EDUR 9131: Exploratory Factor Analysis
25 March 2023

Exploratory Factor Analysis (EFA): Brief Overview with Illustrations

1. Logic of EFA

EFA is designed to determine whether a set of variables can be reduced to a smaller number of factors due to clustering or correlation among variable scores. If two variables correlate highly, for example, it is possible they represent the same construct; this is expected if these items were designed to measure the same construct.

EFA uses correlations among variables to determine whether factors are present. For example, assume there are responses to 6 items on an instrument; Table 1 presents the resulting correlations.

Table 1: Patterns of Correlations Demonstrated

	1	2	3	4	5	6
Item 1	---					
Item 2	**.59**	---				
Item 3	**.64**	**.72**	---			
Item 4	.02	.06	.08	---		
Item 5	-.05	-.14	.12	**.43**	---	
Item 6	.10	.02	.05	**.68**	**.55**	---

Note the bold correlations in green and blue. The correlations among items 1 to 3; these items seem to correlate well together and therefore may form a common measure if those items were designed to measure the same construct. The same may be applied to items 4 to 6. The correlations among the two sets of items, however, are weak and show that the two sets of items appear to be unrelated.

When analyzing data from scales, for example, we assume participants respond to items because the construct measured leads them to respond in a consistent way. If items 1, 2, and 3, for example, were designed to measure mathematics self-efficacy, then those who have high levels of efficacy should respond similarly to items 1, 2, and 3 (assuming there are no reverse-scaled items), and this pattern of responses would produce moderate to strong correlations like those shown above.

Figure 1 illustrates **reflective** factors. The figure shows that items 1 to 3 are reflective (or indicative, or indicators) of factor 1, and items 4 to 6 are reflective of factor 2. Figure 1 indicates items 1, 2, and 3 correlate because their scores are functions of factor 1, and items 4, 5, and 6 correlate due to factor 2.

Factor analysis is often used to assess the internal structure of scales. EFA can be used to determine whether variables (indicators) group or cluster as expected on certain factors; researchers can use EFA to check on the internal structure of scales to ensure that items load on the constructs (factors) for which they were designed. EFA is a power method for providing evidence for construct validity.

2. Formative vs Reflective Models, and Principal Component Analysis (PCA) vs Exploratory Factor Analysis (EFA)

Many argue that factor analysis and principal component analysis are essentially the same, and it is true that they often produce similar results. Conceptually, however, the two are very different. PCA is designed to produce "a linear combination of variables; Factor Analysis is a measurement model of a latent variable" (Karen, 2018).

With PCA, the model for a component is

$$C = b1\ X1 + b2\ X2 + b3\ X3 \ldots$$

where C is the component, b are the coefficients, and X are the variables or items. With EFA, the model is

$$X1 = b1\ F1 + b2\ F2 + b3\ F3 + \ldots + u1$$

where X is the indicator or item, b are the coefficients, F are the factors, and u is the error term for each X.

An EFA model is illustrated in Figure 1 and a PCA model is illustrated in Figure 2. EFA is for **reflective** constructs and PCA is for **formative** constructions.

Figure 1: Reflective Model with Two Factors (Factor Analysis Model)

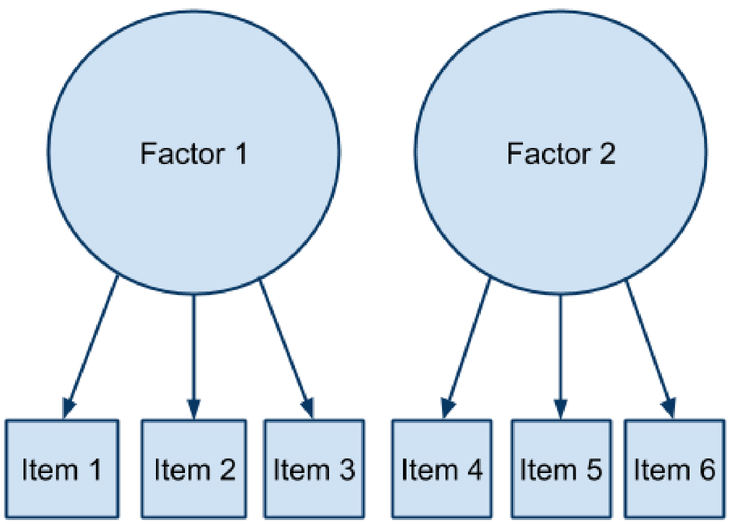
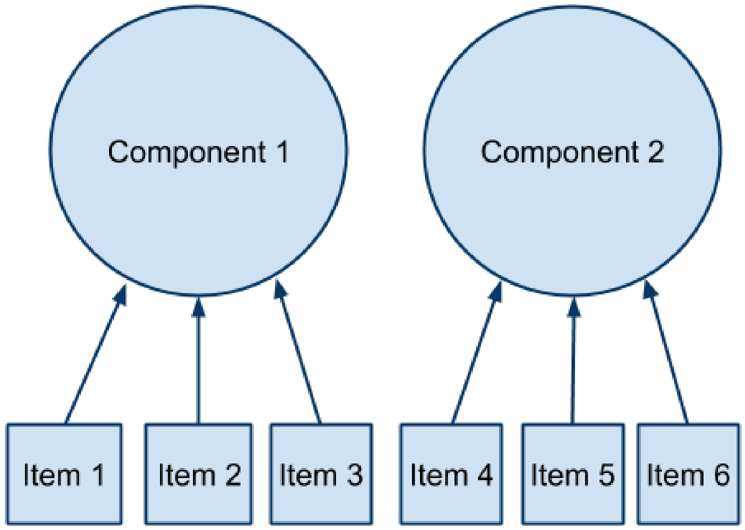


