A Modernized Heuristic Approach to Robust Exploratory Factor Analysis ⬡

Terry O. Gibson Jr. [a] ✉ ⓘ , Jennifer Ann Morrow [b] ⓘ & Louis M. Rocconi [b] ⓘ

[a]Pellissippi State Community College
[b]The University of Tennessee

Abstract ■ Exploratory Factor Analysis (EFA) is a widely used approach for exploring latent traits. The EFA literature contains many inconsistencies regarding appropriate sample sizes, remedies for violation of assumptions, and the use of rotational methods. Given that EFA is an iterative process, which involves varied decision-making, the objective of the current paper is to highlight the use of modernized methods using a heuristic approach for performing robust EFA to achieve optimal results. An example implementing these methods is presented using free online software (FACTOR). The paper concludes with seven recommendations for achieving well-defined factor solutions.

Keywords ■ Robust EFA, Polychoric, Modern EFA, Schwarz's BIC. **Tools** ■ FACTOR.

✉   tgibso10@vols.utk.edu

🔵   10.20982/tqmp.16.4.p295

Introduction

The work of Spearman (1904) on general intelligence laid the groundwork for what has become one of the most utilized statistical methods for latent construct development: Exploratory Factor Analysis (EFA Green, 2000). EFA is a complex statistical process that involves making decisions that are not necessarily objective in nature. However, these decisions can be supported by best practice recommendations in the literature. EFA is an iterative process, which has many applications across fields that allow for the exploration of relationships among observed variables (items) in a survey instrument or questionnaire. Decisions must yield solutions that are "parsimonious, mathematically sound, and theoretically grounded" (Beavers et al., 2013, p. 12).

The development of EFA allowed researchers the ability to discover latent constructs (i.e., those that are unobserved, hidden, and not directly measured), such as attitude, motivation, and personality types that must be assessed through investigations of a set of observed variables in an instrument. This exploration affords researchers the opportunity to interpret the variables of the instrument in the form of a few latent constructs or what researchers in the social and behavioral sciences refer to in some form as factors, factor solutions, factor structures, factor pat-terns (Fabrigar, Wegener, MacCallum, & Strahan, 1999; Fabrigar & Wegener, 2012; Osborne, 2014) or latent traits (Ferrando & Lorenzo-Seva, 2018). These factors are interpretable when they are parsimonious (i.e., sharing little to no variance between factors). Other terms used to describe parsimonious solutions are factor solutions that are well-defined (Beavers et al., 2013) or have achieved simple structure (Fabrigar & Wegener, 2012). In other words, most items have large loadings on one factor but small loading on other factors.

Numerous decisions are made throughout the EFA process from dealing with missing data, interpretation, and the computation of factor score estimates to be used in subsequent analyses. Other issues relate to the violation of assumptions, discrepancies regarding recommended sample sizes, the best methods to use for retaining, extracting, and rotating factors (Basto & Pereira, 2012; Costello & Osborne, 2005; Fabrigar et al., 1999; Osborne, 2014; Zumbo, Gadermann, & Zeisser, 2007). If there were ever a word that would describe the varied decision-making needed, then it would be that EFA is a heuristic approach; a problem-solving approach that involves the process of making decisions to produce optimal results.

The purpose of this paper is to present a heuristic approach to employing statistical techniques found to enhance the EFA process to derive optimal factor solutions

geared towards novice applied researchers. The focus will be to make use of modern statistical techniques that are: 1) alternatives to other methods that tend to violate assumptions; 2) found to produce better factor loadings than the more commonly used techniques; and 3) methods now accessible in readily available free, easy to use, software (e.g. FACTOR). The following sections will describe a heuristic approach specifically geared towards improving the EFA process using oblique methods.

The Process of Exploratory Factor Analysis

Prior to data collection. The heuristic approach begins before any data are collected, cleaned, analyzed, or interpreted. Poor item measurement can limit interpretability of results (DeVellis, 2012). The wording of items, the choice of response options, directions for the instrument, and the organization of items needs careful consideration. Since poor preparation can adversely affect the validity of derived factor solutions, researchers should devote a significant amount of time to item development and make use of best practices during pre-data collection (e.g., using clear and specific terminology, limiting number of concepts, including exhaustive response categories, and avoiding bias in your questions). Minimizing and measuring error (e.g., sampling or systematic) from data collected are the ultimate goals in survey development and validation (Fowler, 2014).

Collection and preparation of data for analysis. While it may seem trivial, every effort must be made to ensure that data are a representative random sample of the target population for minimizing sampling error (Fowler, 2014). It is worth mentioning that reduced sampling error lessens the influence of sample size on achieving simple structure (MacCallum, Widaman, Zhang, & Hong, 1999). Before any analysis can be performed, data preparation must include procedures for dealing with issues of missing data and addressing assumptions in EFA.

Missing data. There are three types of missing data. The probability of data missing completely at random (MCAR) is independent of both the observed and latent variables, while the probability of data missing at random (MAR) depends only on the observed variables, and the probability of data missing not at random (MNAR) depends strictly on the latent variables (Ferrando & Lorenzo-Seva, 2018). The preferred methods for handling missing data are multiple imputation (MI) and maximum likelihood estimation (ML; Enders, 2010). ML estimation uses all available information, complete and incomplete, to identify parameter estimates that have the highest probability of producing the sample data. MI is a stochastic regression method that can be used in conjunction with almost any statistical analysis (Zygmont & Smith, 2014). Imputing data is a process

of filling in missing data values with plausible values, predicted by using an appropriate model that allows for random variation. This process is repeated a certain number of times (multiple imputations) specified by the researcher (i.e., typically at least 20 or as many as 100) after which the EFA is performed on each imputed data set and parameter estimates are pooled across each imputation. MI and ML are better at reducing bias in the data and produce unbiased estimates when data are MAR (Enders, 2010). Both procedures can be implemented to address missing data in EFA (see Lorenzo-Seva & Van Ginkel, 2016; Weaver & Maxwell, 2014). Zygmont and Smith (2014) discussed in more detail an approach for assessing the missingness of data and techniques that can be used in the presence of missing data.

