An empirical Kaiser criterion

Braeken, J.; van Assen, M.A.L.M.

Published in:
Psychological Methods

DOI:
10.1037/met0000074

Publication date:
2017

Document Version
Publisher's PDF, also known as Version of record

Psychological Methods
2017, Vol. 22, No. 3, 450–466

1082-989X/17/$12.00 http://dx.doi.org/10.1037/met0000074

An Empirical Kaiser Criterion

Johan Braeken
University of Oslo

Marcel A. L. M. van Assen
Tilburg University and Utrecht University

In exploratory factor analysis (EFA), most popular methods for dimensionality assessment such as the screeplot, the Kaiser criterion, or—the current gold standard—parallel analysis, are based on eigenvalues of the correlation matrix. To further understanding and development of factor retention methods, results on population and sample eigenvalue distributions are introduced based on random matrix theory and Monte Carlo simulations. These results are used to develop a new factor retention method, the Empirical Kaiser Criterion. The performance of the Empirical Kaiser Criterion and parallel analysis is examined in typical research settings, with multiple scales that are desired to be relatively short, but still reliable. Theoretical and simulation results illustrate that the new Empirical Kaiser Criterion performs as well as parallel analysis in typical research settings with uncorrelated scales, but much better when scales are both correlated and short. We conclude that the Empirical Kaiser Criterion is a powerful and promising factor retention method, because it is based on distribution theory of eigenvalues, shows good performance, is easily visualized and computed, and is useful for power analysis and sample size planning for EFA.

Keywords: exploratory factor analysis, Kaiser criterion, parallel analysis

In exploratory factor analysis, most popular methods for dimensionality assessment such as the screeplot (Cattell, 1966), the Kaiser criterion (Kaiser, 1960), or—the current gold standard—parallel analysis (Horn, 1965), are based on eigenvalues of the correlation matrix. Unfortunately, (a) the link between such methods and statistical theory on eigenvalues is often weak and incomplete, and (b) neither the methods' origin nor the evaluation of their performance is set within the larger context of practical scale development.

These two gaps in research on factor analysis should come as a surprise, because factor analysis is one of the most commonly applied techniques in scale development, and one can argue that the determination of the number of factors to retain is likely to be the most important decision in exploratory factor analysis (Zwick & Velicer, 1986). Specifying too few factors will result in the loss of important information by ignoring a factor or combining it with another (Zwick & Velicer, 1986); specifying too many factors may lead to an overcomplicated structure with many minor factors consisting of one or very few observed variables. Examples of the latter are so-called "bloated specifics," which are factors arising due to artificial overlap between variables, for instance due to similar item phrasing (Cattell, 1961). The consensus is that both underfactoring and overfactoring are likely to result in noninterpretable or unreliable factors and can potentially mislead theory and scale development efforts (Fabrigar, Wegener, MacCallum, & Strahan, 1999; Garrido, Abad, & Ponsoda, 2013; Velicer, Eaton, & Fava, 2000; Zwick & Velicer, 1986).

Dozens of factor retention methods do exist (e.g., Peres-Neto, Jackson, and Somers, 2005), but their use in practice can be quite striking. For instance, despite having been repeatedly shown not to work in simulation studies, the so-called Kaiser criterion or eigenvalue-greater-than-one rule (Kaiser, 1960) continues to be very popular, mostly because of its simplicity, ease of implementation, and it being the default method in many general statistical software packages (e.g., SPSS and SAS). In contrast, parallel analysis, the factor retention method that generally has shown the best performance in simulation studies and gets most recommendations from specialists (for thorough recent reviews, see, e.g., Garrido et al., 2013; Timmerman & Lorenzo-Seva, 2011), is not as well established among practitioners (see, e.g., Dinno, 2009; Fabrigar et al., 1999; Ford, MacCallum, & Tait, 1986). Notwithstanding, in view of its generally good performance and its recommended status, the performance of parallel analysis will be our reference in the current article.

The basic idea of parallel analysis (Horn, 1965) is to use the observed eigenvalues, and not comparing them with a fixed reference value of 1 as in the Kaiser criterion, but instead to reference eigenvalues from generated random data (i.e., independent data without factor structure). In the current article, we use the most recommended variant of parallel analysis suggested by Glorfeld (1995), which retains the first factors that all exceed the 95th percentile of their corresponding distribution of reference eigen-

This article was published Online First March 31, 2016.

Johan Braeken, CEMO: Centre for Educational Measurement, Faculty of Educational Sciences, University of Oslo; Marcel A. L. M. van Assen, Department of Methodology and Statistics, School of Social and Behavioral Sciences, Tilburg University, and Department of Sociology, Utrecht University.

We thank Taya Cohen for sharing the data of the GASP study (Cohen et al., 2011), and Fieke Wagemans for assisting us with obtaining and interpreting the GASP scale.

Correspondence concerning this article should be addressed to Johan Braeken, CEMO: Centre for Educational Measurement, Faculty of Educational Sciences, University of Oslo, UiO, Postboks 1161 Blindern 0318 Oslo, Norway. E-mail: johan.braeken@cemo.uio.no

values. The need for Monte Carlo simulations to generate such reference data, combined with tradition, out-of-date textbooks and educational training, and the lack of default implementation in general commercial software (see, e.g., Dinno, 2009; Hayton et al., 2004), apparently puts a high threshold on the use of parallel analysis in everyday practice.

A first objective of this article is to further understanding and encourage new developments in factor retention methods by bridging the gap between factor retention methods and statistical theory on eigenvalues. Theoretical results on the distribution of sample eigenvalues open up new pathways to develop simple and efficient factor retention rules that are widely applicable and that do not require simulation. A second objective of this article is to propose a new factor retention method that is specifically tailored toward typical research settings in which multiple scales are designed which are desired to be relatively short, but still reliable. The demand for and use of such short(ened) tests has recently become more common (Ziegler, Kemper, & Kruyen, 2014). In personnel selection for instance, there is an increasing tendency to use short tests consisting of, say, five to 15 items, for making decisions about individuals applying for a job (Kruyen, Emons, & Sijtsma, 2012, p. 321). Because of the ubiquity of short tests, factor retention models should particularly perform well in these cases. In this particular setting with correlated factors consisting of only a few variables, the performance of parallel analysis is known to deteriorate significantly (see, e.g., Cho, Li, & Bandalos, 2009; Crawford et al., 2010; De Winter, Dodou, & Wieringa, 2009; Garrido et al., 2013; Green, Levy, Thompson, Lu, & Lo, 2012; Turner, 1998). Thus, this setting would be serviced by having a more suitable alternative factor retention method.

In the next sections we will provide theoretical statistical background for factor retention methods with particular attention to the distinction between population-level and sample-level eigenvalues. These theoretical foundations will be directly linked to the development of a new factor retention method that is easily visualized and very straightforward to apply without requiring Monte Carlo simulation. The new retention method is called the "Empirical Kaiser Criterion:" "Empirical," because the method's series of reference eigenvalues is a function of an application's (a) variables-to-sample-size ratio, and (b) observed eigenvalues; "Kaiser," because, similar to the original Kaiser criterion, it requires eigenvalues to be at least equal to 1, which implies that at the population-level the new and the Kaiser method retrieve the same number of factors. We make analytical predictions under which conditions the Empirical Kaiser Criterion (EKC) will perform well in practically relevant situations, and provide empirical support by targeted simulation studies in which we compare the performance of the newly developed EKC with the performance of parallel analysis and the original Kaiser criterion. For illustration, the methods are applied to data on the Guilt and Shame Proneness Scale (GASP; Cohen, Wolf, Panter, & Insko, 2011), which is a short 16-item scale consisting of four highly correlated subscales. We conclude with a brief discussion and conclusions. An R-Shiny applet, available on our web site *https://cemo.shinyapps.io/EKCapp*, allows the reader to directly implement the EKC, as well as parallel analysis.

Characterizing the Behavior of Eigenvalues

We will first provide an overview of the relevant theoretical background on eigenvalues under the null model assuming no underlying factors. Using results from random matrix theory, we distinguish between eigenvalues at the population level and eigenvalues at the sample level. After better understanding the sample behavior of eigenvalues, we explain why, at the sample level under the null model, Kaiser's greater-than-one rule fails and parallel analysis works well. We continue with results under the factor model, again distinguishing between eigenvalues at the population and eigenvalues at the sample level. We explain why parallel analysis cannot be expected to work well in all situations, and how factor retention methods can and are being adapted to improve upon the performance of parallel analysis under the factor model.

Under the Null Model

Population level. The null model for factor analysis assumes there is no factor structure, that is, all variables are uncorrelated in the population. The null model corresponds to a correlation matrix with all zeros on the off-diagonal and all ones on the diagonal (i.e., the identity matrix). All eigenvalues of an identity matrix are equal to 1.

The "eigenvalues greater than one" rule, often attributed to Kaiser (1960), is implicitly linked to this null model and states that the number of factors to retain should correspond to the number of eigenvalues greater than one (i.e., deviating from the null expectation). Intuitively, one can motivate this rule by stating that an eigenvalue that represents a "true structural dimension" should at least explain more variance than contained in a single variable. A theoretical justification is that for a factor to have positive Kuder–Richardson reliability (cf. Cronbach's alpha), it is necessary and sufficient that the associated eigenvalue be greater than 1 (Kaiser, 1960, p. 145). Hence, the greater-than-one rule is essentially an asymptotical and theoretical lower bound (see, e.g., Guttman, 1954) to the number of true and reliable structural dimensions at the *population level*. Yet at the sample level, Monte Carlo simulation studies showed the rule to have low accuracy in practice (see, e.g., Velicer, Eaton, & Fava, 2000; Zwick & Velicer, 1986).

Sample level. Eigenvalues at the sample level show random variation, with typically about the first half of the eigenvalues above and the latter half below 1 under the null model. Hence, the main reason why the Kaiser criterion underperforms under the null model is that the first sample eigenvalues capitalize on coincidental sampling associations and exceed thereby 1, yielding an overestimation of the number of factors.

Results from random matrix theory (see, e.g., Anderson, Guionnet, & Zeitouni, 2010; Wigner, 1955) show that this wider range of sample eigenvalues is in fact nonrandom, but a direct function of γ, the ratio of the number of variables J to the sample size n (i.e., $\gamma = J/n$). The distribution of sample eigenvalues $\boldsymbol{L} = [l_1, \ldots, l_j, \ldots, l_J]$ under the null model follows asymptotically the Marčenko-Pastur (1967) distribution with density function

$$d(l) = \begin{cases} \dfrac{\sqrt{(l_{up} - l)(l - l_{low})}}{2\pi\gamma l} & \forall l \in [l_{low}, l_{up}] \\ 0 & otherwise, \end{cases}$$

and an additional point mass of $1 - 1/\gamma$ at zero when $\gamma > 1$. Sample eigenvalues can be expected to fall within the range

$[l_{low}, l_{up}] = [(1 - \sqrt{\gamma})^2, (1+\sqrt{\gamma})^2]$, which indicates that when the number of variables J approaches the sample size n, the sample eigenvalues get more and more spread out. Figure 1 shows the Marčenko-Pastur density function for three values of γ, illustrating that dispersion increases in γ. Notice how the range of the sample eigenvalues is considerable, even for a ratio of 25 observations per variable.

