Home Exam: Factor Analysis

Naiara Ferreira Vieira Castello
Department of Psychology, University of Gothenburg

PS2320: Methods II

Fall semester 2022

Methods

Statistical Analysis

Aiming to verify the internal structure of the instrument, an exploratory factor analysis (EFA) was conducted, followed by an assessment of the reliability of the resulting factors as separate subscales. All the analyses were performed using IBM SPSS Statistics 29, except for the parallel analysis, for which an online application was used (Patil et al., 2017).

EFA

Firstly, the data was examined for missing values and distributional shape. No missing values were identified in the dataset. The analysis of histograms and the Shapiro-Wilk test of normality of the initial 23 items indicated that the data was not normally distributed, with the statistic value ranging from W = 798 to W = 913, p-values < .001.

A preliminary analysis was then manually performed using the Spearman's non-parametric correlation matrix. Three items that did not present correlations above .3 with any other variables in the dataset were excluded. This decision is based on the assumption that the items ('Generally, I am not very aware of myself'; 'It's hard for me to work when someone is watching me'; and 'I never take a hard look at myself') were unlikely to measure the underlying concept that the instrument aims to assess (Field, 2018). Two of these items' meanings are opposed to the other items', but the sign of the (rather small) correlations indicate that they had already been correctly recoded in the dataset, with a reversed scale. The same goes for the item 'It's easy for me to talk to strangers', which had one positive correlation above .3 and was kept and further interpreted as a (already) reversed item.

Because the data was not normally distributed, the principal axis factoring was picked. This is an appropriate method for exploring the internal structure of the sample data without generalizing the results to a larger population (Field, 2018). Within psychology, it is mainly assumed that factors might intercorrelate. Thus, I chose the direct oblimin rotation method with Kaiser normalization. The cut-off point to determine relevant factor loadings was .3, a commonly used criterion amongst researchers (Field, 2018; Henson & Roberts, 2009). The Kaiser-Meyer-Olkin measure of sample adequacy was applied and indicated that the sample is suitable for a factor analysis, KMO = .835, χ^2 (190) = 3039.07, p < .001.

A first principal axis factoring was performed with extraction based on Kaiser's rule (eigenvalue > 1). A parallel analysis was then conducted to substantiate the decision about the number of factors that would be kept and indicated a four-factor solution. Considering that parallel analysis is an accurate way to determine the number of factors (Costello & Osborne, 2005) and that the scree test had pointed towards 3–4 factors, I performed another principal axis factoring fixing the number of four factors to extract. Subsequently, willing to explore the statistical and theoretical pertinence of other possible structures, a third principal axis factoring was performed, with the number of factors fixed at three.

Finally, an exploratory internal replicability analysis of the basic factor structure (Osborbe & Fitzpatrick, 2012) was performed by randomly splitting the sample into two groups ($n_1 = 257$, $n_2 = 258$), repeating the same principal axis factoring with oblimin rotation and extraction fixed at 3 factors, and comparing the outcomes between groups.

Reliability Analysis

After the EFA, a reliability analysis was conducted, aiming to verify the reliability coefficient for each of the resulting factors.

Results

EFA

An initial principal axis factoring with extraction based on Kaiser's rule showed five factors explaining a cumulative variance of 44.52%, The four-factor solution suggested by the parallel analysis explained 41.25% of the total variance. Although the parallel analysis is the most reliable method, two of the four clusters could not be substantially differentiated from each other with regards to their meaning and were thence hard to interpret conceptually.

Table 1

Total Variance Explained

Factor	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total
1	4.754	23.768	23.768	4.154	20.770	20.770	3.526
2	2.754	13.769	37.537	2.227	11.134	31.904	2.389
3	1.773	8.866	46.403	1.163	5.816	37.719	3.082
4	1.250	6.251	52.654				
5	1.020	5.102	57.755				
6	0.900	4.502	62.258				
7	0.838	4.190	66.448				
8	0.753	3.764	70.211				
9	0.706	3.529	73.740				
10	0.674	3.372	77.113				
11	0.601	3.005	80.118				
12	0.586	2.932	83.050				
13	0.553	2.763	85.813				
14	0.510	2.548	88.361				
15	0.478	2.390	90.751				
16	0.447	2.237	92.988				
17	0.401	2.004	94.992				
18	0.389	1.947	96.939				

19	0.322	1.609	98.548
20	0.290	1.452	100.000

Note. N = 515. The extraction method was principal axis factoring with an oblique rotation (oblimin with Kaiser normalization).

With a three-factor solution, the combined explained variance decreased to 37.72%, a low value compared to Stevens' (2009) recommendation of 70% or Henson and Roberts' (2006) review's average of 52%. Nevertheless, it was more meaningful in terms of the conceptual association between the items. Hence, the simplicity and theoretical adequacy of the three-factor solution were thought to compensate for the loss of information compared to the four-factor solution and I have decided to keep the three-factor model. The initial eigenvalues and the total variance explained are presented in Table 1. In this online repository, you can find the outputs of previous analyses, namely the five- and four-factor models, that could not be fully reported in this paper due to lack of space.

