

MSA: The Forgotten Index for Identifying Inappropriate Items Before Computing Exploratory Item Factor Analysis

Urbano Lorenzo-Seva¹, Pere J. Ferrando¹

[1] *Department of Psychology, Universitat Rovira i Virgili, Tarragona, Spain.*

Methodology, 2021, Vol. 17(4), 296–306, <https://doi.org/10.5964/meth.7185>

Received: 2021-07-22 • **Accepted:** 2021-11-01 • **Published (VoR):** 2021-12-17

Corresponding Author: Urbano Lorenzo-Seva, Universitat Rovira i Virgili, ctra de Valls s/n, 43007 – Tarragona, Spain. E-mail: urbano.lorenzo@urv.cat

Supplementary Materials: Materials [see [Index of Supplementary Materials](#)]



Abstract

Kaiser's single-variable measure of sampling adequacy (MSA) is a very useful index for debugging inappropriate items before a factor analysis (FA) solution is fitted to an item-pool dataset for item selection purposes. For reasons discussed in the article, however, MSA is hardly used nowadays in this context. In our view, this is unfortunate. In the present proposal, we first discuss the foundation and rationale of MSA from a 'modern' FA view, as well as its usefulness in the item selection process. Second, we embed the index within a robust approach and propose improvements in the preliminary item selection process. Third, we implement the proposal in different statistical programs. Finally, we illustrate its use and advantages with an empirical example in personality measurement.

Keywords

MSA, item selection, item discrimination, sample splitting, replication, exploratory item factor analysis, KMO index, SPSS, R

In the world of factor analysis (FA), [Guttman's \(1956\)](#) image theory is viewed today (at best) as a mathematical curiosity that reflects the spirit of logical positivism of the times when it was proposed ([Ferrando, 2021](#)). However, some basic results of image theory provide useful diagnostic tools and indices for common exploratory FA (EFA). At present, most of these tools have fallen almost totally into disuse, which is only to be expected for at least three reasons. First, the theory on which they are based is considered obsolete. Second, the terminology they use is highly obscure and potentially



confusing: it is derived from models of psychometric inference (the domain-sampling theory) which are no longer in use today (Ferrando, 2021). Finally, they still remain as relics in old programs that are still active but are not generally implemented in modern EFA software.

In this article we review an index of the type discussed above: Kaiser's (1970; Kaiser & Rice, 1974) measure of sampling adequacy (MSA) at the single-variable level (a quite unfortunate name). Furthermore, our review is done within a specific scenario: the preliminary selection of items as part of the item analysis process, when this analysis is based on the EFA model. Within this framework we have several aims. The first is to present the index using modern FA terminology so that the interested practitioner can understand its underlying rationale and why it is of interest for the intended purpose. The second is to improve the use of the index by embedding it within a robust procedure that will efficiently flag the most inappropriate items. The final aim is instrumental and practical: the procedure as developed here is implemented in different statistical programs so that researchers can use the one that is best suited to their purposes.

We shall now discuss the choice of scenario and the potential role that MSA can play. To start with, we consider EFA to be the most appropriate model for item analysis, especially in the initial stages in which the grossly inappropriate items are discarded (Muñiz & Fonseca-Pedrero, 2019). Now, as we discuss in detail below, the inappropriate items that can be flagged with MSA are the types that most frequently give rise to problems when solutions with a different number of factors are tested on the calibration data. So, our interest in using MSA is clear: it can discard the most inappropriate items *before* the FA extraction stage begins, (and so before the number of factors is even specified). This initial cleaning can greatly simplify the subsequent phases of the analysis.

Reviewing MSA: A Modern View

Consider the well-known correlational structure of the EFA model

$$\mathbf{R} = \mathbf{\Lambda}\mathbf{\Phi}\mathbf{\Lambda}' + \mathbf{\Psi}^2 \quad (1)$$

where \mathbf{R} is the $m \times m$ inter-item correlation matrix, $\mathbf{\Lambda}$ is the $m \times r$ pattern, $\mathbf{\Phi}$ is the $r \times r$ inter-factor correlation matrix, and $\mathbf{\Psi}$ is the $m \times m$ diagonal matrix containing the item residual standard deviations. The main result which serves as a basis for MSA is the following (Guttman, 1956): if the set of items under scrutiny behaves according to Model 1, then the inverse of the inter-item correlation matrix should be near diagonal, and should approach a fully diagonal matrix as the number of items per factor increases and the number of common factors remains constant. This limiting result (the number of items indicating a factor increases without bound) is the origin of the unfortunate name "sampling adequacy": in fact, the items under study are viewed as a sample of a potential universe of items that could measure this factor.