Measurement level. Data collected from survey questionnaires are predominately formatted with a Likert scale where participants respond to an item by indicating a level of measurement, commonly defined as a fixed 5-point Likert scale (e.g., a level of agreement from 1=strongly disagree to 3=neutral to 5=strongly agree). Generally, response scales are ordinal and collected data will most likely be skewed in one direction (Basto & Pereira, 2012). Participants may provide responses to items in which they have a high level of agreement or disagreement potentially leading to extreme skewness in a direction. Extreme skewness can lead to violations of assumptions critical in EFA (e.g., univariate and multivariate normality and linearity) with certain extraction methods (e.g., ML extraction). However, ordinal or Likert scale data are also problematic for other extraction methods that do not make distributional assumptions (e.g., Principle Axis Factoring [PAF] and Unweight Least Squares [ULS]) since these methods also assume that data are continuous. An alternative to addressing issues of measurement level, and violation of normality, linearity, and the continuous data assumptions, is to use Polychoric correlations (Basto & Pereira, 2012), which will be discussed further in the section on choosing the appropriate correlation matrix to factor analyze the data.

Assumptions. Violation of assumptions in EFA is nearly inevitable when it involves the analysis of ordinal data. Instruments using response scales are predominately used in the social and behavior sciences to assess unobserved latent traits (Furr & Bacharach, 2014). Table 1 lists recommended procedures for addressing assumptions in EFA.

Choosing the appropriate correlation matrix. In the social and behavior sciences, measurement instruments are predominately developed with Likert-type response scales. These scales produce data that are measured at the ordinal level of measurement. Performing EFA using Pearson's correlation matrix has been the predominate method for factor analyzing data given its availability in

Table 1 ■ Recommended Procedures for Addressing Assumptions in EFA

Procedures	Supporting Information
1) Check pairwise correlations of all measured variables	There must be some degree of correlation between variables in order to detect the existence of any factors. Pairwise correlations should be non-zero with sizable correlations greater than |.3| (Tabachnick & Fidell, 2013).
2) Check for the presence of multicollinearity and singularity	Data that are too highly correlated can distort the nature of the factor solutions. Correlations should be < .90 and the variance inflation factors should be < 5 (Tabachnick & Fidell, 2013).
3) Check Bartlett's Test of Sphericity	The test is a measure of factorability and tests whether the bivariate correlations of all variables, as a whole, differ significantly from zero (i.e., the correlation matrix differ from the identity matrix). The test should be significant given that the null assumes that correlations are equal to zero (Beavers et al., 2013; Tabachnick & Fidell, 2013).
4) Check Kaiser-Meyer-Olkin (KMO) test	The KMO is a test that measures how well data are suited for EFA by measuring the proportion of common variance among the variables. A suitable measure of sampling adequacy is a KMO of at least .80 (Beavers et al., 2013).
5) Assess multivariate normality and linearity	Since multivariate normality implies linearity, violation of linearity is more likely to occur, but not necessarily the reverse. Normality and linearity can be assessed through standardized residual graphics (i.e., histograms and the Normal P-P plot). When the univariate distributions of the variables are asymmetric with excess of kurtosis, Polychoric correlations are preferred, which will eliminate assessing for these assumptions specifically for EFA (Baglin, 2014) or use PAF or ULS.

commonly used commercial statistical software (e.g., SPSS and SAS). Conceptually speaking, Pearson's correlation matrix should not be used to factor analyze ordinal data due to the fact that, to use Pearson's correlation matrix, the data must satisfy the continuous data assumption, which the data must be measured at either the interval or ratio levels of measurement.

An alternative approach, as mentioned previously, is to use the Polychoric correlation matrix; a result of the Underlying Variables Approach (UVA) in ordinal factor analysis (see Ferrando & Lorenzo-Seva, 2013; Moustaki, Joreskog, & Mavridi, 2004). Furthermore, Polychoric correlations are preferred over Pearson's for several reasons: 1) Polychoric correlations take into account the fact that response scale items are ordinal given that the distribution of data will most likely be asymmetric and violate the multivariate normality assumption (Zumbo et al., 2007). 2) Polychoric correlations control for "grouping and transformation error" prevalent under Pearson's correlations provided that latent variables have bivariate normal distribution (Morata-Ramirez & Holgado-Tello, 2013). 3) Polychoric correlations are better at producing optimal fit to the theoretical model than Pearson's, especially when data are in excess of skewness and kurtosis (Gaskin & Happell, 2014). And, 4) Pearson's correlations usually underestimate parameter estimates (Ferrando & Lorenzo-Seva, 2013).

Notably, in Common Factor Model (CFM) theory (i.e., the traditional EFA approach), observed variables are a linear combination of the unobserved latent variables that are assumed to be continuous (MacCallum et al., 1999). On the other hand, this is not the case for ordinal data. The theory supporting the use of Polychoric correlations employs a two-step process for justifying the existence of the observed response variable scores of participants (Lorenzo-Seva & Van Ginkel, 2016). For example, we denote these observed response variable scores as y_{ij}, where i is the case number of cases (participants) and j is the item number of items (observed variables). At the first level, y_{ij} is realized through a "categorization" process of the underlying response variables, which is known as the Underlying Variables Approach (UVA; Ferrando & Lorenzo-Seva, 2013). We denote these variables as $y_j{}^*$. The assumption is that $y_j{}^*$ is continuous and is unobserved and latent in the UVA factor analysis model for ordinal data (Moustaki et al., 2004).

At the second level, $y_j{}^*$ satisfies assumptions posed by classical CFM theory regarding normality and linearity of continuous data. In the categorization process, y_{ij} can be thought of as a function of piece-wise defined constant values that are representative of the response categories in

the scale, which relates y_{ij} to $y_j{}^*$ by threshold parameters (Moustaki et al., 2004). For a 4-point response scale, the following holds:

$$y_{ij} = \begin{cases} 1 \Leftrightarrow & y_j{}^* < \tau_{j1} \\ 2 \Leftrightarrow & \tau_{j1} \leq y_j{}^* < \tau_{j2} \\ 3 \Leftrightarrow & \tau_{j2} \leq y_j{}^* < \tau_{j3} \\ 4 \Leftrightarrow & \tau_{j3} \leq y_j{}^* \end{cases}$$

where τ_j are threshold parameters associated with the observed variable j. It is this "step process" that defines the relationship between y_{ij} and $y_j{}^*$, for there is always one less threshold parameter than response categories. Interestingly, the Polychoric correlations are estimated from the bivariate correlations between $y_j{}^*$ (i.e., the latent unobserved variables in the UVA model; Basto & Pereira, 2012). Ferrando and Lorenzo-Seva (2013) provide more detail on the "reparameterization" of the parameters of the non-linear UVA model that transforms to the two-parameter normal ogive item characteristic curve in Item Response Theory (IRT); an approach used to justify the non-linear UVA model for ordinal data. Noteworthy, when there is clear violation of skewness and kurtosis, Polychoric correlations are recommended (Basto & Pereira, 2012; Ferrando & Lorenzo-Seva, 2018). However, when observed variables are clearly symmetric with skewness and kurtosis values less than one in absolute value, only then are Pearson's correlations suitable (Ferrando & Lorenzo-Seva, 2013; Muthen & Kaplan, 1992). A detailed discussion of the UVA approach for ordinal factor analysis can be found in Choi, Peters, and Mueller (2010) or Ferrando and Lorenzo-Seva (2013).

