Exploratory factor analysis: Current use, methodological developments and recommendations for good practice

David Goretzko[1] · Trang Thien Huong Pham[1] · Markus Bühner[1]

Abstract

Psychological research often relies on Exploratory Factor Analysis (EFA). As the outcome of the analysis highly depends on the chosen settings, there is a strong need for guidelines in this context. Therefore, we want to examine the recent methodological developments as well as the current practice in psychological research. We reviewed ten years of studies containing EFAs and contrasted them with new methodological options. We focused on four major issues: an adequate sample size, the extraction method, the rotation method and the factor retention criterion determining the number of factors. Finally, we present modified recommendations based on these reviewed empirical studies and practical considerations.

Keywords Exploratory factor analysis · Sample size · Retention criteria · Rotation methods · Extraction methods

Introduction

Exploratory factor analysis (EFA) is a frequently used statistical method in psychology. There is hardly any other statistical method shaping the field of test construction as strongly as the EFA, simultaneously causing as many controversial debates about its correct application. Over the years, several publications dealt with recommendations on how to use EFA, trying to familiarize researchers with the most important decisions they have to make.

One of the most influential papers in this context, a meta-analytic review by Fabrigar et al. (1999), investigated the use of EFA in 217 papers published from 1991 through 1995 in the "Journal of Personality and Social Psychology" (JPSP) and the "Journal of Applied Psychology" (JAP). The authors made recommendations for the practical application of EFA regarding an appropriate sample size, the number of items per factor, the extraction method, the factor retention criterion as well as the rotation method and the general applicability of the procedure. In the following, these recommendations will be

discussed briefly and afterwards compared with the current use of EFA in psychological research. This is done by reviewing publications in two highly relevant journals for psychological assessment published over the last decade. The latest developments and empirical findings concerning methodological decisions during the EFA process are presented and merged with those of Fabrigar et al. (1999) to obtain enriched recommendations for EFA. We focus on sample size considerations, the choices of rotation and extraction methods as well as the best way to determine the number of factors.

Theory and Purpose of EFA

EFA is used to explore correlative relations among manifest variables and to model these relations with one or more latent variables. In the common factor model a causal link between latent variable(s) and manifest indicators is assumed ("common cause relation") – an assumption that is comprehensively discussed with all its implications by Borsboom et al. (2003). Based on the common factor model, the covariance matrix Σ of the manifest variables can be decomposed into a part of shared variance $\Lambda^T \Lambda$ (impact of the latent variable(s) or the "common cause") and unique variance Ψ^2 (Jöreskog 1967):

$$\Sigma = \Lambda^T \Lambda + \Psi^2$$

When factor loadings and unique variances are estimated, one faces the problem of rotation indeterminacy which means that the loading matrix can only be defined up to a rotation,

✉ David Goretzko
  david.goretzko@psy.lmu.de

[1] Department of Psychology, Ludwig Maximilians University Munich, Giselastr. 13, 80802 Munich, Germany

because more latent variables have to be estimated than manifest variables are observed (that is why ML estimation for example, uses an iteratively estimation procedure, see Jöreskog 1967). Steiger (1979) discusses the related issue of factor indeterminacy in detail including historical perspectives. Given a rotated solution, there are no unique solutions for factor scores – a problem that should be considered when interpreting EFA results (see also Steiger and Schönemann 1978 for a simple numerical example). When EFA is used as a tool for defining psychological constructs and developing associated questionnaires, rotation indeterminacy might be the predominant issue, so we focus on the related methodological decisions yet inviting the readers to keep in mind the problem of factor indeterminacy especially when considering the results (factor scores) for diagnostic purposes.

Recommendations of Fabrigar et al. (1999)

Fabrigar et al. (1999) established basic guidelines for the general study design, the extraction method, the rotation method and the factor retention criteria. In the following, these recommendations are presented briefly.

Study Design (Number of Items and Sample Size) Besides others, there are two important issues the researcher has to consider when designing a study – the number of variables representing one latent construct and the sample size. Fabrigar et al. (1999) suggest that one should find at least four items with acceptable reliabilities (>.70) for each expected factor. Contrary to former opinions (e.g. Gorsuch 1983; Ford et al. 1986), the authors do not support the idea of a subjects-per-variable ratio as a guiding value for the sample size. In fact, they recommend sample sizes greater than 400 as desirable, as smaller samples might yield invalid results under unfavorable conditions (e.g. low communalities; MacCallum et al. 1999).

Extraction Method When comparing different extraction methods, Fabrigar et al. (1999) conclude that Maximum Likelihood (ML) estimation might be the preferred approach due to the numerous fit indices available for this method. The authors propose three alternatives, when the assumption of multivariate normality is violated: transforming the data, correcting the fit indices or using a different method like principal axis factoring (PAF).

Factor Retention Criteria To determine the number of factors, Fabrigar et al. (1999) recommend to use different criteria and never just one method. They advise to combine fit indices (when using ML EFA) like RMSEA, as proposed by Browne and Cudeck (1992), with common methods such as parallel analysis (PA; Horn 1965). In the case of a sufficiently large sample, they encourage researchers to split the data set and compare the results of the factor retention criteria among the subsets.

Rotation Methods When it comes to the various rotation methods provided by statistical software, Fabrigar et al. (1999) have a strong call for oblique procedures as these can lead to uncorrelated and correlated factors which usually occur in psychology, whereas orthogonal rotation methods force an uncorrelated factor solution. However, they do not give any further recommendations which specific oblique rotation should be favored.

General Recommendations The paper of Fabrigar et al. (1999) has two key learnings. First, it can be seen as a strong call for a more thoughtful application of EFA. The authors also emphasized that EFA and principal component analysis (PCA) are different methods, especially with regard to unique variances, and should not be exchanged unintentionally. Second, the paper draws attention to the fact that the presentation of the method and its results often does not allow for assessing the quality of the analyses. Therefore, Fabrigar et al. (1999) emphasize the importance of transparent and coherent presentations of the whole EFA procedure and criticize the rare documentation of important EFA settings or characteristics of the data. Especially, the lack of information on item communalities is noted by the authors.

