estimates are also influenced by aspects of the statistical model which may compensate for low sample size. MacCallum and Tucker (1991) have defined two primary sources of error in factor analysis: “model error” that results from designing a model that is discrepant from the true model (e.g., to many or too few factors or the wrong factor loadings for variables), and “sampling error” the result of deriving models from a sample that does not exactly mirror the population. This distinction leads to two different methods for improving accuracy: improving statistical rigor in model specification and improving the sampling so as to better reflect the population.

Past efforts have focused on reducing error by improving samplin. One of the quickest ways to make a sample more representative and add statistical power is to increase the sample size. With this in mind a number of researchers have sought to establish proper practices for reducing error by proposing guidelines for appropriate minimum sample sizes for EFA. Among others, Cattell (1978) suggested a minimum of 250, Gorsuch (1983) recommended 100, and Comrey and Lee (1992) provided border recommendations identifying N = 100 as poor, N = 300 as good, and N = 1,000 or more as excellent.

Several authors have noted that the relationship between sample size and accuracy appears to be effected by model complexity. As a result they have attempted to take into account model complexity by recommending a minimum ratio of the number of measured variables to number of participants. Most notably, Cattell (1978) recommended a ratio of 1:2, Gorsuch (1983) recommended 1:5, and Everitte (1975) recommended 1:10. Despite their promise, few of these rules have substantive theoretical or empirical backing. Indeed, results of empirical studies fail to support any rigid rules. Gaudagnoli and Velicer's (1988) examination of the effects of sample size on extraction failed to produce consistent effects of sample size or *n:p* ratio. They found instead that the component saturation (strength of factor loadings) was the most significant predictor of accuracy. These results were supported by Rouquette and Falissard (2011), who were also unable to find empirical support for any consistent effect of the *n:p* ratio.

MacCallum, Widaman, Zhang, and Hong (1999) have questioned the assumption that specific elements (sample size, *n:p* ratio) consistently effect accuracy across all conditions. They asserted that the mathematical formulas for factor analysis consist of many differing interdependent elements that exert influence in a manner which varies across conditions in a systematic and predictable manner. That is to that the mathematical formulas used to estimate factor structure include other elements (primarily strength of model) that moderate the relationship between sample size and accuracy, making the relationship inconsistent across conditions. More specifically, they assert that the accuracy of factor extraction is dependent on the interaction of sample size, overdetermination (ratio between factors and variables), and strength of model (communalities of the variables); with model strength exerting the greatest influence over accuracy.

The typical operational definition for model strength is communality, which is a measure of the variance in a given variable accounted for by the overall EFA model. Each variable within a model has its own communality. MacCallum et al. (1999) conceptualize data sets as compilations of common and unique factors. Common factors represent variance explained by a latent construct (or variance common to that construct) and unique factors represent variance uniquely explained by a given variable. One way to conceptualize this distinction is to consider the common factor an estimate of the latent construct and the unique factor a representation of residual error variance in that given variable.

Common factors, as defined by MacCallum et al. (1999) typically consist of more than one variable and represent a latent construct. Each variable within each common factor is assigned a factor loading (represented as a number between 0 and 1) that represents the amount of variance within this particular variable explained by the latent construct represented by the common factor; communalities then represent the amount of variance within a given variable, explained by all common factors (the EFA model in its entirety). When population values are known, communalities can be calculated by summing the squared loadings for a single variable across all retained factors. As population values are not typically known, communalities are iteratively estimated when using common factors extraction.

Continuing with MacCallum et al.'s (1999) conceptualization, unique factors consist of a single variable and represent the amount of unexplained variance within that variable. Stated another way, loadings on unique factors represent residual variance left unexplained by the main EFA model factors. Each data set will contain one unique factor for each unique measured variable.

The MacCallum et al. (1999) conceptualization provides us with a representation for 100% of the variance in the data; communalities represent variance explained by the model and unique factors representing residual variance. By their very nature, these two quantities must be inversely related. As common factors explain more variance, there is less variance left to be explained by the unique factors and vice-versa. This inverse relationship is the key to the moderating effect of communalities on the relationship between sample size and accuracy.

Theoretically, each unique factor represents a separate and unrelated portion of total variance. Therefore, unique factors should be orthogonal because they are not related to any other factor. However, sampling error creates a shared class of variance which leads to fallacious correlations between common and unique factors as well as within the set of unique factors. MacCallum et al. (1999) noted that the population formula for estimating factor structures is greatly simplified when the orthogonal relationship of unique factors provide zeroes in key locations. They assert that erroneous correlation coefficients, resulting from correlations of sampling error, introduce significant error in model estimates by inserting correlation coefficients where a zero would cancel out an entire set of error. If this were the primary source of error, larger sample sizes would remedy the situation by reducing the shared sampling error. However, as previously stated, empirical studies have failed to verify this type of a simple linear relationship between sample size and accuracy (Guadagnoli & Velicer, 1988; Rouquette & Falissard, 2011).

MacCallum et al. (1999) explain that this difficulty in capturing the relationship between sample size and accuracy is once again due to the mathematical formula used to estimate factor structures. They note that this formula employs a vector of the unique factors in a manner that functions as a weight for the correlation matrix containing the erroneous correlations caused by sampling error. Therefore, the larger the loadings on unique factors, the greater the erroneous correlations are weighted; conversely, the smaller the unique loading, the less the influence of the erroneous correlations. Thus, the impact of sampling error is directly moderated by the unique factor loadings. The larger the unique loading, the larger the effect of sampling error will be. As previously discussed, unique factor loadings are intrinsically inversely related to communalities. Therefore, the influence of sampling error must be inversely related to communalities via the inherently reducing effect of large communalities on unique factor loadings. MacCallum et al. (1999) then conclude that when communalities are high (and therefore unique loading are low) EFA formulas will be robust to error regardless of sample size. On the other hand, they assert that smaller communalities will increase the influence of sampling error and therefore increase the effects of sample size.

MacCallum, et al. (1999) also identified overdetermination (number of variables per factor) as another critical element effecting accuracy of analysis. They noted that increasing the number of variables representing a construct increases the reliability of the estimate because the more variables representing a construct the less susceptible it will be to the effects of random variation in any one given variable. The practice of increasing estimate reliability by using more variables is known as overdetermination. MacCallum et al. (1999) cited overdetermination as the primary method for reducing model error but cautioned that the relationship between model accuracy and overdetermination (ratio of variables to factor) will be moderated by communality strength. They finally assert that the benefit of increasing variables is to increase accuracy without increasing sample size, especially when communalities are low. Thus, MacCallum et al. propose that accuracy of factor analytic models is the result of the interaction of overdetrmination, model strength and sample size, with model strength moderating the effects of sample size and overdetermination. As such, the key focus for increasing accuracy becomes model strength instead of sample size. MacCallum et al. have empirically tested and verified all of these hypothesized relationships using simulated data containing no model error (MacCallum et al., 1999) and simulated data with model error (MacCallum et al., 2001).

The MacCallum et al. (1999) model of EFA indicates the potential for accurate analysis in small samples when communalities are high and variable per factor ratios are adequate. Both of their empirical studies obtained accurate results with samples as small as $N = 60$ (MacCallum et al., 1999; 2001). Other empirical studies have verified accuracy of EFA in even smaller samples. Mundfrom, Shaw, and Ke (2005) examined the performance of factor analysis in simulated data with varying levels of communality, factor to variable ratios, and number of factors. Accuracy of a result was evaluated in terms of the congruence of the solution with known population solutions, defined as the Tucker's coefficient. Excellent accuracy was defined as a Tucker's coefficient of 0.98 or higher, with good accuracy at 0.92 or higher. Recommended sample sizes were established by starting at a small sample size and slowly increasing the sample size until accurate solutions emerged. Final recommendations for excellent accuracy ranged from *N's* of 11 to 30 for one factor with seven variables, to *N's* of 55 to80 for six factors with seven variables each. Good accuracy was achieved with sample sizes as small as 30 in ideal conditions (factors = 3, communalities = .6 - .8, variables:factor = 8) and 35 under less ideal conditions (factors = 3, commonalities = .2-.8, variables:factor = 10).

De Winter, Dodou, and Wieringa (2009) also used MacCaullum et al.'s (1999; 2001) findings to establish minimum sample sizes for accurate extraction. They defined accuracy as a Tucker's Coefficient of at least .95 and used an algorithm that adjusted the sample size based on previous results until a minimum accurate sample size was established for each condition. Accurate results were found in samples as small as $N = 6$ under ideal conditions (communalities = .8, factors = 1, variables = 96). Samples as small as 34 yielded accurate results in more typical conditions (commonalities = .6, factors = 2 variables, = 24). These findings held when distortions (simulated model error) were introduced to the data. However, findings using empirical data were more conservative in their lower range. When applied to empirical data from a Big Five personality inventory, extraction of one factor was accurate with samples as small as 13-17, two factors required 30-50 cases, and full five-factor extraction required 80 – 140. Notably, the average communality for this empirical data set ranged from .37-.42. It is likely that simulated findings would hold more robustly in empirical data with higher communalities. Considered together, the Mudfrom et al. (2005) and De Winter et al. (2009) findings provide proof of concept for EFA accuracy in samples smaller than 50.

Extracting the Proper Number of Factors

Empirical evidence from simulation studies suggests that there exists an ideal number of factors, with under- and over-extraction presenting individual and unique threats to the validity of one's findings. None of the empirical examinations of the MacCallum model (de Winter et al., 2009; MacCallum et al., 2001; MacCallum et al., 1999; Mundfrom et al., 2005) have directly examined the issue of determining the proper number of factors to extract. Complete application of EFA to small samples will require an empirical examination of this issue.

Under-extraction is generally agreed to be the most severe case of miss-extraction (Cattell, 1978; Gorsuch, 1983; Thurstone, 1947), as it creates hybrid factors consisting of collections of loosely associated items that represent a number of different constructs, whereas over-factoring simply splinters factors into multiple factors while maintaining the consistency of each factor. Wood, Tataryn, and Gorsuch (1996) examined the effects of under- and over- extraction in a simulated sample size of $N = 200$ with principal axis extraction and varimax rotation, and found that error in factor loadings significantly increased with each factor under extracted. Increased loading error will increase overall model error. Error is also likely to depress communalities, thus weakening the overall accuracy of the model. Fava and Velicer also examined miss-extraction and found that factor scores significantly changed when factors were under-extracted (1996) and over-extracted (1992). Effects were found to be particularly deleterious when sample size and number of factors were low (Fava & Velicer, 1992, 1996), making proper extraction all the more critical in the proposed small samples. Fava and Velicer (1992) noted that the impact of over-extraction was moderated by model strength, as represented by the strength of item loadings (factor saturation). Given the close relationship between item loadings and communalities, it can be hypothesized that high communalities may serve as a moderator for the effects of miss-extraction on model accuracy.

Statistical Solutions to the Number of Factor Question

Due to the significant impact of extracting an inappropriate number of factors, several different methods for determining the proper number of factors have been established and studied. The Kaiser rule, minimum average partial, parallel analysis and salient loadings criteria appear to be the most widely used and/or promising methods.

Kaiser Rule

The Kaiser Rule (Kaiser, 1960) recommends that any factor with an eigenvalue greater than one be considered significant. Numerous studies have shown this method to consistently over-factor by as many as three to six factors (Gorsuch, 1980; Horn, 1965; Lee & Comrey, 1979; Velicer, Eaton, & Fava, 2000; Zwick & Velicer, 1982, 1986). Despite this problem, the Kaiser rule remains the most commonly used procedure for determining the number of factors to extract, and the default in many software suites (Fabrigar, Wegener, MacCallum, & Strahan, 1999; Hayton, Allen, & Scarpello, 2004). It has been included in this study because of its wide use.

Parallel Analysis (PA)

Horn (1965) noted that sampling introduces error, which in turn inflates eigenvalues. He argued that these circumstances required an inflated cut off score in place of a theoretically based cut off of one (e.g., the Kaiser Rule). PA corrects for sampling error by deriving a new cut off from the averages of eigenvalues derived from random data. The investigator randomly generates at least three datasets of similar dimensions (number of variables and sample size) to the data set that is to be analyzed. The average is calculated for each eigenvalue, and only factors that have eigenvalues larger than the average of randomly generated eigenvalues are considered significant.

PA has been shown to be superior in accuracy to all other methods (Humphreys & Montanelli, 1975; Zwick & Velicer, 1986). Zwick and Velicer (1986) found it accurate 99.6% of the time when factor saturation, measure of communality in simulated data, was .8 and 84.2% of the time at .5 saturation. This was superior to minimum average partial

(97.1% at .8 saturation and 67.5% at .5 saturation). Importantly errors for PA in this study followed the same pattern that MacCallum et al. (1999) identified for accuracy in overall factor analytic models. That is, errors occurred most often when the number of variables per factor was low, sample size was small, or factor saturation was low.