Determining the number of factors to retain. Factor retention has long been a controversial issue (Courtney, 2013; Osborne, 2014). Nevertheless, many researchers currently use the more traditional methods (e.g., Kaiser's eigenvalue greater than one criterion or Cattell's Scree test) rather than one of the more modern methods (e.g., Parallel Analysis[PA], Velicer's Minimum Average Partial[MAP], or Next Eigenvalue Sufficiency Test[NEST]) given the lack of access to the more modern retention methods in popular software packages such as SPSS and SAS. However, one method in particular has been shown to be theoretically suitable for retaining factors derived from ordinal data when using Polychoric correlations (Neath & Cavanaugh, 2012). The method is Schwarz Bayesian Information Criterion (BIC) dimensionality test.

BIC is the more viable option considering it is a more robust simplistic model selection technique. BIC produces more reliable results for smaller sample sizes and is a theoretically justifiable option given that Polychoric correlations can be computed using Bayes method of model selection (Choi et al., 2010). Determining the number of factors remains an unresolved issue in psychometrics given that

no one method is flawless in use (Osborne, 2014). Nonetheless, other methods for determining the number of factors to retain can be used in the event that retained factors advised by BIC does not reflect research or theory. BIC is a feature of FACTOR and will be used to recommend the number of factors to retain.

Choosing appropriate methods of extraction and rotation. Unique to EFA are the extraction and rotation methods. Conceptually, the selection of a rotation method should depend on the type of extraction method used to factor analyze the data. Literature on best practices in EFA encourages researchers to use appropriate extraction and rotation methods in analysis (Beavers et al., 2013; Gaskin & Happell, 2014; Osborne, 2014; Costello & Osborne, 2005).

Methods of extraction. Theoretically, EFA seeks to extract only items with shared variance to generate unique factor solutions (Costello & Osborne, 2005). The most commonly used extraction methods in EFA are principal axis factoring (PAF), unweighted least squares (ULS), and maximum likelihood (ML; Osborne, 2014). When data satisfies multivariate normality, ML tends to produce better factor recovery (Lorenzo-Seva & Ferrando, 2013; deWinter, Dodou, & Wieringa, 2009), while data that violates multivariate normality produces better factor recovery under PAF (Fabrigar et al., 1999) and ULS (Gaskin & Happell, 2014; Lorenzo-Seva & Ferrando, 2013) since these methods make no distributional assumptions regarding the data.

Methods of rotation. Rotational methods (whether orthogonal or oblique) are implemented to aid in the interpretation of factor solutions and retention of factors. Without rotation, it would be difficult to see the underlying pattern in the solution. Orthogonal rotations are varimax, quartimax, and exquimax, while the commonly used oblique rotations are direct oblimin and promax (Beavers et al., 2013). Although these methods were designed for specific types of extractions, resulting factor solutions tend to be similar when there is clear simple structure (Osborne, 2015). While there are other types of rotation methods, the ones listed here are the most commonly used in EFA and are available in popular software packages like SPSS, SAS, and the newly developed FACTOR. Both Osborne (2015) and Basto and Pereira (2012) provide more detail regarding the purpose and mechanics of these methods. Notably, in EFA, no one method of rotation is preferred given that oblique methods tend to produce similar results (Costello & Osborne, 2005). On the other hand, when factors are indeed uncorrelated, which is a rare occurrence in the social and behavior sciences, both orthogonal and oblique methods tend to produce nearly equivalent results (Osborne, 2014).

Interpret resulting factors. Interpretation follows parsimonious solutions by assigning a name to each factor

based on grouped items. When researchers have utilized best practices to derive parsimonious solutions, factors can be interpreted using sound judgement supported by research or from any pre-conceived notions regarding the theoretical relationship between factors (Beavers et al., 2013; Floyd & Widaman, 1995). The pattern matrix is used to interpret the results of factor solutions. Interpretation of these solutions are more understandable because they have been rotated and variable loadings are standardized regression weights; implying that the standard deviation of the common variance is 1 with mean 0 (Fabrigar & Wegener, 2012; Yong & Pearce, 2013). The higher the variable loadings, the stronger the factor solution of a particular common factor. The recommended threshold is >|.32| (Tabachnick & Fidell, 2013).

Choosing a method for computing factor score estimates. Factor scores are computed following EFA analysis. There are several methods for computing factor scores that are both refined and non-refined (Grice, 2001). Distefano, Zhu, and Mindrila (2009) indicate that non-refined scores are simple to compute and involve either summing raw scores or computing averages of item scores for a specific factor, which makes interpretation straight forward. In contrast, the computation of refined scores is more intricate and involves procedures where the scores are usually standardized and generally range from -3 to 3. The more commonly used methods for computing refined scores are: Regression, Bartlett, and Anderson-Rubin (see Distefano et al., 2009). Regardless of the method used, scores are generally used to explore differences among groups.

One caution to researchers is that factor scores are indeterminant (i.e., there are infinite possible solutions, such that factor scores are only estimates and not exact – Distefano et al., 2009). The validity of interpretation using these scores depends on the strength of the model. Assessing indeterminacy should accompany factor score computation. Ferrando and Lorenzo-Seva (2018) recommend researchers use an improved version of Bayes expected a posteriori (EAP) estimate to compute factor scores for ordinal factor analysis because these scores are considered more theoretically justifiable than any other method available. The EAP approach includes the inter-correlations of the factors in the development of the prior distribution necessary to generate the posterior probability distribution so that EAP scores can be computed (i.e., Full-Informative Prior Ordinal EAP scores). The latest version of FACTOR allows researchers to compute the Full-Informative Prior Ordinal EAP scores and run bootstrap analysis to assess the generalizability of factor solutions for robustness and reliability and determinacy indices to assess indeterminacy of factor score estimates (Ferrando & Lorenzo-Seva, 2017b).

Interpretation of EAP score estimates are not as easy as those produced from the non-refined methods that does not use standardized estimates (e.g., summed or averaged scores). The EAP estimate is a standardized estimate that is interpreted in terms of the average deviation that a participants' response score is from the mean with respect to the common factor. For example, a factor score estimate of -1.5 on factor one cannot be interpreted in the same context as a factor score estimate of -1.5 on factor two. A participants' response score of -1.5 on factor one could be relative to an average of low response category scores on the group of observed variables (items) common to factor one, while the participants' response score of -1.5 on factor two could be relative to an average of high response category scores on the group of observed variables common to factor two. This independence of factor score estimate interpretation makes non-refined score estimates (i.e., non-standardized) more appealing. Nonetheless, the purpose of refined methods are to "maximize validity" using predictive techniques that increases the chance of estimates being more accurate and reliable representation of the "true" factor score (Distefano et al., 2009; Grice, 2001).