Review of the Current Use of EFA

Method

As 20 years have passed since the work of Fabrigar et al. (1999), we want to examine what has changed in the meantime and whether the discussed recommendations have been adopted by a broader community. Therefore, we sifted every original article in *Psychological Assessment* and in the *European Journal of Psychological Assessment* (EJPA) from 2007 to 2017. These journals were selected due to their special focus on test construction and the variety of studies using EFA. The database research yielded 993 studies in Psychological Assessment (issues 19(1)-29(4)) and 336 studies in EJPA (issues 23(1)-33(1)). For our analysis, we focused on articles reporting an EFA (e.g. studies on questionnaire construction) and excluded articles which did not report an EFA as a main analysis (e.g. studies only giving hints on EFA results in the footnotes). We analyzed a total of 304 EFAs, 44 from EJPA and 260 from Psychological Assessment (some papers with more than one EFA).

To quantify the current EFA practice, we classified the respective sample sizes, the extraction methods, the rotation methods, the factor retention criteria, the number of variables per factor as well as the average communalities in each EFA.

Articles directly referring to Fabrigar et al. (1999) were also considered separately, as we wanted to examine whether these articles showed a higher compliance to the presented recommendations.

Results

Study Design (Number of Items and Sample Size) Table 1 shows the sample sizes reported for each of the 304 EFAs. About half of the analyses (50.3%) were based on samples larger than 400, while only eight cases (2.6%) had samples smaller than 100. In 1.6% of all cases the sample size was not presented at all.

The ratios of variables per factor are shown in Table 2, reporting the general ratio for each EFA as well as the minimum number of items associated with a factor in each analysis. In 10.5% of the EFAs the general ratio was not provided. In nearly one third of the analyses (31.6%), the smallest number of items of a factor was not listed as well. On the other hand, more than half of the considered EFAs (52.6%) had a general item to factor ratio of five or higher with 22.0% reporting a ratio of 10 or greater. At least three variables associated with the smallest factor were reported for 57.9% of the EFAs, with 11.5% having at least six variables associated with each factor.

We were also interested in the means of communalities for each EFA as those can be seen as indicators for the quality of the measurement (when developing scales and seeking for unidimensional constructs that are represented by several manifest indicators) or rather as measures for the soundness of the extracted factors (Table 3.). The vast majority of studies neither specified the communalities nor gave enough information to calculate them (pattern matrix and correlation among factors). Thus, 87.5% of all EFAs were published neglecting the communalities. When item communalities were reported, they mostly fell between .40 and .70 (10.5% of all EFAs).

Extraction Method The present usage of extraction methods is shown in Table 4. It should be noted that PCAs are excluded from this review. Therefore, only EFAs in the narrow sense, those allowing for unique variances, are included. With 51.3%, the majority of EFAs was based on PAF, followed

Table 2 Item to factor ratio and minimum of variables per factor

Item to factor ratio	N	%
< 3:1	10	3.3
3:1–5:1	102	33.6
5:1–7:1*	47	15.5
> 7:1	113	37.2
Not Reported	32	10.5
Minimum of variables per factor of each EFA		
< 3	32	10.5
3–5	141	46.4
> 5	35	11.5
Not Reported	96	31.6

* The item to factor ratio of exactly five to one is included here

by ML estimation (16.4%). Least-Squares approaches made up less than 10 % of the used extraction methods. In more than 22% of the cases the extraction method was not reported at all.

Rotation Methods Table 5 shows the rotation methods used in the analyzed EFAs. As two of the EFAs were conducted using two different rotation methods for comparison, a total 306 cases are reported. 71.4% of the reported EFAs were implemented with oblique rotation methods, while 20.4% did not report the rotation method. Most researchers chose Promax (32.2%) or Oblimin (14.5%) for oblique rotation. Varimax (8.9%) was the only orthogonal rotation method found in our sample.

More than half of the time researchers did not rely solely on one factor retention criterion to determine the number of factors, but used multiple criteria instead (Please note, that because of the usage of multiple criteria the percentages in Table 6 do not add up to one.) The most common was the Kaiser-Guttman criterion (often referred to as Eigenvalue >1 rule) used in 55.6% of the cases, followed by the Scree-Test (46.4%), PA (42.1%) and theoretical considerations or interpretability of the solution (35.5%). In some cases, only one of these four methods was used as a single criterion. When reporting just one stand-alone criterion, Kaiser-Guttman was the most common (10.5%), followed by Scree-test (9.5%) and PA (8.2%). In total, we found 16 different methods selected as retention criteria (Table 6.).

Studies with References to Fabrigar et al. (1999)

The analyses from articles directly citing Fabrigar et al. (1999) produced quite different results. PAF was prevalently used as the extraction method (88%) while ML estimation was used only once. Every applied rotation method was oblique with Promax being reported the most frequently (82%). 80% of the articles reported multiple criteria and PA was the predominant

Table 1 Sample sizes in EFAs in current psychological research

Sample Size	N	%	% (Fabrigar et al.)
< 100	8	2.6	17.5
100–200	42	13.8	26.7
200–300	44	14.5	15.7
300–400	52	17.1	6.9
> 400	153	50.3	33.2
Not Reported	5	1.6	/

Table 3 Average communalities in EFAs in current psychological research

Average communalities $\overline{h^2}$	N	%
< .40	4	1.3
.40–.49	11	3.6
.50–.59	12	3.9
.60–.69	9	3.0
≥ .70	2	0.7
Not Reported	266	87.5

Communalities are averaged over each EFA, so the range of communalities is left out

retention criterion with 88%, while Kaiser-Guttman (82%), Scree-test (74%) and MAP test (66%) were used at least roughly two-thirds of the time as well. In general, these articles showed a higher tendency to report our variables of interest. At least 40% of them provided both the information about communalities as well as complete information about the other relevant variables. More information on these results can be found in the electronic supplementary material (ESM 1).

Methodological Developments

As there are several new methodological developments in the field of EFA, we want to present an updated review of the methodological questions arising when conducting EFA. The discussed recommendations of Fabrigar et al. (1999) serve as the basis of our overview, which is why the following sections focus primarily on concepts and empiricism published after the year 1999.

Study Design (Number of Items and Sample Size) When it comes to EFA, sample size is a heavily discussed issue. As Fabrigar et al. (1999) point out, recommendations concerning subject to item ratios (N/p) are out of date. In fact, MacCallum et al. (1999) showed in a simulation study that these ratios are not useful, and furthermore, that the communalities of the analyzed variables and the number of items per factor should be considered when searching for an appropriate sample size.

Table 4 Extraction methods in EFAs in current psychological research

Extraction method	N	%
Principal Axis Factoring	156	51.3
Maximum Likelihood	50	16.4
Unweighted Least Squares	11	3.6
Weighted Least Squares*	16	5.3
MinRes	3	1.0
Not Reported	68	22.4

*WLSMV Weighted Least Squares Means and Variance adjusted

Rouquette and Falissard (2011) evaluated the requirements for sample sizes in EFA in the context of psychiatric scales. They found that the subject to item ratio rules did not work appropriately and concluded that it is not necessarily true that shorter scales need smaller samples than larger scales or vice versa. Therefore, they recommended a rule of thumb of 300 subjects or more when using EFA in this specific context.