It has been noted that using the mean of random eigenvalues may be too liberal a cut off, as it allows for a 50% chance that a random value could be considered significant. This can be seen in Zwick and Velicer's (1986) study where nearly two-thirds of PA misses were due to over-extraction. Accordingly, several authors have called for a more conservative cut-off point. The 95th percentile of the randomly generated eigenvalues is the most commonly recommended cut score, since it is above the mean and corresponds to an alpha level of .05, (Buja & Eyuboglu, 1992; Glorfeld, 1995; Longman, Cota, Holden, & Fekken, 1989). Turner (1998) noted that using the first eiganvalue as the cutoff leads to under-extraction and that a more accurate method is to recreate the cut-off for each factor. Accordingly, the common procedure has now been to take the average or 95th percentile of each eigenvalue in the random data and compare it to the eigenvalue of the corresponding factor in the data. Significant factors are those that produce eigenvalues greater than the average or 95th percentile of the random eigenvalues. This study will refer to use PA to indicate analysis with the mean and PA95 to indicate use of the 95th percentile. While there has been a good deal of theoretical discussion on these variations of PA, the authors are not aware of a study that has empirically compared PA to PA95.

Minimum Average Partial (MAP)

The MAP procedure was developed by Velicer (1976) and is based on the theory that the proper number of factors will explain the most systematic variance in a correlation matrix. Removing systematic variance removes the co-variance among items, thus decreasing the correlations among items. Once all the systematic variance has been extracted, removing further variance eliminates noise in the data, causing correlations to increase. Therefore, removing the proper number of factors from a correlation matrix produces the lowest possible correlations in the set of possible correlation matrices derived from partialing out factors. Accordingly, Velicer (1976) recommended that the cut-off for the proper number of factors be the number of factors which produced the smallest average squared partial correlation. Velicer et al. (2000) provide a more in-depth explanation of the mathematical theory behind MAP. Zwick and Velicer (1982, 1986) found MAP to be the second most accurate procedure behind PA. When it was incorrect, MAP tended to under-extract.

Salient Loadings Criteria (SL)

Wrigley (1960) has proposed that the number of factors ought to be determined by examining how the individual variables load on each factor. He purposed that the proper solution is the one in which every extracted and rotated factor contains at least two variables that load highest on that particular factor. This is derived through a series of factor analyses that begins with an intentional over-extraction and ends when the proper solution is found. Howard and Gordon (1963) have provided an applied example of this procedure. While little empirical research has been done to test this method Gorsuch

(1997) has noted its potential and recommended use of this procedure, especially in construction of assessments, as it corrects for low reliability of individual test items. A recent unpublished study conducted by the authors found this method to have comparable accuracy to MAP with less of a tendency to under-extract (Porritt & Boyd, unpublished data).

Velicer, Eaton and Fava (2000) have attempted to improve the accuracy of the MAP procedure by using partial correlations raised to the fourth power (MAP4) instead of squared partial correlations. According to their study, the original MAP procedure was accurate 95.2% of the time, while the MAP4 matched the accuracy of Parallel analysis at 99.6% (Velicer, Eaton, & Fava, 2000). A preliminary study of this method using respondent generated data failed to replicate these findings (Porritt & Boyd, unpublished data).

Porritt and Boyd (unpublished data) used empirical data to examine the performance of Kaiser, MAP, MAP4, PA, PA95 and salient loading criteria in the specific situation of higher order factor analysis, which involves small item to factor ratios. PA was found to be the most accurate procedure, correctly identifying factor structure in 95% of the samples. MAP correctly identified 77% of the factor structures, consistently under factoring when number of items to factors was low. The salient loading criterion had similar performance, correctly identifying 77% of the samples, but showed a relatively even balance between under and over extraction and performed well when items to factors ratio was low. Surprisingly, the Kaiser rule was as consistent as MAP and salient loadings criteria, accurately identifying 77% of the factor structures and over factoring by one in the other 23% of the samples. Consistent with past research, scree plots identified 66% of the proper structures with a tendency to over extract when the number of variables was high. These findings indicate that the majority of these number of factor methods are susceptible to at least one of the major determining causes of overall EFA accuracy (factor to variable ratio). It is likely that these methods will be affected by the same determining causes hypothesized by MacCallum et al. (1999).

Hypotheses

Past research indicates that all other methods, excluding the salient loadings criteria, are susceptible to error in small samples. As such, the salient loadings criteria may prove more suited to small sample EFA. It is hypothesized that the salient loadings criteria will have the highest levels of accuracy in small samples. It is also hypothesized that number of factors procedures will demonstrate the same influence patterns hypothesized by McCallum et al. (1999) namely that strength of model (communalities) will moderate the relationship between sample size and overdetermination, with high communalities compensating for low overdetermination and/or sample size, and overdetermination providing further protection against the effects of small sample size and thus creating a limited set of circumstances under which EFA may be used with small samples. More specifically, it is hypothesized that number of factors criteria will maintain accuracy over 95% when model strength or overdetermination is high.

CHAPTER TWO

METHOD

Data Generation

In order to more fully represent all sources of error (model error and sampling error) data were generated using the method described by Hong (1999). Population correlation matrixes containing model error were produced using the Hong (1999) adaptation of the Tucker-Koopman-Linn procedure (1969), which allows for correlations between all factors. In keeping with Hong's (1999) example, correlations among factors were set at .3 and a minor factor matrix of 50 successively less significant factors was generated using the MacCallum and Tucker (1991) method. Minor factors were scaled to represent 8% of the variance. The Wijsman (1959) transformation as described by Hong (1999) was used to generate sample matrixes. Detailed formulas for these methods can be found in Appendix A.

Following MacCallum et al.'s (1999) lead, nine population matrixes were calculated to cover the combination of three conditions for over determination and three conditions for model strength. Levels of communality included high (communalities = .6, .7, .8) wide (communalities = .2 - .8) and low (communalities = .2, .3, .4) with an approximately equal number of variables being assigned specific communalities within each condition (e.g., 3 variables load at .8, 3 variables load at .7 and 3 variables load at .6). Levels of overdetermination include a low variables condition (12 variables: 3 factors), a stable ratio condition (24 variables:3 factors) and a high factors condition (24 variables:8 factors). Ratios were determined by selecting the values on the De Winter et al. (2009) table that were closest to the original MacCallum et al. (1999) ratios (10:3, 20:3, 20:7), thus allowing for best comparison across both studies. Population matrices and their rotated solutions are displayed in Appendix B.

Sample correlation matrices were calculated using two sample sizes: a "recommended size" of 500 chosen because it is in agreement with most recommended cut offs for sample size, and a minimum sample size of 60 that mirrors MacCallum et al.'s (1999) lower bound. All data were generated using the R statistical package (R_Core_Team, 2012). An annotated version of the syntax used to perform all procedures can be found in Appendix C.

Procedures

All procedures were carried out using the R statistical Package (R Core Team, 2012). *The Kaiser Rule* was implemented by counting the number of eigenvalues derived from the sample matrix that were greater than one. *Parallel Analysis* values for the 50th and 95th percentiles were derived using the "parallel" function found in the 'nFactors' package (Raiche, 2010). These were then compared to the sample eigenvalues and the proper determination was made at the point the random eigenvalues exceeded sample eigenvalues. Partial correlations for Velicer's *Minimum Average Partial* procedures were obtained using the "Very Simple Structure" (VSS) function found in the "psych" package (Revelle, 2013). Fourth power correlations were calculated by multiplying the matrix of squared partial correlations by itself. Both matrixes (squared and fourth power) were then evaluated to determine the lowest correlation

Original R code was written for the *Salient Loadings Criteria.* The algorithm performs a series factor analyses using maximum likelihood extraction with varimax rotation. The series begins by extracting twice the known number of factors, and continues by iteratively reducing the number of factors by one until the analysis yields a satisfactory structure. Satisfactory structures were defined as a set of factors each of which contained at least two items loading at .5 or higher, or three items loading at .4 or higher. Also, none of the salient items were allowed to cross-load on another factor within .1 of the salient loading. If the initial extraction (twice the known number of factors), was satisfactory then the number of factors was iteratively increased by one until an unsatisfactory solution was reached. The last number of factor producing a correct solution was considered correct. An annotated version of the syntax used to perform all procedures can be found in Appendix B.

Evaluation Criteria

The results were evaluated using similar criteria as Velicer, Eaton, and Fava (2000) who used percent correct, and mean difference to measure overall accuracy, and amount of discrepancy, respectively. Deviation of the difference was not included in this study as its non-normal distribution violated the ANOVA techniques assumption of normality and standardization would render the data uninformative for a comparison of means.

Percent Correct indicates the overall accuracy of a technique. Solutions were considered correct when they identified the known number of factors and incorrect when they failed to do so. Individual answers were scored as 1 for correct and 0 for incorrect. The average of this variable represents the percentage of correct answers.

Difference from Correct examines the amount of discrepancy in a technique as defined by how far off the technique estimates which is determined by subtracting the known number of factors from the proposed number of factors. Negative values indicate under-extraction, positive values indicate over-extraction and values of zero indicate a correct answer. Annotated R syntax for the creation of these outcome variables can be found in Appendix B.

CHAPTER THREE

RESULTS

Effects on each outcome variables were examined using a separate 2 x 6 x 3 x 3 between-subjects analysis of variance procedures for each outcome measure. Independent variables were entered in the following order: sample size (low and recommended), method choice (Kaiser, MAP, MAP4, PA, PA95, and salient loading), strength of the model (low, wide, and high), and overdetermination (variable to factor ratios including: 10:3, 24:3, and 24:8). The large sample size provided outsized power and all differences were statistically significant. As a result practical significance, as represented by effect size, will be the focus for this study. All results representing more than 1% or the variance in the data, as defined by an eta squared greater than .01, were considered interpretable.

When examining the outcome variable of percent correct, all main effects and interactions were statistically significant. Model strength (*Means*: Low = 6.12, Wide = 38.83, High = 58.92; F (1, 47612) = 28660.57, $p < .000$, $\eta^2_p = 0.376$) and sample size (*Means:* Low = 21.54, High = 47.71; F (1, 47612) = 3882.03, $p < .000$, $\eta^2_p = 0.075$) were the only main effects with interpretable effect sizes (as defined by $\eta^2_p \geq .01$); full ANOVA results can be seen in Table 1, individual cell means are displayed in Table 2.

Table 1

Results of Factorial ANOVA for Percent Correct

	df	Sum Sq	Mean Sq	*F*	*p*	η^2_p
Sample Size	1	508	508	3882	<.000	0.075*
Method	5	28	28	214	<.000	0.004
Model Strength	2	3751	3751	28661	<.000	0.376*
Overdetermination	2	8	8	63	<.000	0.001
Sample Size x Method	5	100	100	763	<.000	0.016*
Sample Size x Model Strength	2	134	134	1022	<.000	0.021*
Method x Model Strength	10	10	10	74	<.000	0.002
Sample Size x Overdetermination	2	27	27	203	<.000	0.004
Method x Overdetermination	10	162	162	1240	<.000	0.025*
Model Strength x Overdetermination	4	46	46	348	<.000	0.007
Sample Size x Method x Model Strength	10	15	15	114	<.000	0.002
Sample Size x Method x Overdetermination	10	43	43	330	<.000	0.007
Sample Size x Model Strength x Overdetermination	4	30	30	227	<.000	0.005
Method x Model Strength x Overdetermination	20	21	21	159	<.000	0.003
Sample Size x Method x Model Strength x Overdetermination	20	27	27	205	<.000	0.004
Residuals	47520	6231	13			0.559

Note. * = Interpretable effect size, $\eta^2_p > .01$.

Table 2

Individual Cell Means for Percent Correct

		Low Model Strength			Wide Model Strength			High Model Strength			
Overdetermination		12:3	24:3	24:8	12:3	24:3	24:8	12:3	24:3	24:8	Total
High											
Sample Size	Kaiser	0	0	0	14.51	0	2.04	100	100	100	35.17
	PA	50.57	56.23	9.52	96.15	99.77	24.94	100	100	100	70.80
	PA95	4.54	6.35	2.94	83.45	97.96	42.85	100	100	100	59.79
	MAP	0	0	0	0	100	0	98.87	100	0	33.21
	MAP4	0	0	0	0	100	0	98.87	100	0	33.21
	SL	0	0	0	87.3	100	0	99.55	99.77	100	54.07
Low											
Sample Size	Kaiser	0	0	0	1.13	0	5.44	90.7	48.07	65.76	23.46
	PA	8.16	0	8.84	44.44	0.002	1.81	90.02	11.34	1.13	18.42
	PA95	0	0	0	10.43	0	6.12	92.29	0	30.83	15.52
	MAP	0	0	0	1.14	98.41	0	82.09	100	0	31.29
	MAP4	0	0	0	1.14	98.41	0	82.09	100	0	31.29
	SL	0.01	1.81	0	43.99	93.88	0	95.69	100	36.51	41.32

Three small yet interpreteable interactions were observed. Sample size and method ($F(1, 47612) = 763.46$, $p < .001$, $\eta^2_p = 0.016$), interacted such that MAP and MAP4 showed little to no effect of sample sizes, salient loadings criteria and Kaiser showed a moderate effect of sample size (approximately 12% decrease in percent correct when sample size was low), and PA and PA95 showed significant effect of sample size (approximately 45% decreases in percent correct when sample size was low), see figure 1.