Factor retention rules based upon eigenvalues should incorporate this random sample variation. Parallel analysis does exactly that by approximating the distribution of sample eigenvalues under the null model by means of simulating samples from a multivariate normal distribution of J variables, all with a variance of 1 and a zero-correlation between the variables. Figure 2 illustrates the close relation between the results of parallel analysis and the quantiles of the Marčenko-Pastur distribution. The gray points are eigenvalues of 1,000 datasets ($n = 300$, $J = 10$) under the null model, with the gray dashed lines representing their 5% percentile, mean, and 95% percentile, respectively. The black horizontal dashed lines demarcate the asymptotical expected first and last eigenvalue l_{up} and l_{low}. The black straight line represents the quantiles for l_j from the Marčenko-Pastur distribution (Wachter, 1976). Notice that this black line and the middle gray dashed line (i.e., the mean eigenvalues of the simulated data) are practically indistinguishable.

Although the distributional result is an asymptotical result under regularity conditions of a correlation matrix arising from large data matrices (i.e., $n, J \to \infty$, with γ constant) consisting of independently normally distributed variables, this assumption is nonessential in practice; distributions of eigenvalues of correlation matrices of non-normal variables are well approximated by the theoretical distributions, even in small datasets (see, e.g., Johnstone, 2001). This corresponds to findings for parallel analysis where the performance of the procedure is assessed as being robust to the exact univariate distributions of the variables (see, e.g., Buja & Eyuboglu, 1992; Dinno, 2009) and practically feasible for even small datasets.

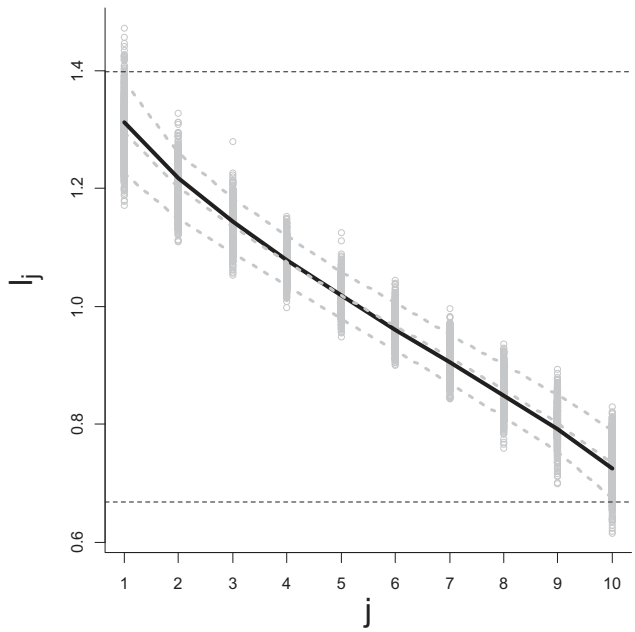


Figure 2. Results of parallel analysis (gray, 1,000 iterations) and quantiles of the Marčenko-Pastur distribution (black) for an example with $n = 300$ and $J = 10$ under the null model.

Under the Factor Model

Population level. Under a factor model with K factors the population eigenvalues will be separated in a structural part consisting of the first K population eigenvalues that absorb the shared variance in the variables accounted for by the common factors, and a residual part consisting of the remaining eigenvalues that will reflect the unique variance. Specific results on population eigenvalues can be straightforwardly derived from

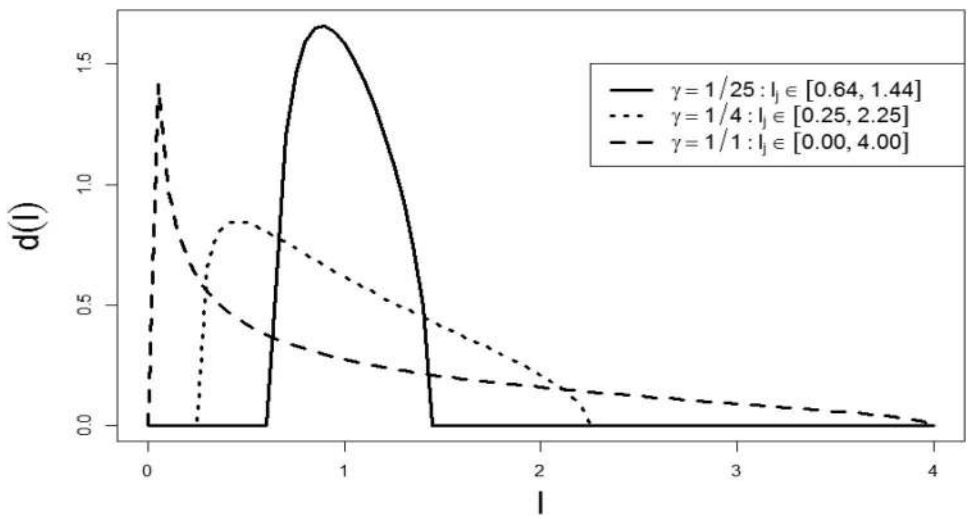


Figure 1. Marčenko-Pastur density function for three values of γ (i.e., the ratio of the number of variables J to the sample size n).

the factor model structure. Consider a simple structure factor model with K correlated factors with homogeneous interfactor correlation ρ, and for each factor, J variables with common factor loading a. The corresponding population eigenvalues are given by:

$$\begin{aligned} \lambda_1 &= 1 + (J-1)a^2 + (K-1)J\rho a^2 \\ \lambda_2 = \ldots = \lambda_K &= 1 + (J-1)a^2 - J\rho a^2 \\ \lambda_{K+1} = \ldots = \lambda_J &= (1-a^2) \end{aligned} \quad (1)$$

In Equation 1, the first term for the first eigenvalue λ_1 reflects that it will necessarily account for the variance of at least one variable; the second term represents the communality with the other variables loading on the same factor, and the third term represents the common share of variance in variables loading on other correlated factors. For the second to Kth eigenvalues, a similar reasoning holds for the first two terms in the equation, but the third term now corrects for the common share of variance that is already accounted for in the first eigenvalue. The second equation implies that the 2nd to Kth population eigenvalues are typically small for highly correlated short factors (i.e., small J and high ρ), and may even be smaller than 1. The last few eigenvalues are then equal to a variable's unexplained variance. Note that the sum of all eigenvalues equals JK.

Sample level. Because of random variation at the sample level, sample eigenvalues will deviate from the population eigenvalues derived above. As far as we know no useful theory on the distribution of empirical eigenvalues exists when there is an underlying factor structure, that is, when at least some variables are correlated in the population. However, a sampling dispersion effect similar as under the null model can also be expected to apply.

An extract of Monte Carlo simulation results is shown in Table 1 to illustrate that two mechanisms play a role: A structural dispersion effect due to the factor model and a residual dispersion effect as under the null model. In general, the first half of eigenvalues in each part (i.e., structural and residual) is pulled upward, whereas the second half of eigenvalues in each part is pulled downward. This is apparent from the results of 2,000 simulations on a reference model presented in the first columns of Table 1. The reference model is based on two uncorrelated factors, four items per factor with $a = .8$, and sample size $n = 100$. The first population eigenvalue is overestimated; the second one is underestimated, whereas the sum of these two sample eigenvalues is approximately equal to the sum of population eigenvalues (small bias, last row). As can be expected, increasing the sample size to get a better γ ratio, reduces the sampling bias in both the factor and the residual part (Condition 1 in Table 1). Decreasing the factor loadings (Condition 2) degrades the separation between the structural and the residual parts, because eigenvalues of the structural part decrease whereas those of the residual part increase, and the two dispersion biases can get mixed together for the latter half of the K factors. Consequently, the fuzzy boundaries between the structural and residual part will make it increasingly more difficult to correctly identify multiple factors, and bias for the structural part increases (last row). Increasing the correlation between factors appears to reduce the structural dispersion effect for strong factor structures, but not the residual dispersion effect (Condition 3). Combining decreased factor loadings and increased interfactor correlations blurs the boundaries again between the structural and residual part, with again higher bias for the structural part (Condition 4).

Toward Factor Retention Under the Factor Model

In parallel analysis, all reference eigenvalues are simulated under the null model of no-structure (i.e., independence). Although this procedure has been shown to perform well in a whole range of conditions, parallel analysis underestimates the number of factors in conditions with oblique factors that highly correlate, particularly when each factor is assessed with few variables (Beauducel, 2001; Cho et al., 2009; Crawford et al., 2010; De Winter, Dodou, & Wieringa, 2009; Garrido et al., 2013; Green et al., 2012; Turner, 1998; Zwick & Velicer, 1986). Harshman and Reddon (1983) were among the first to give an intuition about why parallel analysis can break down, and what should be done to fix this. The problem is an instance of an

Table 1
Bias in Sample Eigenvalues Under a Factor Population Model as a Function of Factor Structure

Condition:	Reference model			Condition 1: Increased sample size to $n = 500$			Condition 2: Decrease factor loading to $a = .4$			Condition 3: Increase factor correlation to ρ = .6			Condition 4: ρ = .6 and $a = .4$		
J	l_j	λ_j	Bias	l_j	λ_j	Bias	l_j	λ_j	Bias	l_j	λ_j	Bias	l_j	λ_j	Bias
1	3.18	2.92	.26	3.03	2.92	.11	1.71	1.48	.23	4.46	4.46	.00	1.94	1.86	.08
2	2.67	2.92	−.25	2.81	2.92	−.11	1.42	1.48	−.06	1.40	1.38	.02	1.26	1.10	.16
3	.51	.36	.15	.42	.36	.06	1.10	.84	.26	.50	.36	.14	1.07	.84	.23
4	.43	.36	.07	.39	.36	.03	.96	.84	.12	.43	.36	.07	.95	.84	.11
5	.37	.36	.01	.37	.36	.01	.85	.84	.01	.37	.36	.01	.84	.84	.00
6	.32	.36	−.04	.35	.36	−.01	.75	.84	−.09	.32	.36	−.04	.74	.84	−.01
7	.28	.36	−.08	.33	.36	−.03	.65	.84	−.19	.28	.36	−.08	.65	.84	−.19
8	.23	.36	−.13	.30	.36	−.06	.54	.84	−.30	.23	.36	−.13	.54	.84	−.30
1 + 2	5.86	5.84	.02	5.84	5.84	.00	3.13	2.96	.17	5.86	5.84	.02	3.20	2.96	.24

Note. Sample eigenvalue l_j, population eigenvalue λ_j; Sample data has sample size $n = 100$ for $J = 8$ items under a reference model of simple structure with $K = 2$ orthogonal factors (factor correlation ρ = .0), with four items loading on each factor with loading $a = .8$.

ill-defined reference. In principle the null model only applies as an adequate reference to the very first observed eigenvalue. The second eigenvalue is conditional upon the structure in the data that is captured by the first eigenvalue. In case of oblique factors, particularly when scales are short, the first eigenvalue is relatively very large, whereas the succeeding eigenvalues will be necessarily much smaller because of the total variance constraints in the eigenvalue decomposition (i.e., sum of eigenvalues = total variance). Hence, a more accurate reference for the second observed eigenvalue is the second eigenvalue of a conditional null model, that is, assuming independence of residuals, conditional on the previous factor. To conclude, a well-behaved factor retention procedure should consider taking into account the serial nature of eigenvalues.

