

Robust factor analysis in the presence of normality violations, missing data, and outliers: Empirical questions and possible solutions

Conrad Zygmont ^a, Mario R. Smith ^b

^a Psychology Department, Helderberg College, South Africa

^b Psychology Department, University of the Western Cape

Abstract ■ Although a mainstay of psychometric methods, several reviews suggest factor analysis is often applied without testing whether data support it, and that decision-making process or guiding principles providing evidential support for FA techniques are seldom reported. Researchers often defer such decision-making to the default settings on widely-used software packages, and unaware of their limitations, might unwittingly misuse FA. This paper discusses robust analytical alternatives for answering nine important questions in exploratory factor analysis (EFA), and provides R commands for running complex analysis in the hope of encouraging and empowering substantive researchers on a journey of discovery towards more knowledgeable and judicious use of robust alternatives in FA. It aims to take solutions to problems like skewness, missing values, determining the number of factors to extract, and calculation of standard errors of loadings, and make them accessible to the general substantive researcher.

Keywords ■ Exploratory factor analysis; analytical decision making; data screening; factor extraction; factor rotation; number of factors; R statistical environment

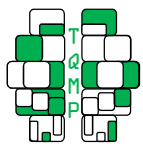
 zygmontc@hbc.ac.za

Introduction

Exploratory factor analysis (EFA) entails a set of procedures for modelling a theoretical number of latent dimensions representing a parsimonious approximation of the relationship between real-world phenomena and measured variables. Confirmatory factor analysis (CFA) implements routines for evaluating model fit and factorial invariance of postulated latent dimensions (MacCallum, Browne, & Cai, 2007; Thompson, 2004; Tucker & MacCallum, 1997). Factor analytic methods trace their history to Spearman's (1904) seminal article on the structure of intelligence, and were eagerly adopted and further developed by other intelligence theorists (e.g. Thurstone, 1936). In celebration of a century of factor analysis research, Cudek (2007) proclaimed "factor analysis has turned out to be one of the most successful of the multivariate statistical methods and one of the pillars of behavioral research" (p. 4). Kerlinger (1986) describes factor analysis as "the queen of analytic methods ... because of its power, elegance, and closeness to the core of scientific purpose" (p. 569). Systematic reviews report that between 13 and 29 percent of research articles in some psychology journals make use of EFA, CFA or principal components analysis (PCA) with this number continuing to increase

(Fabrigar, Wegener, MacCallum, & Strahan, 1999; Russell, 2002; Zygmont & Smith, 2006). This popularity is partly due to the advent of personal computers and increased accessibility to FA calculations afforded substantive researchers by statistical software allowing complex calculations to be done "in only moments, and in a user-friendly point-and-click environment" (Thomson, 2004, p. 4). Nedler (1964) predicted that " 'first generation' programs, which largely behave as though the design did wholly define the analysis, will be replaced by new second-generation programs capable of checking the additional assumptions and taking appropriate action" (p. 245). This has not taken place – the onus still rests on researchers to make judicious choices between analytical procedures at their disposal. Yuan and Lu (2008) caution against relying solely on default output of popular software packages for FA. However, researchers are often unaware of powerful robust alternatives to inefficient analytical options appearing as defaults in standard statistical packages or modern trends in the judicious use of statistical procedures (Erceg-Hurn & Mirosevich, 2008; Preacher & MacCallum, 2003).

Reviews of articles in prominent psychology journals (Fabrigar, Wegener, MacCallum & Strahan,



1999; Russell, 2002; Zygmunt & Smith, 2006), animal behavior research (Budaev, 2010), counseling (Worthington & Wittaker, 2006), education (Schönrock-Adema, Heinje-Penninga, van Hell, & Cohen-Schotanus, 2009), and medicine (Patil, McPherson, & Friesner, 2010) have all noted that FA options being used in substantive research are often inconsistent with statistical literature, and authors often fail to adequately report on the methods being used. Numerous powerful robust procedures are available, but often remain in the realm of academic curiosities (Horsewell, 1990). Dinno (2009) implores “as there are a growing number of fast free software tools available for any researcher to employ, the bar ought to be raised” (p. 386).

Towards this end this paper presents a sequence of nine empirical questions, together with suggested alternatives for exploring answers, which can be used by researchers in the process of conducting robust EFA under a wide range of circumstances. The authors' intention is not to provide detailed expositions on each method, but rather to present options, allowing for researchers to make informed decisions regarding their analysis. Together with the theoretical discussion and example, an R script is provided allowing for replication of these analyses using the R statistical environment. R provides FA relevant functions and the largest collections of statistical tools of any software – all for free (Klinke, Mihoci, & Härdle, 2010; R Development Core Team, 2008).

Question 1: Is my sample size adequate?

Generally methodologists prioritize a large sample when designing a factor analytic study, especially for recovery of weak factor loadings (Ximénez, 2006). A sufficient sample size for factor analysis is generally considered to be above 100, with 200 being considered a large sample size although more is always better, and 50 an absolute minimum (Boomsma, 1985; Gorsuch, 1983). However, absolute rules for sample size are not appropriate, seeing as adequate sample size is partly determined by sample-variable ratios, saturation of factors, and heterogeneity of the sample (Costello & Osborne, 2005; de Winter, Dodou, & Wieringa, 2009). Proposed sample-variable ratios range from 5:1 as an absolute minimum to 10:1 as the commonly used standard (Hair, Anderson, Tatham, and Grablovsky, 1995; Kerlinger, 1986). An inverse relationship between commonalities of variables and sample size exists (Fabrigar et al., 1999). High commonalities (\geq

.70) suggest adequate factor saturation for which sample sizes as low as 60 could suffice. Low commonalities ($\leq .50$) suggest inadequate factor saturation for which sample sizes between 100 and 200 are recommended (MacCallum, Widaman, Zhang, and Hong, 1999). However, these values are typically not available prior to conducting EFA and are difficult to estimate. Item reliability coefficients could provide a useful guideline. Kerlinger (1986) recommend sample ratios of 10:1 or more when item reliability and item inter-correlations are low.

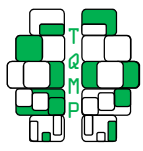
Question 2: Does the data support factor analysis?

Data should be screened prior to analysis so that informed decisions can be made regarding the most appropriate statistics and data cleaning (for example, scrubbing obvious input errors). Important properties to examine include distribution assumptions, impact of outliers, and missing values.

Distribution assumptions.

The assumption of multivariate normality (MVN) forms the basis for correlational statistics upon which FA and various procedures (e.g. χ^2 goodness-of-fit) used in maximum-likelihood (ML) analysis rests (Rowe & Rowe, 2004). In testing this assumption, first examine for univariate normality (UVN). Violation of UVN increases the likelihood that MVN has been violated. However, MVN can be violated even though no individual variables were found to be non-normal. The Skewness and Kurtosis statistics – with critical values for maximum likelihood (ML) methods set at 2 and 7 respectively (Curran, West & Finch, 1996; Ryu, 2011) – and Kolmogorov-Smirnov statistic are most commonly used to investigate UVN. Erceg-Hurn and Mirosevich (2008) caution that these tests can be susceptible to heteroscedasticity. Srivastava and Hui (1987) recommended the Shapiro-Wilk W-test as a more powerful alternative, and rated it as possibly the best test for UVN. Keeping in mind that one test is unlikely to detect all possible variations from normality, Looney (1995) suggested that decisions regarding normality should be based on the aggregate results of a battery of different tests with relatively high power.

Mecklin and Mundfrom (2005) categorised MVN tests into four groups: Graphical and correlational approaches (e.g. chi-squared plot), Skewness and kurtosis approaches (e.g. Mardia's tests of skewness and kurtosis), Goodness of fit approaches (e.g. Anderson-Darling and Shapiro-Wilk multivariate



omnibus tests), and Consistent approaches (e.g. Henze-Zirkler test utilizing the empirical characteristic function). Of the fifty or so procedures available, Mecklin and Mundfrom (2005) recommended two for their high power across a wide range of non-normal situations: Royston's (1995) revision of a goodness of fit multivariate extension to the Shapiro-Wilks W test for smaller samples and the Henze-Zirkler (1990) consistent test for larger samples. The former estimates the straightness of the normal quantile-quantile (Q-Q) probability plot whereas the latter measures the distance between the hypothesized MVN distribution and the observed distribution (Farrell, Salibian-Barrera, & Naczk, 2006). As recommended above, the results of these and other MVN test statistics should be interpreted in unison to make meaningful decisions about normality. It is also advisable to look for outliers, and see whether they may be impacting on normality of your data.

Impact of outliers.