Table 2 presents all the information contained in the structure matrix, the pattern matrix, and the factor correlation matrix from the EFA, as well as the labels that were given to each factor after the interpretation of the results. There weren't any cross-factor loadings in the pattern matrix, but one of the items, 'I generally pay attention to my inner feelings', was not relevantly loaded into any of the factors. Despite the absence of cross-factor loadings, the structure matrix shows that various items correlate simultaneously with factors 1 and 3.

Table 2

Results from an Exploratory Factor Analysis of the Self-Scrutiny Questionnaire

Item	Pattern Matrix (Beta Weights)			Structure Matrix (Correlations)		
	Factor			Factor		
	1	2	3	1	2	3
Factor 1: Perceptions of Oneself in Social Situations						
I'm self-conscious about the way I look	**.78**	-.01	-.10	**.72**	.02	.28
I'm usually aware of my appearance	**.74**	.04	.06	**.77**	.08	**.43**
Before I leave my house, I check how I look	**.62**	.02	-.06	**.59**	.05	.25
I usually worry about making a good impression	**.57**	-.10	.04	**.59**	-.07	**.32**
I'm concerned about what other people think of me	**.57**	.12	.07	**.61**	.16	**.35**
I care a lot about how I present myself to others	**.57**	.13	.02	**.58**	.16	**.31**
I'm concerned about my style of doing things	**.37**	.01	.22	**.47**	.04	**.40**
Factor 2: Discomfort in Social Situations						
It takes me time to get over my shyness in new situations	-.01	**.76**	.10	.08	**.77**	.13
I feel nervous when I speak in front of a group	.16	**.70**	.00	.19	**.71**	.10
Large groups make me nervous	.09	**.70**	.15	.20	**.71**	.22
I get embarrassed very easily	.12	**.64**	.01	.16	**.65**	.10
It's easy for me to talk to strangers	-.19	**.51**	-.15	-.24	**.49**	-.22

Factor 3: Self-Examination						
I think about myself a lot	-.08	.13	**.71**	.28	.16	**.68**
I'm always trying to figure myself out	-.13	.06	**.64**	.19	.08	**.58**
I'm quick to notice changes in my mood	.02	.01	**.61**	.32	.04	**.62**
I'm constantly thinking about my reasons for doing things	.12	-.05	**.50**	.37	-.02	**.56**
I sometimes step back (in my mind) in order to examine myself from a distance	.07	.03	**.42**	.27	.05	**.45**
I know the way my mind works when I work through a problem	.06	-.12	**.40**	.25	-.10	**.42**
I often daydream about myself	.01	.02	**.39**	.20	.03	**.39**
Not retained in any factor						
I generally pay attention to my inner feelings	.18	-.17	.26	.30	-.15	**.34**
Correlations among factors						
1	—					
2	.05	—				
3	**.49**	.04	—			

Note. The same extraction conditions described in Table 1 apply. The pattern matrix values correspond to factor loadings, while the structure matrix presents correlations between variables and factors. Values above .30 are in bold.

The factor correlation matrix confirmed a correlation between factors 1 and 3, which corroborates the choice of an oblique rotation method. On the other hand, results suggest that factor 2 is not associated with either factor 1 or factor 3, since both coefficients lie below the threshold of .1 that would indicate a small correlation (Field, 2018).

Three variables had been deleted from the original 23 variable pool after the preliminary analysis for not presenting a correlation above .3 with any other variable. However, there were still numerous items with most of their intercorrelations below .3, which according to Field (2018) suggests that the data might not be fit for a factor analysis. One of these items was 'I generally pay attention to my inner feelings', that had one correlation slightly above .3 and ended up not having a relevant loading into any factor. The EFA correlation matrix is available as a supplemental file.

Table 3 shows a high frequency of communalities smaller than .4, categorized as low (Field, 2018). The reason for the low correlations and communalities might be associated with aspects of language and context that might make items unreliable, or with conceptual unfitness of the variables (Fabrigar et al., 1999). In any case, low communalities can distort significantly the results of EFA and the results should be analysed cautiously.

Table 3

Communalities of the Self-Scrutiny Scale Variables

Item	Communalities

	Initial	Extraction
I'm always trying to figure myself out	.35	.35
I'm concerned about my style of doing things	.28	.26
It takes me time to get over my shyness in new situations	**.49**	**.60**
I think about myself a lot	**.44**	**.48**
I care a lot about how I present myself to others	.39	.36
I often daydream about myself	.22	.15
I get embarrassed very easily	.39	**.43**
I'm self-conscious about the way I look	**.50**	**.53**
It's easy for me to talk to strangers	.29	.33
I generally pay attention to my inner feelings	.20	.17
I usually worry about making a good impression	.36	.36
I'm constantly thinking about my reasons for doing things	.31	.33
I feel nervous when I speak in front of a group	**.48**	**.53**
Before I leave my house, I check how I look	.37	.35
I sometimes step back (in my mind) in order to examine myself from a distance	.22	.21
I'm concerned about what other people think of me	.38	.39
I'm quick to notice changes in my mood	.38	.39
I'm usually aware of my appearance	**.58**	**.60**
I know the way my mind works when I work through a problem	.23	.20
Large groups make me nervous	**.49**	**.55**

Note. The same extraction conditions described in Table 1 apply. Values above .4 are in bold.