Figure 2: Formative Model with Two Components (Principal Component Model)



Reflective models assume that the factor is the causal agent leading to scores obtained for the indicators; the factor predicts or causes variation in the indicators, so the factor is the independent variable and the indicators are the dependent variables. With this model one assumes that the factor exists independent of the indicators; we use indicators to help us measure the factor. The factor is the causal agent and results in variation observed in the indicators. Example: The greater your math self-efficacy (factor), the (a) more time you spend on difficult problems (indicator), the (b) more interest you have in math (indicator), and the (c) more confidence you have with math problems (indicator).

Formative models represent a different causal assumption compared with reflective models. With formative models, the indicators are predictors or causal agents for variation in the component. Indicators are the independent variables and the component is the dependent variable. It is also possible to view this model not as cause and effect, but simply as a mathematical structure such that the indicators are used to form a composite variable called a component. In either view, the component is formed by combining indicators; this suggests the component may not exist independent of the indicators, although that is not the case in every situation (e.g., see cyber-harassment example below – victim experience exists independent of the indicators). Example: The greater one's (a) wealth (indicator), (b) education (indicator), and (c) occupational prestige (indicator), the greater one's socio-economic status (SES; component).

Coltman et al. (2008) explain that with reflective models we expect to see strong correlations among items and thus high internal consistency for each factor; with formative models items may be independent and uncorrelated since the component is a composite; there is no need for items to correlate (although if there are correlations, the items must correlate positively otherwise reverse scoring is needed because failure to reverse score means items are both adding and subjecting from the composite variable score). Internal consistency is expected and assessed with reflective models, but not necessary for formative models.

Example of Reflective and Formative Models: Cyber-harassment

Cyberbullying exists as both reflective and formative models. Suppose we ask the following three questions.

1. Visual harassment – electronically posting images or videos with the intent to embarrass, threaten, intimidate, offend, manipulate, harass, or otherwise make someone experience negative reactions.

1V. How many times has this **happened to you** in the past 3 years?	1B. How many times have **you done this to someone else** in the past 3 years?
0. Never 1. 1 time 2. 2 times 3. 3 times 4. 4 or more times	0. Never 1. 1 time 2. 2 times 3. 3 times 4. 4 or more times

2. Written harassment – electronically posting written message with the intent to embarrass, threaten, intimidate, offend, manipulate, harass, or otherwise make someone experience negative reactions.

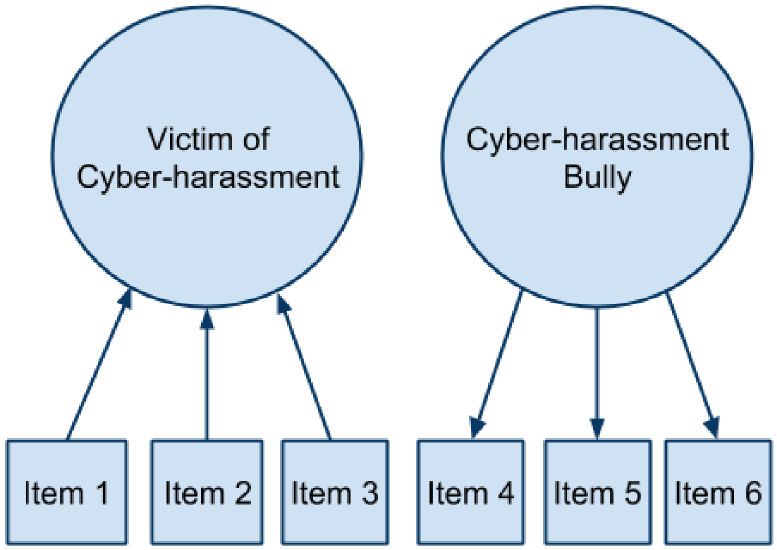
2V. How many times has this **happened to you** in the past 3 years?	2B. How many times have **you done this to someone else** in the past 3 years?
0. Never 1. 1 time 2. 2 times 3. 3 times 4. 4 or more times	0. Never 1. 1 time 2. 2 times 3. 3 times 4. 4 or more times