Consider now, the following transformation of the inverse of \mathbf{R} :

$$\begin{aligned} \mathbf{S}^2 &= [\text{diag}(\mathbf{R}^{-1})]^{-1} \\ \mathbf{P} &= 2\mathbf{I} - \mathbf{S}\mathbf{R}^{-1}\mathbf{S} \end{aligned} \quad (2)$$

Clearly, if \mathbf{R}^{-1} is near diagonal, then the \mathbf{P} matrix in Equation 2 will have to be, too. However, the transformed matrix \mathbf{P} has a far clearer interpretation than \mathbf{R}^{-1} . \mathbf{P} is the partial inter-item correlation matrix. So, its diagonal elements are 1's and its non-diagonal elements are the correlations between the corresponding pairs of items after the influence of the remaining $m - 2$ items have been partialled out. Furthermore, \mathbf{P} is a (finite) estimate of the correlation matrix between the unique factors, which should be identity according to Model 1 (see Ferrando et al., 2021). So, to summarize: if the EFA Model 1 holds, then the partial correlations between any pair of items once the influence of the remaining items have been partialled out (i.e., the nondiagonal elements of \mathbf{P}) should approach zero.

The most popular (or the least forgotten) index derived from the results just described is the Kaiser-Meyer-Olkin (KMO) measure of overall adequacy (Kaiser, 1970; Kaiser & Rice, 1974). The KMO is still used in some item factor analysis (IFA) applications as a test for deciding whether the inter-item correlation matrix is suitable for being factored (Ferrando & Lorenzo-Seva, 2017). It is obtained as:

$$KMO = \frac{\sum_{j \neq k}^m \sum_k^m r_{jk}^2}{\sum_{j \neq k}^m \sum_k^m r_{jk}^2 + \sum_{j \neq k}^m \sum_k^m p_{jk}^2} \quad (3)$$

where j and k are indices referring to individual items, r are the first-order correlations between pairs of items (i.e., non-diagonal elements of \mathbf{R}), and p are the corresponding partial correlations (i.e., non-diagonal elements of \mathbf{P}).

Equation 3 is a relative measure, bounded between 0 and 1, and intended to reflect higher suitability as its value approaches unity. It can be seen from Equation 3 that KMO values are higher as the first-order inter-item correlations become larger and the corresponding partial correlations approach zero. Conceptually then, KMO values are larger when 1) the items are strongly correlated with one another (i.e., high internal consistency) and 2) these correlations do not reflect shared specificity (i.e., there are no unique correlated factors). Minimum cut-off values for considering the correlation acceptable for FA purposes were proposed by Kaiser (1974): .9 (*Very good*), .8 (*Good*), .7 (*Fair*), .6 (*Mediocre*), .5 (*Bad*), and lower than .5 (*Unacceptable*). Finally, simulation studies (Cerny & Kaiser, 1977; Meyer et al., 1977; Shirkey & Dziuban, 1976) suggest that KMO increases as the overall inter-item correlation (internal consistency) increases, the num-

ber of items increases, and the number of factors decreases. Note that the last two determinants can be expected from [Guttman's \(1956\)](#) asymptotic image results.

We turn now to the index considered in this paper. Like KMO, individual-item MSA is a relative index that compares the magnitude of the partial correlations in which the item under study is involved to the corresponding first-order correlations. The final version considered here ([Kaiser & Rice, 1974](#)) is

$$MSA_j = \frac{\sum_{k \neq j}^m r_{jk}^2}{\sum_{k \neq j}^m r_{jk}^2 + \sum_{k \neq j}^m p_{jk}^2} \quad (4)$$

Like KMO, [Equation 4](#) is a normed index bounded between 0 and 1, and the closer to 1 it is, the more appropriate it is for factor analysis. [Kaiser \(1970\)](#), however, was not too explicit or objective when defining what this index measures exactly. So, we are told that low MSA values flag those items that do not “belong to the same family as the other items”, or that “do not sample the same content domains measured by the remaining items”. In our view, these statements are too vague.