Two recently proposed factor retention procedures take into account the serial nature of eigenvalues. Both computer-intensive procedures first estimate factor models with 1 up to X factors on the observed dataset, and then simulate new datasets according to each of these estimated models (i.e., parametric bootstrap) to serve as reference base factor retention decisions. Green, Levy, Thompson, Lu, and Lo (2012) suggest using a procedure in which the sampling distribution of the reference eigenvalues is constructed sequentially: The *j*th eigenvalue of the real data is compared to the Monte Carlo sampling distribution of the *j*th eigenvalue based upon simulated datasets generated in correspondence with the model with $(j - 1)$ factors that was estimated upon the real data. This means, for instance, that for the second eigenvalue the reference sampling distribution is based upon the estimated 1-factor model, whereas for the third eigenvalue it is based upon the estimated 2-factors model. Ruscio and Roche (2012) proposed to compute the discrepancy between the simulated eigenvalues under each factor model and the observed eigenvalues, and to assess by means of sequential model comparison which of the data-generating models fits the observed data best. Both procedures were shown to yield a considerable improvement in performance over parallel analysis in the case with oblique factors.

Yet, the two aforementioned procedures have some disadvantages in common. First, they are not linked to statistical theory on eigenvalues, which prevents deriving conditions when the procedures can be expected to perform well. Second, both procedures require extensive Monte Carlo simulations which hinder widespread application in practice. Nevertheless, because both procedures take the serial nature of eigenvalues into account and are shown to perform better than parallel analysis in applications with short correlated scales, we examine their factor retention performance later on in the Comparison to Computer-Intensive Methods and Goodness-of-Fit Tests section.

An Empirical Kaiser Criterion (EKC)

The EKC also takes into account the serial nature of eigenvalues, but does not have the disadvantages of the procedures of Green et al. (2012) and Ruscio and Roche (2012): The EKC is both linked to statistical theory and researchers' practice to obtain reliable scales, does not require simulations, and is straightforward to apply. The development of the EKC is based on three theoretically motivated ingredients that together formulate an adaptive sequence of reference eigenvalues $l^{EKC} = [l_1^{EKC}, \ldots, l_J^{EKC}]$. The first ingredient in the EKC makes use of the known sampling behavior of eigenvalues under the null model and starts by setting the first reference eigenvalue to the asymptotic maximum sample eigenvalue under the null model: $l_1^{EKC} = l_{up} = (1+\sqrt{\gamma})^2$. Hence, this first reference value will be a direct function of the variables-to-sample-size ratio (i.e., $\gamma = j/n$) in the dataset as given in the Marčenko-Pastur distribution.

The second ingredient in the EKC is an expression for calculating reference values for subsequent eigenvalues that takes into account the serial nature of eigenvalues by means of a proportional empirical correction of the first reference value as a function of prior observed sample eigenvalues:

$$l_j^{REF} = \frac{J - \sum_{j=0}^{j-1} l_j}{J - j + 1}(1 + \sqrt{\gamma})^2, \text{with } l_0 = 0.$$

The correction factor $\frac{J-\sum_{j=0}^{j-1} l_j}{J-j+1}$ has three interpretations: It is (a) the average remaining variance after accounting for the first up to the $(j - 1)^{th}$ observed eigenvalue, (b) the theoretical minimum value of l_j, and (c) the population value of λ_j if the null model of conditional independence were true after accounting for $(j - 1)$ factors.

The third and final ingredient of the EKC is the requirement that the observed eigenvalue should exceed 1. We include this restriction into the factor retention procedure for three reasons. First, a theoretical justification is that for a scale to have positive reliability, it is necessary and sufficient that the associated eigenvalue be greater than 1 (Kaiser, 1960, p. 145). Second, a practical justification is that we want to prevent the procedure to retrieve correlated residuals (corresponding to bloated specifics) or unique factors (single variables with negligible to small loadings on all factors) as factors. Third, this restriction ensures that, at the population level, the EKC is equivalent to the original Kaiser criterion (i.e., for infinite *n*, all reference eigenvalues would be 1).

Putting all ingredients together, the expression for reference eigenvalues of the EKC becomes:

$$l_j^{EKC} = \max\left(\frac{J - \sum_{j=0}^{j-1} l_j}{J - j + 1}(1 + \sqrt{\gamma})^2, 1\right), \text{ with } l_0 = 0. \quad (2)$$

Applying the EKC then implies to retain all factors 1 up to *K* for which $l_j > l_j^{EKC}\ \forall j \in [1, K]$.

Illustration

Consider a factor model with four factors, each consisting of five variables with loadings equal to .564 (i.e., corresponding to a [sub]scale reliability of .7; see next section). The correlation between factors is .6, and two of the scales have bloated specifics, that is, a pair of variables with a correlation between their residuals, here equal to .4. Figure 3 illustrates how the EKC works at the population level (infinite *n*, panel b) and at the sample level ($n = 200$, panel c). The fourth column of panel a presents the "observed" eigenvalues of one dataset generated using the specified factor structure.

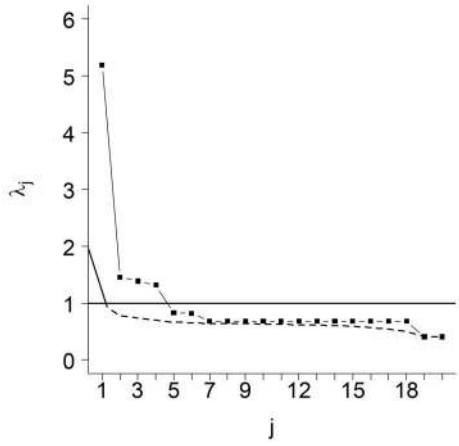
EKC retrieves the four factors with eigenvalues greater than 1, both at the sample and population level, as indicated by the

(a) Eigenvalues

J	λ_j	λ_j^{REF}	l_j	$l_j^{Parallel}$	l_j^{REF}
1	**5.19**	**1.00**	**6.07**	**1.71**	**1.73**
2	**1.46**	**.78**	**1.50**	1.57	**1.27**
3	**1.39**	**.74**	**1.42**	1.46	**1.20**
4	**1.32**	**.70**	**1.15**	1.39	**1.12**
5	.83	.67	.97	1.32	1.07
6	.82	.65	.89	1.25	1.03
7	.68	.64	.85	1.19	.99
8	.68	.64	.83	1.14	.95
9	.68	.64	.79	1.09	.91
10	.68	.63	.71	1.03	.87
11	.68	.63	.65	.99	.84
12	.68	.62	.62	.94	.80
13	.68	.61	.60	.90	.77
14	.68	.60	.55	.85	.73
15	.68	.59	.53	.80	.70
16	.68	.57	.51	.76	.65
17	.68	.55	.45	.71	.59
18	.68	.50	.38	.67	.53
19	.41	.41	.29	.62	.47
20	.41	.41	.25	.57	.43

$\lambda_j^{EKC} = \max(\lambda_j^{REF}, 1)$; $l_j^{EKC} = \max(l_j^{REF}, 1)$

(b) Population-level



(c) Sample-level

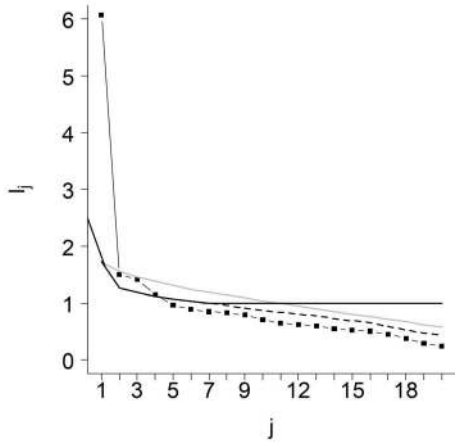


Figure 3. Example illustrating the Empirical Kaiser Criterion (EKC) with and without the greater-than-1 restriction. Note. Successive columns of (a) present the population eigenvalues (λ_j), population reference eigenvalues using the unrestricted EKC(λ_j^{REF}), sample eigenvalues (l_j), and its reference eigenvalues using parallel analysis ($l_j^{Parallel}$) and the unrestricted EKC (l_j^{REF}). The bold reference eigenvalues shows that EKC selects four factors in the example, whereas parallel analysis selects one factor. Panels (b) and (c) depict the same information for the population and sample respectively, with eigenvalues (black squares) and reference eigenvalues (black [dashed] solid line for [unrestricted] EKC and gray solid line for parallel analysis). EKC is identical to Kaiser's greater-than-one rule at the population-level in panel (b).

four eigenvalues above EKC's reference values (black solid lines). Note that without the eigenvalue-greater-than-one restriction, EKC retrieves 14 factors at the population level, because all first 14 population eigenvalues are larger than the average of subsequent population eigenvalues (black dashed line in panel b), with population eigenvalues 15 and 16 corresponding to the bloated specific item pair.

Panel c and the last three columns of panel a of Figure 3 illustrate the performance of the EKC at the sample level and contrast it with parallel analysis using 100 iterations, column-wise row permutations, and employing the 95th percentile as reference value. There are four times five items for a sample size of 200 in this example, so the first reference eigenvalue under EKC amounts to $(1+\sqrt{\gamma})^2 = (1 + \sqrt{20/200})^2 = 1.73$. We observe that, as expected, EKC and parallel analysis, which has 1.71 as the first reference value, have a similar starting point. Panels a (last two columns) and c show that subsequent reference values of the EKC are lower than those of parallel analysis, because EKC accounts for the large first sample eigenvalues. As a result, parallel analysis fails to pick up the multidimensional structure, whereas EKC correctly retrieves all four factors in the sample data. In the next sections we show that we can relatively accurately predict when the EKC will correctly retrieve the number of factors and when it likely goes wrong.

Research Design

Acceptable Scales

Earlier, we have stressed the importance of looking at performance of factor retention methods under practically relevant conditions, that is, we presuppose that researchers are aiming to identify factors from which acceptable scales can be constructed. Following others, we argue that only factors or scales containing at least three variables are viable (e.g., Glorfeld, 1995; Velicer & Fava, 1998; Zwick & Velicer, 1986). Moreover, we consider a scale acceptable if corrected-item total correlations are at least .3 (see, e.g., Nunnally & Bernstein, 1994) and the scale is sufficiently reliable. We consider sufficient reliability values of .6 to .9 in multiples of .1. Using a factor model, we can now derive requirements on factor loading a to obtain acceptable scales of J items with reliability α.

The population reliability of a scale consisting of J homogenous (i.e., parallel) variables, which equals Cronbach's alpha of that scale, with variance 1 and factor loading a can be expressed as:

$$\alpha = \frac{J}{J-1}\frac{Cov(X)}{Var(X)} = \frac{J}{J-1}\frac{J(J-1)a^2}{J(J-1)a^2+J} = \frac{Ja^2}{(J-1)a^2+1}.$$

From here, we can derive the value of loading a to obtain a certain population α:

$$a = \sqrt{\frac{\alpha}{J-(J-1)\alpha}}. \quad (3)$$

Table 2 tabulates a as a function of J (rows) and α (columns). For instance, to obtain a population reliability α equal to .7 for a scale consisting of 20 parallel items, factor loadings of .323 are required. Notice that the factor loading required to obtain a given reliability decreases in J, because α increases with the number of items while keeping a constant.