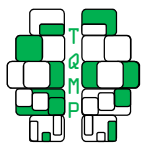
A single outlier can potentially distort correlation estimates (Stevens, 1984), measures of item-factor congruence such as Cronbach's alpha (Christmann & Van Aest, 2006), and FA model parameters and goodness-of-fit estimators (Mavridis & Moustaki, 2008). Outliers may eventually lead to incorrect models being specified (Bollen, 1987; Pison et al., 2003). Conversely, good leverage points – outliers with very small residuals from the model line despite lying far from the center of the data cloud – can actually lower standard errors on estimates of regression coefficients (Yuan & Zhong, 2008). Start investigating the impact of outliers by examining univariate distributions (e.g. box-plots or values furthest from the mean), then bivariate distributions (e.g. standardized residuals more than three absolute values from the regression line), and finally scores that stray significantly from the multivariate average of all scores.

Mahalanobis' D^2 (distance of a score from the centroid of all cases) and Cooks distance (estimate of an observation's combined influence on both predictor and criterion spaces expressed as the change in the regression coefficient attributable to each case) are the most common statistics used to identify multivariate outliers (Stevens, 1984). Despite their popularity they suffer from masking (the presence of outliers makes it difficult to estimate location and scatter), are vulnerable to heteroscedasticity, and distributional variations (Wilcox & Keselman, 2004). Improved

multivariate outlier detection methods that utilize robust estimations of location and scatter, have high breakdown points (can handle more outliers before estimates are compromised), and are differentially sensitive to good and bad leverage points have been developed (Mavridis & Moustaki, 2008; Pison, Rousseeuw, Filzmoser, & Croux, 2003; Rousseeuw & van Driessen, 1999; Yuan & Zhong, 2008). Examples of affine-equivariant estimators (invariant under rotations of the data) that achieve a breakdown point of approximately .05 include: 1) the minimum-volume ellipsoid (MVE) estimator, which attempts to estimate the smallest ellipsoid to encapture half of the available data; 2) the minimum-covariance determinant (MCD), which searches for the subset of half of the data with the smallest generalized variance; 3) the translated-biweight S -estimator (TBS), which seeks to empirically determine how much data should be trimmed and minimize the value of scale of the data; 4) the minimum generalized variance (MGV), which iteratively moves the data between two sets working out which points have the highest generalized variance from the center of the cloud, and 5) projection methods, which consider whether points are outliers across a number of orthogonal projections of the data (Wilcox, 2012). Of the robust procedures available, no single method works best in all situations – their performance varies depending on where a given outlier is located relative to the data cloud and other outliers, how many outliers there happen to be, and the sample size and number of variables (Wilcox, 2008). MVE works well if the number of variables is less than 10, MCD and TBS when there are at least 5 observations per dimension, and MGV that has the advantage of being scale invariant. When there are 10 or more variables, MGV or projection algorithms with simulations used to adjust the decision rule to limit the number of outliers identified to a specified value are suggested (Wilcox, 2012).

Missing values.

Burton and Altman (2004) found that few researchers consider the impact of missing data on their models, viewing it as a non-issue or merely a nuisance best ignored. Best practice guidelines suggest that every quantitative study should report the extent and nature of missing data, as well as the rationale and procedures used to handle missing data (Schlomer, Bauman, & Card, 2010). Little and Rubin (2002) propose three possibilities regarding the nature of

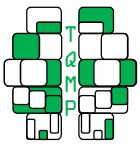


missing data: Completely random missing data (MCAR), where missing data are unrelated to predicted or observed values; Randomly missing values (MAR), where missing values may be related to other observed values, but not to missing values; or Non-random missing data (MNAR), where missing data are dependent on the value which would have been observed. The mechanism by which data is missing is very important when determining the efficacy and appropriateness of imputation strategies. The default techniques for dealing with missing values in most statistical packages are listwise and pairwise deletion. Listwise excludes the entire case and will lead to unbiased parameter and standard error estimates if data are MCAR, but may yield biased parameter estimates in MAR, and is likely to result in reductions to power. Pairwise deletion estimates moments for all pairs of cases in which all data is present. Although allowing for greater power, pairwise analysis may result in more sampling variance than listwise deletion, produce biased standard error estimates, and a covariance matrix that is not positive definite (Allison, 2003; Jamshidian & Mata, 2007).

A few missing values need not signal the decimation of your degrees of freedom, these values can often be imputed. The simplest method is simply imputing the mean for that variable, although this method is almost never appropriate as it leads to severely underestimated variance (Jamshidian & Mata, 2007; Little & Rubin, 2002). Nonstochastic regression methods are easily computed, but should be avoided as biases in variance and covariance estimates may result, and accurate standard errors cannot be calculated (Lumley, 2010; Schlomer, Bauman, & Card, 2010). If the missing data mechanism is not modeled, Yuan and Lu (2008) recommend a two stage ML procedure. However, when samples sizes are small to moderate and the asymptotic assumptions of ML are violated, Bayesian approaches are favored over EM based ML estimates (Tan, Tian, & Ng, 2010). The preferred approach at present is multiple imputation (MI), which can be used in almost any situation (Allison, 2003; Ludbrook, 2008). MI works by constructing an initial model to predict the missing data that has good fit to the observed data. The missing data are then sampled a number of times from the predicted distribution resulting in a number of potential complete datasets (higher numbers result in better estimates of imputation variance). The same analysis can then be run on each imputed dataset, and an average of all

analyses used for the overall estimate. A special formula is used to estimate variance from the imputed data, as these tend to have smaller variance than actual data (Rubin, 1987). It is important to realize that MI will not remove bias completely, but will reduce bias to a greater extent than listwise deletion or mean imputation, simply because non-responders are likely to be different (Lumley, 2010).

There are a number of packages available for performing imputation in R (Horton & Kleinman, 2007). For example, *Amelia II* (Honaker, King, & Blackwell, 2006) can impute combinations of both cross-sectional and time series data using a bootstrapping-based EM algorithm, and does provide a user-friendly GUI. Multiple imputation of mixed-type categorical and continuous data using different methods is available in the *mix* package (Schafer, 1996). Similarly *missForest* (Stekhoven & Buehlmann, 2012) allows for imputation of mixed-type data and is useful when MVN is violated as it uses non-parametric estimators. The *mi* package, and associated *mitools* package (Su, Gelman, Hill, & Yajima, 2010), impute missing data using an iterative regression approach and calculate Rubin's standard errors respectively. Multivariate Imputation by Chained Equations (*MICE*) allows for imputation of multivariate data using multiple imputation methods including predictive mean matching, Bayesian linear regression, logistic and polytomous regression, and linear discriminant analysis (van Buuren & Groothuis-Oudshoorn, in press). Fully conditional specification (FCS), as implemented in *MICE*, has demonstrated better performance than two-way imputation in maintaining structure among items and the correlation between scales under the MCAR assumption, and should work well under the MAR assumption (van Buuren, 2010). Allison (2003) recommends a sensitivity analysis following imputation to explore the consequences of different modeling assumptions. Seeing as *MICE* allows users to program their own imputation functions, this theoretically allows for sensitivity analysis of different missingness models (Horton & Kleinman, 2007). This can be done after choosing a model and estimation method by 1) calculating parameter estimates with complete cases (n_c), 2) sample n_c cases randomly from the complete imputed dataset, calculating sample estimates each time, 3) repeat step 2 a number of times to capture variation in parameter estimates, 4) compare the complete case parameter estimate to those obtained from subsamples. If the parameter estimates



vary significantly, the missingness mechanism is unlikely to be MCAR (Jamshidian & Mata, 2007).

Researchers should carefully evaluate, and report to readers, their decision-making process in dealing with distributional assumptions, outliers, and missing data. Gao, Mokhtarian, and Johnston (2008) suggest that researchers identify and remove outliers that most impact on a sample's multivariate skewness and kurtosis; finding an appropriate balance between full data that could generate an untrustworthy model, and a trustworthy model with limited generalizability due to excluded values. Various estimation methods should be used when trying to identify outliers, and triangulated analysis is recommended when potential outliers are identified not resulting from gross human error involving: analysis of data as collected, analysis using a scalable robust covariance matrix with high breakdown point, and analysis in which suspected outliers are excluded. Furthermore, when distributional assumptions have been violated FA estimators with greater robustness like the Minimal Residuals (MINRES), Asymptotically Distribution Free (ADF) generalized least-squares for large sample sizes, or Continuous/Categorical Variable Methodology (CVM) techniques should be compared to the performance of the default ML procedure (Jöreskog, 2003; Muthén & Kaplan, 1985).

Question 3: Are separate analyses on different groups indicated?