The relevance of sample size. Recommendations of minimum sample sizes are incompatible and erroneous at best (Preacher & MacCallum, 2002). According to Hogarty, Hines, Kromrey, Ferron, and Mumford (2005), the relevance of sample size is reduced when certain conditions regarding commonality and overdetermination are met. Commonality is the proportion of variance accounted for by the common factor and overdetermination refers to the number of observed variables needed to define a common factor (typically at least three with factor loadings > |.32|; Tabachnick & Fidell, 2013). When commonalities and factor loadings are moderate to high (at least .60; Zhao, 2009), common factors are relatively strong thereby reducing the influence of sample size (Fabrigar & Wegener, 2012). Therefore, inquires of minimum sample sizes a priori are not necessary given care has been taken in the development of the item set. Otherwise, larger sample sizes will be necessary to produce more stable results (Beavers et al., 2013; MacCallum et al., 1999; Osborne & Costello, 2004).

Application in FACTOR

The Data

The Motivation Assessment Scale (MAS-12) is a scale developed by the first author and used as part of a program evaluation project designed to measure the extent to which a student-centered learning environment was effective. MAS-12 is a 3-Factor questionnaire that measures the extent to which students' basic psychological needs (BPN)

Table 2 ■ Bivariate Polychoric Correlations

	1	2	3	4	5	6	7	8	9	10	11	12
AUTO1	1.00											
AUTO2	0.77	1.00										
AUTO3	0.77	0.77	1.00									
AUTO4	0.77	0.81	0.73	1.00								
AUTO5	0.76	0.75	0.75	0.84	1.00							
AUTO6	0.75	0.71	0.65	0.71	0.73	1.00						
EXTR1	0.18	0.06	0.00	0.12	0.14	0.26	1.00					
EXTR2	0.26	0.21	0.15	0.31	0.28	0.41	0.55	1.00				
EXTR3	0.41	0.32	0.26	0.39	0.36	0.51	0.72	0.54	1.00			
PERS2	0.52	0.38	0.39	0.47	0.49	0.49	0.40	0.39	0.46	1.00		
PERS1	0.64	0.55	0.47	0.58	0.56	0.57	0.24	0.26	0.44	0.71	1.00	
PERS3	0.44	0.34	0.30	0.40	0.41	0.40	0.36	0.30	0.41	0.77	0.64	1.00

were satisfied over the course of the learning experience rooted in Self-Determination Theory (SDT; Ryan & Deci, 2017). The data consist of 228 learning support math students at a community college in the southeastern United States.

The first two factors measured participants' agreement with the items on a 5-point Likert response scale (e.g., 1=strongly disagree to 3=neither agree nor disagree to 5=strongly agree). The first factor consisted of six items designed to measure high quality autonomous motivation labeled AUTO. A sample item states: "The program helped me increase my problem-solving abilities." The second factor consisted of three items designed to measure extrinsic regulatory motivation. These items were considered to be the least form of autonomous motivation and were labeled EXTR. A sample item states: "When I was in the program, my main concern was getting a good grade." And, the third factor consisted of three items designed to measure participants' motivation to persist labeled PERS. A sample item states: "Rate your motivation to complete your individualized curriculum?" These items were measured on a 6-point response scale with categories ranging from 1=none to 3=moderate to 6=very high.

The latest release of FACTOR is capable of reporting reliability indices (e.g., Greatest Lower Bound [GLB], McDonald's ordinal Omega, and Cronbach's alpha). For the current paper, only the ordinal Omega coefficients are reported for each factor. Following EFA, researchers can separate derived factor variables by moving other variables to the "Excluded" column in the "Configure Analysis" step of the process. Within that display, click on "Other specifications of factor model" to compute the Omega coefficients (Figure 2). Notably, all Omega coefficients exceeded the recommended threshold (> .70; Tabachnick & Fidell, 2013). These were: AUTO [0.948], EXTR [0.828], and PERS [0.880]. Data used in the current paper are available on the jour-

nal's web site for testing the analyses.

Prior to Analysis. Little's MCAR test for missingness was significant, $\chi^2(85) = 111.6$, $p = 0.028$. Results indicated that the data were either MAR or MNAR. Nonetheless, the chosen method for handling missing data in FACTOR is MI. An advantage to using MI is that it makes no assumptions about the missingness of data (Tabachnick & Fidell, 2013). Furthermore, less than 5% of data is missing, the distribution is asymmetric, and the sample size is modest. These are conditions favorable to the use of MI (Zygmont & Smith, 2014). The bivariate Polychoric correlations between the observed variables shown in Table 2 did not show signs of multicollinearity and singularity. While not all bivariate correlations were greater than |.3|, many were sizable correlations and less than .90. Bartlett's test of sphericity was significant, $\chi^2(66) = 1639.2$, and $p = .0001$. The proportion of common variance among the variables were suitable for EFA with a KMO test value = 0.90 (very good). A measure of the robustness of this value to generalize across samples resulted in a 95% precise Bias-corrected (BC) Bootstrap CI of the KMO [0.894, 0.921].

For the MAS-12 data, not all skewness and kurtosis values were less than one in absolute value (i.e., EXTR1, EXTR2, PERS2, and PERS3 – See Table 3). These univariate results uniquely suggested the use of Polychoric correlations. Furthermore, FACTOR produced Mardia's multivariate test of skewness and kurtosis to determine multivariate normality. The test results showed that kurtosis was significant, p < 0.01, while skewness was not, p = 1. Again, a significant kurtosis finding further supported the use of Polychoric correlations to factor analyze the data.