If only corrected item-total correlations of at least .3 are deemed satisfactory in practice, then 8, 16, 32, 81 parallel items with loadings of at least .397, .357, .333, .316 (printed in bold in Table

Table 2
Factor Loading a Required to Obtain Scale Reliability α as a Function of Number of Items J

Number of items J	Scale reliability			
	$\alpha = .6$	$\alpha = .7$	$\alpha = .8$	$\alpha = .9$
3	.577	.661	.756	.866
4	.522	.607	.707	.832
5	.480	.564	.667	.802
6	.447	.529	.632	.775
7	.420	.500	.603	.750
8	**.397**	.475	.577	.728
10	.361	.435	.535	.688
12	.333	.403	.500	.655
16	.293	**.357**	.447	.600
20	.264	.323	.408	.557
25	.238	.292	.371	.514
32	.212	.261	**.333**	.469
50	.171	.211	.272	.391
81	.135	.167	.217	**.316**

2), respectively, yield a population reliability of at least .6, .7, .8, .9, respectively.[1,2] In other words, longer scales yielding these reliability values are considered unsatisfactory in practice because the interitem correlations are (too) weak (with corrected item-total correlations smaller than .3). On the other hand, shorter scales yielding these reliabilities contain stronger indicator variables (with higher corrected item-total correlations) and are therefore satisfactory in practice.

Monte Carlo Experiments

The performance of the EKC is evaluated through a series of Monte Carlo simulation experiments. Experiments are defined by their data-generating population factor model, which are the null model, unidimensional factor models, and orthogonal and oblique multidimensional factor models. Across the set of experiments, the following design properties are manipulated: sample size n = (100, 250, 500), number of items per factor J = (3, 8, 16, 32, 81), scale reliability α = (.6, .7, .8, .9), number of factors K = (0, 1, 2, 3, 4, 5), and correlation between factors ρ = (0, .2, .4, .6, .8). Exact combinations of experimental factors and levels depend on the experiment. For the null model only sample size and number of items are considered, and for multidimensional factor models $K > 1$ and $J < 81$, for orthogonal and unidimensional models $\rho = 0$. Additionally, we examined the performance of the retention criteria in an experiment focusing on short scales (J = 3–6).

For all experiments, 2,500 datasets are simulated in each condition, with variables being multivariate normally distributed with a correlation matrix defined by the data-generating population model. For factor models we assume simple structure with homogenous factor loadings as derived in Equation 3. The population factor models that are in line with our definition of an acceptable scale correspond to conditions for which $J \leq 8$ & $\alpha = .6$, $J \leq 16$ & $\alpha = .7$, $J \leq 32$ & $\alpha = .8$, and $J \leq 81$ & $\alpha = .9$.

Analytical Predictions

For each experiment, analytical predictions are made on under which conditions the EKC can be expected to successfully retrieve the number of factors of the data-generating model. The analytical predictions are based on a comparison between the population eigenvalues of the data-generating model (see Equation 1) and the EKC's reference eigenvalues for the condition's sample size and number of variables (see Equation 2), where we plug in the population eigenvalues to the correction factor as proxy for the sample eigenvalues. We expect our predictions to be conservative (i.e., too restrictive), as the first sample eigenvalue(s) will be typically larger than the corresponding population eigenvalues. Specific details of the experimental design and analytical predictions for each simulation can be found in the corresponding Results section.

Performance Evaluation

For each experimental condition, the percentage of datasets for which the number of factors of the data-generating model is correctly identified is computed per factor retention method. This percentage is referred to as "hit rate" or "power," because it corresponds to the probability of correctly specifying the true "hypothesis" or number of factors. In the further evaluation of these results, a distinction will be made between relevant conditions with acceptable scales and less relevant conditions with unacceptable scales. Performance of the EKC will be classified as successful if it reaches a hit rate of at least 80% (cf., common power value) in conditions with acceptable scales where the method is predicted to work.

EKC's performance was compared to that of parallel analysis. For parallel analysis, a version was employed based on 100 iterations using column-wise row permutations and the 95th percentile as reference value (Glorfeld, 1995). This version of parallel analysis performs well in many studies (Buja & Eyuboglu, 1992; Dinno, 2009; Garrido et al., 2013; Hayton et al., 2004; Peres-Neto et al., 2005; Ruscio & Roche, 2012; Velicer, Eaton, & Fava, 2000). Given that EKC is an empirical version of the original Kaiser criterion, the performance of the latter was also evaluated, but only in the context of short correlated scales, because it may perform well in this context whereas it is well-known to perform very badly in most other conditions. Other factor retention methods were not included in our analyses because they either perform worse than parallel analysis or are not easily applicable.

Results

The Null Model: Zero Factors

Theoretical expectations. By definition of the procedure parallel analysis was expected to have a power of about 95% to detect zero factors. EKC was also predicted to have high power since

[1] Population corrected item-total correlation

$$R_{X_j X_{-j}} = \frac{Cov(X_j, X_{-j})}{\sqrt{Var(X_j)Var(X_{-j})}} = \frac{(J-1)a^2}{\sqrt{1 \times [(J-1)(J-2)a^2 + (J-1)]}}.$$

[2] Scales with satisfactory corrected item-total correlations amount to interitem correlations equal to at least .158 (a = .397, α = .6, J = 8), .127 (a = .357, α = .7, J = 16), .111 (a = .333, α = .8, J = 32), and .100 (a = .316, α = .9, J = 81).

$L_1 = (1+\sqrt{\gamma})^2$ is the asymptotically expected first sample eigenvalue under the null model.

Monte Carlo results. The Monte Carlo results summarized in Table 3 supported our expectations. The hit rate of parallel analysis was 95% in all 15 conditions; hit rate of the EKC even surpassed 95% for up to 32 variables, whereas power was somewhat lower than 95% for 81 variables. For the record, the traditional Kaiser criterion was too liberal: It retrieved more than two factors in 49% of the iterations in the $J = 3$ conditions, and 100% in all other conditions.

Unidimensionality: One Factor

Theoretical expectations. We predict the EKC to correctly identify the single factor whenever the population eigenvalue λ_1 exceeds the asymptotically expected first sample eigenvalue under the null model $(1+\sqrt{\gamma})^2$. Hence, the analytical predictions (contrasting Equation 1 and 2) are based on whether it holds that:

$$\frac{J}{J-(J-1)\alpha} > (1+\sqrt{J/n})^2.$$

Solving for sample size n, we predict that EKC will work for all 60 conditions except for three of the practically irrelevant conditions: ($J = 32$, $\alpha = .6$, $n < 108$), ($J = 81$, $\alpha = .6$, $n < 252$), and ($J = 81$, $\alpha = .7$, $n < 127$). Given that the first reference eigenvalue under parallel analysis is the simulated counterpart of the first reference eigenvalue for EKC, we do not anticipate large differences between the performances of the two methods.

Monte Carlo results. In Table 4 we provide the percentage of correct identifications by both EKC as well as parallel analysis as a function of reliability, number of variables, and sample size. These results can be summarized as follows: (a) For relevant conditions corresponding to acceptable scales for which EKC is predicted to work (upper right of Table 4, normal font), the single factor corresponding to the acceptable scale was correctly identified in all conditions, and this by both methods; with a hit rate of at least 93% ($M = 97.8\%$) for EKC and at least 97.5% ($M = 99.8\%$) for parallel analysis. Parallel analysis slightly outperformed EKC in most of these conditions. (b) For irrelevant conditions with a practically unacceptable scale but for which EKC is still anticipated to work (lower left of Table 4, italic font), EKC still had a high hit rate (at least 93.5%, $M = 95.9\%$), whereas parallel analysis showed more variable and generally weaker performance (at least 78.6%, $M = 92.3\%$). (c) Finally, for irrelevant conditions with a practically unacceptable scale for which EKC does not give theoretical guarantees (Table 4, bold font), the anticipated underperformance is confirmed. EKC's performance was still good (89.7, 92.2, and 95.0%) in conditions almost satisfying $\lambda_1 > (1+\sqrt{\gamma})^2$, but considerably worse when this was clearly not satisfied (62.2%; $J = 81$, $\alpha = .6$, $n = 100$). In contrast, parallel analysis' performance was under the 80% threshold in all these conditions (48.4%–77.5%).

Multidimensionality: K Orthogonal Factors

Theoretical expectations. Given that the K population eigenvalues are all equal to $\frac{J}{J-(J-1)\alpha}$ under this design, we predict the EKC to correctly identify the K orthogonal factors whenever the first population eigenvalue λ_1 exceeds the asymptotically expected first sample eigenvalue under the null model $(1+\sqrt{\gamma})^2$. The second to Kth population eigenvalue will also exceed their corresponding reference values as EKC corrects the starting reference eigenvalues downward for each subsequent factor. Hence, the analytical predictions are based on whether it holds that:

$$\frac{J}{J-(J-1)\alpha} > (1+\sqrt{JK/n})^2.$$

We are aware that the sample structural eigenvalues differ systematically from their corresponding population eigenvalues, and that this may affect the accuracy of our predictions. In the orthogonal case the first half of the structural eigenvalues has a positive bias, whereas the second half of structural eigenvalues has a negative bias (Table 1, columns 2 and 3, row 4). This implies that the first half of the structural sample eigenvalues will be more easily retrieved than the latter half. Yet, this positive bias also results in even lower subsequent reference eigenvalues and we expect that this downward adjustment will compensate for the slightly downward bias in the latter half of structural sample eigenvalues. Our predictions' accuracy in the simulation study will shed light on this issue.

Table 5 presents the predictions on EKC's performance as a function of sample size, reliability, and number of factors. Scale lengths from $J = 3$ up to $J = 70$ items were examined. Each cell presents the maximum scale length at which EKC is still predicted to accurately retrieve the number of factors. For instance, the number "28" for $n = 100$, $\alpha = .7$, $K = 2$ means that EKC is predicted to accurately retrieve two factors if each of the two equally long scales consists of up to 28 homogenous items with factor loadings resulting in a scale reliability of .7. Using the fact that scales are acceptable up to length $J = 8, 16, 32$, for reliabilities $\alpha = .6, .7, .8$, respectively, it follows from Table 5 that EKC is

Table 3
Percentage of Correct Identifications of Zero Factors as a Function of Sample Size n and the Number of Variables J

K = 0	EKC			Parallel analysis		
J	*n* = 100	*n* = 250	*n* = 500	*n* = 100	*n* = 250	*n* = 500
3	99.2	99.4	99.2	94.3	95.6	95.2
8	97.9	98.2	97.6	94.7	95.5	95.0
16	97.1	97.0	96.6	95.5	95.2	95.0
32	96.2	96.0	95.4	95.2	95.4	94.8
81	94.4	94.4	93.6	95.4	95.4	95.6

Table 4
Percentage of Correct Identifications of One Factor by the Empirical Kaiser Criterion (EKC) and Parallel Analysis as a Function of Reliability (α), Number of Variables (J), and Sample Size (N)

K = 1		EKC				Parallel analysis			
J	*n*	α = .6	α = .7	α = .8	α = .9	α = .6	α = .7	α = .8	α = .9
3	100	98.9	100.0	100.0	100.0	99.7	100.0	100.0	100.0
	250	100.0	99.9	99.8	100.0	100.0	100.0	100.0	100.0
	500	99.9	99.9	100.0	99.9	100.0	100.0	100.0	100.0
8	100	98.2	98.8	98.7	97.6	98.0	99.9	100.0	100.0
	250	99.2	98.8	98.2	97.5	99.9	100.0	100.0	100.0
	500	98.8	97.4	98.0	97.6	100.0	100.0	100.0	100.0
16	100	*96.1*	98.1	97.8	96.2	*91.6*	97.8	99.9	100.0
	250	*98.6*	98.0	97.5	96.6	*98.1*	99.8	100.0	100.0
	500	*97.6*	97.4	96.8	95.8	*99.3*	100.0	100.0	100.0
32	100	***89.7***	*96.6*	96.0	94.6	***77.5***	*90.4*	97.5	100.0
	250	*96.6*	*96.8*	96.0	95.6	*89.8*	*96.6*	99.9	100.0
	500	*97.0*	*96.4*	95.9	95.6	*94.7*	*98.9*	100	100.0
81	100	***62.2***	***92.2***	*94.4*	93.0	***48.4***	***75.3***	*88.3*	98.7
	250	***95.0***	*94.8*	*94.9*	94.0	***75.7***	*84.2*	*94.5*	99.9
	500	*93.5*	*94.9*	*93.8*	94.4	*78.6*	*89.3*	*97.7*	100

Note. The 18 conditions with practically unacceptable scales are printed in italics; the conditions for which EKC does not give theoretical guarantees are also printed in bold.

predicted to retrieve up to five acceptable scales for sample sizes of 250 and 500. For $N = 100$, EKC is not predicted to perform well for three or more acceptable scales with $\alpha = .6$, for four or five acceptable scales with $\alpha = .7$, and for five scales with $\alpha = .8$. We excluded the theoretical predictions for $\alpha = .9$ from Table 5 because EKC is predicted to perform well in all these conditions.