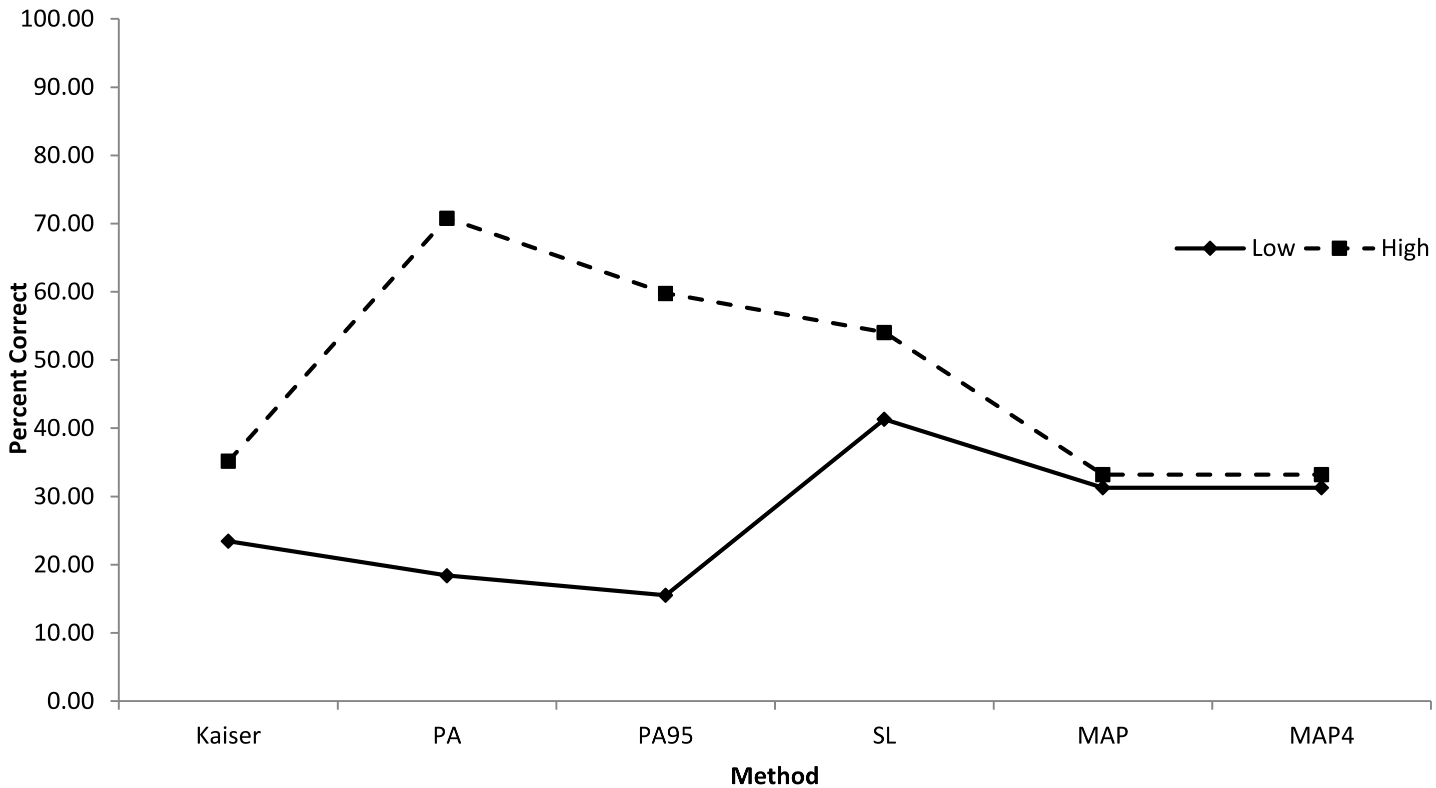

Figure 1. Interaction of sample size and method for percent correct

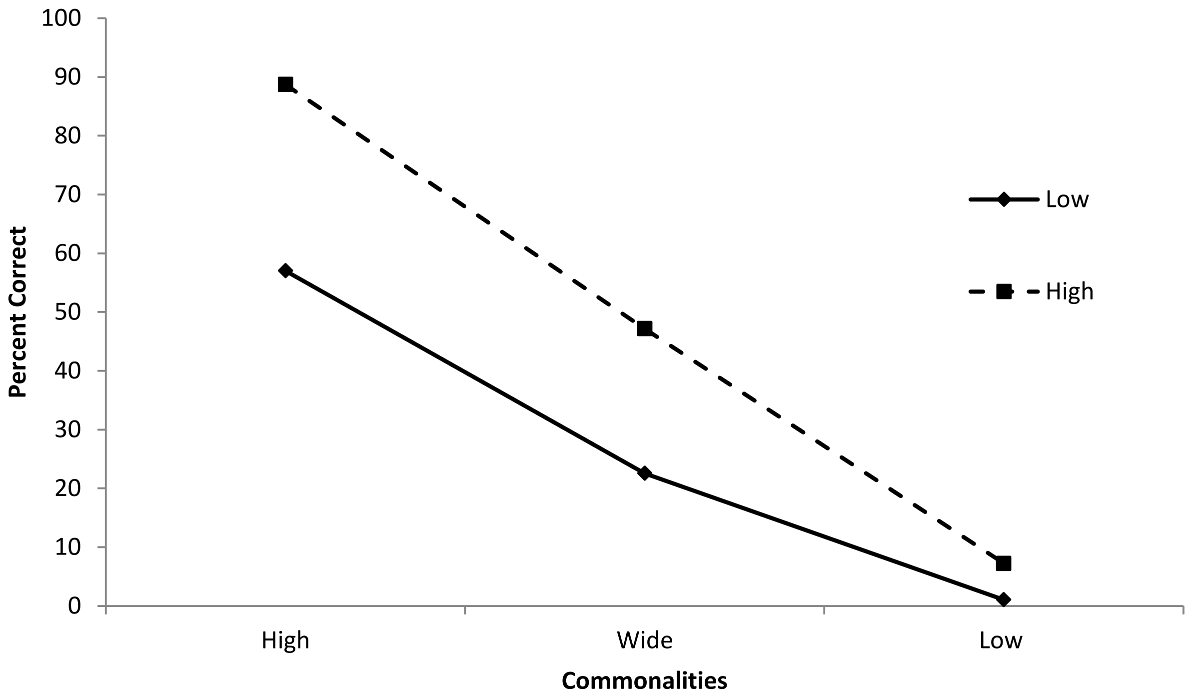

Figure 2. Interaction of sample size and model strength for percent correct

Sample size and model strength (F(1, 47612) =1022.24, $p < .001$, $\eta^2_p = 0.021$), interacted such that the deleterious effect of low sample size was reduced as model strength decreased, see figure 2.

Finally, method and overdetermination (F(1, 47612) =1240.23, $p < .001$, η^2_p = 0.026) interacted such that Kaiser, PA and PA95 were most correct with low variables and moderate factors (12:3), and MAP, MAP4, and Salient Loadings Criteria were most accurate with high overdetermination (24:3), see Figure 3. While statistically significant all other interactions were not interpretable ($\eta^2_p < .01$), see table 1.

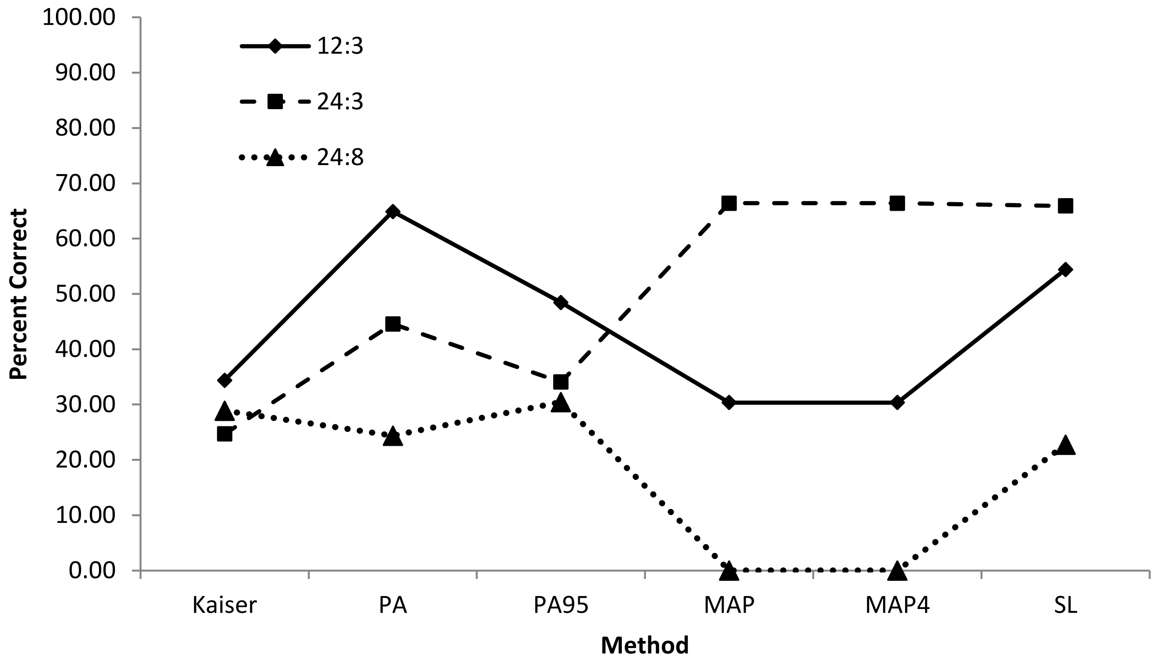

Figure 3. Interaction of overdetermination and method

Effects on difference from correct were examined using a 2 x 6 x 3 x 3 factorial ANOVA. Independent variables were entered in the following order: sample size (low and recommended), method (Kaiser, MAP, MAP4, PA, PA95, and salient loading), strength of the model (low, wide, and high), and overdetermination (10:3, 24:3, and 24:8). All main effects and interactions were statistically significant. Main effects were interpretable for sample size (*Means:* low = -0.02, High = -0.55; F (1, 47612) = 510, $p < .000$, $\eta^2_p = 0.011$), method (*Means*: Kaiser = 2.18, PA = 0.79, PA95 = 3.46, MAP = -3.01, MAP4 = -3.01, SL = -2.12; F (1, 47612) = 15100, $p < .000$, $\eta^2_p = 0.241$), model strength (*Means:* Low = -0.75, Wide = 0.05, High = -0.16; F (1, 47612) = 972, $p < .000$, $\eta^2_p = 0.020$), and overdetermination (*Means:* 12:3 = -0.10, 24:3 = 2.12, 24:8 = -2.88; F (1, 47612) = 978, $p < .000$, $\eta^2_p = 0.020$). Full ANOVA results can be seen in table 3 individual cell means are displayed in table 4.

Table 3

Results of Factorial ANOVA for Difference from Correct

	df	Sum Sq	Mean Sq	*F*	*P*	η^2_p
Sample Size	1	5500	5500	510	<.000	0.011*
Method	5	162909	162909	15100	<.000	0.241*
Model Strength	2	10492	10492	972	<.000	0.020*
Overdetermination	2	10546	10546	978	<.000	0.020*
Sample Size x Method	5	1683	1683	156	<.000	0.003
Sample Size x Model Strength	2	1886	1886	175	<.000	0.004
Method x Model Strength	10	68414	68414	6341	<.000	0.118*
Sample Size x Overdetermination	2	4614	4614	428	<.000	0.009
Method x Overdetermination	10	8927	8927	827	<.000	0.017*
Model Strength x Overdetermination	4	3460	3460	321	<.000	0.007
Sample Size x Method x Model Strength	10	644	644	60	<.000	0.001
Sample Size x Method x Overdetermination	10	1424	1424	132	<.000	0.003
Sample Size x Model Strength x Overdetermination	4	560	560	52	<.000	0.001
Method x Model Strength x Overdetermination	20	3938	3938	365	<.000	0.008
Sample Size x Method x Model Strength x Overdetermination	20	296	296	27	<.000	0.001
Residuals	47520	513673	11			0.643

Note. *= Interpretable effect size, $\eta^2_p > .01$.

Table 4

Individual Cell Means for Difference from Correct

Overdetermination		Low Model Strength			Wide Model Strength			High Model Strength			Total
		12:3	24:3	24:8	12:3	24:3	24:8	12:3	24:3	24:8	
High											
Sample Size	Kaiser	1.91	6.46	2.64	0.93	2.78	1.5	0	0	0	1.80
	PA	0.21	0.56	-2.39	0.04	7.48	-1.07	0	0	0	0.54
	PA95	2.70	2.40	5.34	0.17	11.44	0.03	0	0	0	2.45
	MAP	-2.00	-2.00	-7.00	-2.00	0.10	-7.00	-0.02	0	-7.00	-2.99
	MAP4	-2.00	-2.00	-7.00	-2.00	0.10	-7.00	-0.02	0	-7.00	-2.99
	SL	-3.00	-2.98	-8.00	-0.13	0.34	-4.97	-0.01	0	0	-2.08
Low											
Sample Size	Kaiser	2.37	7.21	2.30	1.61	4.97	1.43	0.09	0.55	-0.14	2.27
	PA	2.18	10.32	2.46	-0.07	4.44	-3.54	-0.10	1.00	-3.28	1.49
	PA95	7.95	14.57	13.87	2.22	8.54	4.15	0.04	2.84	-0.65	5.95
	MAP	-2.00	-1.98	-6.99	-1.81	-0.02	-6.83	-0.28	0	-6.42	-2.93
	MAP4	-2.00	-1.98	-6.99	-1.81	-0.02	-6.83	-0.28	0	-6.42	-2.93
	SL	-2.52	-2.23	-7.05	-0.6	0.06	-5	0.04	0	-0.96	-2.03

Significant interactions were observed for method and model strength (*F* (1, 47612) = 6341, $p < .000$, $\eta^2_p = 0.118$), and method and overdetermiantion (*F* (1, 47612) = 827, $p < .000$, $\eta^2_p = 0.017$). Method interacted with model strength such that Kaiser, PA, and PA95 tended to overestimate the number of factors while MAP, MAP4 and salient loadings criteria tended to underestimate the number of factors, see figure 4.