Other studies followed the findings of MacCallum et al. (1999). Hogarty et al. (2005) reported a strong influence of item communalities on the accuracy of EFA solutions. Especially when overdetermination was strong (e.g. three factors represented by 20 variables) and communalities were high (h^2 between .60 and .80), sample factor loadings and population factor loadings corresponded vastly. Quite similar results were obtained in simulations by Mundfrom et al. (2005): the higher the item communalities were and the stronger overdetermination was, the smaller the sample could have been to find accurate factor solutions. Thus, even samples smaller than 100 observations could be appropriate when communalities are sufficiently high and factors are represented by a great number of items.

Contrary to EFA, there are some methods to determine sample size for CFA which go beyond common rules of thumb (Schmitt 2011). One of them is a method based on Monte Carlo simulations evaluating the minimum sample size for a particular model and a desired power for the Likelihood ratio test (Muthén and Muthén 2002). This process determining the sample size analogue to sample size planning for other analyses (e.g. ANOVA) seems to be a practicable solution for CFA, but will not fit in the context of EFA as necessary assumptions about the factor structure and the size of loadings cannot be made in advance (otherwise CFA should be used).

As there is often little or no evidence in advance about the concrete size of the item communalities, one has to come back to rough rules of thumb. We therefore recommend to (highly) overdetermine the expected factors and stick with an item to factor ratio of at least 4, better 5, so that samples of approximately 400 subjects will promise trustworthy results (see,

Table 5 Rotation methods in EFAs in current psychological research

Rotation method	N	%	% (Fabrigar et al.)
Varimax	27	8.9	53.0
Promax	98	32.2	1.8
Oblimin	44	14.5	/
Geomin	23	7.6	/
Equimax	5	1.6	/
GeoMax	2	0.7	/
Varimax (oblique)	11	3.6	/
Other oblique rotations	34	11.2	13.8*
Not Reported	62	20.4	15.2

* in 13.8% of the cases Direct Quartimin was used as the rotation method

Table 6 Factor retention criteria in current psychological research

Factor retention criterion	N	%
Kaiser-Guttman*	169	55.6
Scree-test*	141	46.4
Parallel Analysis*	128	42.1
Theory/Interpretability*	108	35.5
AIC	1	0.3
BIC	8	2.6
χ^2-test	1	0.3
Comparison Data	1	0.3
Eigenvalue > .70	1	0.3
Variance accounted for	24	7.9
Variables per factor ≥ 3	16	5.3
MAP test	59	19.4
RMSEA	3	1.0
SRMR	1	0.3
Standard Error Scree	16	5.3
Very Simple Structure	1	0.3
Not Reported	13	4.3

Percentages do not add up to 1, because multiple criteria were used among the majority of EFAs. * Criteria, that were used stand-alone to determine the number of factors in at least one study

Mundfrom et al. 2005). Hogarty et al. (2005) likewise recommended overdetermination to limit the need of excessive sample sizes due to potentially low item communalities. Increasing the item to factor ratio can be harmful though, when the content validity is not regarded. Artificial duplication of items can lead to violations of local independence. The item to factor ratio should therefore be increased carefully.

The number of observations which allows for stable estimations of correlations (as EFA is based on the correlative relations among variables) might be another reference value for a desirable sample size. Schönbrodt and Perugini (2013) demonstrate at which sample sizes Pearson correlations stabilize depending on different levels of confidence and definitions of stability. As secondary loadings in EFA are often based on rather small correlations more than 300 observations seem to be necessary to achieve reasonably stable correlations in this context. Therefore, this rough assessment is an additional indicator that the rule of thumb of Rouquette and Falissard (2011) with sample sizes greater than 300 might be a good lower bound when planning the sample for an EFA. We would go further and suggest to researchers to surpass this number following Fabrigar et al. (1999) with samples containing at least 400 observations. Even though there are some methods especially designed for small samples (e.g. Jung and Takane 2008, see Extraction methods), using those should be exceptional and reserved for cases in which strong ethical or resource-related objections

can be made. In general, researchers should collect greater samples so that factor loadings and therefore factor scores are estimated more precisely – especially when tests are designed for clinical diagnostics.

Current Practice Against this background, it is encouraging to see that sample sizes in our review tend to be higher than in the study of Fabrigar et al. (1999) twenty years ago. This might be an effect of the differing journals we used for our review, but it could also indicate real improvements in current practice. As the sample size can be judged only when communalities and item-to-factor ratios are known, one has to be cautious with results of studies based on extremely small samples when these measures are not reported. Thus, we recommend to provide this information within every article. Sample sizes of more than 400 observations are therefore still an essential base for the conduction of EFA and should not be smaller. The tendency of studies directly referring to Fabrigar et al. (1999) showing higher sample sizes than the average of the considered studies, can be seen as a confirmation that methodological education can help to improve psychological research.

Extraction Method

Another central decision when performing EFA is the choice of an appropriate extraction method. It has been repeatedly stated that PCA is not the same as EFA and therefore PCA is not an equivalent alternative when dealing with latent variables measured by manifest items (Costello and Osborne 2005; Fabrigar et al. 1999; Gorsuch 1990, 1997). A short introduction on the differences between EFA and PCA is presented by Suhr (2005). When item communalities are close to one both methods yield similar results while results can differ heavily when communalities decrease. The decision between EFA and PCA should be linked directly to the purposes of the analysis – when exploring latent constructs that are measured (measurement error!) via manifest indicators common EFA should be preferred.

Even when excluding PCA from the set of possible extraction methods, researchers are confronted with various different options: ML estimation, Minres introduced by Harman and Jones (1966), different least squares approaches, Minimum Rank Factor Analysis (MRFA) and PAF, just to name the most common ones. Jöreskog et al. (2016) point out that ML estimation can be described as an iteratively reweighted least squares approach (for more detail, see Browne 1974). So, the framework of the weighted least squares family (WLS) covers ML, unweighted least squares (ULS) as well as generalized least squares (GLS) as special cases.