Monte Carlo results. Table 6 summarizes the simulation results by presenting the hit rate across the 192 conditions as a function of scale quality (i.e., acceptable or not) and theoretical predictions. (a) Both EKC and parallel analysis perform at a very high standard in conditions containing acceptable scales for which EKC is predicted to correctly identify the number of factors (upper left). Only in the two conditions with acceptable scales where the number of variables was larger than the sample size (cf. γ *ratio* $>$ 1), EKC showed a hit rate under 90% (i.e., 82% and 87%). Noteworthy is that EKC accurately retrieved the number of factors when it was predicted to work even in the 18 conditions with unacceptable scales, whereas parallel analysis broke down (lower left). (b) Both EKC and parallel analysis showed bad performance in the 30 conditions where we were unable to give theoretical guarantees that the EKC would work (right column). In the 12 conditions containing acceptable scales (upper right) there is a large variability in performance as indicated by the large difference between minimum and mean hit rate. Not surprisingly, all 12 conditions are characterized by low sample sizes ($n = 100$) and the worst performing of these combine small sample size with low scale reliability ($\alpha = .6$) and many factors ($K \geq 4$; i.e., all ingredients for a small signal-to-noise ratio).

Multidimensionality: K Oblique Factors

Theoretical expectations. We predict EKC to correctly retrieve the number of factors if $\lambda_k > \frac{KJ - \sum_{j=0}^{k-1}\lambda_j}{KJ - K + 1}(1 + \sqrt{KJ/n})^2 = L_k^*$, for all values $k = 1, \ldots, K$. Using population structural eigenvalues we derived a range for the correlation between factors (ρ) for which this condition is satisfied. This yields three conditions for ρ, as a function of sample size, number of factors, reliability of scales (or factor loading), and number of items.

The first condition is that the first population eigenvalue $\lambda_1 \geq L_1$, which is satisfied if

$$\rho \geq \frac{(1 + \sqrt{\gamma})^2 - 1 - (J-1)a^2}{(K-1)Ja^2}. \quad (4)$$

Table 5
Maximum Scale Length at Which EKC is Predicted to Accurately Retrieve the Number of Factors as a Function of Sample Size (N), Reliability (α), and Number of Factors (K)

	n = 100			*n* = 250			*n* = 500		
K	α = .6	α = .7	α = .8	α = .6	α = .7	α = .8	α = .6	α = .7	α = .8
2	12	28	68	+	+	+	+	+	+
3	X	16	43	23	51	+	52	+	+
4	X	10	30	16	37	+	37	+	+
5	X	6*	22	12	28	68	29	62	+

Note. + = performance guarantee for scale length $J = 3$ up to $J = 70$; X = no performance guarantee given; * = no performance guarantee for scale length $J = 3$.

Table 6
Monte Carlo Experiment for Orthogonal Factors: Hit Rate (i.e., Correctly Identified Factors) as a Function of Scale Quality and Theoretical Predictions for the Empirical Kaiser Criterion

	EKC Predicted to work			EKC No guarantee provided		
#C = 192	#C = 144	EKC	PAR	#C = 12	EKC	PAR
Acceptable scales	Mean	.98	.99	Mean	.67	.72
	Min	.82	.84	Min	.11	.46
	#C = 18	EKC	PAR	#C = 18	EKC	PAR
Unacceptable scales	Mean	.97	.75	Mean	.47	.37
	Min	.95	.33	Min	.00	.11

Note. #C indicates the number of represented experimental conditions. EKC = Empirical Kaiser Criterion; PAR = parallel analysis.

It is obtained by equating λ_1 to L_1. The first condition reflects the fact that the first eigenvalue increases with the correlation between factors. The second condition is that $\lambda_2 \geq L_2^*$. Note that if the first factor is retrieved, the first sample eigenvalue exceeds L_1, and hence it follows from Equation 2 that L_2^* exceeds L_2. Therefore, if $\lambda_2 \geq L_2^*$ it also exceeds L_2. The second condition holds if

$$\rho \leq \frac{1 + (J-1)a^2(KJ-1) - (1+\sqrt{\gamma})^2(KJ-1-(J-1)a^2)}{Ja^2(KJ-1) - (1+\sqrt{\gamma})^2(K-1)Ja^2} \quad (5)$$

which is obtained by equating λ_2 to L_2^*. Because $\lambda_2 = \ldots = \lambda_K$ and $L_2^* > \ldots > L_K^*$, we get $\lambda_k \geq L_k^*$ for all $2 < k \leq K$. The second condition reflects that the 2nd to *K*th eigenvalues decrease with the correlation between factors; if this correlation is too high, the 2nd to *K*th eigenvalues will not exceed their corresponding reference values. The third and final condition is that $\lambda_2 = \ldots = \lambda_K \geq 1$. It directly follows from Equation 1 that the third condition holds if

$$\rho \leq \frac{J-1}{J}. \quad (6)$$

The third condition reflects that, if the correlation between factors is too high, the remaining structural eigenvalues will be smaller than 1. Again, we are aware that the sample structural eigenvalues differ systematically from their corresponding population eigenvalues, and that this may affect the accuracy of our predictions.

Table 7 presents the correlation ranges for the conditions corresponding to acceptable scales ($\alpha = .6$, $J = 3, 8$), ($\alpha = .7$, $J = 3, 8, 16$), and ($\alpha = .8$, $J = 3, 8, 16, 32$). For instance, the ".325" in row "$K = 2$, $J = 16$" and column "$N = 100$, $\alpha = .7$" means that in this condition EKC is predicted to accurately retrieve the two

Table 7
Range for the Correlation Between Factors (ρ) as a Function of Sample Size n, Number of Factors K, Reliability of Scales α, and Number of Items J per Factor, for Which EKC is Predicted to Perform Well

	n = 100			*n* = 250			*n* = 500		
	α = .6	α = .7	α = .8	α = .6	α = .7	α = .8	α = .6	α = .7	α = .8
K = 2									
J = 3	.469	.658	.667*	.667*	.667*	.667*	.667*	.667*	.667*
J = 8	.264	.527	.724	.576	.727	.841	.714	.816	.875*
J = 16	X	.325	.606	X	.616	.776	X	.743	.850
J = 32	X	X	.410	X	X	.671	X	X	.783
K = 3									
J = 3	.012–.203	.488	.667*	.571	.667*	.667*	.667*	.667*	.667*
J = 8	X	.352	.622	.439	.639	.790	.628	.761	.860
J = 16	X	.089	.469	X	.498	.707	X	.669	.807
J = 32	X	X	.204	X	X	.570	X	X	.721
K = 4									
J = 3	X	.311	.598	.455	.650	.667*	.650	.667*	.667*
J = 8	X	.185	.524	.316	.560	.743	.552	.712	.832
J = 16	X	X	.337	X	.391	.645	X	.604	.769
J = 32	X	X	X	X	X	.478	X	X	.665
K = 5									
J = 3	X	.01–.12	.486	.342	.577	.667*	.586	.667*	.667*
J = 8	X	.012–018	.427	.199	.485	.700	.482	.667	.806
J = 16	X	X	.209	X	.290	.586	X	.543	.734
J = 32	X	X	X	X	X	.390	X	X	.613

Note. Only the upper bound of the range is given; the lower bound is equal to zero, unless mentioned otherwise. X = no performance guarantee can be given for EKC in this condition; * = the upper bound is set to $(J-1)/J$, the third condition of the EKC such that $\lambda_2 = \ldots = \lambda_K \geq 1$.

factors if their correlation is in the interval [0, .325]. Note that there are three cells (e.g., $K = 3$, $J = 3$, $\alpha = .6$, $N = 100$) with a correlation range excluding $\rho = 0$, suggesting that in some conditions EKC may perform better if factors are slightly correlated than when they are uncorrelated. This occurs if $\rho = 0$ and $L_1 \geq \lambda_1 = \lambda_2 \geq L_2^*$; because λ_1 is increasing in ρ, and λ_2 is decreasing in ρ, a slight increase in ρ may result in $\lambda_1 \geq L_1 \geq \lambda_2 \geq L_2^*$. A further increase of ρ ultimately yields $\lambda_2 < L_2^*$. Hence, EKC is not predicted to perform well when factors are strongly correlated, particularly so for smaller sample size, more factors, lower reliability, and more variables (conditional on reliability). Some correlation ranges are indexed with *, which reflects that the upper bound of the range is equal to $(J - 1)/J$ (see Equation 6). This also reflects that EKC may detect the K factors, after omitting the restriction $\lambda_j \geq 1$.

Monte Carlo results. Table 8 summarizes the simulation results by presenting the hit rate across the 768 conditions as a function of scale quality (i.e., acceptable or not) and theoretical predictions. We predicted EKC to correctly retrieve the number of factors in 407 conditions with acceptable scales (upper left cell). The average hit rate across all 407 conditions was .95, and the hit rate exceeded .8 in 384 of these conditions (94.3%), generally corroborating the good performance of EKC. However, EKC did not correctly retrieve the number of factors with hit rate larger than .8 in 23 conditions (5.7%), with a minimum hit rate of .15. All these conditions had in common that the scales consisted of $J = 3$ items, whereas they differed in number of scales, scale reliability, sample size, and correlation between factors. Closer inspection of these cases revealed that the sample structural eigenvalues, based on three variables each, was higher than their corresponding reference value, but not higher than 1. Dropping the eigenvalue-greather-than-1 restriction boosted EKC's performance dramatically. Average hit rate increased to 97%, whereas the hit rate exceeded .8 in all but three conditions (99.3%), with a minimal hit rate across all conditions equal to .71. Turning to the unacceptable scales where EKC was also predicted to work (lower left cell), EKC indeed performed well in all 23 conditions; the minimum hit rate was .94, with average hit rate equal to .97.

In the conditions where EKC was predicted to work (left column), parallel analysis' performance failed to match EKC's performance. Parallel analysis' hit rate did not exceed .8 in 89 conditions with acceptable scales (21.9%), with a minimum hit rate of 0% (attained for 18 conditions), whereas average hit rate was .83. Closer inspection of conditions where parallel analysis failed, confirmed that it mainly failed to detect strongly correlated scales. This was the case even for conditions with scale reliability as high as .9 (e.g., hit rate of 0.00 for five strongly correlated ($\rho = .8$) scales with 16 items each and sample size $n = 250$). Concerning unacceptable scales, the hit rate of parallel analysis did not exceed .8 in 5 of 23 conditions (22%), with the average hit rate equal to .85.