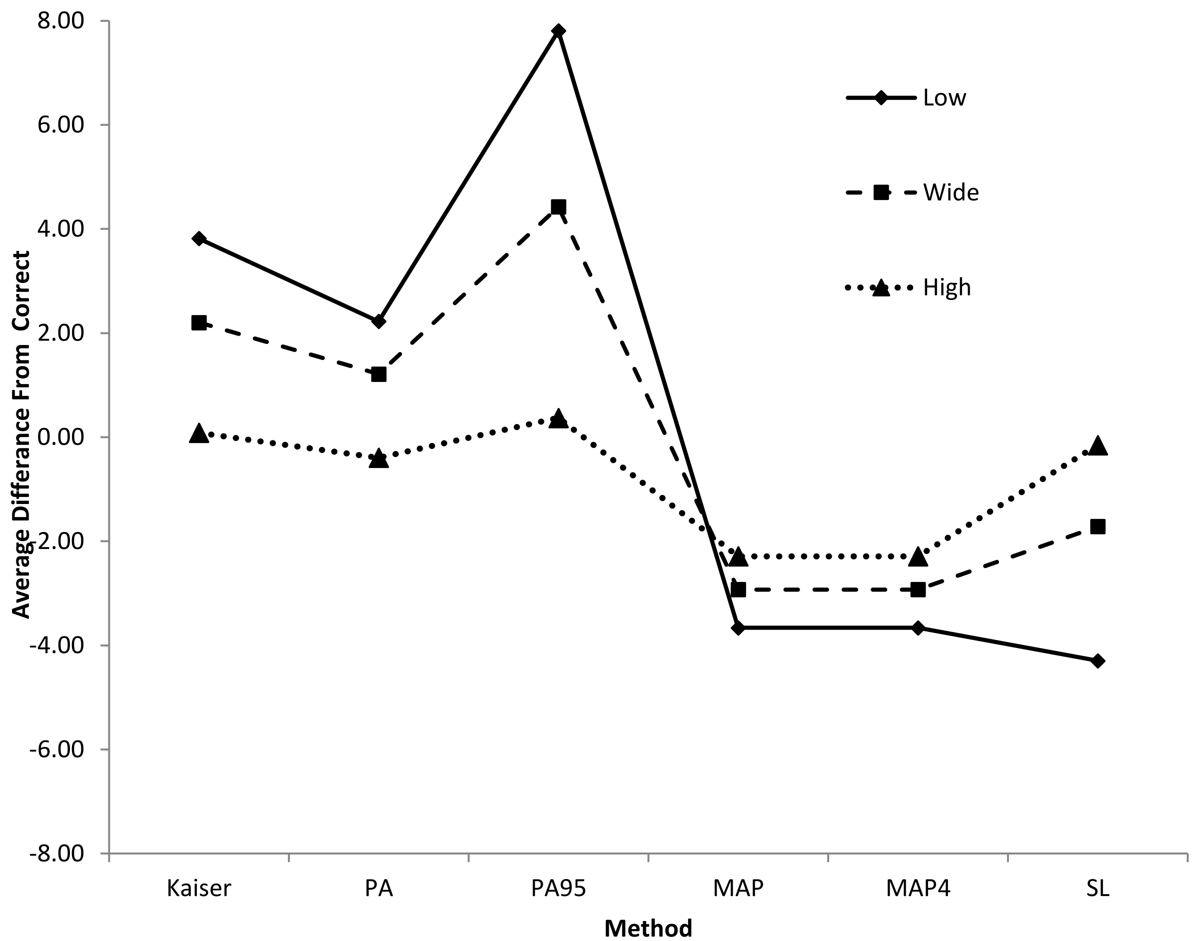

Figure 4. Interaction of method and model strength for difference from correct. Negative difference represent underestimations.

Method and overdetermination interacted such that high overdetermination was detrimental to discrepancy with Kaiser, PA, PA95, ware as MAP, MAP4, and salient loading criteria provided the most accurate estimates with high overdetermination and saw the greatest errors when the number of factors increased in relation to the number of variables, see figure 5.

Post-hoc pairwise planned comparisons with Bonferroni correction of familywise error were used to examine differences between specific methods, all contrasts were significant with the exception of MAP versus MAP4 (difference = 0, *SE* = .013, *p* = 1.00), see table 5. The results of PA were the closest to correct, followed by

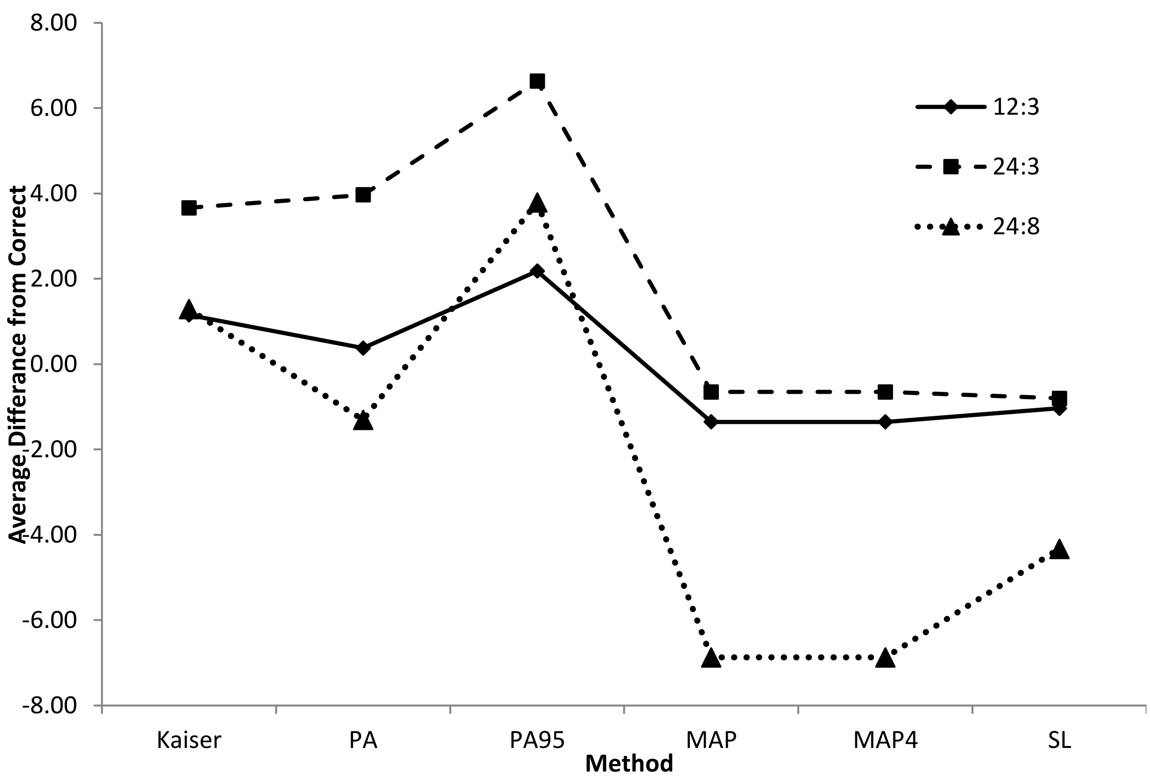

Figure 5. Interaction of method and overdetermination for average difference. Negative difference represent underestimations.

salient loading criteria, Kaiser, both versions of MAP and PA95. Additionally, PA, PA95 and Kaiser tended to over factor while MAP, MAP4 and salient loadings tended to under extract, see means in Table 4.

Table 5

Pairwise Comparison of Methods for Difference from Correct

	diff.	*SE*	*p*
Kaiser – PA	1.47	0.013	<.000
PA - PA95	2.97	0.013	<.000
PA95 – MAP	6.5	0.013	<.000
MAP - MAP4	0	0.013	1
MAP4 - Salient Loading	-0.89	0.013	<.000

Note. Negative differences represent underextractions

CHAPTER FOUR

DISCUSSION

Sample size remains a major predictor of technique accuracy and seeking the largest possible sample size is still the best practice. Nevertheless, it would appear that accurate factor extraction in samples as small as 60 is possible given the proper conditions. Several studies have demonstrated accurate factor extraction in small sample sizes (de Winter et al., 2009; MacCallum et al., 2001; MacCallum et al., 1999; Mundfrom et al., 2005); however, none of these studies examined the accuracy of methods for determining the appropriate number of factors to extract. The results of the current study indicate that a subset of these methods may be cautiously applied in small sample size situations.

It would appear that, at least in terms of model strength, methods for determining the proper number of factors to extract are subject to similar influences as those observed for factor analysis on the whole. More specifically, increases in model strength lead to increases in the accuracy of all methods for determining the number of factors to extract. The most effective thing a researcher can do to yield proper results is to ensure that he/she specifies the most accurate model possible.

The effects of overdetermination are more complex to determine as they appear to be moderated by method choice. Item level methods such as MAP, MAP4 and salient loadings criteria, demonstrated the expected relationship in which overdetermiantion improves accuracy and protects against the deleterious effects of small sample sizes. On the other hand, eigenvalue-dependant methods such as Kaiser, PA, and PA95, performed best when item to factor ratio was moderate, with number of factors low in relation to number of items. Contrary to the expected pattern, accuracy of these methods decreased as number of variables increased in relation to factors. It is unclear what the specific mechanism for this relationship is, however; it is likely that the increase in number of variables increases the probability that a group of variables belonging to one factor will create a subfactor that will be produce a significant eigenvalues. This theory is corroborated by the fact that the methods in question tended to over-extract by at least two factors under the high overdetermination condition. Use of eigenvalue-dependent methods under low sample conditions cannot be recommended at this time, as high overdetermination is necessary to ensure accuracy when sample size is low. With that caveat it is important to recognize that eigenvalue methods, PA in particular, are highly accurate when the sample size complies with current recommendations. Indeed, PA was the most accurate method under most conditions and the only method that performed satisfactorily when model strength was moderate and variables were low in relationship to factors.

Accuracy was also examined by looking at technique discrepancy or the difference between the provided answer and the correct answer. Method was the most important predictor of discrepancy. Examination of individual methods indicates that PA provided the least discrepant estimates tending to over-extract by an average of one factor. Kaiser and salient loading criteria, were discrepant by an average of approximately 2 and -2, respectively, indicatin that errors with Kaiser are likely to be less detrimental to overall accuracy of estimate. MAP and MAP4 (off by an average of - 3 factors) and PA95 (off by an average of approximately 4 factors) were the most discrepant methods.

Method was also a significant moderator of the other effects. Eigenvalue methods and item level methods again. In general eigenvalue methods tended to over factor and item level methods tending to under factor. This is of particular importance to researchers using item level methods when sample sizes are low, as this tends to exacerbate the negative effects of under-extracting (Wood et al., 1996) which are already considered to be worse than the effects of over-extracting. The negative impact of overdetermination on performance of eigenvalue methods was also apparent with this analysis as discrepancies for these methods tended to be greatest when overdetermination was high.

Another point of interest for this study was to elucidate the difference in performance between PA and PA95. PA appears to outperform PA95. PA was more accurate and less discrepant across all conditions, and was applicable to more conditions than PA95. However, more replication is needed before concrete recommendations are made.

The focus of this study has been to provide a proof of concept for the small sample size use of empirical methods for determining the proper number of factors. It would appear that a select set of these methods, namely the item level methods of MAP, MAP4, and salient loadings criteria, can produce accurate results when sample sizes are low. However it must be emphasized that all methods performed best under higher sample sizes conditions and that small sample size analysis should be viewed as a special exception to the general rule of large sample analysis. It should also be noted that no one method is a “silver bullet” which will constantly provide the correct answer. As such, researchers are advised to seek guidance from multiple methods at once. Specific details and recommendations for appropriate points of application for each method are displayed in table 6.

Table 6

Recommendations for Use of Methods

		Wide Model Strength			High Model Strength		
Overdetermination		12:3	24:3	24:8	12:3	24:3	24:8
High							
Sample Size	Kaiser				P	P	P
	PA	P	R		P	P	P
	PA95		R		P	P	P
	MAP		P		R	P	
	MAP4		P		R	P	
	SL		P		R	R	P
Low							
Sample Size	Kaiser						
	PA						
	PA95						
	MAP		P			P	
	MAP4		P			P	
	SL		R		P	P	

Note. Low model strength condition was not included as none of the methods met minimum requirement of 95% correct

R = Recommended (accuracy > 95%)

P = Preferred (Highest accuracy for conditions)

The primary limitation of this study is the use of Monte Carlo data. While we have taken every known step to generate complex data similar to that encountered in respondent generated data, it is likely that there are additional nuances and complexities we were unable to capture. It must also be mentioned that caution should be used when attempting to apply these findings to data outside of the bounds of the study parameters.