Fabrigar et al. (1999) suggest that the sample should be heterogeneous in order to avoid inaccurate low estimates of factor loadings. However, reduced homogeneity attributable largely to group differences may artificially inflate the variance of scores. Researchers should examine for significant differences in performance between homogeneous groups within the sample, and perform separate factor analyses for significantly different groups before attempting FA on the entire sample group. When distributional assumptions have been met, an analysis of variance (ANOVA) may be performed with different groupings. Erceg-Hurn and Mirosevich (2008) recommend the ANOVA-type statistic (ATS), also called Brunner, Dette, and Munk (BDM) method, as a robust alternative when distribution assumptions are violated. ATS tests the null hypothesis that the groups being compared have identical distributions, and that their relative treatment effects are the same (Wilcox, 2005). McKean (2004), and Terpstra and McKean (2005), suggest R routines

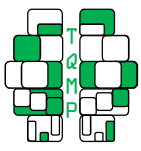
for the *weighted Wilcoxon techniques* (WW) providing a useful option for testing linear models when normality assumptions are violated or there are outliers in both the x- and y-spaces. When the question of a priori group analysis has been resolved adequately, the ensuing FA will be more robust and empirically supported.

Question 4: Do correlations support factor analysis?

The correlation matrix should give sufficient evidence of mild multicollinearity to justify factor extraction before FA is attempted. Mild multicollinearity is demonstrated by significant moderate correlations between each pair of variables. Field (2009) suggests that if two variables correlate higher than .80 one should consider eliminating one from the analysis. The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy for the R-matrix can be used to examine whether the variables are measuring a common factor as evidenced by relatively compact patterns of correlation. The KMO provides an index for comparing the magnitude of observed correlation coefficients to the magnitude of partial correlation coefficients with acceptable values ranging from 0.5 to 1 (Hutcheson & Sofroniou, 1999). Bartlett's test of sphericity is used to test whether the correlation matrix resembles an identity matrix, where off diagonal components are non-collinear. A significant Bartlett's statistic (χ^2) suggests that the correlation matrix does not resemble an identity matrix, that is correlations between variables are the result of common variance between variables. Good practice suggests that the correlation matrix should routinely be used as a prerequisite indicator for factor extraction. Though many researchers already include FA as the method of data analysis at the proposal stage, it remains a theoretical supposition that has to be supported empirically by the data. Using this particular guiding question will assist researchers in applying FA more judiciously.

Question 5: Is FA or PCA more appropriate?

Principle components analysis (PCA) is one of the most popular methods of factor extraction, appearing as the default procedure in many statistical software packages. However, PCA and FA are not simply different ways of doing the same thing. FA has the goal of accurately representing off-diagonal correlations among variables as underlying latent dimensions, has indeterminate factor scores, and generates parameter estimates that should remain stable even if batteries of



manifest variables vary across studies. PCA, on the other hand, has the goal of explaining as much of the variance in the matrix of raw scores in as few components as possible, has determinate component scores, systemically uses overestimates of communality (i.e. unity, all standardized variance), and emphasizes differences in the qualities of scores for individuals on components rather than parameters, which in PCA do not generalize beyond the battery being analyzed (Widaman, 2007). They may produce similar results when the number of manifest variables and pairwise differences between unique variances relative to the lengths of the loading vectors are small (Schneeweiss, 1997). But empirical evidence suggests they often lead to considerably different numerical representations of population estimates (Widaman, 1993). In most psychological studies researchers are interested in defining latent variables generalizable beyond the current battery, and acknowledge that latent dimensions are likely to covary in the sample even if not in the population; in such cases FA is more appropriate than PCA (Costello & Osborne, 2005; Preacher & MacCallum, 2003; Widaman, 2007).

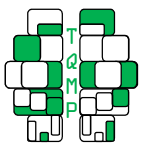
Question 6: Which factor extraction method is best suited?

Factor analysis models are approximations of reality susceptible to some degree of sampling and model error. Different models have different assumptions about the nature of model error, and therefore perform differently relative to the circumstances under which they are used (MacCallum, Browne, & Cai, 2007). The ML method of factor extraction has received good reviews as it is largely generalizable, gives preference to larger correlations than weaker ones, and the estimates vary less widely around the actual parameter values than do those obtained by other models (Fabrigar et al., 1999). However, ML is sensitive to skewed data and outliers (Briggs & MacCallum, 2003). Ordinary Least Squares (OLS) and Alpha factor analysis (extracts factors that exhibit maximum coefficient alpha) have a systematic advantage over ML in being proficient in recovering weak factors even when the degree of sampling error is congruent with ML assumptions, or when the amount of such error is large, and produce fewer Heywood cases [borderline estimations] (Briggs & MacCallum, 2003; MacCallum, Tucker, & Briggs, 2001; MacCallum et al., 2007). Two other methods that have received favorable reviews for coping with small sample sizes and many variables

while not being as limited by distributional assumptions are Minimum Residuals (MINRES) and Unweighted Least Squares (ULS), which are in most accounts equivalent (Jöreskog, 2003). The MINRES algorithm is similar in structure to ULS except that it is based on the principle of direct minimization of the least squares, rather than the minimization of eigenvalues of the reduced correlation matrix in ULS. Finally, image analysis is useful when factor score indeterminacy is a problem, and reduces the likelihood of factors that are loaded on by only one measured variable (Thompson, 2004). Multiple analyses should be performed using different extraction techniques, and differences in outcomes interpreted based on the assumptions and statistical properties of each method. However, avoid data torturing - selecting and reporting only those results that meet favored hypothesis (Mills, 1993).

Question 7: How many dimensions should I retain?

This question has possibly generated the most heated critique and comment by factor analytic theorists, and is often implemented using poor decision-making criteria (Thompson, 2004). Kaiser is cited by Revelle (2006) as saying “solving the number of factors problem is easy, I do it everyday before breakfast. But knowing the right solution is harder.” The most common methods for deciding the number of factors to extract are “Kaiser’s little jiffy” and the scree test. “Kaiser’s little jiffy”, or the eigenvalue greater than one rule, became the default option on many statistical software packages because it performed well with several classic data sets and because of its easy programmability on the first generator computer, Illiac (Gorsuch, 1990; Widaman, 2007). It is unreliable, sometimes leading to over-extraction and at other times under-extraction (Thompson, 2004). Cattell (1966) proposed the “scree test” as a subjective method of identifying the number of factors to extract. A scree plot graphs eigenvalue magnitudes on the vertical axis and factor numbers on the horizontal axis. The values are plotted in descending sequence and typically consist of a slope that levels out at a certain point. The number of factors is determined by noting the point above a corresponding factor number at which the line on the scree plot makes a sharp demarcation or ‘elbow’ towards horizontal. It has been criticized mostly for poor reliability, as even among experts, interpretations have been found to vary widely (Streiner, 1998). In an effort to remedy this Nasser,



Benson, and Wisenbaker (2002) suggested regression analyses as a less subjective method of determining the position of the elbow on the scree plot.

A number of statistically based alternatives for determining the number of factors are available. Parallel Analysis, originally proposed by Horn (1965), has been described by several authors as one of the best methods of deciding how many factors to extract, particularly with social science data (Hoyle & Duvall, 2004). Parallel analysis creates eigenvalues that take into account the sampling error inherent in the dataset by creating a random score matrix of exactly the same rank and type of variables in the dataset. The actual matrix values are then compared to the randomly generated matrix. The number of components, after successive iterations, that account for more variance than the components derived from the random data are taken as the correct number of factors to extract (Thompson, 2004). Velicer's Minimum Average Partial (MAP) test has also been received well (Stellefson & Hanik, 2008). It progresses through a series of loops corresponding to the number of variables in the analysis less one. Each time a loop is completed, one more component is partialled out of the correlation between the variables of interest, and the average squared coefficient in the off-diagonals of the resulting partial correlation matrix is computed. The number of factors to be extracted equals the number of the loop in which the average squared partial correlation was the lowest. As the analysis steps through each loop it retains components until there is proportionately more unsystematic variance than systematic variance (O'Connor, 2000). These procedures are complementary in that MAP averts over-extraction (Gorsuch, 1990), while Parallel Analysis avoids under-extraction (O'Connor, 2000). Another approach is to maximize interpretability of the solution. The Very Simple Structure (VSS) criterion works by comparing the original correlation matrix to one reproduced by a simplified version of the original factor matrix containing the greatest loadings per variable for a given number of factors. VSS tends to peak when the solution produced by the optimum number of factors is most interpretable (Revelle & Rocklin, 1979). Lastly, calculating and comparing the goodness-of-fit statistics calculated for FA models from 1 to the theoretical threshold number of factors provides a *post hoc* method of determining the best number of factors to extract (Friendly, 1995; Moustaki, 2007). There are currently a number of well supported model fit indexes

available (Hu & Bentler, 1998). This approach can also be used to select variables for factor analysis models (Kano, 2007). Fabrigar et al. (1999) argue that many of the model fit indexes currently available have been extensively tested using more general covariance structure models, and there is a compelling logic for their use in determining number of factors in EFA.

Gorsuch (1983) recommended that several analytic procedures be used and the solution that appears consistently should be retained. To this end, Parallel Analysis, Velicer's MAP test, the VSS criterion, and *post hoc* analysis of the goodness-of-fit statistics should be used side-by-side to determine the appropriate number of factors to extract.