A close scrutiny of [Equation 4](#) suggests that MSA is expected to flag two main types of poor items: first, and above all, “noisy” items that behave almost at random, and, therefore, which lack discriminating power (see [Ferrando, 2012](#)); and second, “redundant” items that share specific content with other items in the pool. As far as the first type is concerned, consider that the expected values of both the first-order and the partial correlations for a random item are zero. So, for both types of correlation, all the observed departures from zero reflect only sampling error, so it follows that the expected value for the MSA in this case is 0.50. As for the second type, the partial correlations between items that share specific content are expected to increase faster than the corresponding first-order correlations ([Ferrando et al., 2021](#)) so the MSA for an item of this type is expected to decrease.

[Kaiser and Rice \(1974\)](#) proposed .50 as a cut-off value for discarding items that do not conform to the EFA [Model 1](#). As discussed above, this cut-off seems reasonable for “noisy” items. Furthermore, simulation results based on random or almost-random items repeatedly show that their expected MSA values across different conditions are indeed of about .50 ([Cerny & Kaiser, 1977](#); [Meyer et al., 1977](#); [Shirkey & Dziuban, 1976](#)). For “redundant” items, however, this cut-off is less clear. Our preliminary research suggests that an item of this type will be flagged with this cut-off only if it 1) has low loadings on the common factors, and 2) has several strong residual correlations with a small group of items within the pool. More generally, our research suggests that direct inspection of the partial correlation matrix **P** in [Equation 2](#), or indices based only on this matrix, provide more powerful methods than MSA for detecting “redundant” items ([Ferrando et al., 2021](#)). Even so, we believe it should be pointed out that MSA is also expected to flag

redundant items under certain conditions even when it is not a very sensitive index for this type of inappropriateness.

Noisy, low discriminating items and redundant items are amongst those expected to give rise to more problems when an EFA solution is fitted for purposes of item analysis. Noisy items are those that do not show substantial loadings on any factor when multiple (usually rotated) solutions are tried. Faced with this result, the researcher, who does not know whether these items measure a different factor or are pure noise, tries solutions with an increasing number of common factors. Usually this ends up in over-factoring, and with some of the non-discriminating items having non-negligible, totally artifactual, loadings on some of the obtained factors. As for redundant items, the problems are discussed in detail in [Ferrando et al. \(2021\)](#) but can be summarized in three points: 1) bad model-data fit, 2) spurious evidence of multidimensionality, and 3) biased parameter estimates.

From the discussion above, it should be clear that, if items with MSA estimates below .50 are discarded *before* starting the IFA process, researchers can save themselves a lot of trouble. Furthermore, the overall suitability of the debugged inter-item correlation matrix as measured by the KMO would necessarily increase. Indeed, an inspection of [Equation 3](#) and [Equation 4](#) clearly suggests that the overall KMO is some sort of average of the item MSAs. More specifically (proof can be obtained from the authors), the KMO is a weighted average (a linear composite) of the MSAs. This result is used in the proposal that follows.

A limitation of the 'original' MSA discussed so far is that it is a purely descriptive index, subject to sampling fluctuation and so potentially affected by capitalization on change. To overcome these shortcomings, we propose below a robust procedure that 1) provides confidence intervals for MSA point estimates, and 2) minimizes the risk of capitalization on change by using a cross-validation assessment schema.

Robust MSA

As stated above, our proposal for assessing MSA within a robust context is a double one. On the one hand, we propose using bootstrap re-sampling to estimate confidence intervals (CIs) for MSA. If the lowest end of the CI is above Kaiser's .50 threshold, then the corresponding item can be retained in the analysis; otherwise, the item should be removed. On the other hand, we also propose to assess the replicability of the decisions obtained from a calibration sample on the basis of further analyses in a second sample in order to avoid capitalization on chance. When the available sample is large enough to be split into two subsamples, the first subsample can be used to decide which items are to be discarded, and the second sample to assess whether the increase in the KMO value observed in the first subsample can also be replicated.