Having established the viability of empirical methods for determining the proper number of factors to extract, there are a number of directions for future research. Past research ((de Winter et al., 2009; Mundfrom et al., 2005)) has established lower bounds for accurate factor analysis that are often far lower than N = 60. Further research is needed to establish similar lower bounds for the accuracy of methods used to determining the proper number of factors to extract. Further exploration is also needed to determine the reason overdetermination is detrimental to eigenvalue methods and to explore options for ameliorating these effects. Based upon our hypothesis regarding subfactors, one may consider exploring the possibility of creating a combination between an eigenvalue and item level method, perhaps PA with a minimum item limit for a factor to be considered significant. Finally, it would be beneficial to replicate these findings with respondent-generated data.

REFERENCES

Buja, A., & Eyuboglu, N. (1992). Remarks on parallel analysis. *Multivariate Behavioral Research, 27*(4), 509-540.

Cattell, R. B. (1978). *The scientific use of factor analysis in behavioral and life sciences*. New Yourk, NY US: Plenum Press.

Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis (2nd ed.)*. Hillsdale, NJ England: Lawrence Erlbaum Associates, Inc.

De Winter, J. C. F., Dodou, D., & Wieringa, P. A. (2009). Exploratory factor analysis with small sample sizes. *Multivariate Behavioral Research, 44*(2), 147-181. doi: 10.1080/00273170902794206

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

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

Fava, J. L., & Velicer, W. F. (1992). The effects of overextraction on factor and component analysis. *Multivariate Behavioral Research, 27*(3), 387-415.

Fava, J. L., & Velicer, W. F. (1996). The effects of underextraction in factor and component analyses. *Educational and Psychological Measurement, 56*(6), 907-929.

Glorfeld, L. W. (1995). An improvement on Horn's parallel analysis methodology for selecting the correct number of factors to retain. *Educational and Psychological Measurement*, *55*(3), 377-393.

Gorsuch, R. L. (1980). Factor score reliabilities and domain validities. *Educational and Psychological Measurement*, *40*(4), 895-897.

Gorsuch, R. L. (1983). *Factor Analysis (2nd Ed.).* Hillsdale, NJ US: Lawrence Erlbaum Associates, Inc., Publishers.

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 to sample size to the stability of component patterns. *Psychological Bulletin, 103*(2), 265-275. doi: 10.1037/0033-2909.103.2.265

Hayton, J. C., Allen, D. G., & Scarpello, V. (2004). Factor Retention Decisions in Exploratory Factor Analysis: A Tutorial on Parallel Analysis. *Organizational Research Methods*, *7*(2), 191-205.

Hong, S. (1999). Generating correlation matrices with model error for simulation studies in factor analysis: A combination of the Tucker-Koopman-Linn model and Wijsman's algorithm. *Behavior Research Methods, Instruments & Computers, 31*(4), 727-730. doi: 10.3758/bf03200754

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*(2), 179-185.

Howard, K. I., & Gordon, R. A. (1963). Empirical note on the 'number of factors' problem in factor analysis. *Psychological Reports*, *12*(1), 247-250.

Humphreys, L. G., & Montanelli, R. G. (1975). An investigation of the parallel analysis criterion for determining the number of common factors. *Multivariate Behavioral Research, 10*(2), 193-205.

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

Lee, H. B., & Comrey, A. L. (1979). Distortions in a commonly used factor analytic procedure. *Multivariate Behavioral Research, 14*(3), 301-321.

Longman, R. S., Cota, A. A., Holden, R. R., & Fekken, G. C. (1989). A regression equation for the parallel analysis criterion in principal components analysis: Mean and 95th percentile eigenvalues. *Multivariate Behavioral Research, 24*(1), 59-69.

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. doi: 10.1037/0033-2909.109.3.502

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*(4), 611-637. doi: 10.1207/s15327906mbr3604_06

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, *4*(1), 84-99. doi: 10.1037/1082-989x.4.1.84

Mundfrom, D. J., Shaw, D. G., & Ke, T. L. (2005). Minimum Sample Size Recommendations for Conducting Factor Analyses. *International Journal of Testing*, *5*(2), 159-168. doi: 10.1207/s15327574ijt0502_4

R_Core_Team. (2012). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing ISBN 3-900051-07-0, URL http://www.R-project.org/.

Raiche, G. (2010). *nFactors: an R package for parallel analysis and non graphical solutions to the Cattell scree test*: R package version 2.3.3.

Revelle, W. (2013). *psych: Procedures for Personality and Psychological Research, Northwestern University*, . Evanston, Illinois, USA,: http://CRAN.R-project.org/package=psych Version = 1.3.2.

Rouquette, A., & Falissard, B. (2011). Sample size requirements for the internal validation of psychiatric scales. *International Journal of Methods in Psychiatric Research, 20*(4), 235-249. doi: 10.1002/mpr.352

Thurstone, L. L. (1947). *Multiple-factor analysis; a development and expansion of The Vectors of Mind*. Chicago, IL US: University of Chicago Press.

Tucker, L. R., Koopman, R. F., & Linn, R. L. (1969). Evaluation of factor analytic research procedures by means of simulated correlation matrices. *Psychometrika, 34*(4), 421-459.

Turner, N. E. (1998). The effect of common variance and structure pattern on random data eigenvalues: Implications for the accuracy of parallel analysis. *Educational and Psychological Measurement, 58*(4), 541-568.

Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika, 41*(3), 321-327.

Velicer, W. F., Eaton, C. A., & Fava, J. L. (2000). Construct explication through factor or component analysis: A review and evaluation of alternative procedures for determining the number of factors or components. In R. D. Goffin & E. Helmes (Eds.), *Problems and solutions in human assessment: Honoring Douglas N. Jackson at seventy.* (pp. 41-71). New York, NY US: Kluwer Academic/Plenum Publishers.

Widaman, K. F. (2012). Exploratory factor analysis and confirmatory factor analysis. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf & K. J. Sher (Eds.), *APA handbook of research methods in psychology, Vol 3: Data analysis and research publication.* (pp. 361-389). Washington, DC US: American Psychological Association.

Wijsman, R. A. (1959). Applications of a certain representation of the Wishart matrix. *Annals of Mathematical Statistics, 30*(2), 597-601. doi: 10.1214/aoms/1177706276

Wood, J. M., Tataryn, D. J., & Gorsuch, R. L. (1996). Effects of under- and overextraction on principal axis factor analysis with varimax rotation. *Psychological Methods, 1*(4), 354-365.

Wrigley, C. (1960). A procedure for objective factor analysis. *Paper presented at the Society of Multivariate Experimental Psychology., 41*(3), 321-327.

Zwick, W. R., & Velicer, W. F. (1982). Factors influencing four rules for determining the number of components to retain. *Multivariate Behavioral Research, 17*(2), 253-269.

Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin, 99*(3), 432-442.

APPENDIX A

FORMULAS FOR DATA GENERATION

Formulas for Generation of Population Correlation Matrix

Hong (1999) described the Tucker-Koopman-Linn (1969) method as follows:

$$P = VV' + D + WW'$$

Where:
P = Population correlation matrix (order k, where k is # of variables)
V = Major factor loading matrix (k X r, where r is # of factors)
D = Diagonal matrix of unique factor variances (k X k)
W = Unique factor loading matrix (k X q, where q = # of minor factors)

Hong (1999) made the following alterations in order to include correlations between factors (both minor and major). This alter version of the formula is the one used in this study.

$$P = JBJ' + D$$

Where:
J = Super Loading Matrix [V,W] (k X (r + q))
B = Matrix of Factor Correlations ((r + q) X (r + q))
D = Diagonal Matrix of Unique Factor Variances (k X k)

The MacCallum and Tucker (1991) method for generating a W matrix is as follows: Random factor loadings for the first factor are gained from a random distribution with a mean of 0 and a standard deviation of 1the standard deviation of the population is successively reduced by .8 for each factor. The matrix is than rescaled by row so that the minor factors account for the desired amount of variance.

Formulas for Sample Correlation Matrix Generation

Population correlation matrixes were calculated using the Wijsman (1959) procedures. Hong (1999) described this procedure as follows:

Generate a matrix A as follows:

$$A = FGG'F'$$

Where:
F = k X k factor matrix of population correlation matrix such that FF' = P
G = Randomly generated lower right triangle matrix. Off diagonal elements are random diviates drawn from a normal distribution with a mean of 0 and variance of 1.Diagonal elements are positive square roots of random values drawn from chi-square distributions with degrees of freedom of n – j where n= sample size[1] and j= column number.

Calculate Sample Covariance Matrix :

$$C = (1/n)A$$

Where
C = Sample Covariance Matrix
n = Sample size
A = As calculated above

Calculate Sample Correlation Matrix:

$$R = D^{-1/2}CD^{-1/2}$$

Where:
R = Sample correlation matrix
D = Diagonal matrix containing diagonal elements of C (sample variances)

[1] (N * 2) was substituted in instances where number of variables exceeded n. (low sample conditions for the high communality of the 12:3 and 24:3 item to factor ratios)

APPENDIX B

ANNOTATED SYNTAX

All syntax is original work of Marc Porritt, please cite this dissertation for any reuse. For questions contact the author at marc.porritt@gmail.com

Comments found after the “//”

Syntax used to generate population correlation matrixes currently set for 12 variables into three factors with low communality:

```
nVars = 12 // Set the number of files
popFile = "C:/data/PopulationCorrelations/12-3-L.txt"//destination for output

// creates B matrix for use Hong's alteration to Tucker-Koopman-Linn
B= matrix(.3, nrow=53, ncol = 53)
diag(B) = 1

// Input theoretically established major factor loading matrix
V=matrix(c(.4, .3, .3, .2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, .4, .3, .3, .2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, .4, .3, .3,
.2 ), nrow = 12)

// Randomly generates minor factor loading matrix of 50 factors
W = matrix(c(rnorm(nVars,mean=0,sd=1)))
sd=1
for(i in 1:49)//first loop creates matrix with random variables
{
  W2 = matrix(c(rnorm(nVars,mean=0,sd= sd)))
  W = cbind(W,W2)
  sd= sd*.8
}
for(i in 1:nVars)// second loop rescales the rows
{
  Rsum = sum(abs(W[i,]))
  for(j in 1:50){W[i,j]= (W[i,j]/Rsum)*.08}
}

// uses matrixes instantiated above to carry out Hong's version of Tucker-//Coopman-Linn

J =  cbind(V, W)

JBJ= J%*%B%*%t(J)

D = diag(nVars)
diag(D) = diag(D)-diag(JBJ)

P= JBJ+D

//writes matrix to file specified above
write.table(P,popFile, sep="\t")
```