Analyzing data in FACTOR. FACTOR is a "user-friendly" downloadable software program available from http://psico.fcep.urv.es/utilitats/factor/index.html. FACTOR was designed specifically to run both traditional and modern procedures of EFA (Ferrando & Lorenzo-Seva, 2017b;

Table 3 ■ Univariate Descriptive Statistics

Variable	Mean (95%) CI[a]	Variance	Skewness	Kurtosis
AUTO1	3.61 [3.42, 3.79]	1.14	-0.69	-0.15
AUTO2	3.28 [3.06, 3.49]	1.51	-0.31	-0.84
AUTO3	3.32 [3.12, 3.53]	1.42	-0.35	-0.73
AUTO4	3.55 [3.35, 3.75]	1.32	-0.59	-0.38
AUTO5	3.47 [3.26, 3.68]	1.43	-0.53	-0.62
AUTO6	3.71 [3.52, 3.90]	1.18	-0.79	0.21
EXTR1	4.21 [4.08, 4.35]	0.64	-0.99	**1.04**
EXTR2	3.78 [3.61, 3.96]	1.02	-0.65	0.11
EXTR3	4.10 [3.93, 4.27]	0.95	**-1.28**	**1.48**
PERS2	5.05 [4.86, 5.24]	1.17	**-1.52**	**2.77**
PERS1	4.51 [4.26, 4.76]	2.07	-0.94	0.16
PERS3	5.14 [4.94, 5.33]	1.25	**-1.72**	**3.23**

Note. [a]: CI=Confidence Interval, AUTO = Autonomous Motivation, EXTR = Extrinsic Regulatory Motivation, PERS = Motivation to Persist

Lorenzo-Seva & Ferrando, 2006). Baglin (2014) provided steps for performing EFA in FACTOR with emphasis on ordinal data as well as provided information on downloading FACTOR. However, there have been several releases of the FACTOR software since Baglin (2014). The latest version (10.10.01) was released October 2019 and used in this paper. The site provides more information on the latest statistical techniques that can be performed in FACTOR. The current paper will discuss using FACTOR to explore the general factorability of the scale MAS-12 using modern methods of EFA.

FACTOR involves a three-step process. These are: Read Data, Configure Analysis, and Compute; accessible from the main menu of FACTOR (Figure 1). For specifics regarding the three-step process, readers are referred to the Baglin (2014) article on getting started with FACTOR. The latest version of FACTOR includes options for estimating robust factor analysis using bootstrap resamples, handling missing data, and assessing factor score estimates and replicability of factor solutions. The maximum number of bootstrap resamples were generated in FACTOR (n = 3000) for MAS-12. The missing value code used in FACTOR was 99, which was previously set during the data-cleaning phase. The configuration of analysis included the following selections: 1) the Polychoric correlation matrix to factor analyze the data, 2) the BIC dimensionality test to determine the number of factors to retain, 3) the factor model (i.e., the selection of three factors, robust factor analysis, ULS extraction method), and 4) Promax rotation. It is worth noting that FACTOR can take a long time to produce results. The amount of time it takes will depend on the configuration of specific analyses and active applications on the electronic device for which FACTOR was downloaded.

The number of factors to retain was hypothesized a priori (i.e., three factors) and were based on theoretical underpinnings that were the foundation of initial item development. Table 4 displays the results from the BIC dimensionality test, which supported a 3-Factor solution. Retained factors were determined by the smallest BIC factor (269.87). The numerical value is a criterion used to penalize the number of parameters in the statistical model that was derived, in part, from the likelihood function – such that, the smaller the BIC value the more probable the statistical model is an accurate fit for the given data (Neath & Cavanaugh, 2012). The consistency of the method to accurately choose the correct model is a strength of the test (Neath & Cavanaugh, 2012).

To aid in the assessment of optimal factor solutions and replicability of results, the Bias-corrected and Accelerated (BCa) bootstrap approach proposed by (Lambert, Wildt, & Durand, 1991) was used to generate CIs for goodness of fit indices, factor loadings and inter-factor correlations between variables. The BCa method corrected for bias and adjusted for skewness. To support the accuracy and computation of factor score estimates, the necessary minimal selections are: EAP factor scores, assess construct replicability, and assess quality of factor scores (Figure 2). Additionally, to generate factor score estimates, a file name must be entered. FACTOR will output a .dat file.

Interpretation of results. For comparison purposes, both Pearson's and Polychoric correlations were used to factor analyze the data. EFA results of MAS-12 achieved parsimonious solutions when using both Pearson's and Polychoric correlations. It is clear from Table 5 that using Pearson's correlations for this set of data underestimated the factor solutions of MAS-12. While parsimonious, factor loadings derived using Polychoric correlations are stronger and

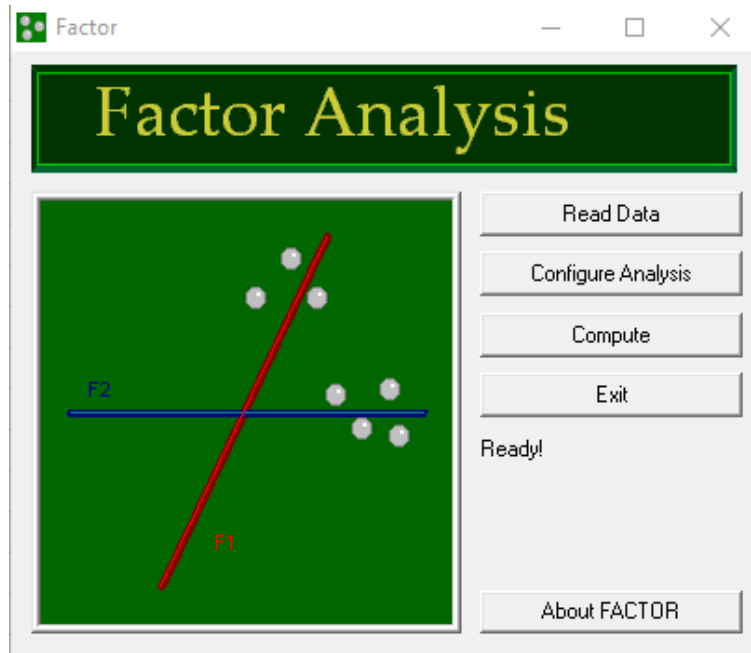**Figure 1 ■** The Main Window of FACTOR

Table 4 ■ Schwarz's BIC Dimensionality Test

Factors	BIC
0	4338.97
1	508.86
2	335.07
3	269.87*
4	330.88
5	398.63

Note. *Advised number of factors is 3

more likely to be an accurate representation of the factor structure given the apparent asymmetry of the observed variables and kurtosis issues (Ferrando & Lorenzo-Seva, 2017a). The positive factor loading differences (FLD) illustrate the strength of loadings produced using Polychoric correlations. All factor loadings from the Polychoric correlations are at least 0.6 and indicate well-defined factor solutions that are more likely to produce replicable results.

Indeterminacy. One major benefit of FACTOR is the capability to determine whether factor score estimates are accurate. The Factor Determinacy Index (FDI) is a measure of the accuracy of the factor score estimates given that these estimates in EFA are not unique (Ferrando & Lorenzo-Seva, 2017a). The development of the FDI is based on indeterminacy assessment criteria proposed by Grice (2001). FDI values > 0.90 are an indication of estimates

that are an accurate measure of individuals' "true" score response. The Overall Reliability of fully-Informative prior Oblique N-EAP scores (ORION) is an assessment of the reliability of the factor score estimates. ORION (also known as marginal reliabilities) values > 0.80 indicate precise measures of reliability of the factor score estimates (Ferrando & Lorenzo-Seva, 2018).