Question 8: Which type of rotation is most appropriate?

Rotation is used to simplify or clarify the unrotated factor loading matrix, which allows for theoretical interpretation but does not improve the statistical properties of the analysis in any way (Lorenzo-Seva, 1999). Orthogonal rotation methods, such as Varimax, Quartimax and Equamax, do not allow factors to correlate (even if items do in reality load on more than one factor). They produce a simple, statistically attractive and more easily interpreted structure that is unlikely to be a plausible representation of the complex reality of social science research data (Costello & Osborne, 2005). Oblique rotation approaches, such as Direct Quartimin, Geomin, Promax, Promaj, Simplimax, and Promin, are more appropriate for social science data as they allow inter-factor correlations and cross-loadings to increase, resulting in relatively more diluted factor pattern loadings (Schmitt & Sass, 2011). As an artifact of the days of performing rotation by hand, some oblique procedures, such as Promax, attempt to indirectly optimize a function of the reference structure by first carrying out a rotation to a simple reference structure using an approach such as Varimax. Such orthogonal-dependant procedures struggle when there is a high correlation between factors in the true solution. Other approaches, such as Direct Quartimin and Simplimax, are able to rotate directly to a simple factor pattern, can deal with varying degrees of factor correlation, and give good results even with complex solutions (Browne, 2001). Two of the most powerful of these are Simplimax and Promin (Lorenzo-Seva, 1999). Jennrich (2007) suggests that to a large extent the rotation problem has been solved, as there are very simple, very general, and reliable algorithms for orthogonal and oblique rotation. He states "In a sense

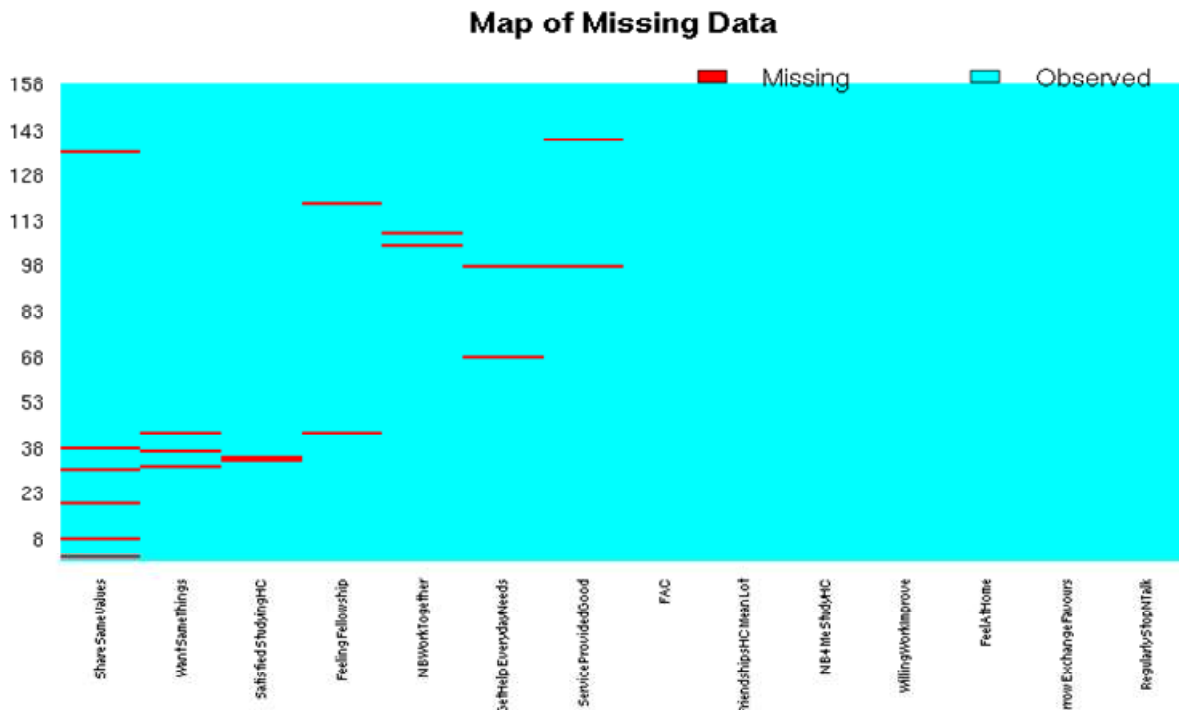


Figure 1 ■ Map of missing data in the original dataset

the Browne and Cudeck line search and the Jennrich gradient projection algorithms solve the rotation problem because they provide simple, reliable, and reasonably efficient algorithms for arbitrary criteria” (p. 62). Seeing as several orthogonal and oblique rotation objective functions from several different approaches are available, and different rotation criteria inversely affect cross-loadings and inter-factor correlations, researchers should investigate and compare results from several rotation methods (Bernaards & Jennrich, 2005; Schmitt & Sass, 2011).

Question 9: How should I interpret the factors, what should I name them?

The process of naming factors involves an inductive translation from a set of mathematical rules within the FA model into a conceptual, grammatical, linguistic form that can be constitutive and explanatory of reality. The common FA model allows for an infinite number of latent common factors, none of which is mathematically incorrect, and is therefore fundamentally indeterminate. Most factor-solution strategies have been specifically developed to detect structure which can be interpreted as explaining common sources (Rozeboom, 1996). For some this process is reminiscent of the most suggestive practices in

psychometrics (Maraun, 1996), while others describe it as a poetic, theoretical and inductive leap (Prett, Lackey, & Sullivan, 2003). Tension between these camps can be significantly reduced when researchers understand and use language that explains factors as similes, rather than metaphors, of reality. Researchers must be aware that factors are not unobservable, hypothesized, or otherwise causal underlying variables, but rather explanatory inductions that have a particular set of relationships to the manifest variates.

Factor names should be kept short, theoretically meaningful, and descriptive of the relationships they hold to the manifest variates. The factor loadings of the known indicators are used to provide a foundation for interpreting the common properties or attributes that these indicators share (McDonald, 1996). The items with the highest loadings from the factor structure matrix are generally selected and studied for a common element or theme that represents the theoretical or conceptual relationship between those items. Rules of thumb suggest suggest between 0.30 and 0.40 for the minimum loading of an item, but such heuristics fail to take the stability and statistical significance of estimated factor pattern loadings into account (Schmitt & Sass, 2011). For this reason standard errors and confidence intervals of rotated loadings should be