Syntax used to generate sample matrixes and carry out methods on them:

```
Nvars = 12			// number of variables
Nfacs = 3		// number of factors
sampleN=353			// Sample size
strength = 1		//codes for strength of communality
Nsets = 441			// designates the number of samples to be generated


//destination file for generated data set
Master =  "C:/Users/mporritt/Documents/Dissertation/data/12-3-L-L.txt"
//destination file for generated sample correlation matrixes
dataFile = "C:/Users/mporritt/Documents/Dissertation/data/12-3-L-L/12-3-L-L-"
// extension for sample matrix files
extension = ".txt"

//instantiates the matrix that will hold the generated data ( ultimately this //data set will contain variables for
the number of factors, number of //variables, strength of communalities, specific sample data came from,
//method used to get the answer, and the answer produced by the method)
dataFrame = matrix(0,6, nrow=1)
colnames(dataFrame) = c("Nvars", "Nfactors", "LoadStrength","setNumber","Method", "Solution")

//F matrix as determined by Maximum Likelihood extraction of 24 factors from //the population correlation
matrix
F = matrix(c(
.222,.385,.065,.664,-.545,-.223,-.019,.005,-.014,.101,-.022,-.011,
.235,-.236,.308,.494,.671,-.293,.018,-.004,-.010,.020,.055,.102,
.246,-.136,-.355,.392,.100,.788,-.013,-.010,.027,.007,.084,-.054,
.331,-.286,.373,-.102,-.249,.129,.690,-.031,-.020,.046,.158,.277,
.357,-.278,.359,-.128,-.215,.103,-.653,.047,.063,.054,.208,.337,
.353,.449,.075,-.175,.188,.094,.053,.679,.014,.357,-.031,-.008,
.362,.443,.073,-.186,.176,.078,-.024,-.665,.015,.390,-.023,.006,
.363,-.162,-.418,-.070,-.033,-.280,.060,.005,.689,.042,.286,-.148,
.389,-.156,-.402,-.094,-.031,-.233,-.020,.009,-.649,.059,.383,-.164,
.413,.472,.079,-.090,.103,.042,.013,-.019,.008,-.748,.130,.056,
.441,-.165,-.427,-.047,-.017,-.136,.001,.003,-.058,-.059,-.641,.397,
.450,-.276,.355,-.070,-.096,.042,-.048,-.006,-.007,-.076,-.355,-.667
), ncol = 12, byrow =TRUE)




// this loop computes a sample correlation matrix, performs all six methods, //and saves the answers from
each method along with identifying information.
//the loop will iterate as many times as was specified above with the Nsets //variable
for(setNum in 1:Nsets)
{
	//Generates a G matrix that meets Wijsman criteria.
	G=matrix(0, Nvars,Nvars)
	for(j in 1:Nvars){G[j,j] = abs(sqrt(rchisq(1,df=(sampleN-j))))}
	r=1
	for(k in 1:(Nvars-1))
	{
```

```
        r=r+1
    for(c in 1:k){G[r,c]=rnorm(1,mean=0,sd=1)}
}

    //Computes A matrix according to Wijsman method
    A=F%*%G%*%t(G)%*%t(F)

    //Sample covariance matrix
    CovA = (1/sampleN)*A

    //Scale covariance to correlation
    D=diag(diag(CovA)^(-1/2))
    Corr = D%*%CovA%*%D

    //corrects for difference due to rounding – ensures the correlation // matrix is indeed a mirror of
    itself
    for (i in 1:Nvars){for (j in 1:Nvars){ Corr[j,i] = Corr[i,j] }}

    //generates a unique file name for each sample matrix by concatanating //the name, location, and
    extension specified above with the counter //for this loop (setNum), writes the matrix to the file
    and prints the //file name – so as to provide a way for the user to keep track of //computations
    while the syntax is running
    fileName= paste(dataFile, setNum, extension, sep="")
    write.table(Corr,fileName, sep="\t")
    print(fileName)

    //this matrixcontains the identifying information that is common to all //elements of thi particular
    set of data (# variables, #factors, //communality strength, the specific sample matrix the data came
    //from.)an indentifyer for a method and the solution provided by that //method will be
    concatenated to this matrix before it is appended to //the master data matrix.
    dataSeed1 = matrix(c(Nvars,Nfacs,strength,setNum), nrow =1)

    //retrieves sample matrix eigenvalues and stores them to vector EV
    eigen = eigen(Corr, only.values=TRUE)
    EV = eigen$values


    //Performs Kaiser Rule
    K = 0
    for (i in 1:Nvars){if (EV[i] >= 1) {K = K + 1}}

    //adds Kaiser specific values to dataSeed and appends to data matrix
    dataSeed2 = matrix(c(1,K), nrow=1)
    Krow = cbind(dataSeed1,dataSeed2)
    dataFrame = rbind(dataFrame,Krow)

    //Retrieves Parallel Analysis values and stores mean values in EV_PAM //and 95th percentile
    values in EV_PA95
    PA = parallel(subject=sampleN,var=Nvars,rep=100,cent=.05)
    EV_PAM = PA$eigen$mevpea
    EV_PA95 = PA$eigen$qevpea

    //performs mean PA
    PAM = 0
    for (i in 1:Nvars){if (EV[i] >= EV_PAM[i]) {PAM = PAM + 1}}
```

```
	//adds mean PA specific values to dataSeed and appends to data matrix
	dataSeed2 = matrix(c(2,PAM), nrow=1)
	PAMrow = cbind(dataSeed1,dataSeed2)
	dataFrame = rbind(dataFrame,PAMrow)

	//performs PA95 and adds values to data
	PA95 = 0
	for (i in 1:Nvars){if (EV[i] >= EV_PA95[i]) {PA95 = PA95 + 1}}
	dataSeed2 = matrix(c(3,PA95), nrow=1)
	PA95row = cbind(dataSeed1,dataSeed2)
	dataFrame = rbind(dataFrame,PA95row)

	//obtains squared average partial correlations and calculates forth //power average partial
	correlations – stores each in a separate vector
	AP1 = VSS(Corr, n = Nvars-1, n.obs=sampleN, plot = FALSE)
	APs = AP1$map
	APs4= APs*APs

	//Loop determines minimum average partial correlation and breaks at the //first rise in values.
	Remaining code appends values to data
	MAP = 1
	for (i in 1:(length(APs)-1))
	{
		if (APs[i] > APs[i+1]){MAP = MAP+1}
		else{break}
	}

	dataSeed2 = matrix(c(4,MAP), nrow=1)
	MAProw = cbind(dataSeed1,dataSeed2)
	dataFrame = rbind(dataFrame,MAProw)

	//performs identical functions using fourth power average partails
	MAP4 = 1
	for (i in 1:(length(APs4)-1))
	{
		if (APs4[i+1] <= APs4[i]){MAP4 = MAP4+1}
		else{break}
	}

	dataSeed2 = matrix(c(5,MAP4), nrow=1)
	MAP4row = cbind(dataSeed1,dataSeed2)
	dataFrame = rbind(dataFrame,MAP4row)

	//performs Salient Loading Criteria and stores data
	SLanswer = 0

	SalientLoading()

	dataSeed2 = matrix(c(6,SLanswer), nrow=1)
	SLrow = cbind(dataSeed1,dataSeed2)
	dataFrame = rbind(dataFrame,SLrow)
}
//when the loop is complete this writes the data matrix to the master file //specified above with tab
delimitation
write.table(dataFrame, Master, sep="\t")
```

Salient Loading Criteria Syntax
Intellectual property of Marc Porritt – please cite this work for any reuse.

```
//sets the number of factors that will initially be extracted – currently set to twice the known factors
START = (round(Nfacs*2))

//Loop performs successive factor analysis and evaluates the results against Salient Loading Criteria it will
start by extracting START factors and continue by decrementing that number by 1 until criteria are met
for(g in START:1)
{
    if(SLanswer>0){break}//Stops loop and returns answer if previous answer was satisfactory

    //Performs maximum likelihood FA with verimax rotation and stores rotated solution in LDs
    FA = factanal(covmat = Corr, factors= g, n.obs = sampleN, rotation = "varimax")
    LDs = FA$loadings

    SalFacs = 0 // This is a counter for the number of factors that meet criteria

        //This loop evaluates the saved rotated loadings matrix to determine if it meets criteria
        for (i in g:1)
        {
            // this variable evaluates salience of >= 4 is threshold for meeting criteria
            factorValue=0

            //This loop evaluates each individual factor for salience on a variable by variable basis
            starting at the last factor extracted and working to the first

            for(j in Nvars:1)
            {
                crossLoad= 0 // True if a cross loads bellow threshold of .1
                CheckMe= LDs[j,i]//pointer for a given loading

                 //evaluates for cross loading sets crossLoad to true/false
                for (k in 1:(length(LDs)/Nvars)){if(abs(LDs[1,k]- CheckMe)<.1){crossLoad = crossLoad
                  + 1}}

                //determines strength of variables loading and updates factorValue according
                if(crossLoad<=1)
                {
                    if(CheckMe >= 0.4 && CheckMe <0.5) {factorValue = factorValue + 1.5}
                    if(CheckMe >= 0.5) {factorValue = factorValue + 2}
                }
            }

            // if any factors are determined to be non salient the evaluation stops and the next set of
                //factors are extracted and evaluated
            if(factorValue<4) {break}

            //otherwise the count of salient factors is incremented and the set of factors are evaluated
             ELSE
            {
                SalFacs = SalFacs+1
                if(SalFacs == g)// if all factors in the set are salient
                {
                    if(g== START)//if this is the first set of factors extracted
```

```
{
//the same process is repeated only this time the number of factors is incremented
//instead of decremented – this functions essentially like a recursive function //call –
therefore redundant code will be replaced by pseudo function calls
    for(g2 in (START):(START*2))
    {
        extractAndEvaluateFactorMatrix()

        //if the answer is still correct the loop stops and we move forward to the
        following evaluations
        if(SLanswer>0){break}

        //if the factor is not salient than the answer is no longer correct and //the
        previous number of factors (the last successful solution) is //returned and
        the process stops
        if(factorValue<4)
        {
            SLanswer=(g2-1)
            break
        }
        //otherwise if we have not evaluated the whole matrix but the factor is
        //salient than we increment the number of salient factors and evaluate //the
        next factor
        Else
        {
            SalFacs = SalFacs+1
            //if all factors have been evaluated than the number of factors is
            incremented by //1 and the process continues
            if(SalFacs == g2){break}
        }
    }
}

}
//otherwise(if the entire matrix is salient AND this is not the first set of factors
extracted) than the current number of factors is the solution and the entire process stops
Else
{
    SLanswer=SalFacs
    break
}
}
}
}
}
```

APPENDIX C

POPULATION CORRELATION MATRICES

Twelve Variables, Three Factors, High Communalities

1.000											
0.559	1.000										
0.569	0.488	1.000									
0.484	0.415	0.422	1.000								
0.194	0.159	0.168	0.142	1.000							
0.177	0.146	0.153	0.128	0.560	1.000						
0.166	0.136	0.143	0.120	0.550	0.487	1.000					
0.143	0.117	0.124	0.104	0.472	0.418	0.411	1.000				
0.197	0.162	0.171	0.143	0.188	0.171	0.161	0.139	1.000			
0.170	0.140	0.147	0.124	0.162	0.148	0.138	0.119	0.556	1.000		
0.168	0.138	0.146	0.123	0.161	0.146	0.137	0.118	0.555	0.484	1.000	
0.163	0.133	0.141	0.118	0.155	0.141	0.132	0.114	0.493	0.430	0.428	1.000

Varimax Rotated Maximum Likelihood Solution

	1	2	3
V9	.783	.113	.112
V10	.683	.097	.095
V11	.682	.095	.094
V12	.601	.099	.097
V1	.121	.790	.118
V3	.104	.690	.101
V2	.095	.680	.093
V4	.086	.587	.084
V5	.113	.112	.779
V6	.105	.105	.688
V7	.095	.094	.679
V8	.082	.082	.582

Twelve Variables, Three Factors Wide Communalities

CORR	1.00											
CORR	0.49	1.00										
CORR	0.33	0.25	1.00									
CORR	0.18	0.13	0.09	1.00								
CORR	0.21	0.15	0.11	0.06	1.00							
CORR	0.16	0.12	0.08	0.05	0.49	1.00						
CORR	0.11	0.08	0.06	0.03	0.33	0.25	1.00					
CORR	0.05	0.04	0.02	0.01	0.16	0.12	0.08	1.00				
CORR	0.22	0.16	0.11	0.06	0.21	0.16	0.11	0.05	1.00			
CORR	0.16	0.12	0.08	0.05	0.15	0.12	0.08	0.04	0.50	1.00		
CORR	0.09	0.07	0.05	0.03	0.09	0.07	0.05	0.02	0.32	0.24	1.00	
CORR	0.05	0.04	0.02	0.01	0.05	0.04	0.03	0.01	0.16	0.12	0.08	1.00

Varimax Rotated Maximum Likelihood Solution

	1	2	3
V1	.790	.121	.123
V2	.594	.088	.086
V3	.399	.060	.066
V4	.215	.035	.038
V9	.136	.794	.127
V10	.098	.598	.089
V11	.050	.386	.050
V12	.029	.192	.032
V5	.125	.116	.787
V6	.097	.091	.594
V7	.068	.064	.399
V8	.029	.028	.194

Twelve Variables, Three Factors, Low Communalities

1.000											
0.119	1.000										
0.121	0.093	1.000									
0.074	0.056	0.056	1.000								
0.050	0.039	0.041	0.019	1.000							
0.041	0.032	0.034	0.016	0.129	1.000						
0.036	0.028	0.029	0.014	0.123	0.095	1.000					
0.021	0.017	0.017	0.009	0.079	0.060	0.058	1.000				
0.054	0.042	0.045	0.021	0.059	0.050	0.043	0.024	1.000			
0.035	0.028	0.030	0.014	0.039	0.033	0.028	0.016	0.127	1.000		
0.031	0.024	0.025	0.012	0.033	0.027	0.024	0.014	0.120	0.087	1.000	
0.020	0.015	0.015	0.009	0.021	0.017	0.016	0.010	0.079	0.057	0.056	1.000

Varimax Rotated Maximum Likelihood Solution

	1	2	3
V5	.401	.060	.060
V6	.306	.055	.052
V7	.295	.043	.042
V8	.190	.021	.023
V9	.077	.405	.071
V10	.047	.296	.043
V11	.034	.284	.034
V12	.022	.188	.020
V1	.059	.055	.387
V3	.050	.048	.298
V2	.047	.043	.295
V4	.018	.018	.184