Construct replicability. As a measure of multidimensionality, Ferrando and Lorenzo-Seva (2017a) proposed the G-H index, where G stands for generalized and H refers to both H-latent and H-observed (i.e., H is a measure of the multiple correlations between a factor and the respective items common to the factor). H-latent is a measure of how well the factor can be measured by the unobserved (latent) variables that underlie the observed response variable scores, while H-observed is a measure of how well the

Figure 2 ■ Indices for the Assessment of the Factor Model
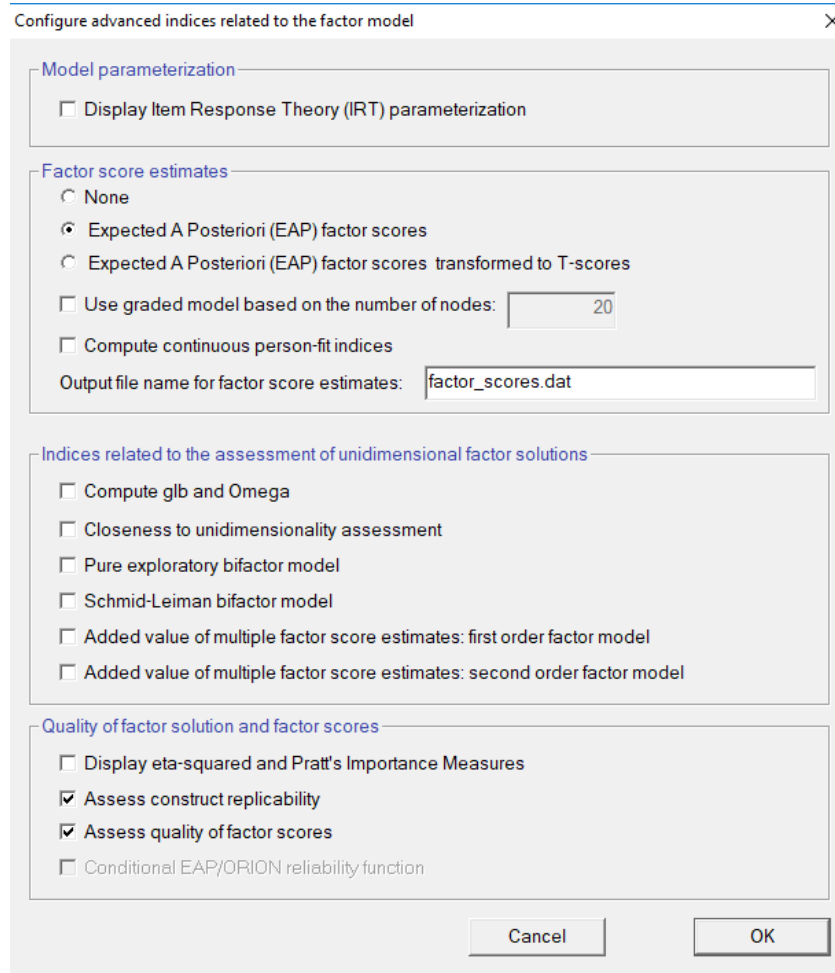
factor can be measured by the observed variables. High G-H values >.80 indicate well-defined factor solutions with stable replicable results. Table 6 consist of a list of the assessment indices for factor score determinacy and reliability and construct replicability for MAS-12.

The FDI for each factor is at least 0.94, which is a strong indication that the factor score estimates are excellent representations of the latent factor and will be highly correlated to each other with respect to the factor that each are related. Table 7 is a list of the first four participants (cases) and their predicted factor score estimates per factor.

In terms of replicability of factor solutions, the G-H indices in Table 6 are at least marginally acceptable results. The H-latent values for each factor (0.977, 0.941, and 0.952 respectively) are an indication that the unobserved latent variables that underlie the observed variables in the UVA model for ordinal factor analysis are strong representa-

tions of the respective factors (AUTO, EXTR, and PERS). However, only the H-observed value for AUTO (0.897) is evidence that the observed variables are a strong representation of the intrinsic motivation factor, while the H-observed variables for both EXTR and PERS were the same (0.789) and were marginally acceptable values for extrinsic motivation and motivation to persist. Overall, the accompany BC 95% CI supports the stability of the MAS-12 3-Factor solution to be replicable and potentially generalizable across samples.

Conclusion

The purpose of this paper was to provide novice researchers with an introduction to modern EFA approaches. The following are seven recommendations for implementing a modern heuristic approach to EFA.

1. Consider designing a study exploring fewer factors. De-

Table 5 ■ MAS-12 Comparison between the Pearson and Polychoric Correlations in FACTOR

| | Pearson (n = 228) | | | | Polychoric (n = 228) | | | | |
| | | Factor Loadings | | | | | Factor Loadings | | | |
Items	Communality	F1	F2	F3	Communality	F1	F2	F3	FLD*
AUTO1	0.729	0.808			0.790	0.823			0.015
AUTO2	0.742	0.910			0.789	0.939			0.029
AUTO3	0.669	0.877			0.737	0.919			0.042
AUTO4	0.749	0.851			0.808	0.892			0.041
AUTO5	0.741	0.852			0.780	0.864			0.012
AUTO6	0.658	0.714			0.709	0.743			0.029
EXTR1	0.679		0.853		0.791		0.924		0.071
EXTR2	0.356		0.567		0.435		0.631		0.064
EXTR3	0.639		0.760		0.724		0.790		0.030
PERS2	0.731			0.838	0.830			0.896	0.058
PERS1	0.592			0.576	0.673			0.649	0.073
PERS3	0.615			0.832	0.707			0.884	0.053
Eigen.		5.921	1.886	1.135		6.455	2.048	1.079	
Max.	0.749				0.830				
Min.	0.356				0.435				

Note. *FLD=Factor Loading Differences (positive values favor Polychoric correlations). Both Pearson and Polychoric correlations were run using ULS extraction with Promax rotation. F1=AUTO, F2=EXTR, F3=PERS

Table 6 ■ Construct Replicability of Factor Solutions and Accuracy of Factor Score Estimates MAS-12

Index	AUTO	EXTR	PERS
[a]FDI	0.977	0.941	0.952
[b]MR	0.955	0.885	0.906
	Latent	Latent	Latent
[c]G-H	0.955 [0.935 0.963]	0.885 [0.823 0.928]	0.906 [0.847 0.948]
	Observed	Observed	Observed
	0.897 [0.864 0.916]	0.786 [0.722 0.817]	0.786 [0.726 0.819]

Note. [a]: FDI = Factor Determinacy Index, [b]: MR = Marginal Reliability, [c]: G-H = Construct Replicability

signing a study with fewer factors is the only way to "manipulate" the number of factors to retain given that exploring fewer factors contributes to parsimonious solutions (Preacher & MacCallum, 2002).