Despite these methodological similarities among them, the choice of an extraction method can have a severe impact on the concrete EFA solution and the literature lacks advice which exact extraction method should be used under which conditions (Costello and Osborne 2005; Fabrigar et al. 1999). Numerous researchers (e.g. Conway and Huffcutt 2003; Costello and Osborne 2005) follow Fabrigar et al. (1999) preferring ML estimation when multivariate normality is given. Again, the main reason for this preference is the variety of fit indices one can use for model evaluation and comparison. In addition, ML estimation is implemented in all major statistical programs (e.g. SPSS, FACTOR, R, MPLUS).

However, using Likert type items multivariate normality might be questionable. When multivariate normality is violated, Costello and Osborne (2005) recommend PAF, while Yong and Pearce (2013) suggest to conduct PCA at first to reduce the dimensionality of the data and subsequently perform a "real" factor extraction using one of the methods above.

Accordingly, PAF is often used as an alternative extraction method. De Winter and Dodou (2012) compared PAF and ML estimation via simulations and showed that ML estimation was more likely to produce Heywood Cases throughout all conditions, but outperformed PAF when loadings were unequal and underextraction was given. PAF, on the other hand, performed better when the factor structure was orthogonal and when overextraction was present.

So, neither PAF nor ML estimation can be seen as preferable in general. Barendse et al. (2015) compared ML with WLS and robust WLS for different response scales (continuous, dichotomous and polytomous) and found robust WLS with polychoric correlations to yield better results when discrete data was evaluated – findings comparable to those that have been made in the field of confirmatory factor analysis (CFA). Beauducel and Herzberg (2006), for example, compared ML estimation to weighted least squares means and variance adjusted (WLSMV) estimation for CFA by simulating data sets based on variables with different response scale formats. They found that WLSMV performed better for variables with two or three categories which are situations where normality assumptions might be questionable anyway. Comparable results were reported by Rhemtulla et al. (2012), who showed that ML estimation can be used when variables have five or more categories yielding results of equal quality as WLSMV. Both simulation studies revealed a slight greediness of WLSMV estimation for greater sample sizes. These findings might not be applicable directly to EFA, but they can give some evidence which conditions might be suitable for either ML estimation or WLS approaches.

All these estimation algorithms require a minimum sample size (another reason for rather big samples, see section sample size) and do not provide reliable results with small samples. Therefore, a regularized EFA for small sample sizes has been proposed (Jung and Takane 2008). Contrary to common estimation methods (e.g. ML), it does not estimate the unique variances for each item and the factor loadings iteratively, but rather estimates a single regularization parameter[1] λ to avoid improper solutions. The regularization parameter λ shrinks the initial estimates of the unique variances, while the factor loadings are estimated as usual with common ML, ULS or GLS estimations. Initially, the unique variances are either assumed to be constant across all variables, proportional to the anti-image[2] covariance (see, e.g. Kaiser 1976) or proportional to the Ihara-Kano estimates (see, Ihara and Kano 1986). Jung and Lee (2011) showed in a simulation study that this procedure works better for small samples (less than 50 observations) with ML estimation and anti-image assumption than common ML estimation or PCA. Nevertheless, these assumptions for the unique variances are hardly ever met in psychological studies, so this procedure is reserved for situations where common extraction procedures are not feasible for a given sample size.

Current Practice For the majority of studies PAF was used, a tendency which was even stronger for those referring to Fabrigar et al. (1999). Yet, there are several advantages of ML and the Least-Squares approaches as mentioned above. EFA results should be cross-validated with CFA, so we recommend to use ML or LS approaches instead of PAF as these estimation methods are available for CFA as well and therefore yield comparable outcomes among the analyses. For normally distributed data, one should rely rather on ML estimation, whereas WLS estimation should be preferred for nonnormal and ordinal data (especially when Likert type items with less than five categories are used). Extracting via PAF should rather be restricted to cases where the other extraction methods suffer from non-convergence or improper solutions. Depending on the particular data, more than one method can be tried though and results can be examined for matching patterns as suggested by Widaman (2012).

Factor Retention Criteria

Determining the number of factors is a very decisive issue in the EFA process because of its influential power within the

---

[1] Regularization means that an additional term is added to an objective function to solve an otherwise not solvable problem. Here instead of estimating several unique variances which can be infeasible when the sample size is too small, a so-called regularization parameter is selected that adjusts the initial estimates of the unique variances.

[2] The anti-image can be pictured as the negative of the image of a matrix. The image covariance matrix contains the variation of each variable that can be explained by the other variables (partial covariance coefficients), the respective anti-image consists of the negatives which can be described as the unique components. For more detail, have a look at Kaiser (1976) or detailed EFA textbooks as the anti-image correlation matrix is a commonly used tool to evaluate whether an EFA is applicable to the data (see also Measuring Sampling Adequacy (MSA), Kaiser 1970).

exploratory analysis. While in many articles authors write about the true number of factors and the problem to find this exact number, Preacher et al. (2013) argue that there is no true factor model and researchers rather have to approximate the data generating process. The authors describe an error framework which covers two different directions in the factor retention issue.

Preacher et al. (2013) explain that one has to choose the aim of the EFA - approximating the "true" factor structure (approximation goal) or finding the most replicable solution (replicability goal) which is a decision analogue to the bias-variance tradeoff. They conclude that different factor retention criteria are best for these different goals. In simulation studies the authors focused on fit indices based on ML estimation and found the RMSEA (to be more precise: its confidence interval's lower bound) to perform the best for the approximation goal while AIC and BIC were far less accurate especially in great sample size scenarios. Contrary, for the replicability goal and in cases of small samples BIC performed best.

Often the approximation goal has priority in EFA research. There is a broad range of evidence that in this case PA produces the best results when comparing the most common criteria (see, Fabrigar et al. 1999; Peres-Neto et al. 2005; Zwick and Velicer 1986). The generally good performance of PA might be based on its robustness against varying distributional assumptions (Dinno 2009). Timmerman and Lorenzo-Seva (2011) evaluated different extraction methods within the PA and recommended to use MRFA instead of PCA or PAF for ordered polytomous items which are usually used in psychological questionnaires.

PA has become some kind of gold standard for factor retention criteria, but promising alternatives have been proposed recently. Lorenzo-Seva et al. (2011) developed the so-called hull method. This method is based on four major steps. First the researcher chooses a range of possible numbers of factors, then an arbitrary fit index is evaluated for each number of factors (CFI performed best in simulations). Afterwards the degrees of freedom of this set of factor solutions is assessed and finally the values of the chosen fit index are plotted against the respective degrees of freedom. The higher boundary of the convex hull of the plotted data points shows an elbow which defines the factor number to retain. The authors showed a superiority of their method to PA and the minimum average partial test (MAP) in simulations and for a real data set. This reported superiority of the hull method was based on cases with an extremely high item to factor ratio (items/factor = 20). In cases of smaller ratios PA yielded equivalent or even better results.