Implementation of Robust MSA

We implemented the Robust MSA procedure in three different statistical programs, and made it available (see [Supplementary Materials](#)). The utilities developed are:

1. The R script “RobustMSA.r”. This script uses only native functions in R, so no packages need to be downloaded. In order to use it, researchers have to store participants’ responses in a text file, update the name of the input file, and execute the script. The number of bootstrap samples, the confidence interval, and the threshold MAS value can also be configured.
2. The SPSS script “RobustMSA.sps”. Again, to use this script, researchers must have participants’ responses in a SPSS data file, and execute the script. The same parameters as the R script can be configured.
3. The Matlab function “RobustMSA.m”. To use this script, researchers must have participants’ responses in a Matlab matrix, and execute the script. The same parameters as the R script can be configured.
4. Finally, we implemented the Robust MSA method in our program to compute factor analysis that can be downloaded free from the site <https://psico.fcep.urv.cat/utilitats/factor/>. MSA is computed by default when the quality of the correlation matrix to be analyzed using factor analysis is assessed. If bootstrap sampling is active in the program, the program computes the 95% CI.

Method: Illustrative Analysis of a Real Dataset

In this section, we illustrate how robust MSA can be used to decide whether some items need to be removed from the item pool before an exploratory IFA is performed for purposes of item analysis.

Participants

The sample consisted of 1,156 participants (37.2% females), aged between 16 and 53 years ($M = 21.2$, $SD = 4.2$). This is the sample that was used to validate the test for the Spanish culture (Piera et al., 1993).

Instruments

The sample responded to the Spanish version of the Reducer-Augmenter Scale (Piera et al., 1993), which has 61 binary items. The test is intended to be unidimensional. In the Spanish adaptation, the original 61 items were translated, and the authors decided to maintain them all, even though some showed a low loading on the factor that was retained. The estimated reliability of the sum scores (Cronbach’s alpha) was .847, and sum scores correlated .542 with Extraversion.

Data Analysis

The aim of the present analysis is to reanalyze the original dataset, and to assess if some of the 61 items could be removed. In order to study replicability, the sample ($N = 1,156$) was split into two equivalent subsamples ($N = 578$) using Solomon method (Lorenzo-Seva, *in press*). This method improves the representativeness of the subsamples (i.e., all possible sources of variance are contained in the subsamples). The first subsample was analyzed using Robust MSA, with 3,000 bootstrap samples, and a confidence interval of 95%. As a threshold value to decide whether an item could be removed, Kaiser's proposal of .50 was used: the items that should remain were the ones that presented a 95% confidence interval above .50. The Kaiser-Meyer-Olkin (KMO) statistic (Kaiser, 1970; Kaiser & Rice, 1974) was computed to assess whether the quality of the reduced correlation matrix actually increased. Finally, the items removed from the first subsample were also removed from the second so that the replicability of the outcomes obtained in the first subsample could be inspected.

Results

The KMO index for the correlation matrix between the 61 items in the first subsample was .8219. MSA sample indices and 95% confidence intervals are shown in Table 1. As can be seen in Table 1, when only the point-estimated sample MSA value was evaluated, only one item was proposed for removal: Item 17, with an MSA value of .467. However, when the 95% confidence intervals obtained with bootstrap sampling were considered, 19 items were proposed for removal (i.e., lower interval ends below .50). When these 19 items were removed, the KMO of the trimmed correlation matrix (i.e., the correlation matrix for the 42 remaining items) was .8664.

In order to inspect the replicability of this outcome, the KMO index for the second subsample was inspected. When all the items were present, the KMO value was .8201, while the KMO of the trimmed correlation matrix was .8663. The conclusion is that the increment in the KMO of the first subsample when the 19 items were removed was reproduced in the second subsample. This suggests that the 19 discarded items were not contributing substantially to the overall adequacy across samples taken from the target population, and that this result is also to be expected in the population.