Both EKC and parallel analysis showed bad performance in the 338 conditions where we were unable to give theoretical guarantees that the EKC would work (right column); average hit rates did not exceed .26. Noteworthy, dropping the eigenvalue-greater-than-1 restriction did not improve EKC's performance much; average hit rate increased to .39 and .23 for acceptable and unacceptable scales, respectively. This implies that the signal-to-noise ratio in these conditions is just too small to allow for accurate factor retrieval, and that our derivations accurately predicted this.

Short Scales

We set up a fully factorial Monte Carlo simulation design with 720 conditions in which we manipulated sample size $n = (100, 250, 500)$, number of items per factor $J = (3, 4, 5)$, scale reliability $\alpha = (.6, .7, .8, .9)$, number of factors $K = (2, 3, 4, 5)$, and interfactor correlations $\rho = (0, .2, .4, .6, .8)$. All population models are in line with our definition of an acceptable scale.

Theoretical expectations. Applying our analysis using the three conditions in Equation 4–6 yields 499 conditions in which EKC is predicted to work, and 221 where our analytical predictions would not guarantee EKC to work. We evaluate performance of the EKC and parallel analysis, but now also explicitly include the original Kaiser criterion. Kaiser's criterion is expected to perform better than parallel analysis for short correlated scales, particularly for larger ρ, for two reasons. First, the reference eigenvalues of Kaiser (equal to 1) are smaller than those of parallel analysis (larger than 1). Second, it follows from Equation 1 that the 2nd to Kth population eigenvalues decrease in ρ and are especially small when scales are short. Hence, particularly for larger ρ, the 2nd to Kth population eigenvalue will likely be larger than 1 but not larger than the corresponding reference eigenvalue of parallel analysis.

Monte Carlo results. The simulation results (see Table 9) mirror those of the previous sections. EKC performs well when it

Table 8
Monte Carlo Experiment for Oblique Factors: Hit Rate (i.e., Correctly Identified Factors) as a Function of Scale Quality and Theoretical Predictions for the Empirical Kaiser Criterion

	EKC Predicted to work			EKC No guarantee provided		
#C = 768	#C = 407	EKC	PAR	#C = 217	EKC	PAR
Acceptable scales	Mean	.95	.83	Mean	.26	.15
	Min	.15	.00	Min	.00	.00
	#C = 23	EKC	PAR	#C = 121	EKC	PAR
Unacceptable scales	Mean	.97	.85	Mean	.23	.24
	Min	.94	.59	Min	.00	.00

Note. #C = number of represented experimental conditions; EKC = Empirical Kaiser Criterion; PAR = parallel analysis.

Table 9
Monte Carlo Experiment for Short Scales: Hit Rate (i.e., Correctly Identified Factors) Across the 720 Conditions as a Function of Theoretical Predictions for the Empirical Kaiser Criterion

	EKC Predicted to work		EKC No guarantee provided	
	#C = 499		#C = 221	
	EKC	PAR	EKC	PAR
Mean	.97	.83	.18	.07
Min	.17	.00	.00	.00
	EKC(un)	Kaiser	EKC(un)	Kaiser
Mean	.98	.89	.46	.25
Min	.74	.59	.00	.00

Note. #C = number of represented experimental conditions; EKC = Empirical Kaiser Criterion; PAR = parallel analysis; Kaiser = eigenvalue-greater-than-one rule; EKC(un) is obtained after dropping the restriction from EKC that factor eigenvalues should exceed 1.

is predicted to work; the average hit rate was .97, with hit rate exceeding .8 in 468 conditions (93.8%), with minimum hit rate equal to .17. Again, performance of the EKC was boosted if the eigenvalue-greater-or-equal-to-1 restriction was dropped; hit rate exceeded .8 in all but six conditions (98.8%), with minimal hit rate across all conditions equal to .74. Performance of parallel analysis was worse than EKC, but also worse than the original Kaiser criterion in conditions when EKC was predicted to work; the average hit rate was .83, but the hit rate did not exceed .8 in 23% of conditions, with 30 conditions having a hit rate smaller than .05 (6%). Finally, all four methods performed poorly in conditions where our analytical predictions would not guarantee EKC to work (average hit rates were .18, .07, .46, and .25 for the EKC, parallel analysis, the unrestricted EKC, and the original Kaiser criterion, respectively).

Comparison to Computer-Intensive Methods and Goodness-of-Fit Tests

Finally, we conclude this series of Monte Carlo experiments with a head-to-head comparison of parallel analysis and the EKC to other methods. These methods include the two computer-intensive methods of Green et al. (2012) and Ruscio and Roche (2012). For these two methods 100 parametric bootstrap samples were estimated per estimated factor model. Two other methods are based on goodness-of-fit tests within the structural equation modeling framework, that is, the chi-square test of exact fit and the RMSEA test for close fit. For these methods a parsimony heuristic was applied that selects the factor model with the least number of factors for which the goodness-of-fit test of exact/close fit was not rejected. Table 10 summarizes the procedures of all methods. Notice that Green et al.'s (2012) method is a more direct logical extension of parallel analysis, whereas Ruscio and Roche's (2012) method is more similar to the goodness-of-fit statistics in the sense that their underlying reference statistic is based on the full eigenvalues series under a simulated model.

The comparison is focused on conditions that pose the greatest challenges to factor retention methods: Few items per factor ($J = 3$) for a relatively large number of factors ($K = 3$) with low scale reliability ($\alpha = .6$, i.e., relatively low factor loadings), and inter-factor correlations varied across four levels $\rho = (0, .25, .5, .75)$ with higher correlations being more challenging for accurate factor retention. All population models are in line with our definition of an acceptable scale and will be tested across three sample size levels $n = (100, 250, 500)$, leading to an experimental design with 12 conditions. We feel it is important to mention here that we did *not* examine other conditions; hence, we did not select these 12 conditions post hoc to give the impression that some methods perform particularly good or bad relative to the EKC. Finally, we do not report the results of the unrestricted EKC because it did not perform substantially better than the EKC.

Theoretical expectations. Our analysis using Equations 4–6 predicts the EKC to accurately predict the number of factors in five out of 12 conditions. These are indicated by bold sample sizes in Table 11. Based on the results of previous sections, we expect the predictions to be correct and a deterioration of the performance of parallel analysis for correlated scales. Given that the chi-square and RMSEA test are asymptotically based, it may be possible that their performance is relatively worse in conditions with small sample size ($n = 100$).

Monte Carlo results. Corroborating our expectations and previous results, the EKC accurately retrieves the number of factors when it is predicted to do so (hit rate $\geq .97$), and fails to do so otherwise (hit rate $\leq .71$; see Table 11). Moreover, parallel analysis performed very bad in one condition where the EKC performed well (hit rates of .17 vs. .97, respectively, for $\rho = .5$ and $n = 500$). Importantly, the methods of Green et al. (2012); Ruscio et al. (2012), and the chi-square test performed well when the EKC was predicted to perform well (hit rates $\geq .99$, $\geq .92$, $\geq .94$, respectively), and also performed worse otherwise (hit rates $\leq .83$, $\leq .65$, $\leq .61$). The RMSEA test did not perform well, at least not as we implemented it; it failed to accurately retrieve the number of factors in two out of five conditions where the other methods performed well. Hence, our general conclusions are that our method for predicting conditions of accurate factor retention also seems to work for three other methods, and in the case of correlated short scales EKC and these three other methods perform about equally well and outperform parallel analysis.

Application: The Guilt and Shame Proness Scale (GASP)

Cohen, Wolf, Panter, and Insko (2011) developed and validated the GASP to measure individual differences in the propensity to experience the related moral emotions guilt and shame. The 16-item GASP consists of four highly correlated subscales called guilt-NBE (negative behavior-evaluations), guilt-repair, shame-NSE (negative self-evaluations), and shame-withdraw, each consisting of four items with seven response alternatives. The GASP was developed in their first study. After about half of their 450 student participants answered 60 potential GASP items (15 for each scale), the 16 GASP items were selected based on both item score analysis and strongest factor loadings obtained by exploratory factor analyses, conducted separately for each of the four subscales. Confirmatory factor analysis (CFA) on the data of remaining participants validated the 16-item GASP scale. CFA also validated GASP's factor structure in their Study 2 with 862

Table 10
Procedural Overview of Factor Retention Methods

Empirical Kaiser Criterion		
Computation	Compute $l_{up} = (1 + \sqrt{J/n})^2$ Compute Define cumulatively summed eigenvalue vector V: $v_j = \sum_{i=1}^{j} l_i$ Omit last element and put a zero upfront: $V = (0, v_1, \ldots, v_{J-1})$ Define reflected eigenvalue order vector $W = (J, J-1, J-2, \ldots, 1)$	
Reference eigenvalues	Set vector of reference eigenvalues as $\max\left(\frac{J-V}{W} l_{up}, 1\right)$	
Decision step	Choose the number of factors K for which the 1st to Kth observed eigenvalue is higher than their corresponding reference eigenvalue	
Goodness-of-Fit tests	Chi-square	RMSEA
Estimation step	Based on the observed dataset, estimate factor model with k factors; Start at k = 0, and proceed onwards until positive decision on K	
Decision step	Choose the number of factors K corresponding to the first model that does not significantly deviate from the null hypothesis of exact fit	Choose the number of factors K corresponding to the first model that does not significantly deviate from the null hypothesis of close fit
Computer intensive simulation methods		
Simulation step 1	Estimate factor model with k factors based on the observed dataset (if k = 0, no estimation required) Repeatedly simulate datasets of size n by J under model with k factors Compute eigenvalues of the correlation matrix of each simulated dataset	
	Parallel Analysis	Green et al., 2012
Simulation step 2	Not required ($k = 0$)	Repeat step 1 until decision has been reached (i.e., K + 1)
Reference eigenvalues	Set reference value for each observed eigenvalue as the value corresponding to the 95% percentile in the simulated distribution for that jth eigenvalue under the null model	Set reference value for the jth observed eigenvalue as the value corresponding to the 95% percentile in the simulated distribution for that jth eigenvalue under the model with $(j-1)$ factors
Decision step	Choose the number of factors K for which the 1st to Kth observed eigenvalue is higher than their corresponding reference eigenvalue	
	Ruscio & Roche, 2012	
Simulation step 2	Repeat step 1 until decision has been reached (i.e., K + 1)	
Reference	For each simulated dataset, compute the root mean square residual eigenvalue: RMSR = $\sqrt{\overline{(l_j - l_j^{Simulated})^2}}$	
Decision step	Choose number of factors K corresponding to the first model not showing significantly lower RMSR compared to the model with one additional factor (nonparametric difference test).	

Note. n = sample size; J = total number of items.

adults from an online subject pool. The four-factor model fitted best in both samples, although both CFAs revealed three strong correlations between factors (e.g., .67, .83, .84 in Study 2). Reliabilities of all the four scales varied from .61 to .69 in Study 1 and .62 to .71 in Study 2. The construct and predictive validity GASP were corroborated in Study 1 and Study 2, as well as in their last Study 3. Based on Cohen et al.'s (2011) findings we assume the GASP indeed measures four separate but highly related concepts.