Twenty Four Variables, Three Factors, High Communality

1	1.000											
2	0.647	1.000										
3	0.653	0.672	1.000									
4	0.567	0.582	0.589	1.000								
5	0.551	0.565	0.570	0.495	1.000							
6	0.484	0.497	0.502	0.436	0.423	1.000						
7	0.485	0.497	0.503	0.437	0.423	0.372	1.000					
8	0.476	0.487	0.493	0.428	0.416	0.365	0.366	1.000				
9	0.188	0.204	0.211	0.180	0.164	0.152	0.153	0.144	1.000			
10	0.194	0.210	0.218	0.185	0.168	0.156	0.158	0.148	0.647	1.000		
11	0.199	0.217	0.224	0.191	0.173	0.161	0.162	0.152	0.653	0.659	1.000	
12	0.168	0.183	0.189	0.161	0.146	0.136	0.137	0.128	0.565	0.571	0.576	1.000
13	0.154	0.167	0.172	0.147	0.134	0.124	0.125	0.118	0.550	0.554	0.559	0.484
14	0.127	0.138	0.142	0.121	0.111	0.103	0.103	0.097	0.466	0.470	0.473	0.410
15	0.141	0.153	0.159	0.136	0.123	0.114	0.115	0.108	0.481	0.486	0.490	0.424
16	0.145	0.157	0.162	0.138	0.126	0.117	0.118	0.111	0.485	0.488	0.493	0.427
17	0.178	0.193	0.199	0.170	0.155	0.143	0.144	0.136	0.183	0.188	0.193	0.163
18	0.178	0.193	0.200	0.171	0.155	0.144	0.145	0.136	0.183	0.189	0.193	0.164
19	0.197	0.215	0.222	0.189	0.171	0.160	0.161	0.150	0.203	0.209	0.215	0.182
20	0.153	0.165	0.171	0.146	0.133	0.123	0.124	0.116	0.157	0.161	0.165	0.140
21	0.174	0.190	0.197	0.168	0.152	0.141	0.142	0.133	0.180	0.185	0.191	0.161
22	0.143	0.155	0.160	0.137	0.124	0.115	0.116	0.109	0.147	0.151	0.155	0.131
23	0.136	0.148	0.153	0.130	0.119	0.110	0.110	0.104	0.140	0.144	0.148	0.125
24	0.132	0.144	0.149	0.126	0.115	0.107	0.107	0.101	0.136	0.140	0.144	0.122
	1	2	3	4	5	6	7	8	9	10	11	12

13	1.000											
14	0.401	1.000										
15	0.413	0.350	1.000									
16	0.415	0.352	0.364	1.000								
17	0.149	0.124	0.137	0.140	1.000							
18	0.150	0.124	0.138	0.140	0.621	1.000						
19	0.166	0.137	0.153	0.156	0.640	0.640	1.000					
20	0.128	0.106	0.117	0.120	0.541	0.541	0.557	1.000				
21	0.146	0.120	0.135	0.138	0.561	0.562	0.581	0.489	1.000			
22	0.120	0.099	0.110	0.113	0.474	0.475	0.490	0.413	0.430	1.000		
23	0.114	0.095	0.105	0.107	0.469	0.469	0.483	0.408	0.424	0.358	1.000	
24	0.111	0.092	0.102	0.104	0.465	0.465	0.479	0.405	0.420	0.355	0.351	1.000
	13	14	15	16	17	18	19	20	21	22	23	24

Varimax Rotated Maximum Likelihood Solution

	1	2	3
V3	.803	.130	.129
V2	.796	.123	.122
V1	.779	.107	.106
V4	.698	.109	.108
V5	.681	.093	.092
V7	.597	.093	.091
V6	.596	.091	.090
V8	.587	.083	.082
V11	.129	.795	.124
V10	.123	.790	.119
V9	.118	.784	.113
V12	.107	.690	.103
V13	.092	.675	.088
V16	.092	.591	.088
V15	.089	.588	.085
V14	.074	.573	.071
V19	.129	.125	.793
V18	.108	.105	.774
V17	.108	.105	.773
V21	.115	.112	.695
V20	.091	.089	.674
V22	.091	.089	.589
V23	.084	.081	.583
V24	.080	.078	.579

Twenty Four Variables, Three Factors, Wide Communalities

1	1.000											
2	0.556	1.000										
3	0.480	0.423	1.000									
4	0.396	0.348	0.301	1.000								
5	0.318	0.280	0.242	0.199	1.000							
6	0.311	0.273	0.236	0.194	0.156	1.000						
7	0.227	0.199	0.173	0.142	0.114	0.112	1.000					
8	0.156	0.138	0.120	0.098	0.079	0.077	0.057	1.000				
9	0.186	0.166	0.147	0.117	0.095	0.088	0.060	0.046	1.000			
10	0.170	0.152	0.133	0.107	0.087	0.080	0.054	0.041	0.564	1.000		
11	0.141	0.127	0.111	0.089	0.072	0.066	0.045	0.034	0.479	0.425	1.000	
12	0.110	0.098	0.087	0.069	0.056	0.051	0.035	0.027	0.391	0.346	0.295	1.000
13	0.089	0.080	0.071	0.056	0.046	0.042	0.029	0.022	0.315	0.278	0.236	0.194
14	0.088	0.079	0.070	0.056	0.045	0.041	0.028	0.021	0.313	0.278	0.236	0.193
15	0.065	0.058	0.052	0.041	0.033	0.030	0.021	0.016	0.234	0.207	0.176	0.144
16	0.054	0.048	0.042	0.034	0.027	0.025	0.017	0.012	0.166	0.148	0.125	0.102
17	0.198	0.177	0.156	0.125	0.101	0.093	0.063	0.048	0.201	0.184	0.153	0.119
18	0.167	0.149	0.132	0.106	0.085	0.079	0.054	0.041	0.169	0.155	0.129	0.100
19	0.135	0.121	0.107	0.085	0.069	0.064	0.044	0.033	0.138	0.126	0.104	0.081
20	0.116	0.103	0.091	0.073	0.059	0.054	0.037	0.028	0.117	0.107	0.089	0.070
21	0.099	0.089	0.079	0.062	0.051	0.046	0.032	0.024	0.101	0.092	0.077	0.060
22	0.096	0.086	0.076	0.061	0.049	0.045	0.031	0.023	0.097	0.089	0.074	0.058
23	0.070	0.062	0.056	0.044	0.036	0.033	0.022	0.017	0.071	0.065	0.054	0.042
24	0.054	0.049	0.043	0.034	0.028	0.026	0.017	0.013	0.056	0.051	0.042	0.033
	1	2	3	4	5	6	7	8	9	10	11	12

13	1.000											
14	0.155	1.000										
15	0.116	0.115	1.000									
16	0.082	0.082	0.061	1.000								
17	0.096	0.096	0.070	0.058	1.000							
18	0.081	0.081	0.059	0.049	0.573	1.000						
19	0.066	0.065	0.048	0.039	0.483	0.417	1.000					
20	0.056	0.056	0.041	0.034	0.405	0.350	0.295	1.000				
21	0.049	0.048	0.036	0.029	0.332	0.287	0.242	0.203	1.000			
22	0.046	0.047	0.034	0.028	0.328	0.284	0.239	0.201	0.164	1.000		
23	0.034	0.034	0.025	0.021	0.244	0.211	0.178	0.149	0.122	0.121	1.000	
24	0.027	0.026	0.019	0.016	0.171	0.148	0.125	0.105	0.086	0.084	0.063	1.000
	13	14	15	16	17	18	19	20	21	22	23	24

Varimax Rotated Maximum Likelihood Solution

	1	2	3
V17	.796	.122	.119
V18	.690	.101	.098
V19	.583	.079	.076
V20	.489	.068	.066
V21	.399	.061	.060
V22	.395	.058	.057
V23	.294	.041	.040
V24	.204	.036	.035
V9	.116	.782	.110
V10	.109	.691	.103
V11	.089	.588	.084
V12	.066	.482	.062
V13	.054	.387	.051
V14	.053	.386	.050
V15	.039	.288	.036
V16	.036	.203	.034
V1	.115	.111	.778
V2	.105	.101	.684
V3	.094	.091	.590
V4	.073	.070	.487
V5	.060	.058	.391
V6	.052	.050	.384
V7	.033	.032	.282
V8	.028	.027	.193

24 Variables, 3 factors, Low Communalities

1	1.000											
2	0.151	1.000										
3	0.149	0.156	1.000									
4	0.110	0.114	0.112	1.000								
5	0.118	0.125	0.123	0.090	1.000							
6	0.076	0.079	0.078	0.057	0.063	1.000						
7	0.078	0.082	0.081	0.059	0.065	0.041	1.000					
8	0.078	0.082	0.081	0.059	0.066	0.041	0.043	1.000				
9	0.045	0.054	0.052	0.035	0.049	0.027	0.030	0.031	1.000			
10	0.039	0.046	0.044	0.030	0.041	0.023	0.026	0.026	0.166	1.000		
11	0.044	0.053	0.050	0.034	0.047	0.026	0.029	0.030	0.174	0.164	1.000	
12	0.031	0.038	0.036	0.024	0.034	0.019	0.021	0.021	0.128	0.121	0.126	1.000
13	0.032	0.039	0.036	0.025	0.034	0.019	0.021	0.022	0.129	0.122	0.127	0.094
14	0.015	0.017	0.016	0.011	0.015	0.009	0.009	0.009	0.075	0.072	0.075	0.055
15	0.020	0.025	0.023	0.016	0.022	0.012	0.014	0.014	0.085	0.080	0.084	0.062
16	0.020	0.023	0.022	0.015	0.020	0.011	0.013	0.013	0.083	0.079	0.082	0.061
17	0.040	0.048	0.045	0.031	0.043	0.024	0.027	0.027	0.056	0.048	0.054	0.039
18	0.039	0.047	0.044	0.030	0.042	0.024	0.026	0.026	0.055	0.046	0.053	0.038
19	0.036	0.042	0.040	0.027	0.037	0.021	0.023	0.023	0.049	0.041	0.047	0.034
20	0.029	0.035	0.033	0.022	0.031	0.018	0.019	0.019	0.041	0.034	0.039	0.028
21	0.030	0.036	0.034	0.024	0.032	0.018	0.020	0.021	0.042	0.035	0.041	0.029
22	0.019	0.022	0.021	0.015	0.020	0.011	0.012	0.012	0.027	0.022	0.025	0.018
23	0.018	0.021	0.020	0.014	0.019	0.011	0.012	0.012	0.025	0.021	0.024	0.017
24	0.023	0.028	0.026	0.018	0.025	0.014	0.016	0.016	0.032	0.027	0.031	0.022
	1	2	3	4	5	6	7	8	9	10	11	12

13	1.000											
14	0.055	1.000										
15	0.063	0.036	1.000									
16	0.061	0.036	0.040	1.000								
17	0.040	0.017	0.026	0.024	1.000							
18	0.038	0.017	0.024	0.023	0.160	1.000						
19	0.034	0.016	0.022	0.021	0.155	0.154	1.000					
20	0.028	0.013	0.018	0.017	0.120	0.119	0.116	1.000				
21	0.030	0.013	0.019	0.018	0.122	0.120	0.116	0.090	1.000			
22	0.019	0.009	0.012	0.011	0.079	0.078	0.077	0.059	0.059	1.000		
23	0.017	0.008	0.011	0.011	0.077	0.077	0.076	0.058	0.058	0.039	1.000	
24	0.023	0.010	0.015	0.014	0.084	0.084	0.081	0.063	0.063	0.041	0.041	1.000
	13	14	15	16	17	18	19	20	21	22	23	24

Varimax Rotated Maximum Likelihood Solution

	1	2	3
V9	.407	.071	.070
V11	.403	.067	.067
V10	.389	.054	.054
V13	.301	.048	.048
V12	.299	.046	.047
V15	.198	.029	.030
V16	.194	.027	.027
V14	.180	.015	.015
V17	.060	.393	.059
V18	.056	.391	.056
V19	.046	.381	.047
V21	.045	.295	.044
V20	.042	.293	.042
V24	.037	.203	.036
V22	.027	.194	.026
V23	.023	.191	.024
V2	.057	.055	.390
V3	.051	.051	.386
V1	.040	.041	.375
V5	.056	.055	.305
V4	.032	.032	.283
V8	.034	.033	.201
V7	.034	.033	.200
V6	.028	.028	.195

Twenty Four Variables, Eight Factors, High Communality

1	1.000											
2	0.552	1.000										
3	0.480	0.436	1.000									
4	0.183	0.180	0.161	1.000								
5	0.162	0.159	0.142	0.573	1.000							
6	0.124	0.123	0.109	0.475	0.417	1.000						
7	0.178	0.174	0.156	0.198	0.176	0.135	1.000					
8	0.154	0.151	0.136	0.172	0.153	0.117	0.559	1.000				
9	0.137	0.135	0.120	0.153	0.135	0.104	0.485	0.423	1.000			
10	0.188	0.186	0.166	0.211	0.186	0.143	0.204	0.178	0.158	1.000		
11	0.154	0.152	0.135	0.172	0.152	0.117	0.167	0.145	0.129	0.569	1.000	
12	0.127	0.125	0.111	0.141	0.125	0.097	0.137	0.120	0.106	0.482	0.413	1.000
13	0.179	0.176	0.157	0.200	0.177	0.136	0.194	0.169	0.150	0.206	0.168	0.139
14	0.155	0.153	0.136	0.173	0.153	0.119	0.168	0.146	0.130	0.178	0.146	0.121
15	0.129	0.126	0.113	0.144	0.127	0.097	0.140	0.121	0.107	0.148	0.121	0.099
16	0.181	0.178	0.159	0.202	0.179	0.137	0.196	0.171	0.152	0.208	0.170	0.141
17	0.157	0.155	0.138	0.175	0.155	0.120	0.170	0.147	0.132	0.180	0.148	0.122
18	0.141	0.140	0.124	0.158	0.139	0.108	0.153	0.133	0.119	0.163	0.133	0.110
19	0.174	0.170	0.152	0.194	0.172	0.132	0.188	0.164	0.145	0.199	0.163	0.134
20	0.165	0.164	0.146	0.185	0.163	0.126	0.179	0.156	0.139	0.191	0.156	0.128
21	0.126	0.123	0.110	0.140	0.124	0.095	0.137	0.119	0.105	0.144	0.118	0.097
22	0.177	0.175	0.156	0.199	0.176	0.135	0.193	0.167	0.149	0.204	0.168	0.138
23	0.150	0.148	0.131	0.167	0.148	0.114	0.162	0.141	0.126	0.172	0.141	0.117
24	0.131	0.128	0.115	0.145	0.129	0.099	0.141	0.123	0.109	0.150	0.122	0.101
	1	2	3	4	5	6	7	8	9	10	11	12