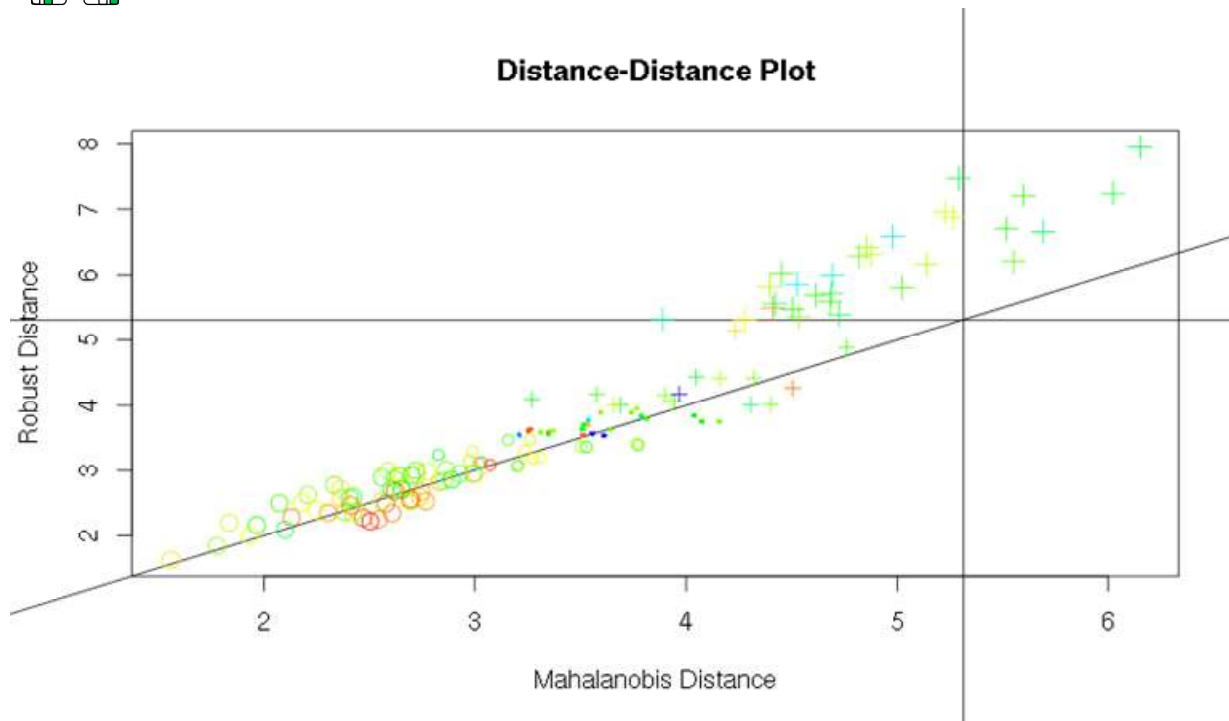


Figure 2 ■ Distance-Distance plot used to identify multivariate outliers

calculated when interpreting (Browne, Cudeck, Tateneni, & Mels, 2008). Standard errors of rotated loadings can be used in EFA to perform hypothesis tests on individual coefficients, test whether orthogonal or oblique rotations fit data best, and compute confidence intervals for parameters (Cudeck & O'Dell, 1994). For example it is possible for a larger loading derived using a rotation criteria producing small cross-loadings to be statistically non-significant (could be 0 in the population) but a smaller loading on a criterion favoring smaller inter-factor correlations to be statistically significant (Schmitt & Sass, 2011). A number of asymptotic methods based on linear approximations exist for producing standard errors for rotated loadings (Jennrich, 2007). Work is also underway in developing algorithms without alignment issues using bootstrap and Markov-chain-Monte-Carlo (MCMC) methods (eg. Zientek & Thompson, 2007). When using MI, either the EFA model can be calculated on the pooled correlation matrix of imputations, or separate EFA loading estimates are calculated for each imputation, and these estimates then pooled together. The standard errors calculated from these parameter estimates must be corrected to take into account the variation introduced through imputation (van Ginkel & Kiers, 2011). Although a highly subjective process, interpretation is guided by both the statistical, and theoretical or conceptual context of the analysis.

A Research Example

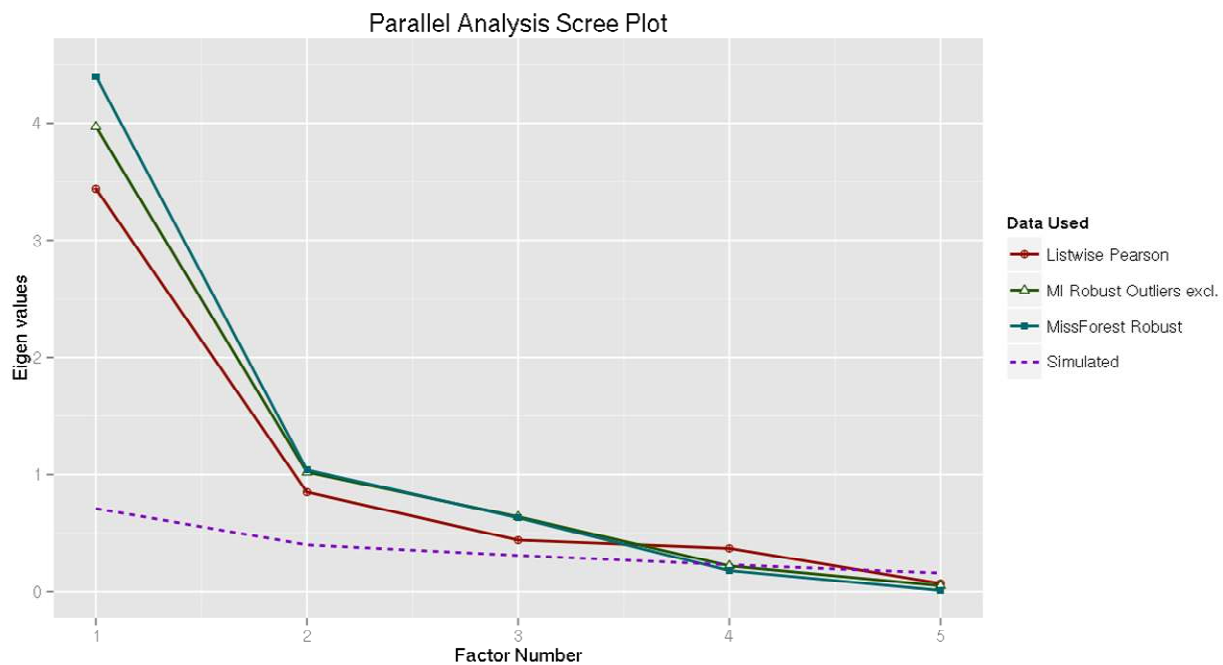
The data used in this example was collected by community psychology students using a self-report questionnaire designed during an intervention aimed at increasing the sense of community among students at a small Christian College. A selection of thirteen seven-point Likert-type items from the survey used to measure sense of community and one demographic variable were used for this example. The distribution of responses on a number of items was significantly skewed, prejudicing the use of parametric statistics. As is common in social science research there were a number of questionnaires with a few missing responses. The greatest fraction missing for any one variable was 0.037, and seven of the fourteen variables had absolutely no missing values. Listwise deletion would result in a sample size of 141, compared to 158 when missing values are imputed. Figure 1 demonstrates the pattern of missing data across participants and variables.

In addition to missing values, a number of multivariate outliers were detected. Using various methods the number of outliers identified ranged from 1 to 32. Seeing as MCD, MVE and similar methods break down and overestimate the number of outliers with high dimension data, a projection algorithm was used with restrictions on the rate of outliers identified.

Table 1 ■ Suggested number of factors to retain

Method	Pearson listwise	MI Robust outliers excl.	Forest Robust
“Little Jiffy” *	1	2	2
PA	4	3	3
MAP	1	2	3
VSS	3	1	1
RMSR	3 Factors = 0.05	3 Factors = 0.05	3 Factors = 0.05

* Number of factors with eigenvalues greater than 1 (Not recommended)


Figure 3 ■ Comparison of scree plots produced by parallel analysis using correlations from different methods

After comparing a number of methods, seven outliers were identified and correlation matrices computed using Pearson correlation coefficients with listwise deletion, polychoric and robust estimators using imputed data, and the same set of estimators with imputed data where outliers had been excluded prior to imputation. Pearson correlation matrices correlated strongly with polychoric correlations using imputed data ($r = 0.98$), but not as strongly with robust correlation estimates for imputations using *mice* and *mi* ($r = 0.83$) or *missForest* ($r = 0.81$). All methods resulted in stronger correlation estimates on average than Pearson listwise estimates, with robust procedures using data with missing values imputed using the non-parametric *missForest* being strongest

(mean difference of 0.06). For example, between variables nine and eleven the Pearson correlation was only slightly larger in the imputed datasets ($r = 0.32$, $p < 0.001$) than when listwise deletion was used ($r = 0.28$, $p < 0.001$), but did increase significantly when a robust estimator was used ($r = 0.59$, $p < 0.001$). Using these alternatives resulted in a slight improvement in the overall measure of sampling adequacy ($KMO = 0.82$ vs 0.79).

If run using defaults in most software, namely “little jiffy”, one would be tempted to only extract one factor when using listwise data. However, analytical tools suggest more factors should be retained. The RMSR fit index suggested a poorer fit for the imputed datasets (0.08 at 2 factors) than the listwise estimate (0.07 at 2

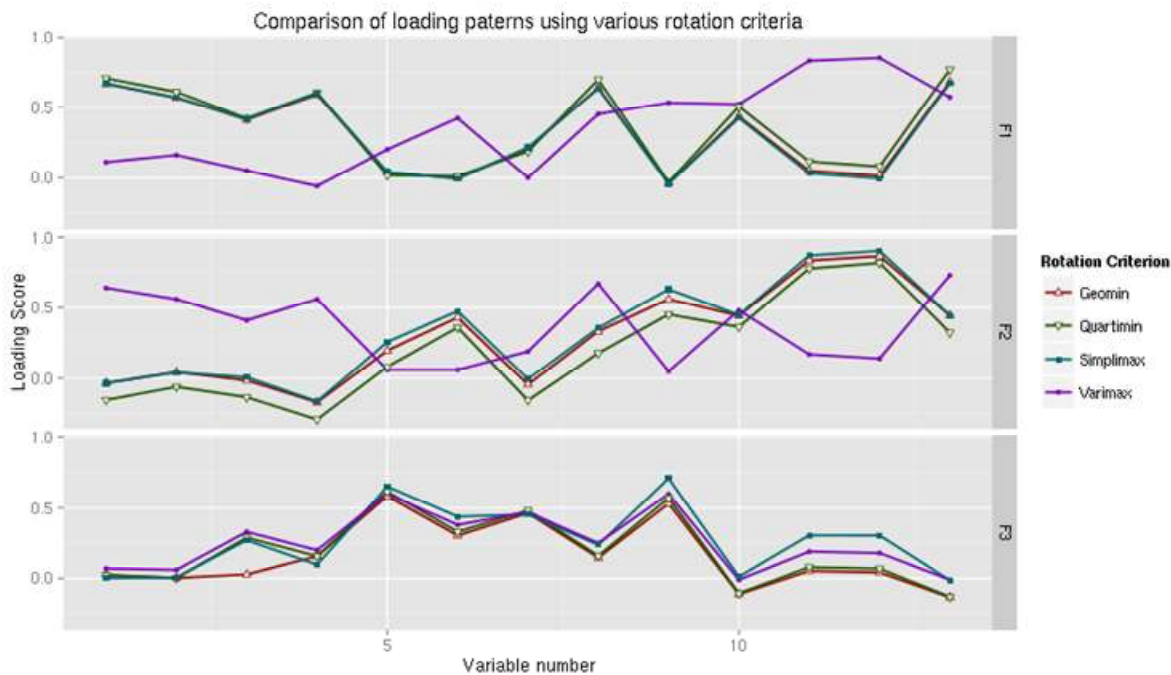


Figure 4 ■ Rotated factor loadings compared across four rotation criteria

factors) when the number of factors was 2 or less, but equal fit when 3 or more factors are retained ($RMSR = 0.05$). Estimates of the numbers of factors to extract using three correlation matrix estimates provided varying solutions summarized in table 1 and figure 3.