Table 1*MSA Indices Related to the 61 RAS Items*

Item	MSA	95% CI	Item	MSA	95% CI
1	.654 ^a	[.483, .700]	32	.601 ^a	[.473, .646]
2	.701	[.558, .731]	33	.897	[.814, .898]
3	.786	[.674, .807]	34	.871	[.707, .866]
4	.599 ^a	[.453, .667]	35	.591 ^a	[.481, .629]
5	.901	[.775, .893]	36	.775	[.528, .788]
6	.674 ^a	[.446, .703]	37	.918	[.845, .914]
7	.808	[.636, .822]	38	.655 ^a	[.490, .715]
8	.568 ^a	[.412, .621]	39	.871	[.752, .872]
9	.917	[.827, .910]	40	.871	[.748, .876]
10	.866	[.733, .862]	41	.747	[.546, .768]
11	.578 ^a	[.402, .665]	42	.675 ^a	[.493, .723]
12	.827	[.648, .833]	43	.848	[.733, .860]
13	.618 ^a	[.431, .675]	44	.847	[.683, .849]
14	.661	[.561, .696]	45	.868	[.734, .869]
15	.838	[.731, .850]	46	.549 ^a	[.400, .645]
16	.571 ^a	[.430, .628]	47	.802	[.659, .809]
17	.467 ^a	[.355, .595]	48	.716 ^a	[.488, .749]
18	.781	[.590, .794]	49	.636 ^a	[.468, .689]
19	.669	[.558, .697]	50	.908	[.831, .906]
20	.889	[.816, .893]	51	.836	[.662, .840]
21	.616 ^a	[.424, .689]	52	.865	[.715, .865]
22	.861	[.752, .861]	53	.800	[.608, .817]
23	.868	[.764, .871]	54	.677 ^a	[.471, .718]
24	.755	[.586, .757]	55	.657	[.564, .685]
25	.845	[.709, .847]	56	.527 ^a	[.385, .636]
26	.735 ^a	[.474, .762]	57	.743	[.588, .767]
27	.866	[.760, .875]	58	.920	[.853, .919]
28	.855	[.772, .864]	59	.760	[.627, .790]
29	.878	[.797, .880]	60	.839	[.737, .851]
30	.699	[.509, .747]	61	.790	[.573, .795]
31	.900	[.836, .903]			

^aItems proposed for removal.

Discussion

The authors of this article often review manuscripts dealing with psychometric applications, most of which, as expected, include some type of item EFA for screening

or selection purposes (Muñiz & Fonseca-Pedrero, 2019). Now, item selection is not so straightforward, particularly in multidimensional solutions. However, our view is that too much time and effort is often spent on tasks that could be solved in a much simpler way. Indeed, our first recommendation along these lines is to “clean up” the data and discard the most offending items before starting to fit different FA solutions.

In this article we have adopted a multi-faceted approach to rescue an old and forgotten index that, in our view, is quite suited to the initial debugging process mentioned above. We first discussed the rationale behind the index and why it is of interest for the task at hand using a more up-to-date FA perspective. Next, we proposed an improved procedure for using the MSA index, which is based on a cross-validation schema and provides confidence intervals around the point estimated value. In this way, the MSA becomes more of an inferential statistic than a purely descriptive index. Thirdly, we implemented our proposal in a variety of statistical programs. And finally we illustrated its usefulness with real data. We feel that practitioners now have a useful new tool in their panoply. All that remains to be seen now is to what extent it will be used.

Like any proposal of this type, ours has its shares of limitations and points that deserve further study, of which we shall discuss two before we close. First, the .50 cut-off value is the expected MSA value for an item that behaves totally at random (i.e., a totally inappropriate item with zero discrimination). However, further research on alternative cut-offs and their practical interest is warranted. A less lenient criterion may well be more useful. Second, our proposal (and the initial MSA proposal for that matter) is solely intended for product-moment correlation matrices, which implies fitting the linear FA model. In principle, the whole procedure could also be applied to tetrachoric/polychoric matrices (and so, to the nonlinear IFA model). However, some preliminary checks suggest that its use in this case might lead to results that are not so interpretable. Polychoric matrices are not product-moment matrices and their elements are estimated on a pairwise basis and have different amounts of sampling error. So, a careful study is needed to assess the behavior of the index in this case. This is also left for future research.