The pattern matrix presented in Table 2 was used as a reference for the final three-factor model and the model was interpreted with respect to the content of the grouped items.

The first factor was called Perceptions of Oneself in Social Situations and gathers items that represent people's concerns with the way they present themselves and are perceived by others in social situations. They involve preoccupations with physical appearance, as in 'I'm self-conscious about the way I look', and a more general worry with the way one is seen, as in 'I'm concerned about what other people think of me'.

Factor 2 was named Discomfort in Social Situations and depicts variables as 'I feel nervous when I speak in front of a group', and 'It's easy for me to talk to strangers' (reversed). It depicts nervousness in dealing with groups and new scenarios.

The third factor assembles items that represent involvement with thoughts about oneself. It shows acts of introspection focused on one's own inner characteristics, as reasoning ('I'm constantly thinking about my reasons for doing things') and state of mind ('I'm quick to notice changes in my mood'), and was called Self-Examination.

The three factors describe different facets of a broader phenomenon that I have named Self-Scrutiny. While Self-Examination refers to private processes of assessing one's own thoughts, emotions and behaviours, Perceptions of Oneself in Social Situations adds a social element to these private processes. Although the cognition still focuses on oneself, Perceptions of Oneself in Social Situations assumes the presence of an external observer who is thought to assess and judge ('I'm concerned about what other people think of me'). Discomfort in Social Situations' items have a different construction. They take place in concrete social situations (e.g., speaking in front of a group, talking to strangers) and involve not only thoughts and reasoning, but also emotions and behaviours.

The lack of correlation between Discomfort in Social Situations and the other two factors indicates that intense self-assessment and concerns about how one is seen by others are not necessarily accompanied by social uneasiness in a behavioural and emotional level. Likewise, social awkwardness and discomfort doesn't always take overthinking. Nonetheless, they are thought to be expressions of the larger construct of Self-Scrutiny.

An internal replicability analysis of the basic factor structure showed that, in addition to the previously problematic 'I generally pay attention to my inner feelings', the item 'I'm concerned about my style of doing things' failed to be loaded consistently into one factor. This implies that it possibly needs to be revised or deleted (Osborne & Fitzpatrick, 2012). This was performed as an exploratory procedure and details are in supplemental material.

Reliability Analysis

The factors that resulted from the EFA were submitted to reliability analyses as three independent subscales: Perceptions of Oneself in Social Situations, consisting of 7 items (α = .81); Discomfort in Social Situations, with 5 items (α = .79); and Self-Examination, containing 7 items (α = .73). No variable removals would enhance reliability, except for 'It's easy for me to talk to strangers', whose deletion would increase the Cronbach's α of Discomfort in Social Situations to .81. The complete output is in the supplemental material.

Because Cronbach's α is sensible to the number of items, it should not be taken as an absolute measure, but rather interpreted with regards to the context (Field, 2018). That being said, the three scales presented acceptable values for Cronbach's α, which generally indicates that they are reliable.

Conclusions

The EFA of the Self-Scrutiny Scale resulted in 3 meaningful factors that explained a total of 37.72% of the variance and the subscales had acceptable Cronbach's α values. Out of the 23 original items, 5 had their suitability questioned at some point of the analyses. This, along with the presence of low communalities, may indicate a need for revision of the items or for new studies with larger samples. Considering the exploratory characteristic of this study, I recommend a confirmatory factor analysis to investigate more accurately the replicability of structure and magnitude of factor loadings.

References

Costello, A. B., & Osborne, J. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research, and Evaluation*, *10*(7). https://doi.org/10.7275/JYJ1-4868

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the Use of Exploratory Factor Analysis in Psychological Research. *Psychological Methods*, *4*(3), pp. 272-299. https://doi.org/10.1037/1082-989X.4.3.272

Field, A. (2018). *Discovering Statistics Using IBM SPSS Statistics* (5th ed.). SAGE.

Henson, R. K., & Roberts, J. K. (2006). Use of Exploratory Factor Analysis in Published Research: Common Errors and Some Comment on Improved Practice. *Educational and Psychological Measurement*, *66*(3), 393–416. https://doi.org/10.1177/0013164405282485

Osborne, J. W., & Fitzpatrick, D. C. (2012). Replication analysis in exploratory factor analysis: What it is and why it makes your analysis better. *Practical Assessment, Research, and Evaluation*, *17*. https://doi.org/10.7275/h0bd-4d11

Patil, V. H., Surendra, N. S., Sanjay, M., & Donavan, D. T. (2017). *Parallel Analysis Engine to Aid in Determining Number of Factors to Retain using R* [Computer software]. https://analytics.gonzaga.edu/parallelengine/

Stevens, J. P. (2009). *Applied Multivariate Statistics for the Social Sciences* (5th ed.). Taylor & Francis Group. ProQuest Ebook Central. https://ebookcentral.proquest.com/lib/gu/detail.action?docID=1046791