A three factor solution was chosen and the results of various rotation criteria inspected. As shown in Figure 4 below, the three oblique solutions produce a very similar loading pattern, but differ from orthogonal Varimax rotation that is set as a default in many statistical software programs (Varimax switches F1 and F2). When the performance of the rotation criteria are inspected by means of sorted absolute loading (SAL) plots (Jennrich, 2006; Fleming, 2012) as shown in figure 5, it appears that Simplimax delivers the best performance.

Although a three factor solution is suggested by MAP and PA and produces the highest fit indices, bootstrap standard error estimates across a number of missing value imputations suggest that loadings produced by the variables loading highly on this factor are not stable. The standard errors for the two variables loading highest on this factor were approximately 0.22 and absolute sample to population deviations over 0.15. All the other variables with a loading higher than .32 on factor one and/or two (except “ShareSameValues”) had standard errors lower than 0.153 and absolute sample to population

deviations smaller than 0.08.

Conclusion

This paper provides substantive researchers, even those without advanced statistical training, guidance in performing robust exploratory factor analysis. These analysis can easily be replicated using the R script provided. The theoretical discussion emphasizes the importance of approaching statistical analysis using an informed reasoned approach, rather than relying on the default settings and output of statistical software. The consensus arrived at in the literature reviewed is that a triangulated approach to analysis is of value. In the example provided, it was shown that while imputation had only a slight effect on the estimated correlations, using robust estimators with imputed data did increase correlation estimates overall, resulted in better sampling adequacy, a different model being specified, and a superior model fit. Combining this with estimates of rotated loading standard errors allowed the researchers to identify inconsistent structure not evident in the initial sample statistics.

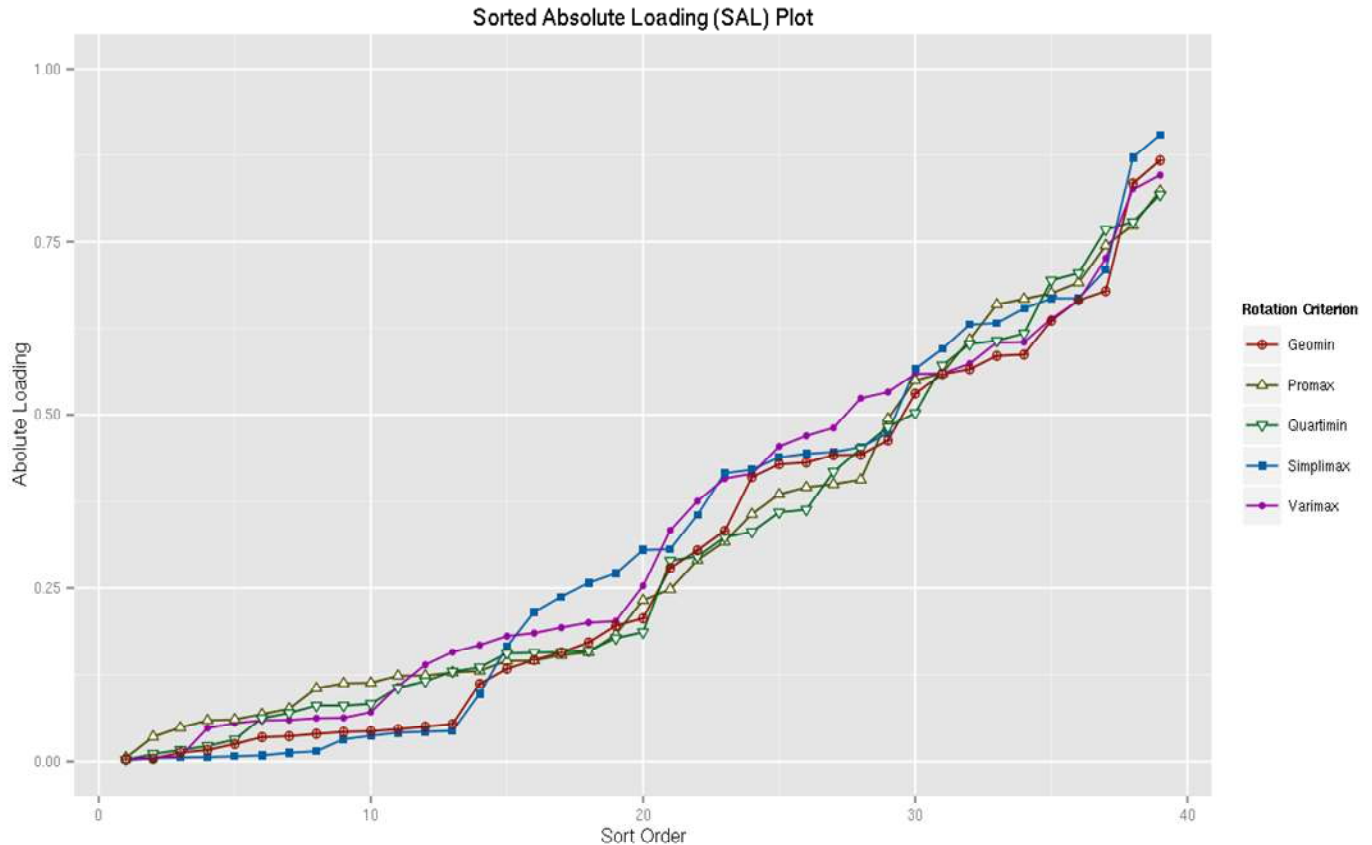
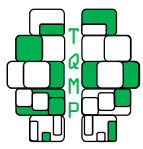


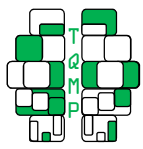
Figure 5 ■ Sorted absolute loading plot comparing loading patterns for five rotation criteria

References

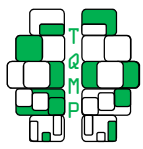
- Allison, P.D. (2003). Missing data techniques for Structural Equation Modeling. *Journal of Abnormal Psychology, 112*(4), 545-557. doi: 10.1037/0021-843X.112.4.545
- Bernaards, C.A., & Jennrich, R.I. (2005). Gradient Projection Algorithms and software for arbitrary rotation criteria in factor analysis. *Educational and Psychological Measurement, 65*, 676-696.
- Bollen, K. A. (1987). Outliers and improper solutions: A confirmatory factor analysis example. *Sociological Methods and Research, 15*, 375-384.
- Boomsma, A. (1985). Nonconvergence, improper solutions, and starting values in LISREL maximum likelihood estimation. *Psychometrika, 50*(2), 229-242.
- Box, G.E.P., & Cox, D.R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society, Ser. B, 26*, 211-252.
- Briggs, N.E., & MacCallum, R.C. (2003). Recovery of weak common factors by Maximum Likelihood and Ordinary Least Squares Estimation. *Multivariate Behavioral Research, 38*(1), 25-56.
- Browne, M.W. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research, 36*(1), 111- 150.
- Browne, M.W., Cudeck, R., Tateneni, K., & Mels, G. (2008). *CEFA: Comprehensive Exploratory Factor Analysis, Version 2.00* [Computer Software]. Retrieved from <http://faculty.psy.ohio.state.edu/browne/software.php>.
- Budaev, S.Y. (2010). Using principle components and factor analysis in animal behaviour research: Caveats and Guidelines. *Ethology, 116*, 472-480. doi: 10.1111/j.1439-0310.2010.01758.x
- Burton, A., & Altman, D. G. (2004). Missing covariate data within cancer prognostic studies: A review of current reporting and proposed guidelines. *British Journal of Cancer, 91*, 4-8.
- Cattell, R.B. (1966). The scree test for the number of factors. *Multivariate Behavioural Research, 1*, 245-276.
- Christmann, A., & Van Aeist, S. (2006). Robust estimation of Cronbach's alpha. *Journal of Multivariate Analysis, 97*(7), 1660-1674.
- Costello, A.B., & Osborne, J.W. (2005). Best practices in exploratory factor analysis: Four recommendations