13	1.000											
14	0.562	1.000										
15	0.476	0.415	1.000									
16	0.198	0.171	0.142	1.000								
17	0.171	0.149	0.123	0.565	1.000							
18	0.155	0.135	0.110	0.493	0.430	1.000						
19	0.190	0.164	0.137	0.192	0.166	0.149	1.000					
20	0.181	0.157	0.130	0.183	0.159	0.143	0.567	1.000				
21	0.137	0.118	0.099	0.139	0.120	0.108	0.470	0.420	1.000			
22	0.195	0.169	0.140	0.196	0.171	0.154	0.188	0.180	0.136	1.000		
23	0.164	0.142	0.117	0.166	0.144	0.130	0.158	0.152	0.114	0.555	1.000	
24	0.142	0.123	0.102	0.144	0.125	0.112	0.138	0.132	0.100	0.478	0.413	1.000
	13	14	15	16	17	18	19	20	21	22	23	24

Varimax rotated Maximum Likelihood Solution

	1	2	3	4	5	6	7	8
V16	.765	.095	.094	.094	.095	.094	.094	.093
V17	.669	.081	.081	.081	.082	.081	.081	.081
V18	.579	.075	.075	.075	.075	.074	.075	.074
V7	.094	.762	.093	.092	.093	.092	.092	.091
V8	.082	.666	.080	.080	.081	.080	.079	.079
V9	.074	.574	.072	.072	.072	.072	.072	.072
V4	.098	.097	.767	.097	.097	.096	.096	.095
V5	.087	.087	.672	.086	.086	.086	.085	.085
V6	.064	.063	.564	.063	.063	.062	.063	.062
V10	.103	.102	.101	.770	.101	.100	.100	.099
V11	.081	.080	.080	.666	.080	.080	.079	.079
V12	.066	.065	.064	.566	.065	.064	.064	.064
V1	.083	.082	.081	.081	.750	.082	.081	.081
V2	.087	.085	.085	.086	.671	.084	.085	.085
V3	.079	.078	.078	.078	.580	.077	.077	.076
V19	.091	.090	.089	.089	.090	.760	.089	.088
V20	.091	.089	.089	.090	.090	.673	.088	.088
V21	.064	.064	.064	.063	.064	.566	.063	.063
V13	.096	.095	.094	.094	.094	.093	.763	.093
V14	.083	.081	.081	.081	.081	.080	.666	.080
V15	.067	.067	.066	.066	.067	.066	.567	.065
V22	.095	.094	.094	.093	.094	.093	.093	.762
V23	.079	.077	.077	.078	.078	.077	.077	.662
V24	.069	.068	.067	.067	.068	.068	.067	.569

Twenty Four Variables, Eight Factors, Wide Communalities

1	1.000											
2	0.402	1.000										
3	0.158	0.098	1.000									
4	0.192	0.117	0.044	1.000								
5	0.131	0.080	0.030	0.406	1.000							
6	0.044	0.028	0.010	0.154	0.099	1.000						
7	0.206	0.126	0.047	0.198	0.137	0.046	1.000					
8	0.124	0.076	0.029	0.119	0.083	0.028	0.409	1.000				
9	0.050	0.031	0.011	0.049	0.033	0.011	0.164	0.101	1.000			
10	0.195	0.119	0.045	0.187	0.128	0.043	0.202	0.121	0.049	1.000		
11	0.128	0.079	0.029	0.123	0.084	0.029	0.132	0.080	0.032	0.405	1.000	
12	0.051	0.031	0.012	0.049	0.033	0.011	0.052	0.031	0.013	0.162	0.103	1.000
13	0.204	0.126	0.047	0.197	0.135	0.046	0.213	0.129	0.051	0.200	0.131	0.051
14	0.132	0.081	0.030	0.127	0.087	0.029	0.137	0.082	0.033	0.129	0.085	0.034
15	0.044	0.028	0.010	0.043	0.029	0.010	0.046	0.028	0.011	0.044	0.029	0.011
16	0.195	0.120	0.045	0.188	0.129	0.044	0.203	0.123	0.049	0.191	0.126	0.050
17	0.118	0.073	0.027	0.114	0.078	0.026	0.122	0.074	0.030	0.116	0.076	0.030
18	0.049	0.030	0.011	0.047	0.032	0.011	0.050	0.030	0.012	0.047	0.032	0.012
19	0.207	0.126	0.048	0.199	0.137	0.046	0.215	0.130	0.052	0.202	0.133	0.052
20	0.117	0.072	0.027	0.113	0.078	0.027	0.122	0.074	0.029	0.115	0.076	0.029
21	0.049	0.030	0.011	0.048	0.033	0.011	0.051	0.031	0.012	0.048	0.031	0.012
22	0.202	0.123	0.047	0.194	0.134	0.045	0.210	0.127	0.051	0.198	0.130	0.051
23	0.133	0.081	0.030	0.128	0.088	0.029	0.138	0.083	0.033	0.130	0.085	0.034
24	0.057	0.035	0.013	0.055	0.037	0.012	0.059	0.035	0.014	0.056	0.036	0.015
	1	2	3	4	5	6	7	8	9	10	11	12

13	1.000											
14	0.416	1.000										
15	0.158	0.099	1.000									
16	0.201	0.130	0.044	1.000								
17	0.122	0.078	0.027	0.396	1.000							
18	0.050	0.032	0.011	0.160	0.099	1.000						
19	0.213	0.137	0.046	0.204	0.123	0.050	1.000					
20	0.122	0.078	0.027	0.116	0.070	0.029	0.403	1.000				
21	0.051	0.032	0.011	0.049	0.029	0.012	0.163	0.099	1.000			
22	0.208	0.134	0.045	0.199	0.120	0.049	0.211	0.120	0.050	1.000		
23	0.137	0.088	0.030	0.130	0.079	0.032	0.138	0.079	0.033	0.415	1.000	
24	0.058	0.037	0.013	0.055	0.034	0.014	0.059	0.033	0.014	0.169	0.108	1.000
	13	14	15	16	17	18	19	20	21	22	23	24

Varimax Rotated Maximum Likelihood Solution

	1	2	3	4	5	6	7	8
V7	.527	-.465	-.348	-.197	.035	-.063	-.030	-.031
V19	.526	-.022	.460	-.400	.039	-.082	-.037	-.035
V13	.524	.480	-.319	-.215	.025	-.067	-.032	-.026
V22	.513	-.004	.057	.233	-.495	-.268	-.081	-.062
V1	.503	.008	.072	.337	.442	-.266	-.066	-.061
V10	.490	.002	.040	.131	-.008	.477	-.366	-.125
V14	.335	.300	-.199	-.134	.016	-.041	-.020	-.016
V23	.334	-.003	.036	.148	-.313	-.169	-.051	-.039
V8	.321	-.292	-.218	-.124	.022	-.040	-.018	-.019
V11	.319	.002	.025	.083	-.005	.300	-.229	-.079
V2	.309	.006	.045	.211	.277	-.167	-.041	-.038
V20	.307	-.013	.289	-.253	.025	-.052	-.023	-.022
V24	.141	-.001	.015	.060	-.125	-.067	-.021	-.015
V9	.129	-.117	-.086	-.048	.009	-.016	-.007	-.007
V21	.127	-.006	.116	-.101	.010	-.021	-.009	-.008
V12	.126	.000	.011	.034	-.002	.120	-.092	-.031
V3	.118	.002	.018	.084	.110	-.067	-.016	-.015
V15	.118	.120	-.080	-.054	.006	-.017	-.008	-.007
V16	.490	.004	.033	.098	-.019	.215	.534	-.229
V17	.299	.003	.020	.063	-.011	.136	.337	-.145
V18	.122	.002	.008	.025	-.004	.055	.135	-.058
V4	.480	-.002	.030	.081	-.010	.116	.108	.604
V5	.324	-.001	.018	.050	-.007	.073	.068	.381
V6	.113	.000	.007	.019	-.003	.029	.027	.152

Twenty Four Variables, Eight Factors, Low Communalities

1	1.000											
2	0.115	1.000										
3	0.075	0.056	1.000									
4	0.043	0.031	0.020	1.000								
5	0.030	0.022	0.014	0.115	1.000							
6	0.019	0.014	0.008	0.075	0.055	1.000						
7	0.047	0.034	0.021	0.048	0.033	0.020	1.000					
8	0.032	0.023	0.014	0.032	0.022	0.014	0.119	1.000				
9	0.021	0.016	0.010	0.022	0.015	0.009	0.080	0.058	1.000			
10	0.042	0.030	0.019	0.043	0.029	0.019	0.047	0.032	0.021	1.000		
11	0.038	0.028	0.017	0.039	0.027	0.017	0.043	0.029	0.019	0.122	1.000	
12	0.029	0.021	0.012	0.029	0.020	0.013	0.032	0.022	0.014	0.085	0.069	1.000
13	0.039	0.029	0.018	0.040	0.028	0.017	0.043	0.029	0.020	0.038	0.035	0.026
14	0.035	0.025	0.016	0.036	0.025	0.015	0.039	0.026	0.018	0.035	0.032	0.024
15	0.022	0.016	0.010	0.022	0.016	0.010	0.024	0.016	0.011	0.022	0.020	0.015
16	0.041	0.030	0.018	0.042	0.029	0.018	0.045	0.031	0.021	0.041	0.037	0.028
17	0.037	0.027	0.016	0.038	0.026	0.016	0.041	0.027	0.019	0.037	0.034	0.025
18	0.027	0.020	0.012	0.028	0.019	0.012	0.030	0.020	0.014	0.027	0.026	0.019
19	0.045	0.032	0.020	0.046	0.031	0.020	0.050	0.034	0.022	0.045	0.041	0.031
20	0.036	0.026	0.016	0.037	0.025	0.016	0.040	0.027	0.018	0.036	0.033	0.024
21	0.023	0.017	0.010	0.024	0.016	0.011	0.026	0.018	0.012	0.023	0.022	0.017
22	0.047	0.034	0.021	0.048	0.033	0.021	0.053	0.036	0.023	0.047	0.043	0.033
23	0.033	0.025	0.015	0.034	0.024	0.015	0.037	0.025	0.017	0.033	0.030	0.023
24	0.025	0.018	0.011	0.025	0.017	0.011	0.027	0.019	0.012	0.025	0.023	0.017
	1	2	3	4	5	6	7	8	9	10	11	12

13	1.000											
14	0.116	1.000										
15	0.076	0.060	1.000									
16	0.038	0.034	0.021	1.000								
17	0.034	0.030	0.019	0.120	1.000							
18	0.025	0.023	0.015	0.082	0.066	1.000						
19	0.041	0.037	0.023	0.044	0.039	0.029	1.000					
20	0.033	0.030	0.018	0.035	0.031	0.023	0.122	1.000				
21	0.021	0.019	0.012	0.023	0.021	0.016	0.081	0.062	1.000			
22	0.042	0.040	0.024	0.046	0.041	0.031	0.050	0.040	0.026	1.000		
23	0.031	0.027	0.018	0.032	0.029	0.022	0.035	0.028	0.019	0.121	1.000	
24	0.022	0.021	0.012	0.024	0.022	0.016	0.026	0.021	0.014	0.084	0.061	1.000
	13	14	15	16	17	18	19	20	21	22	23	24

Varimax Rotated Maximum Likelihood Solution

	1	2	3	4	5	6	7	8
V10	.375	.042	.042	.042	.040	.039	.038	.038
V11	.296	.042	.042	.043	.040	.040	.038	.037
V12	.203	.034	.033	.034	.032	.030	.029	.029
V22	.053	.386	.050	.050	.048	.047	.045	.045
V23	.035	.284	.034	.035	.033	.033	.032	.031
V24	.028	.195	.027	.027	.026	.025	.024	.024
V19	.048	.047	.382	.047	.045	.044	.041	.042
V20	.039	.038	.289	.038	.037	.037	.035	.035
V21	.027	.025	.191	.026	.023	.023	.022	.022
V16	.041	.041	.040	.373	.039	.039	.037	.037
V17	.041	.039	.040	.293	.038	.038	.036	.036
V18	.033	.031	.031	.199	.030	.029	.028	.027
V7	.053	.051	.051	.051	.384	.048	.045	.046
V8	.034	.034	.032	.033	.282	.030	.028	.029
V9	.022	.021	.022	.022	.188	.022	.021	.020
V13	.038	.037	.038	.039	.036	.372	.036	.036
V14	.040	.039	.038	.038	.037	.285	.034	.035
V15	.024	.023	.023	.024	.022	.189	.021	.021
V4	.047	.046	.046	.047	.045	.044	.377	.042
V5	.031	.030	.030	.031	.030	.030	.279	.028
V6	.019	.019	.018	.019	.018	.018	.184	.018
V1	.047	.046	.045	.045	.044	.042	.040	.375
V2	.033	.032	.032	.033	.030	.030	.029	.280
V3	.019	.018	.019	.019	.018	.018	.018	.184