2. Focus on developing well-defined items and utilizing techniques for achieving optimal response rates. Use best practices regarding survey design and construction to avoid introducing bias into the development of items (Colton & Covert, 2007; Fowler, 2014). Bias increases sources of measurement error that can contribute to low commonality coefficients (Fabrigar & Wegener, 2012). This can lead to unwanted issues (Haywood cases, cross-loadings, and variables loading on the wrong factors). Otherwise, larger sample sizes will be necessary to obtain simple structure (Fabrigar et al.,

1999; Osborne & Costello, 2004; Tabachnick & Fidell, 2013).

3. Plan to develop 7 to 10 items per factor. There is a possibility that some items may not be a good fit for the factor and must be deleted during the pilot testing phase of extracting factors. When recommendations #1 and #2 are considered, parsimonious solutions are achievable with a moderate to high degree of overdetermination and moderate to high variable loadings (>.60) even with less than recommended sample sizes.

4. Acquire basic knowledge of CFM theory. CFM aids in informed decision-making in EFA and provides conceptual understanding of the relationship between the common and unique factors to the observed variables as well as allow one to understand why error is a major

Table 7 ■ First Four Participants and Respective Factor Score Estimates per Factor

Participants	AUTOa	EXTRa	PERSa
Case1	1.411	-0.012	0.910
Case2	0.090	0.039	0.606
Case3	-1.241	-0.377	-1.351
Case4	-0.175	-0.538	-0.154

Note. [a]: Standardized estimates computed using the Fully-Informative Prior Oblique EAP scores

influence of sample size. Readers interested in learning more about CFM can find a mathematical, conceptual, or geometrical interpretation of EFA and CFM in suggested resources (e.g., Fabrigar & Wegener, 2012; MacCallum et al., 1999; Yong & Pearce, 2013).

5. Ignore sample size recommendations "a priori". Simulation studies, supported by empirical evidence, suggest that minimum sample size recommendations are useless and should be ignored given that other indicators could reduce the relevance of sample size (MacCallum et al., 1999; MacCallum, Widaman, Preacher, & Hong, 2001). Some indicators are high commonalities and variable factor loadings (at least .6; Zhao, 2009) and overdetermination (i.e., at least three but four are more variables with high factor loadings, which diminishes the relevance of sample size recommendations in general; Fabrigar et al., 1999; Hogarty et al., 2005), which is why the next recommendation is critical.

6. Choose exploratory methods appropriate for the given data. Given the asymmetric nature of ordinal data and the fact that data in the social and behavior sciences will most likely be correlated, the recommended methods for exploring the factorability of data in EFA should be carried out using oblique methods and the appropriate correlation matrix (e.g., Polychoric). Orthogonal methods are suggested only when item distributions of ordinal data are symmetric and not in excess of skewness and kurtosis (Lorenzo-Seva & Ferrando, 2013).

7. Include means for assessing factor score indeterminacy and factor stability. Given the infinite number of possible solutions of factor score estimates, there must be means for assessing the reliability and determinacy of estimates. Bootstrap resampling is an effective way to assess whether the factor structure is replicable across other samples or populations (Thompson, 2004) especially when it is not feasible to collect new data to assess the stability of factor solutions due to specific constraints faced by researchers (Bamberger, Rugh, & Mabry, 2012).

References

Baglin, J. (2014). Improving your exploratory factor analysis for ordinal data: A demonstration using factor.

Practical Assessment, Research & Evaluation, *19*, 5. Retrieved from https://pareonline.net/getvn.asp?v=19&n=5

Bamberger, M., Rugh, J., & Mabry, L. (2012). *Real world evaluation: Working under budget, time, data, and political constraints, 2nd edition*. Thousand Oaks, CA: Sage.

Basto, M., & Pereira, J. M. (2012). An spss r-menu for ordinal factor analysis. *Journal of Statistical Software*, *46*, 4. doi:10.18637/jss.v046.i04

Beavers, A. S., Lounsbury, J. W., Richards, J. K., Huck, S. W., Skolits, G. J., & Esquivel, S. L. (2013). Practical considerations for using exploratory factor analysis in educational research. *Practical Assessment, Research & Evaluation*, *18*, 6. Retrieved from http://pareonline.net/getvn.asp?v=18&n=6

Choi, J., Peters, M., & Mueller, R. O. (2010). Correlational analysis of ordinal data: From pearson's r to Bayesian polychoric correlation. *Asia Pacific Educational Review*, *11*, 459–466. doi:10.1007/s12564-010-9096-y

Colton, D., & Covert, R. W. (2007). *Designing and constructing instruments for social research and evaluation*. San Francisco, CA: John Wiley & Sons, Inc.

Costello, A. B., & Osborne, J. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment Research & Evaluation*, *10*, 7. Retrieved from http://pareonline.net/getvn.asp?v=10&n=7

Courtney, M. G. R. (2013). Determining the number of factors to retain in efa: Using the spss r-menu v2.0 to make more judicious estimations. *Practical Assessment, Research & Evaluation*, *18*, 8. Retrieved from http://pareonline.net/getvn.asp?v=18&n=8

DeVellis, R. F. (2012). *Scale development: Theory and applications (3rd ed.)* Los Angeles, CA: SAGE.

deWinter, 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

Distefano, C., Zhu, M., & Mindrila, D. (2009). Understanding and using factor scores: Considerations for the applied researcher. *Practical Assessment, Research and Evaluation, 14*(20), 1–11. Retrieved from http://pareonline.net/pdf/v14n20.pdf

Enders, C. K. (2010). *Applied missing data analysis*. New York: The Guilford Press.

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. doi:1082-989X/99/S3.00

Fabrigar, L. R., & Wegener, D. T. (2012). *Exploratory factor analysis*. New York: Oxford University Press.