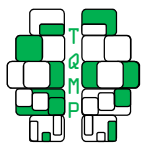
- for getting the most from your analysis. *Practical Assessment, Research and Evaluation [online]*, 10(7). Retrieved from <http://pareonline.net/getvn.asp?v=10&dn=7>
- Cudeck, R. (2007). Factor analysis in the year 2004: Still spry at 100. In R. Cudeck & R. C. MacCallum (Eds.), *Factor analysis at 100: Historical developments and future directions*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.
- Cudeck, R., & O'Dell, L. L. (1994). Application of standard error estimates in unrestricted factor analysis: Significance tests for factor loadings and correlations. *Psychological Bulletin*, 115(3), 475-487.
- Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods*, 1, 16-29.
- de Winter, J. C. F., & Dodou, D., & Wieringa, P. A. (2009). Exploratory factor analysis with small sample sizes. *Multivariate Behavioral Research*, 44, 147-181. doi: 10.1080/00273170902794206
- Dinno, A. (2009). Exploring the sensitivity of Horn's Parallel Analysis to the distributional form of random data. *Multivariate Behavioral Research*, 44, 362-388. doi: 10.1080/00273170902938969
- Erceg-Hurn, D.M., & Mirosevich, V.M. (2008). Modern robust statistical methods: An easy way to maximize the accuracy and power of your research. *American Psychologist*, 63(7), 591-601. doi: 10.1037/0003-066X.63.7.591
- 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.
- Farrell, P.J., Salibian-Barrera, M., & Naczk, K. (2006). On tests for multivariate normality and associated simulation studies. *Journal of Statistical Computation and Simulation*, 0(0), 1-14.
- Field, A. (2009). *Discovering Statistics using SPSS*. Thousand Oaks, CA: SAGE.
- Fleming, J. S. (2012). The case for Hyperplane Fitting Rotations in Factor Analysis: A comparative study of simple structure. *Journal of Data Science*, 10, 419-439.
- Gao, S., Mokhtarian, P. L., & Johnston, R.A. (2008). Nonnormality of data in structural equation models. *Transportation Research Journal*, 2082, 116-124. doi: 10.3141/2082-14
- Gorsuch, R.L. (1983). *Factor analysis* (2nd Ed.). Hillsdale, NJ: Erlbaum.
- Gorsuch, R.L. (1990). Common factor analysis versus component analysis: Some well and little known facts. *Multivariate Behavioral Research*, 25(1), 33-39.
- Hair, J.F. Jr., Anderson, R.E., Tatham, R.L., & Grablowsky, B.J. (1979). *Multivariate data analysis*. Tulsa: Petroleum Publishing Company.
- Henze, N., & Zirkler, B. (1990). A class of invariant consistent tests for multivariate normality. *Communications in Statistics - Theory and Methods*, 19, 3595-3617.
- Honaker, J., King, G., & Blackwell, M. (2006). *Amelia Software* [Web Site]. Retrieved from <http://gking.harvard.edu/amelia>
- Horn, J.L. (1965). A rational and test for the number of factors in factor analysis. *Psychometrika*, 30, 179-185.
- Horsewell, R. (1990). *A Monte Carlo comparison of tests of multivariate normality based on multivariate skewness and kurtosis*. Unpublished doctoral dissertation, Louisiana State University.
- Horton, N. J., & Kleinman, K. P. (2007). Much ado about nothing: A comparison of missing data methods and software to fit incomplete data regression models. *The American Statistician*, 61(1), 79-90. doi: 10.1198/000313007X172556
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1-55.
- Hutcheson, G., & Sofroniou, N. (1999). *The multivariate social scientist*. London: Sage.
- Hoyle, R.H., & Duvall, J.L. (2004). Determining the number of factors in exploratory and confirmatory factor analysis. In D. Kaplan (Ed.), *The SAGE handbook of quantitative methodology for the social sciences* (pp. 301-315). London: SAGE Publications.
- Jamshidian, M., & Mata, M. (2007). Advances in analysis of mean and covariance structure when data are incomplete. In S-Y. Lee (Ed.), *Handbook of latent variable and related models* (pp. 21-44). doi: 10.1016/S1871-0301(06)01002-X
- Jennrich, R. I. (2006). Rotation to simple loadings using component loss functions: the oblique case. *Psychometrika* 71, 173-191.
- Jöreskog, K.G. (2003). *Factor analysis by MINRES: To the memory of Harry Harman and Henry Kaiser*. Retrieved from www.ssicentral.com/lisrel/techdocs/minres.pdf
- Jöreskog, K.G., & Sörbom, D. (2006). *LISREL 8.8 for*



- Windows. [Computer Software]. Lincolnwood, IL: Scientific Software International, Inc.
- Kano, Y. (2007). Selection of manifest variables. In S-Y. Lee (Ed.), *Handbook of Latent Variable and Related Models* (pp. 65-86). doi: 10.1016/S1871-0301(06)01004-3
- Kerlinger, F.N. (1986). *Foundations of behavioral research* (3rd Ed.). Philadelphia: Harcourt Brace College Publishers.
- Klinke, S., Mihoci, A., & Härdle, W. (2010). Exploratory factor analysis in MPLUS, R and SPSS. *Proceedings of the Eighth International Conference on Teaching Statistics*, Slovenia. Retrieved from http://www.stat.auckland.ac.nz/~iase/publications/icots8/ICOTS8_4F4_KLINKE.pdf
- Little, R.J.A., & Rubin, D.B. (2002). *Statistical analysis with missing data* (2nd Ed.). New York: Wiley.
- Looney, S.W. (1995). How to use tests for univariate normality to assess multivariate normality. *The American Statistician*, 49(1), 64-70.
- Lorenzo-Seva, U. (1999). Promin: A method for oblique factor rotation. *Multivariate Behavioral Research*, 34(3), 347-365.
- Ludbrook, J. (2008). Outlying observations and missing values: How should they be handled? *Clinical and Experimental Pharmacology and Physiology*, 35, 670-678. doi: 10.1111/j.1440-1681.2007.04860.x
- Lumley, T. (2010). *Complex surveys: A guide to analysis using R*. Hoboken, NJ: Wiley.
- MacCallum, R. C., Browne, M. W., & Cai, L. (2007). Factor analysis models as approximations. In R. Cudeck & R. C. MacCallum (Eds.), *Factor analysis at 100: Historical developments and future directions*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.
- MacCallum, R.C., Tucker, L.R., & Briggs, N.E. (2001). An alternative perspective on parameter estimation in factor analysis and related methods. In R. Cudeck, S. du Toit, and D. Sörbom (Eds), *Structural equation modeling: Present and future* (pp. 39-57). Lincolnwood, IL: Scientific Software International, Inc.
- MacCallum, R.C., Widaman, K.F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, 4(1), 84-99.
- Maraun, M.D. (1996). Metaphor taken as math: Indeterminacy in the factor analysis model. *Multivariate Behavioral Research*, 31(4), 517-538.
- Mavridis, D., & Moustaki, I. (2008). Detecting outliers in factor analysis using the forward search algorithm. *Multivariate Behavioral Research*, 43, 543-475. doi: 10.1080/00273170802285909
- McDonald, R.P. (1996). Consensus emerges: A matter of interpretation. *Multivariate Behavioral Research*, 31(4), 663-672.
- McKean, J.W. (2004). Robust analysis of linear models. *Statistical Science*, 19, 562-570.
- Mecklin, C.J., & Mundfrom, J.D. (2005). A Monte Carlo comparison of the Type I and Type II error rates of tests of multivariate normality. *Journal of Statistical Computation and Simulation*, 75, 93-107. doi: 10.1080/0094965042000193233
- Mills, J.L. (1993). Data torturing. *New England Journal of Medicine*, 329, 1196-1199.
- Moustaki, I. (2007). Factor analysis and latent structure of categorical and metric data. In R. Cudeck & R. C. MacCallum (Eds.), *Factor analysis at 100: Historical developments and future directions*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.
- Muthén, B., & Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology*, 38, 171-189.
- Nasser, F., Benson, J., & Wissenbaker, J. (2002). The performance of regression-based variations of the visual scree for determining the number of common factors. *Educational and Psychological Measurement*, 62, 397-419.
- Nelder, J.A. (1964). Discussion on paper by professor Box and professor Cox. *Journal of the Royal Statistical Society, Series B*, 26(2), 244-245.
- O'Connor, B.P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behaviour Research Methods, Instruments, and Computers*, 32, 396-402.
- Patil, V.H., McPherson, M.Q., & Friesner, D. (2010). The use of exploratory factor analysis in public health: A note on Parallel Analysis as a factor retention criterion. *American Journal of Health Promotion*, 24(3), 178-181. doi: W.4278/ajhp.08033131
- Pison, G., Roussneuw, P.J., Filzmoser, P., & Croux, C. (2003). Robust factor analysis. *Journal of Multivariate Analysis*, 84, 145-172. doi:10.1016/S0047-259X(02)00007-6
- Preacher, K.J., & MacCallum, R.C. (2003). Repairing Tom Swift's electric factor analysis machine. *Understanding Statistics*, 2(1), 13-43.
- Prett, M.A., Lackey, N.R., & Sullivan, J.J. (2003). *Making sense of factor analysis: The use of factor analysis for*