Another method is the comparison data (CD) approach (Ruscio and Roche 2012). CD can be framed as an extended PA which reproduces the observed correlation matrix instead of using random data. The researcher specifies the upper bound for the possible number of factors. Then data of

populations with one, two, etc. factors (up to this predefined upper bound) are simulated each reproducing the given empirical covariance structure as closely as possible. Samples (the authors suggest 500) of the same size as the empirical data are drawn from each population and the respective eigenvalues of the item correlation matrix are compared to the observed eigenvalues via the Root-Mean-Square-Error (RMSE).[3] One gets as many RMSE values as samples drawn from each population. These values of each factor solution are then compared to those of the next factor solution by a nonparametric Mann-Whitney U test (the one factor solution against the two factor solution, the two factor solution against the three factor solution and so on). The iterative procedure stops when no significant improvement is indicated (Ruscio and Roche 2012).

In simulation studies the authors showed that an α-level of .30 seems adequate (note that α-error means possible overextraction, while β-error means underextraction) and that CD outperformed PA and other minor retention criteria under several conditions.[4]

As this method (and similar approaches using simulated data) can be computationally intensive, Braeken and van Assen (2017) proposed the Empirical Kaiser Criterion (EKC) which makes use of the statistical properties of eigenvalues and does not require any simulations.[5] It is based on the so-called Marčenko-Pastur distribution, which asymptotically describes the distribution of sample eigenvalues under the null model (no underlying factor structure) and is therefore closely related to the results of PA, and the idea of the Kaiser criterion that only eigenvalues greater than one should be taken into account. The theoretically expected eigenvalues are corrected by a factor including the remaining variance after the respective higher eigenvalues are accounted for. The authors were able to show superiority over PA for oblique structures and found comparable results to CD and other simulation based approaches. However, this evaluation was based on simple structure assumptions, so little is known so far about the performance of EKC when cross-loadings are present.

Current Practice In current research, more than 50% of the EFAs are based on multiple factor retention criteria, whereas Fabrigar et al. (1999) reported just about 20% of studies to do so. In articles referring to their article, the percentage rises to 80%. That speaks in favor of the current research practice,

---

[3] The RMSE is defined as the root of the MSE which is the averaged squared distance between parameters and its estimates. In this case, the differences between the given eigenvalues and the eigenvalues obtained of the simulated data sets of the specific k-factor population are computed.

[4] They varied the number of factors (one to five), the number of response categories (two to 20), used correlated and uncorrelated solutions and sample sizes between 200 and 1000.

[5] It only requires the number of items and the sample size, so it can be applied without knowing much about the structure of the data – for example when evaluating published results.

although the frequent use of invalid methods such as Kaiser-Guttman rule or Scree-test (even as a single criterion) has to be criticized. There are even tutorial papers for EFA recommending these methods (Maroof 2012) or completely ignoring more appropriate tools (Beavers et al. 2013). Instead it should become scientific standard to avoid MAP-test or Kaiser-Guttman rule as a (stand-alone) factor retention criterion in common factor analysis as these methods are created for PCA and therefore associated with different assumptions.

As there is enough evidence demonstrating problems with some of the commonly used criteria, we want to encourage researchers to use the whole spectrum of methods determining the number of factors and whenever feasible to split the sample and evaluate the subsamples separately. A practical solution could be using PA and CD in combination with a descriptive measure like the explained variance or theoretical considerations. Nevertheless, this decision still remains the most difficult to make within EFA. Thus, it is inevitable to be aware of its consequences and to report every consideration concerning this issue.

Rotation Method

After extraction, researchers almost always decide to rotate the factor solution to get results that are easier to interpret. It has become common understanding in literature on EFA methods that oblique rotation is preferable (e.g. Fabrigar et al. 1999; Costello and Osborne 2005; Conway and Huffcutt 2003; Baglin 2014), but it is also stated that it is not clear which oblique rotation has to be used. Browne (2001) gives a detailed overview of the different rotation methods and highly recommends a multimethod approach. He argues that using various complexity functions[6] might be an appropriate way to handle a situation in which no solution is undoubtedly superior to others. One could use a method from the Crawford-Ferguson (CF) family, plus Infomax rotation and Geomin rotation, for example. The CF family (Crawford and Ferguson 1970) covers several well-known rotation methods by formulating the complexity function as a function of row complexity (items) and column complexity (factors):

$$f(\Lambda) = (1-\kappa) \sum_{i=1}^{p} \sum_{j=1}^{k} \sum_{l \neq j}^{k} \lambda_i^2 \lambda_{il}^2 + \kappa \sum_{j=1}^{k} \sum_{i=1}^{p} \sum_{k \neq i}^{p} \lambda_{ij}^2 \lambda_{kj}^2$$

with p indicating the number of variables, k indicating the number of factors and κ being an arbitrary constant weighting the row-wise (first part of the equation) or column-wise complexity. Some values of κ lead to common criteria. $\kappa = 0$, for example, corresponds to the Quartimin-criterion and $\kappa = \frac{1}{p}$ to

the Varimax rotation (Browne 2001). Browne explains that in cases of almost perfect cluster patterns most complexity functions work perfectly fine, but when complexity in the factor patterns increases, one has to weigh up stability of the solution against its accuracy.

When complex structures are expected (higher amount and amplitude of cross-loadings), rotation methods like CF-Equamax or CF-Facparsim should be used. When fewer or smaller cross-loadings are expected, common techniques like Geomin or CF-Quartimin might be more appropriate (Sass and Schmitt 2010; Schmitt and Sass 2011).

Browne (2001) states that a standardization like the Cureton-Mulaik (CM) weighting (for more details, see Cureton and Mulaik 1975) can improve the solution. Nonetheless, if there are only a few complex variables among many perfectly discriminative ones (only loadings on one factor), weighting procedures might focus too much on these variables. The advantages of CM weighting were empirically shown by Lorenzo-Seva (2000) comparing weighted Oblimin with Direct Oblimin, Promaj, Promin and weighted Promax.