Funding: The authors have no funding to report.

Acknowledgments: This project has been made possible by the support of the Ministerio de Ciencia e Innovación, the Agencia Estatal de Investigación (AEI) and the European Regional Development Fund (ERDF; PID2020-112894GB-I00).

Competing Interests: The authors have declared that no competing interests exist.

Supplementary Materials

For this article source code in R, SPSS and Matlab are available via the PsychArchives repository (for access see [Index of Supplementary Materials](#) below).

Index of Supplementary Materials

Lorenzo-Seva, U., & Ferrando, P. J. (2021). *Supplementary materials to: MSA: The forgotten index to identify undiscriminating items before computing exploratory factor analysis* [Code]. PsychOpen GOLD. <https://doi.org/10.23668/psycharchives.5300>

References

- Cerny, B. A., & Kaiser, H. F. (1977). A study of a measure of sampling adequacy for factor-analytic correlation matrices. *Multivariate Behavioral Research, 12*(1), 43-47. https://doi.org/10.1207/s15327906mbr1201_3
- Ferrando, P. J. (2012). Assessing the discriminating power of item and test scores in the linear factor-analysis model. *Psicológica, 33*(1), 111-134.
- Ferrando, P. J. (2021). Seven decades of factor analysis: From yela to the present day. *Psicothema, 33*(3), 378-385. <https://doi.org/10.7334/psicothema2021.24>
- Ferrando, P. J., & Lorenzo-Seva, U. (2017). Program FACTOR at 10: Origins, development and future directions. *Psicothema, 29*(2), 236-240. <https://doi.org/10.7334/psicothema2016.304>
- Ferrando, P. J., Lorenzo-Seva, U., & Hernández-Dorado, A. (2021). Detecting correlated residuals in exploratory factor analysis: New proposals and a comparison of procedures [Manuscript submitted for publication]. *Structural Equation Modeling*.
- Guttman, L. (1956). "Best possible" systematic estimates of communalities. *Psychometrika, 21*, 273-285. <https://doi.org/10.1007/BF02289137>
- Kaiser, H. F. (1970). A second generation little jiffy. *Psychometrika, 35*, 401-415. <https://doi.org/10.1007/BF02291817>
- Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika, 39*(1), 31-36. <https://doi.org/10.1007/BF02291575>
- Kaiser, H. F., & Rice, J. (1974). Little jiffy, mark IV. *Educational and Psychological Measurement, 34*(1), 111-117. <https://doi.org/10.1177/001316447403400115>
- Lorenzo-Seva, U. (in press). SOLOMON: A method for splitting a sample into equivalent subsamples in factor analysis. *Behavior Research Methods*.
- Meyer, E. P., Kaiser, H. F., Cerny, B. A., & Green, B. F. (1977). MSA for a special Spearman matrix. *Psychometrika, 42*, 153-156. <https://doi.org/10.1007/BF02293753>
- Muñiz, J., & Fonseca-Pedrero, E. (2019). Diez pasos para la construcción de un test. *Psicothema, 31*(1), 7-16. <https://doi.org/10.7334/psicothema2018.291>
- Piera, P. J. F. I., Colet, A. V. I., Pallarés, J. T. I., & Seva, U. L. I. (1993). Spanish adaptation of the Reducer-Augmenter Scale: Relations with EPI-A scales. *Personality and Individual Differences, 14*(4), 513-518. [https://doi.org/10.1016/0191-8869\(93\)90143-Q](https://doi.org/10.1016/0191-8869(93)90143-Q)

Shirkey, E. C., & Dziuban, C. D. (1976). A note on some sampling characteristics of the measure of sampling adequacy (MSA). *Multivariate Behavioral Research*, 11(1), 125-128.

https://doi.org/10.1207/s15327906mbr1101_9



Methodology is the official journal of the European Association of Methodology (EAM).



leibniz-psychology.org

PsychOpen GOLD is a publishing service by Leibniz Institute for Psychology (ZPID), Germany.