Ferrando, P. J., & Lorenzo-Seva, U. (2013). *Unrestricted item factor analysis and some relations with item response theory*. Technical Report. Retrieved from http://psico.fcep.urv.es/utilitats/factor/

Ferrando, P. J., & Lorenzo-Seva, U. (2017a). Assessing score determinacy, measurement quality, and closeness to unidimensionality in exploratory item factor analysis. *Educational and Psychological Measurement, 78*(5), 762–780. doi:10.1177/0013164417719308

Ferrando, P. J., & Lorenzo-Seva, U. (2017b). Program factor at 10: Origins, development and future directions. *Psicothema*, *29*, 236–240. doi:10.7334/psicothema2016.304

Ferrando, P. J., & Lorenzo-Seva, U. (2018). Assessing the quality and appropriateness of factor solutions and factor score estimates in exploratory item factor analysis. *Educational and Psychological Measurement*, *78*, 762–780. doi:10.1177/0013164417719308

Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment*, *7*(3), 286–299. doi:10.1037/1040-3590.7.3.286

Fowler, F. J. (2014). *Survey research methods (5th ed.)* Los Angeles, CA: SAGE.

Furr, R. M., & Bacharach, V. R. (2014). *Psychometrics: An introduction (2nd ed.)* Los Angeles, CA: SAGE.

Gaskin, C. J., & Happell, B. (2014). On exploratory factor analysis: A review of recent evidence, an assessment of current practice, and recommendations for future use. *International Journal of Nursing Studies*, *51*(3), 511–521. doi:10.1016/j.ijnurstu.2013.10.005

Green, C. D. (2000). *Classics in the history of psychology*. Retrieved from http://g0o0o0gle.com/documents/2014/02/spearman-c-general-intelligence-objectively-determined-measured-1904.pdf

Grice, J. W. (2001). Computing and evaluating factor scores. *Psychological Methods*, *6*(4), 430–450. doi:10.1037/1082-989X.6.430

Hogarty, K. Y., Hines, C. V., Kromrey, J. D., Ferron, J. M., & Mumford, K. R. (2005). The quality of factor solutions in exploratory factor analysis: The influence of sample size, communality, and overdetermination. *Educational and Psychological Measurement*, *65*(2), 202–206. doi:10.1177/0013164404267287

Lambert, Z. V., Wildt, A. R., & Durand, R. M. (1991). Approximating confidence intervals for factor loadings. *Multivariate behavioral research*, *26*(3), 421–434. doi:10.1207/s15327906mbr2603_3

Lorenzo-Seva, U., & Ferrando, P. J. (2006). Factor: A computer program to fit the exploratory factor analysis model. *Behavior Research Methods*, *38*(1), 88–91. doi:10.3758/BF03192753

Lorenzo-Seva, U., & Ferrando, P. J. (2013). A comprehensive program for fitting exploratory and semiconfirmatory factor analysis and irt models. *Applied Psychological Measurement*, *37*(6), 497–498. doi:10.1177/0146621613487794

Lorenzo-Seva, U., & Van Ginkel, J. R. (2016). Multiple imputation of missing values in exploratory factor analysis of multidimensional scales: Estimating latent trait scores. *Anales de Psicología/Annals of Psychology*, *32*(2), 596–608. doi:10.6018/analysis.32.2.215161

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

Morata-Ramirez, M. A., & Holgado-Tello, F. P. (2013). Construct validity of likert scales through confirmatory factor analysis: A simulation study comparing different methods of estimation based on pearson and polychoric correlations. *International Journal of Social Science Studies*, *1*(1), 54–61. doi:10.11114/ijsss.v1i1.27

Moustaki, I., Joreskog, K. G., & Mavridi, D. (2004). Factor models for ordinal variables with covariate effects on the manifest and latent variables: A comparison of lisrel and irt approaches. *Structural Equation Modeling*, *11*(4), 487–513. doi:10.1207/s15328007sem1104_1

Muthen, B., & Kaplan, D. (1992). A comparison of some methodologies for the factor analysis of non-normal likert variables: A note on the size of the model. *British Journal of Mathematical and Statistical Psychology*, *45*, 19–30. doi:10.1111/j.2044-8317.1992.tb00975.x

Neath, A. A., & Cavanaugh, J. E. (2012). The Bayesian information criterion: Background, derivation, and applications. In *Interdisciplinary reviews: Computational statistics, 4(2* (pp. 199–203). doi:10.1002/wics.199

Osborne, J. W. (2014). *Best practices in exploratory factory analysis*. Scotts Valley, CA: CreateSpace Independent Publishing.

Osborne, J. W. (2015). What is rotating in exploratory factor analysis? *Practical Assessment, Research & Evalua-*

tion, *20*, 2. Retrieved from http://pareonline.net/getvn.asp?v=20&n=2
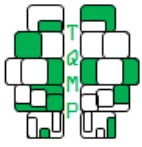
Osborne, J. W., & Costello, A. B. (2004). Sample size and subject to item ratio in principal components analysis. *Practical Assessment, Research & Evaluation*, *9*(11), 1–9. Retrieved from http://pareonline.net/getvn.asp?v=9&n=11

Preacher, K. J., & MacCallum, R. C. (2002). Exploratory factor analysis in behavior genetics research: Factor recovery with small sample sizes. *Behavior genetics*, *32*(2), 153–161. doi:10.1023/A:1015210025234

Ryan, R. M., & Deci, E. L. (2017). *Self-determination theory: Basic psychological needs in motivation, development, and wellness*. New York, NY: The Guildford Press.

Spearman, C. (1904). "General intelligence" objectively determined and measured. *American Journal of Psychology*, *15*(2), 201–292. doi:10.2307/1412107

Tabachnick, B., & Fidell, L. (2013). *Using multivariate statistics (6th ed.)* New York, NY: Routledge.

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

Weaver, B., & Maxwell, H. (2014). Exploratory factor analysis and reliability analysis with missing data: A simple method for spss users. *The Quantitative Methods for Psychology*, *10*(2), 143–152. doi:10.20982/tqmp.10.2.p143

Yong, A. G., & Pearce, S. (2013). A beginner's guide to factor analysis: Focusing on exploratory factor analysis. *Tutorials in Quantitative Methods for Psychology*, *9*(2), 79–94. doi:10.20982/tqmp.09.2.p079

Zhao, N. (2009). *The minimum sample size factor analysis*. Retrieved from https://www.encorewiki.org/display/~nzhao/The+Minimum+Sample+Size+in+Factor+Analysis

Zumbo, B. D., Gadermann, A. M., & Zeisser, C. (2007). Ordinal versions of coefficients alpha and theta for likert rating scales. *Journal of Modern Applied Statistical Methods*, *6*(1), 21–29. doi:10.22237/jmasm/1177992180

Zygmont, 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. doi:10.20982/tqmp.10.1.p040

Open practices

The *Open Material* badge was earned because supplementary material(s) are available on the journal's web site.

Citation