Another interesting, yet different rotation method is the rotation to target procedure, where the factor matrix is rotated in a way that a partially specified target matrix (some coefficients of the factor pattern matrix are defined in advance) is replicated as closely as possible (Myers et al. 2015). It seems to be an appropriate rotation method when additional information is available as it has some similarities to exploratory structure equation modelling (ESEM; for further readings, see Marsh et al. 2014). Therefore, it might be the right rotation method when theoretical or empirical information is given and when many factor cross loadings tend to be zero, because this seems to be the most reasonable specification a researcher can make in advance. Browne (2001) suggests to apply this procedure iteratively, updating the target matrix in every step.

As the choice of the best rotation method appears to be arbitrary to some degree, we want to present a totally different approach: the penalized factor analysis (Hirose and Yamamoto 2014). Instead of conducting EFA in the classical two steps – extracting k factors and afterwards rotating the solution to increase interpretability – this new method obtains sparsity in the pattern matrix through penalizing the likelihood. The penalty[7] is analogue to the complexity function discussed before, but it is now integrated into the estimation process, so that cross-loadings get shrunk towards zero in the first place.

---

[6] The so-called complexity function is the objective function which is minimized with regard to specific constraints to achieve a particular rotation of the pattern matrix. We recommend the article of Browne (2001) explaining the link between constraints and rotation criterion in more detail.

[7] In common ML estimation an objective function (that is derived from the log-likelihood) is minimized. Here a so-called penalty term is added to this function. It penalizes a high number of parameters (in this case loadings, especially cross-loadings). The more parameters are estimated to be non-zero, the higher this term gets and it "becomes harder" to achieve a minimum, so in turn adding this penalty yielding more small (or even zero-) loadings (depending on the type of penalty). You can read about penalizing the likelihood in the EFA estimation process in more detail in Jin, Moustaki and Yang-Wallentin (2018).

First simulations revealed some promising results for wide data with many variables and sparse loading matrices (Hirose and Yamamoto 2015) as well as for a real data set (Hirose and Yamamoto 2014). The latter was analyzed with both the new approach with a MC+ penalty and the common ML estimation with Promax rotation. The penalized factor analysis produced quite similar yet sparser and well interpretable results. Nevertheless, the penalized factor analysis still has to be evaluated under a broader range of conditions to investigate whether it will be an appropriate tool for psychological research questions.

In general, it is appropriate to use different rotation methods and to choose the one with the most reasonable solution as all rotated solutions are mathematically equal[8] (in case of the two-step process,[9] not the penalized ML estimation). For replication purposes, it is necessary to name the chosen rotation procedure.

Current Practice For one out five cases in our review, the rotation method was not reported, so the current research practice clearly lacks transparency. Only two studies used different rotation methods and compared different solutions – a procedure highly recommended by Browne (2001). However, a positive aspect is that more than 70% of all EFAs used oblique rotation methods. A number that increased to 98% for studies referring to Fabrigar et al. (1999), where 53% of the examined studies had used the orthogonal Varimax rotation. Accordingly, psychological research seems to be on the right track regarding this issue.

When using different rotation methods on the same sample or on subsamples as suggested by Browne (2001), Fabrigar et al. (1999) or Preacher et al. (2013), it might be helpful to evaluate the similarity of different solutions. Lorenzo and Ferrando (1996, 1998), for example, developed the FACOM/NFACOM library which allows for comparison of different factor solutions based on different methodological decisions. A decision can be made by weighing up the mathematical interpretability and the theoretical plausibility of the respective solution.

Further Recommendations

EFA is often applied to questionnaire items which are not normally distributed but rather skewed. Investigating this problem, Holgado-Tello et al. (2010) found EFA based on polychoric correlations to reproduce the true factor model more accurately than EFA based on Pearson correlations. Baglin (2014) nicely illustrated this issue and also

recommended polychoric correlations for these cases. Hence, item skewness should be evaluated before conducting EFA and in case of severely skewed variables polychoric correlations should be chosen when using ML estimation. When extracting via WLS approaches, polychoric or tetrachoric correlations are used instead of Pearson correlation anyway, so the type of correlation does not have to be selected in these cases. Other approaches worth considering for ordinal data are those that are based on response patterns (IRT models) instead of approximating the correlation matrix assuming underlying normal distributed latent variables (e.g. Jöreskog and Moustaki 2001). As full-information item factor analysis (full information maximum likelihood, FIML) can be computationally challenging (IRT approach as well as when assuming an underlying continuous variable) and problematic with small samples in particular, Katsikatsou et al. (2012) proposed a pairwise likelihood (PML) estimation approach that closely matched the results of FIML and can be seen as a practical alternative.

When conducting EFA, researchers should specify their research goals precisely and select the best suited methods. We want to clarify that the presented methods and related recommendations are designed for the researcher's goal to approximate the data generating process as precisely as possible. Often interpretability and theoretical considerations can be equally important. In particular, for test construction purposes content validity should be first priority.

Researchers therefore should report transparently which objectives they have, which methodological decisions they take and all outcomes they collect. This ensures that the quality of a solution can be evaluated and implications of particular studies can be weighted. The Journal of the Society for Social Work and Research has taken a leading role in demanding certain reporting guidelines for EFA (see Cabrera-Nguyen 2010). Other journals should follow this example and call for openness in reporting EFAs. Especially in the light of the current discussion about the replication crisis in psychology (e.g. Shrout and Rodgers 2018), transparency with regard to data, research material and methodological decisions is essential (for further readings: OSF Guidelines for Transparency, Klein et al. 2018). Furthermore, we encourage researchers to consider various procedures in this context, instead of performing a standard practice based on default settings or personal routines.

Summary

As pointed out, EFA is a very complex analysis and it is therefore not easy to make general recommendations on how to conduct it properly. Each case should be evaluated individually, so this paper tries to sensitize researchers for careful decisions and transparent reporting. Nevertheless, we want to formulate some "default" settings which can be seen as a

---

[8] Problem of rotation indeterminacy (see introduction section)

[9] The optimization process is done with respect to different constraints, but apart from that equivalent for all rotation methods. Therefore, theoretical considerations must be taken into account to make a reasonable decision (are cross-loadings consistent with theoretical assumptions, etc.).

basis for further considerations. Samples for EFA should be greater than 400 participants to get reliable factor patterns and precisely estimated factor scores. One should use ML or WLS estimation as extraction method depending on the respective item distributions and the response format, because these methods allow for evaluations of model fit and cross-validation with CFAs. To determine the number of factors, we recommend combining PA and CD (or maybe EKC) with a descriptive measure (e.g. explained variance) and theoretical considerations. Latter should be included for test construction purposes, but should be ignored when the data generating process of the specific data is approximated. In any case, multiple retention criteria should be applied and reported later on to provide the full picture. As different rotation methods yield mathematically indeterminate factor solutions, researchers should compare factor patterns between different methods and choose the solution that fits theoretical considerations best. Again, it is necessary to report the chosen method to enable other researchers to replicate the respective solution.