Will parallel analysis and the EKC retrieve the four GASP factors in the data of Study 1 and Study 2 of Cohen et al. (2011)? Based on previous findings and our analysis showing bad performance of parallel analysis in contexts with short correlated subscales, we expected parallel analysis to retrieve too few factors. Because the EKC performed well in our analysis, and also in contexts with short correlated subscales, we expected EKC to accurately retrieve four factors. Figure 4 shows the results of EKC and parallel analysis on the data of the second half of Study 1 (panel a) and Study 2 (panel b). The data of the first half of Study 1 were initially not included, because these data were used to create the four subscales. Parallel analysis suggested extracting only two factors in both data sets. EKC retrieved four factors in the data of the second half of Study 1, with the fourth sample eigenvalue just above the EKC reference value, and three factors in Study 2, with the fourth sample eigenvalue being equal to .981, just below the EKC reference value. Finally, we note that EKC retrieved four factors when only applied to the data of the first half of Study 1, and when applied to the dataset combining the data of the second half of Study 1 and Study 2. Thus, all in all, the EKC provides evidence in favor of the four factors of the GASP with four short highly correlated scales. Yet, the results also illustrate that the factor structure of the GASP is relatively noisy, as the forth structural eigenvalue is not well separated from the residual eigenvalues.

Discussion

We developed a new factor retention method, the Empirical Kaiser Criterion (EKC), which is directly linked to statistical theory on eigenvalues and to researchers' goals to obtain reliable

Table 11
Monte Carlo Experiment Comparing the Hit Rate (i.e., Correctly Identified Factors) Between the Empirical Kaiser Criterion and Alternative Methods That are Either Computer-Intensive or Rely on Asymptotical Goodness-of-Fit Tests

Condition		Factor retention method					
$K = 3, J = 3, \alpha = .6$	n	Parallel analysis	Empirical Kaiser Criterion	Green et al. 2012	Ruscio Roche 2012	χ^2 Exact fit	RMSEA Close fit
$\rho = .00$	**100**	77	69	83	65	52	19
	250	100	100	99	94	95	81
	500	100	100	99	94	94	100
$\rho = .25$	**100**	30	35	54	35	34	8
	250	90	100	99	92	94	44
	500	100	100	100	95	95	90
$\rho = .50$	**100**	0	2	7	3	10	1
	250	3	71	81	61	61	2
	500	17	97	99	93	95	6
$\rho = .75$	**100**	0	0	0	0	1	0
	250	0	0	1	1	4	0
	500	0	0	27	13	23	0

Note. The sample sizes (n) of the conditions for which EKC does not give theoretical guarantees are printed in bold.

scales. EKC is easily visualized, and easy to compute and apply (no specialized software or simulations are needed). EKC can be seen as a sample-variant of the original Kaiser criterion (which is only effective at the population level), yet with a built-in empirical correction factor that is a function of the variables-to-sample-size ratio and the prior observed eigenvalues in the series. The links with statistical theory and practically relevant scales allowed us to derive conditions under which EKC accurately retrieves the number of acceptable scales, that is, sufficiently reliable scales and strong enough items.

Our simulations verified our derivations, and showed that (a) EKC performs about as well as parallel analysis for data arising from the null, 1-factor, or orthogonal factors model; and (b) clearly outperforms parallel analysis for the specific case of oblique factors, particularly whenever interfactor correlation is moderate to high and the number of variables per factor is small, which is

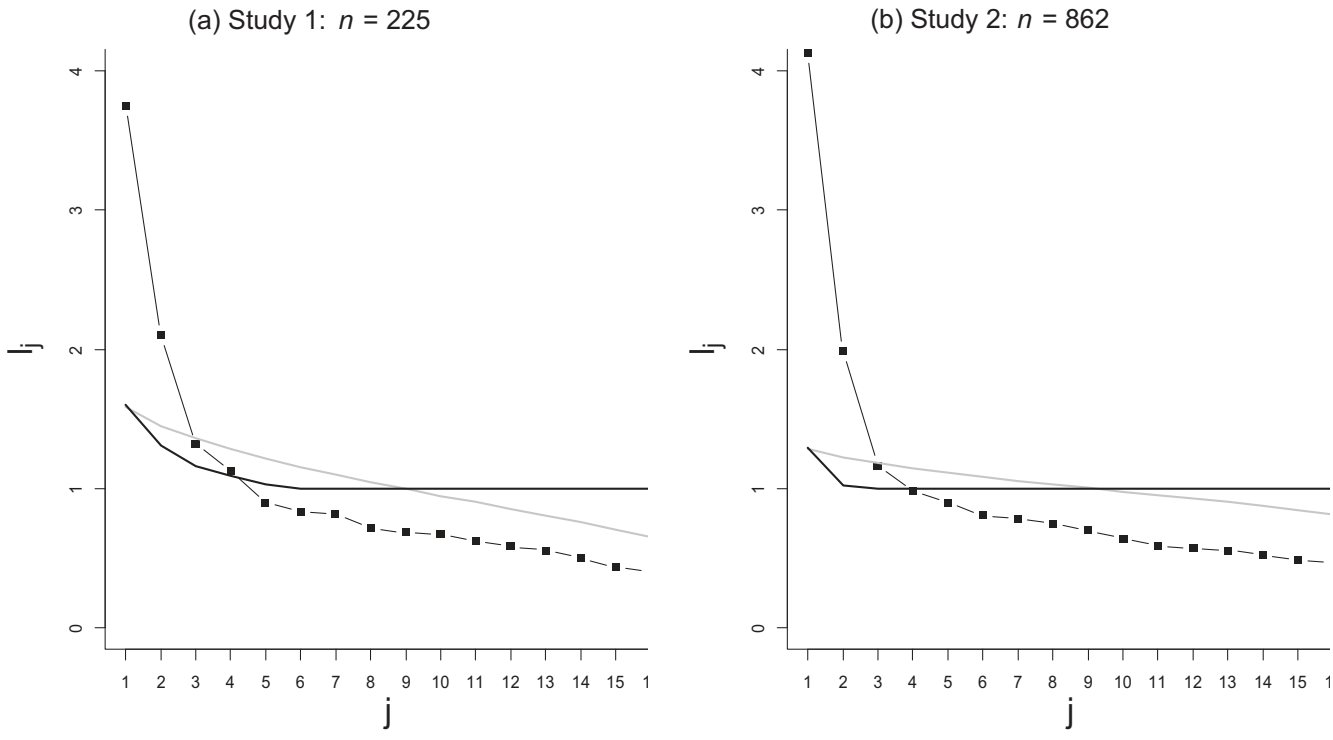


Figure 4. Annotated screeplots for the two GASP studies. Note. Black squares represent sample eigenvalues, whereas reference values are represented by solid black lines (EKC) and gray lines (parallel analysis, 95% percentile).

characteristic of many applications these days. Moreover, additional simulations suggest that our method for predicting conditions of accurate factor retention also work for the more computer-intensive methods of Green et al. (2012) and Ruscio et al. (2012). The GASP, a scale consisting of four highly correlated subscales of four variables each, was used as an illustration. The ease-of-use and effectiveness of EKC make this method a prime candidate for replacing parallel analysis, and the original Kaiser criterion that, although it empirically does not perform too well, is still the number one method taught in introductory multivariate statistics courses and the default in many commercial software packages. Furthermore, the link to statistical theory opens up possibilities for generic power curves and sample size planning for exploratory factor analysis studies.

The overall pattern of results for all Monte Carlo experiments is, unsurprisingly, in line with previous simulation studies, showing that accuracy of factor retention improves as number of variables per scale increases, sample size increases, item strengths (factor loadings) increase, number of scales decrease, and the interfactor correlation decreases. In other words, performance deteriorates with less information and a noisier factor structure. The results also indicate that it will likely be impossible to propose universal factor retention rules that always work, because the rules' performance is highly dependent on aforementioned application characteristics. Hence, to achieve accurate factor retention in an application of exploratory factor analysis, we recommend defining a potential set of expected factor structures and predefining requirements for the scales, and then conducting targeted power studies. These power studies identify the minimum sample size needed to accurately retrieve the number of factors given the predefined factor structure and scale requirements.

From the perspective of power studies, an important result of our simulations is that the formally derived predictions on performance of the EKC were confirmed by the simulations. Generally, the EKC accurately retrieved the number of factors in conditions whenever it was predicted to work well, and its performance was worse when it was not predicted to work well. More precisely, hit rate or power exceeded .8 in accordance with predictions under the null model, 1-factor model, the orthogonal factor model, and the oblique factor model with more than three variables per scale. Only in the case of minimal scales, that is, with three items per scale, did EKC sometimes not accurately retrieve the number of factors as predicted; dropping the restriction that eigenvalues should exceed 1 then mended EKC's performance. A general guideline for application that can be derived from our results (and would not need a study-specific power study), is that EKC will accurately retrieve the number of factors in samples of at least 100 persons, when there is no factor, one practically relevant scale, or up to five practically relevant uncorrelated scales with a reliability of at least .8.

More generally, our analytic and simulation results improve understanding of the role of sample size in factor retention. There are many rules of thumb that prescribe minimum sample sizes for exploratory factor analysis (see, e.g., Steger, 2006, p. 268), but that lack clear foundations. This is for instance illustrated by de Winter, Dodou, and Wieringa (2009) that show that in some cases sample sizes below 50 can be sufficient. Hence, it appears that in the current state, there is no solid advice available for sample size planning for an exploratory factor analysis study. Yet, our technique for making theoretical predictions shows promise and can potentially provide the basis for generic power curves and sample size recommendations based on hypothesized factor structure or acceptable-scale requirements.

Our predictions and results on the performance of EKC also enable improving our understanding of findings of previous studies on factor retention. We provide a few examples, particularly relevant for the practice of using correlated short scales. Cho, Li, and Bandalos (2009) and Garrido et al. (2013) examined the performance of parallel analysis with ordinal variables in a simulation study. They found that "with highly correlated factors, parallel analysis tends to moderately underfactor with the mean eigenvalue criterion and to severely underfactor with the 95th percentile criterion" (Garrido et al., 2013, p. 13), and "the performance of the P[arallel] A[analysis] procedure with highly correlated factors improved somewhat as the number of variables increased. . . . also interesting to note that increases in the [factor] loading size did little or nothing to ameliorate the effects of the high interfactor correlations." (Cho et al., 2009, p. 757). Explaining first Garrido et al.'s (2013) finding, the population 2nd to *K*th eigenvalues decrease as a function of the interfactor correlation (see Equation 1). As a result, parallel analysis tends to underfactor, and factor retention methods with lower reference values perform better, such as parallel analysis with the mean eigenvalue criterion, and even more so the original Kaiser criterion and EKC as shown in our study. Note, however, that parallel analysis with the mean eigenvalue criterion and the original Kaiser criterion perform poorly in other cases, such as no or orthogonal factors. Explaining Cho et al.'s (2009) findings, loading size is irrelevant whenever interfactor correlation is higher than $(J - 1)/J$ (see Equation 6); whatever the loading size, population eigenvalue will be smaller than 1, and the corresponding sample eigenvalue lower than its reference value. Finally, increasing the number of variables J will improve performance of factor retention methods when factors are correlated; the 2nd to *K*th population eigenvalues will increase in J linearly by a factor of $(1 - \rho)a^2$ (see Equation 1), whereas reference eigenvalues will increase less than linearly in J.