- instrument development in health care research. London: SAGE Publications, Inc.
- R Development Core Team. (2008). *A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Available at <http://www.r-project.org>.
- Revelle, W. (2006). *Very simple structure*. Retrieved from <http://personality-project.org/r/r.vss.html>
- Revelle, W., & Rocklin, T. (1979). Very Simple Structure: An alternative procedure for estimating the number of interpretable factors. *Multivariate Behavioral Research, 14*, 403-414.
- Rogers, J.L. (2010). The epistemology of mathematical and statistical modeling. *American Psychologist, 65*(1), 1-12. doi: 10.1037/a0018326
- Rousseeuw, P.J. & Van Driessen, K. (1999). A fast algorithm for the minimum covariance determinant estimator. *Technometrics, 41*, 212-223.
- Rowe, K.J., & Rowe, K.S. (2004). *Developers, users and consumers beware: Warnings about the design and use of psycho-behavioral rating inventories and analyses of data derived from them*. International Test Users' Conference, Melbourne.
- Royston, P. (1995). Remark AS R94: A remark on algorithm AS 181: The W-test for normality. *Journal of the Royal Statistical Society. Series C (Applied Statistics), 44*(4), 547-551.
- Rozeboom, W.W. (1996). What might common factors be? *Multivariate Behavioral Research, 31*(4), 555-570.
- Russell, D.W. (2002). In search of underlying dimensions: The use (and abuse) of factor analysis in Personality and Social Psychology Bulletin. *Personality and Social Psychology Bulletin, 28*(12), 1629-1646.
- Ryu, E. (2011). Effects of skewness and kurtosis on normal-theory based maximum likelihood test statistic in multilevel structural equation modeling. *Behavioral Research Methods, 43*, 1066-1074. doi: 10.3758/s13428-011-0115-7
- Schmitt, T. A. & Sass, D. A. (2011). Rotation criteria and hypothesis testing for Exploratory Factor Analysis: Implications for factor pattern loadings and interfactor correlations. *Educational and Psychological Measurement, 71*(1), 95-113. doi:10.1177/0013164410387348
- Schafer, J.L. (1996). *Analysis of incomplete multivariate data*. London: Chapman and Hall
- Schlomer, G.L., Bauman, S., & Card, N.A. (2010). Best practices for missing data management in counseling psychology. *Journal of Counseling Psychology, 57*(1), 1-10. doi: 10.1037/a0018082
- Schneeweiss, H. (1997). Factors and principle components in the near spherical case. *Multivariate Behavioral Research, 32*(4), 375-401.
- Schönrock-Adema, J., Heinje-Penninga, M., van Hell, E.A., & Cohen-Schotanus, J. (2009). Necessary steps in factor analysis: Enhancing validation studies of educational instruments. The PHEEM applied to clerks as an example. *Medical Teacher, 31*, 266-232. doi: 10.1080/01421590802516756
- Spearman, C. (1904). General intelligence, objectively determined and measured. *American Journal of Psychology, 15*, 201-293.
- Srivastava, M.S., & Hui, T.K. (1987). On assessing multivariate normality based on the Shapiro Wilk W statistic. *Statistics and Probability Letters, 5*, 15-18.
- Stekhoven, D.J., & Buehlmann, P. (2012). MissForest - nonparametric missing value imputation for mixed-type data. *Bioinformatics, 28*(1), 112-118. doi: 10.1093/bioinformatics/btr597
- Stellefson, M., & Hanik, B. (2008). *Strategies for determining the number of factors to retain in Exploratory Factor Analysis*. Paper presented at the annual meeting of the Southwest Educational Research Association, New Orleans. Retrieved from <http://www.eric.ed.gov/PDFS/ED500003.pdf>
- Stevens, J.P. (1984). Outliers and influential data points in regression analysis. *Psychological Bulletin, 95*, 334-344.
- Streiner, D.L. (1998). Factors affecting reliability of interpretations of scree plots. *Psychological Reports, 83*, 687-694.
- Su, Y. S., Gelman, A., Hill, J., & Yajima, M. (2010) Multiple Imputation with Diagnostics (mi) in R: Opening Windows into the Black Box. *Journal of Statistical Software, 45*(2), 1-31. Retrieved from <http://www.jstatsoft.org/v45/i02/>
- Tan, M. T., Tian, G., & Ng, K. W. (2010). *Bayesian missing data problems: EM, data augmentation and noniterative computation*. Boca Raton, FL: Chapman & Hall/CRC Biostatistics Series.
- Terpstra, J. T., & McKean, J. W. (2005). Rank-based analysis of linear models using R. *Journal of Statistical Software, 14*(7), 1-26.
- Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications*. Washington, DC: American Psychological Association.
- Thurstone, L.L. (1936). The factorial isolation of



- primary abilities. *Psychometrika*, 1, 175-182.
- Tucker, L.R., & MacCallum, R.C. (1997). *Exploratory Factor Analysis*. Unpublished manuscript, Ohio State University, Columbus.
- Van Buuren, S. (2010). Item imputation without specifying scale structure. *Methodology*, 6(1), 31-36. doi: 10.1027/1614-2241/a000004
- Van Buuren, S., & Groothuis-Oudshoorn, K. (in press). MICE: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*. Retrieved from [http://www.stefvanbuuren.nl/publications/MICE in R - Draft.pdf](http://www.stefvanbuuren.nl/publications/MICE%20in%20R%20-Draft.pdf)
- van Ginkel, J. R., & Kiers, H. A. L. (2011). Constructing bootstrap confidence intervals for principal component loadings in the presence of missing data: A multiple-imputation approach. *British Journal of Mathematical and Statistical Psychology*, 64, 498-515. doi:10.1111/j.2044-8317.2010.02006.x
- Widaman, K.F. (1993). Common factor analysis versus principal component analysis: Differential bias in representing model parameters? *Multivariate Behavioural Research*, 28(3), 263-311.
- Widaman, K. F. (2007). Common factors versus components: Principals and principles, errors and misconceptions. In R. Cudeck, & R. C. MacCallum (Eds.), *Factor analysis at 100: Historical developments and future directions*. Mahway, NJ: Lawrence Erlbaum Associates, Publishers.
- Wilcox, R.R. (2008). Some small sample properties of some recently proposed multivariate outlier detection techniques. *Journal of Statistical Computation and Simulation*, 78(8), 701-712. doi: 10.1080/00949650701245041
- Wilcox, R.R. (2012). *Introduction to robust estimation and hypothesis testing* (3rd ed.). San Diego, CA: Elsevier.
- Wilcox, R.R., & Keselman, H.J. (2004). Robust regression methods: Achieving small standard errors when there is heteroscedasticity. *Understanding Statistics*, 34(4), 349-364.
- Worthington, R.L., & Whittaker, T.A. (2006). Scale development research: A content analysis and recommendations for best practices. *The Counseling Psychologist*, 34, 806-838.
- Ximénez, C. (2006). A monte carlo study of recovery of weak factor loadings in confirmatory factor analysis. *Structural Equation Modeling*, 13(4), 587-614.
- Yuan, K., & Lu, L. (2008). SEM with missing data and unknown population distributions using two-stage ML: Theory and its application. *Multivariate Behavioral Research*, 43, 621-652. doi: 10.1080/00273170802490699
- Yuan, K., & Zhong, X. (2008). Outliers, leverage observations, and influential cases in factor analysis: Using robust procedures to minimize their effect. *Sociological Methodology*, 38(1), 329-368.
- Zientek, L. R., & Thompson, B. (2007). Applying the bootstrap to the multivariate case: Bootstrap component/factor analysis. *Behavior Research Methods*, 39(2), 318-325.
- Zygmunt, C.S., & Smith, M.R. (2006). *Overview of the contemporary use of EFA in South Africa*. Paper presented and the 12th South African Psychology Congress, Johannesburg, Republic of South Africa.

Citation

- Zygmunt, C. & Smith, M. R. (2014). Robust factor analysis in the presence of normality violations, missing data, and outliers: Empirical questions and possible solutions. *The Quantitative Methods for Psychology*, 10(1), 40-55.

Copyright © 2014 Zygmunt and Smith. This is an open-access article distributed under the terms of the *Creative Commons Attribution License (CC BY)*. The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Received: 19/06/13 ~ Accepted: 05/07/13