Compliance with Ethical Standards

Conflict of Interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

Baglin, J. (2014). Improving your exploratory factor analysis for ordinal data: A demonstration using FACTOR. *Practical Assessment, Research & Evaluation, 19*, 5 Retrieved from http://pareonline.net/getvn.asp?v=19&n=5. Accessed 15 Dec 2017

Barendse, M. T., Oort, F. J., & Timmerman, M. E. (2015). Using exploratory factor analysis to determine the dimensionality of discrete responses. *Structural Equation Modeling: A Multidisciplinary Journal, 22*(1), 87–101. https://doi.org/10.1080/10705511.2014.934850.

Beauducel, A., & Herzberg, P. Y. (2006). On the performance of maximum likelihood versus means and variance adjusted weighted least squares estimation in CFA. *Structural Equation Modeling, 13*(2), 186–203. https://doi.org/10.1207/s15328007sem1302_2.

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. Accessed 15 Dec 2017

Borsboom, D., Mellenbergh, G. J., & Van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review, 110*(2), 203–219. https://doi.org/10.1037/0033-295X.110.2.203.

Braeken, J., & Van Assen, M. A. (2017). An empirical Kaiser criterion. *Psychological Methods, 22*(3), 450–466. https://doi.org/10.1037/met0000074.

Browne, M. W. (1974). Generalized least squares estimators in the analysis of covariance structures. *South African Statistical Journal, 8*(1), 1–24 Retrieved from http://hdl.handle.net/10520/AJA0038271X_175. Accessed 28 Mar 2018

Browne, M. W. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research, 36*(1), 111–150. https://doi.org/10.1207/S15327906MBR3601_05.

Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research, 21*(2), 230–258. https://doi.org/10.1177/0049124192021002005.

Cabrera-Nguyen, P. (2010). Author guidelines for reporting scale development and validation results in the journal of the Society for Social Work and Research. *Journal of the Society for Social Work and Research, 1*(2), 99–103. https://doi.org/10.5243/jsswr.2010.8.

Conway, J. M., & Huffcutt, A. I. (2003). A review and evaluation of exploratory factor analysis practices in organizational research. *Organizational Research Methods, 6*(2), 147–168. https://doi.org/10.1177/1094428103251541.

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 & Evaluation, 10*(7), 1–9 Retrieved from http://pareonline.net/getvn.asp?v=10&n=7. Accessed 24 Oct 2017

Crawford, C. B., & Ferguson, G. A. (1970). A general rotation criterion and its use in orthogonal rotation. *Psychometrika, 35*(3), 321–332. https://doi.org/10.1007/BF02310792.

Cureton, E. E., & Mulaik, S. A. (1975). The weighted varimax rotation and the promax rotation. *Psychometrika, 40*(2), 183–195. https://doi.org/10.1007/BF02291565.

De Winter, J. C. F., & Dodou, D. (2012). Factor recovery by principal axis factoring and maximum likelihood factor analysis as a function of factor pattern and sample size. *Journal of Applied Statistics, 39*(4), 695–710. https://doi.org/10.1080/02664763.2011.610445.

Dinno, A. (2009). Exploring the sensitivity of Horn's parallel analysis to the distributional form of random data. *Multivariate Behavioral Research, 44*(3), 362–388. https://doi.org/10.1080/00273170902938969.

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. https://doi.org/10.1037/1082-989X.4.3.272.

Ford, J. K., MacCallum, R. C., & Tait, M. (1986). The application of exploratory factor analysis in applied psychology: A critical review and analysis. *Personnel Psychology, 39*(2), 291–314. https://doi.org/10.1111/j.1744-6570.1986.tb00583.x.

Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale: Lawrence Erlbaum Associates.

Gorsuch, R. L. (1990). Common factor analysis versus component analysis: Some well and little known facts. *Multivariate Behavioral Research, 25*(1), 33–39. https://doi.org/10.1207/s15327906mbr2501_3.

Gorsuch, R. L. (1997). Exploratory factor analysis: Its role in item analysis. *Journal of Personality Assessment, 68*(3), 532–560. https://doi.org/10.1207/s15327752jpa6803_5.

Harman, H. H., & Jones, W. H. (1966). Factor analysis by minimizing residuals (minres). *Psychometrika, 31*(3), 351–368. https://doi.org/10.1007/BF02289468.

Hirose, K., & Yamamoto, M. (2014). Estimation of an oblique structure via penalized likelihood factor analysis. *Computational Statistics & Data Analysis, 79*, 120–132. https://doi.org/10.1016/j.csda.2014.05.011.

Hirose, K., & Yamamoto, M. (2015). Sparse estimation via nonconcave penalized likelihood in factor analysis model. *Statistics and Computing, 25*(5), 863–875. https://doi.org/10.1007/s11222-014-9458-0.

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), 202–226. https://doi.org/10.1177/0013164404267287.

Holgado-Tello, F. P., Chacón-Moscoso, S., Barbero-García, I., & Vila-Abad, E. (2010). Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables. *Quality*

& Quantity, 44(1), 153–166. https://doi.org/10.1007/s11135-008-9190-y.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. Psychometrika, 30(2), 179–185. https://doi.org/10.1007/BF02289447.

Ihara, M., & Kano, Y. (1986). A new estimator of the uniqueness in factor analysis. Psychometrika, 51(4), 563–566. https://doi.org/10.1007/BF02295595.

Jin, S., Moustaki, I., & Yang-Wallentin, F. (2018). Approximated penalized maximum likelihood for exploratory factor analysis: An orthogonal case. Psychometrika, 83(3), 628–649. https://doi.org/10.1007/s11336-018-9623-z.

Jöreskog, K. G. (1967). Some contributions to maximum likelihood factor analysis. Psychometrika, 32(4), 443–482. https://doi.org/10.1007/BF02289658.

Jöreskog, K. G., & Moustaki, I. (2001). Factor analysis of ordinal variables: A comparison of three approaches. Multivariate Behavioral Research, 36(3), 347–387. https://doi.org/10.1207/S15327906347-387.

Jöreskog, K. G., Olsson, U. H., & Yang-Wallentin, F. (2016). Multivariate analysis with LISREL. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-33153-9.

Jung, S., & Lee, S. (2011). Exploratory factor analysis for small samples. Behavior Research Methods, 43(3), 701–709. https://doi.org/10.3758/s13428-011-0077-9.