As a final example, Green et al. (2012) examined seven variants of parallel analysis in a simulation study varying five dimensions. After summarizing their results they state that "readers are likely to wonder what to make of recommendations of seven different methods dependent on conditions of a study" (p. 16). Calculating population eigenvalues in their conditions enables interpreting their results. For instance, they found that the variants of parallel analysis performed very badly when interfactor correlation was .8, and number of variables per factor was three or six, with performance even decreasing in sample size (p. 15). They report these "results were particularly difficult to understand" (p. 15), but calculations using Equation 1 show that the population eigenvalues of these conditions were only a little larger than 1 ($\lambda_2 = 1.03$ for $J = 6$, $a = .4$; $\lambda_2 = 1.1$ for $J = 6$, $a = .7$) or smaller than 1 ($\lambda_2 = .94$ for $J = 3$, $a = .4$; $\lambda_2 = .8$ for $J = 3$, $a = .7$). Factors with these low population eigenvalues, particularly those with values smaller than 1, are difficult to extremely hard to detect with parallel analysis using the 95th percentile criterion; for a population eigenvalue equal to 1, the probability of detection is about 5% by this variant of parallel analysis.

All in all, our theoretical and simulation results show that eigenvalues of a correlation matrix are useful summary statistics

that can be employed to obtain accurate factor retention rules. In the literature there is some controversy about this, because eigenvalues of a correlation matrix also form the basis for principal components analysis (PCA). We agree with Widaman (1993) that PCA is not optimally designed for interpreting the factor structure of a set of variables as it concentrates on extracting the total instead of the common variance. Yet, we disagree with the suggestions that there is no direct relationship between eigenvalues of a correlation matrix and the number of common factors (e.g., Timmerman & Lorenzo-Seva, 2011, p. 210). In fact, our theoretical results show that eigenvalues can be directly derived from a hypothesized population factor model. Furthermore, eigenvalues of a reduced correlation matrix are likely influenced by additional sources of sampling and systematic bias induced by the model used for constructing plug-in estimates for the common variances, whereas eigenvalues of the original correlation matrix are more data-driven. Some studies (e.g., Garrido et al., 2013; Velicer & Fava, 1998) suggest that variants of parallel analysis that use a reduced correlation matrix with an estimate of the common variance on the diagonal (i.e., in line with principal-axis factor analysis), or are based on minimum rank factor analysis, are less accurate than the default parallel analysis variant (i.e., using "PCA"-based eigenvalues).

Further research needs to explore how the proposed EKC performs under less clear-cut factor structures, that is, with different number of variables per factor, cross-loadings, more heterogeneous factor loadings, and including bloated specifics. Bloated specifics, caused by for instance items that are essentially rephrasings of each other, are a commonly observed anomaly in practice, but their impact on factor retention has not yet been thoroughly investigated. Deriving predictions on the performance of EKC given these less clear-cut factor structures is rather straightforward, since their population eigenvalues can directly be calculated and EKC's reference eigenvalues are not dependent on this structure. However, it is yet unclear how well these predictions for these structures will perform. An underlying assumption remains that some factor model underlies the variables' population covariance structure; hence, we need to be cautious with generalizations to conditions with population-model error (see, e.g., MacCallum, 2003).

Another direction of future research is further examining and developing the statistical theory on eigenvalues of correlation matrices. The Marčenko-Pastur distributional result provides an indication of the expected values for the sequence of eigenvalues, but it does not provide an indication about the variability around each individual sample eigenvalue. However, for covariance matrices of identically and independently normally distributed variables, Johnstone (2001) derived the asymptotical sampling distribution of the first eigenvalue to be the Tracy and Widom (1996) distribution of order 1. Unfortunately a similar result that holds for correlation matrices is not available (although, for a potential ad hoc simulated adaptation, see p. 308, Johnstone, 2001), and theoretical results for subsequent eigenvalues or for more complex structural models than the null model are—as far as we know—less developed or absent. Still, there might be other hidden gems or new developments in random matrix theory that are useful for factor analysis or other classical multivariate statistical methods such as MANOVA and canonical correlation analysis.

A practical venue for future research, as suggested by John Ruscio (personal communication, October, 2015), is to apply the EKC retroactively to published EFA results. This is possible because only sample size and number of variables are required to calculate the reference values, and observed eigenvalues are usually reported. Hence, the quality of factor retention decisions in EFA in the literature, and its development over time, can now easily be addressed using the EKC.

As a final thought we want to add that we do not advocate considering factor retention as a one-time event merely determined by a statistical optimality criterion, a yes-or-no outcome in line with current hasty scientific practice. We stress that in practice factor retention should be seen as part of a larger cumulative measurement validation project (as in the empirical illustration of the GASP), benefitting from other than statistical optimality criteria: Substantive interpretation, practical relevance, purpose of the scales, and the extent to which structures replicate for the same target population or generalize across different populations (for a discussion on replicable vs. optimal factors, see, e.g., Preacher, Zhang, Kim, & Mels, 2013) should all form important pieces of the bigger picture.

References

Anderson, G. W., Guionnet, A., & Zeitouni, O. (2010). *An introduction to random matrices*. Cambridge, UK: Cambridge University Press.

Beauducel, A. (2001). On the generalizability of factors: The influence of changing contexts of variables on different methods of factor extraction. *Methods of Psychological Research Online, 6,* 69–96.

Buja, A., & Eyuboglu, N. (1992). Remarks on parallel analysis. *Multivariate Behavioral Research, 27,* 509–540. http://dx.doi.org/10.1207/s15327906mbr2704_2

Cattell, R. B. (1961). Theory of situational, instrument, second order, and refraction factors in personality structure research. *Psychological Bulletin, 58,* 160–174. http://dx.doi.org/10.1037/h0045221

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1,* 245–276. http://dx.doi.org/10.1207/s15327906mbr0102_10

Cho, S.-J., Li, F., & Bandalos, D. (2009). Accuracy of the parallel analysis procedure with polychoric correlations. *Educational and Psychological Measurement, 69,* 748–759. http://dx.doi.org/10.1177/0013164409332229

Cohen, T. R., Wolf, S. T., Panter, A. T., & Insko, C. A. (2011). Introducing the GASP scale: A new measure of guilt and shame proneness. *Journal of Personality and Social Psychology, 100,* 947–966. http://dx.doi.org/10.1037/a0022641

Crawford, A. V., Green, S. B., Levy, R., Lo, W.-J., Scott, L., Svetina, D., & Thompson, M. S. (2010). Evaluation of parallel analysis methods for determining the number of factors. *Educational and Psychological Measurement, 70,* 885–901. http://dx.doi.org/10.1177/0013164410379332

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

Dinno, A. (2009). Exploring the sensitivity of Horn's parallel analysis to the distributional form of random data. *Multivariate Behavioral Research, 44,* 362–388. http://dx.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,* 272–299. http://dx.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,* 291–314. http://dx.doi.org/10.1111/j.1744-6570.1986.tb00583.x

Garrido, L. E., Abad, F. J., & Ponsoda, V. (2013). A new look at Horn's parallel analysis with ordinal variables. *Psychological Methods, 18,* 454–474. http://dx.doi.org/10.1037/a0030005

Glorfeld, L. W. (1995). An improvement on Horn's parallel analysis methodology for selecting the correct number of factors to retain. *Educational and Psychological Measurement, 55,* 377–393. http://dx.doi.org/10.1177/0013164495055003002

Green, S. B., Levy, R., Thompson, M. S., Lu, M., & Lo, W.-J. (2012). A proposed solution to the problem with using completely random data to assess the number of factors with parallel analysis. *Educational and Psychological Measurement, 72,* 357–374. http://dx.doi.org/10.1177/0013164411422252

Guttman, L. (1954). Some necessary conditions for common factor analysis. *Psychometrika, 19,* 149–161. http://dx.doi.org/10.1007/BF02289162

Harshman, R. A., & Reddon, J. R. (1983). Determining the number of factors by comparing real with random data: A serious flaw and some possible corrections. *Proceedings of the Classification Society of North America at Philadelphia*, 14–15.

Hayton, J. C., Allen, D. G., & Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational Research Methods, 7,* 191–205. http://dx.doi.org/10.1177/1094428104263675

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

Johnstone, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Annals of Statistics, 29,* 295–327. http://dx.doi.org/10.1214/aos/1009210544

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement, 20,* 141–151. http://dx.doi.org/10.1177/001316446002000116

Kruyen, P. M., Emons, W. M. H., & Sijtsma, K. (2012). Test length and decision quality in personnel selection: When is short too short? *International Journal of Testing, 12,* 321–344. http://dx.doi.org/10.1080/15305058.2011.643517

MacCallum, R. C. (2003). Working with imperfect models. *Multivariate Behavioral Research, 38,* 113–139. http://dx.doi.org/10.1207/S15327906MBR3801_5

Marčenko, V. A., & Pastur, L. A. (1967). Distribution of eigenvalues for some sets of random matrices. *Mathematics of the USSR-Sbornik, 1,* 457–483. http://dx.doi.org/10.1070/SM1967v001n04ABEH001994

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York, NY: McGraw-Hill.

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,* 974–997. http://dx.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,* 28–56. http://dx.doi.org/10.1080/00273171.2012.710386

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,* 282–292. http://dx.doi.org/10.1037/a0025697

Steger, M. F. (2006). An illustration of issues in factor extraction and identification of dimensionality in psychological assessment data. *Journal of Personality Assessment, 86,* 263–272. http://dx.doi.org/10.1207/s15327752jpa8603_03

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

Tracy, C. A., & Widom, H. (1996). On orthogonal and symplectic matrix ensembles. *Communications in Mathematical Physics, 177,* 727–754. http://dx.doi.org/10.1007/BF02099545

Turner, N. E. (1998). The effect of common variance and structure pattern on random data eigenvalues: Implications for the accuracy of parallel analysis. *Educational and Psychological Measurement, 58,* 541–568. http://dx.doi.org/10.1177/0013164498058004001

Velicer, W. F., Eaton, C. A., & Fava, J. L. (2000). Construct explication through factor or component analysis: A review and evaluation of alternative procedures for determining the number of factors or components. In R. D. Goffin & E. Helmes (Eds.), *Problems and solutions in human assessment: Honoring Douglas Jackson at seventy* (pp. 41–71). Boston, MA: Kluwer. http://dx.doi.org/10.1007/978-1-4615-4397-8_3

Velicer, W. F., & Fava, J. L. (1998). Effects of variable and subject sampling on factor pattern recovery. *Psychological Methods, 3,* 231–251. http://dx.doi.org/10.1037/1082-989X.3.2.231

Wachter, K. W. (1976). Probability plotting points for principal components. In D. Hoaglin & R. Welsch (Eds.), *Ninth Interface Symposium Computer Science and Statistics* (pp. 299–308). Boston: Prindle, Weber and Schmidt. http://dx.doi.org/10.1007/978-1-4615-4397-8_3

Widaman, K. F. (1993). Common factor analysis versus principal component analysis: Differential bias in representing model parameters? *Multivariate Behavioral Research, 28,* 263–311. http://dx.doi.org/10.1207/s15327906mbr2803_1

Wigner, E. (1955). Characteristic vectors of bordered matrices with infinite dimensions. *The Annals of Mathematics, 62,* 548–564. http://dx.doi.org/10.2307/1970079

Ziegler, M., Kemper, C., & Kruyen, P. (2014). Short scales—Five misunderstandings and ways to overcome them. *Journal of Individual Differences, 35,* 185–189. http://dx.doi.org/10.1027/1614-0001/a000148

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

Received June 29, 2015
Revision received December 7, 2015
Accepted December 18, 2015 ■