Jung, S., & Takane, Y. (2008). Regularized common factor analysis. In K. Shigemasu, A. Okada, T. Imaizumi, & T. Hoshino (Eds.), New trends in psychometrics (pp. 141–149). Tokyo: University Academic Press.

Kaiser, H. F. (1970). A second generation little jiffy. Psychometrika, 35(4), 401–415. https://doi.org/10.1007/BF02291817.

Kaiser, H. F. (1976). Image and anti-image covariance matrices from a correlation matrix that may be singular. Psychometrika, 41(3), 295–300. https://doi.org/10.1007/BF02293555.

Katsikatsou, M., Moustaki, I., Yang-Wallentin, F., & Jöreskog, K. G. (2012). Pairwise likelihood estimation for factor analysis models with ordinal data. Computational Statistics & Data Analysis, 56(12), 4243–4258. https://doi.org/10.1016/j.csda.2012.04.010.

Klein, O., Hardwicke, T. E., Aust, F., Breuer, J., Danielsson, H., Hofelich Mohr, A., … Frank, M. C. (2018). A practical guide for transparency in psychological science. Retrieved from https://osf.io/79epu. Accessed 28 Mar 2018

Lorenzo, U., & Ferrando, P. J. (1996). FACOM: A library for relating solutions obtained in exploratory factor analysis. Behavior Research Methods, Instruments, & Computers, 28(4), 627–630. https://doi.org/10.3758/BF03200553.

Lorenzo, U., & Ferrando, P. J. (1998). NFACOM: A new program for relating solutions in exploratory factor analysis. Behavior Research Methods, Instruments, & Computers, 30(4), 724–725. https://doi.org/10.3758/BF03209493.

Lorenzo-Seva, U. (2000). The weighted oblimin rotation. Psychometrika, 65(3), 301–318. https://doi.org/10.1007/BF02296148.

Lorenzo-Seva, U., Timmerman, M. E., & Kiers, H. A. (2011). The Hull method for selecting the number of common factors. Multivariate Behavioral Research, 46(2), 340–364. https://doi.org/10.1080/00273171.2011.564527.

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

Maroof, D. A. (2012). Exploratory factor analysis. In D. A. Maroof (Ed.), Statistical methods in neuropsychology: Common procedures made comprehensible (pp. 23–34). New York: Springer Science + Business Media, LLC. https://doi.org/10.1007/978-1-4614-3417-7_4.

Marsh, H. W., Morin, A. J., Parker, P. D., & Kaur, G. (2014). Exploratory structural equation modeling: An integration of the best features of exploratory and confirmatory factor analysis. Annual Review of Clinical Psychology, 10, 85–110. https://doi.org/10.1146/annurev-clinpsy-032813-153700.

Mundfrom, D. J., Shaw, D. G., & Ke, T. L. (2005). Minimum sample size recommendations for conducting factor analyses. International Journal of Testing, 5(2), 159–168. https://doi.org/10.1207/s15327574ijt0502_4.

Muthén, L. K., & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. Structural Equation Modeling, 9(4), 599–620. https://doi.org/10.1207/S15328007SEM0904_8.

Myers, N. D., Jin, Y., Ahn, S., Celimli, S., & Zopluoglu, C. (2015). Rotation to a partially specified target matrix in exploratory factor analysis in practice. Behavior Research Methods, 47(2), 494–505. https://doi.org/10.3758/s13428-014-0486-7.

Peres-Neto, P. R., Jackson, D. A., & Somers, K. M. (2005). How many principal components? Stopping rules for determining the number of non-trivial axes revisited. Computational Statistics & Data Analysis, 49(4), 974–997. https://doi.org/10.1016/j.csda.2004.06.015.

Preacher, K. J., Zhang, G., Kim, C., & Mels, G. (2013). Choosing the optimal number of factors in exploratory factor analysis: A model selection perspective. Multivariate Behavioral Research, 48(1), 28–56. https://doi.org/10.1080/00273171.2012.710386.

Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. Psychological Methods, 17(3), 354–373. https://doi.org/10.1037/a0029315.

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

Ruscio, J., & Roche, B. (2012). Determining the number of factors to retain in an exploratory factor analysis using comparison data of known factorial structure. Psychological Assessment, 24(2), 282–292. https://doi.org/10.1037/a0025697.

Sass, D. A., & Schmitt, T. A. (2010). A comparative investigation of rotation criteria within exploratory factor analysis. Multivariate Behavioral Research, 45(1), 73–103. https://doi.org/10.1080/00273170903504810.

Schmitt, T. A. (2011). Current methodological considerations in exploratory and confirmatory factor analysis. Journal of Psychoeducational Assessment, 29(4), 304–321. https://doi.org/10.1177/0734282911406653.

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. https://doi.org/10.1177/0013164410387348.

Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize? Journal of Research in Personality, 47(5), 609–612. https://doi.org/10.1016/j.jrp.2013.05.009.

Shrout, P. E., & Rodgers, J. L. (2018). Psychology, science, and knowledge construction: Broadening perspectives from the replication crisis. Annual Review of Psychology, 69, 487–510. https://doi.org/10.1146/annurev-psych-122216-011845.

Steiger, J. H. (1979). Factor indeterminacy in the 1930's and the 1970's some interesting parallels. Psychometrika, 44(2), 157–167. https://doi.org/10.1007/BF02293967.

Steiger, J. H., & Schönemann, P. H. (1978). A history of factor indeterminacy. In S. Shye (Ed.), Theory construction and data analysis in the behavioral sciences (pp. 136–178). San Francisco: Jossey-Bass.

Suhr, D. D. (2005). Principal component analysis vs. exploratory factor analysis (paper 203-30). Paper presented at the meeting of the SAS Users Group International (SUGI 30), Philadelphia, PA. Retrieved

from http://www2.sas.com/proceedings/sugi30/203-30.pdf. Accessed 17 Oct 2017

Timmerman, M. E., & Lorenzo-Seva, U. (2011). Dimensionality assessment of ordered polytomous items with parallel analysis. *Psychological Methods, 16*(2), 209–220. https://doi.org/10.1037/a0023353.

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

Yong, A. G., & Pearce, S. (2013). A beginner's guide to factor analysis: Focusing on exploratory factor analysis. *Tutorial in Quantitative Methods for Psychology, 9*(2), 79–94. https://doi.org/10.20982/tqmp.09.2.079.

Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin, 99*(3), 432–442. https://doi.org/10.1037/0033-2909.99.3.432.