3. Spoken/Verbal harassment – to speak or leave a spoken message electronically with the intent to embarrass, threaten, intimidate, offend, manipulate, harass, or otherwise make someone experience negative reactions.

3V. How many times has this **happened to you** in the past 3 years?	3B. How many times have **you done this to someone else** in the past 3 years?
0. Never 1. 1 time 2. 2 times 3. 3 times 4. 4 or more times	0. Never 1. 1 time 2. 2 times 3. 3 times 4. 4 or more times

Items 1V, 2V, and 3V are indicators for victims cyer-harassment, and items 1B, 2B, and 3B are indicators of cyber-harassment bullying behavior. The wording of items 1V, 2V, and 3V make clear the experience of cyber-harassment was thrust upon the vicitm, and the wording of items 1B, 2B, and 3B make clear these harassment behaviors were caused by the bully. The theoretical model for cyber-harassment is shown in Figure 3.

Figure 3: Formative and Reflective Models for Cyber-harassment



Vicitms are subjected to harassment activities. These experiences are directed toward them; they are not the perpetrator of these actions, so the causal links in Figure 3 must flow from item to componet. This is an example that would be suitable for PCA – a composite indicator of victim experience.

Bullies, on the other hand, initiate and perpetrate cyber-harassing behaviors. These behaviors and actions emanate from the bully – the bully is the causal agent of these behaviors. Given this, the links flow from from factor to item. This is an exampel that would be suitable for EFA – a theoretical measurment model for the bully behavior.

3. EFA Steps, Components, and Concepts

EFA assumes variables are ordinal (~5 or more categories), interval, or ratio. EFA software is typically not designed for nominal or categorical variables. Variables must be able to form a correlation (or covariance) matrix for analysis.

(a) Initial Extraction

With the initial extraction we obtain estimates of amount of variance each factor predicts among all model indicators. We expect this to be high, usually 60% or more.

Eigenvalues are reported; these indicate the amount of factor variance attributed to each factor.

Communalities are also reported; these are R^2 values that indicate the proportion of variance in each indicator that is predicted by the factors. Items with low communalities are not predicted well by factors and perhaps should be removed. They tend to show low loadings with both before and after rotation.

An **extraction** method must be selected. **Principal Axis Factoring** is commonly used and will be recommended here. Other options exist and often results are similar. Examples include Maximum Likelihood, Alpha Factoring, and Generalized Least Squares.

We can learn whether the correlation matrix among indicators is suitable for EFA by examing **Bartlett's test of sphericity** and **Kaiser-Meyer-Olkin (KMO)** measure of sampling adequacy. Note that usually one analzes the **correlation matrix** because the covariance matrix will produce loadings that can be difficult to interpret if indicator variances greatly differ.

(b) Determine Number of Factors to Retain

Examine results of the initial extraction to help determine how many factors should be retained in the measurement model. This can be done several ways:

- **Theoretical Model** – if scale developed for three factors, then one should see three factors
- **Scree Plot** – look for elbow in screen plot (a sharp turn to the right) to determine number of factors
- **Eigenvalues Size** – a default option; select the number of factors with eigenvalues greater than 1.00
- **Percentage Explained** – keep number of factors that account for 70% or 80% of item variance
- **Parallel Analysis** – estimate eigenvalues size expected by chance and select only those eigenvalues larger than what would be expected by chance

(c) Factor Rotation and Interpretation

The goal with this step is to simplify the factor loadings to make factor interpretation easier. A **simple structure** is sought; this means indicators load high on one factor and low on other factors.

Orthogonal rotation means the factors are uncorrelated. This is rarely a reasonable assumption so I don't recommend orthogal rotation options (e.g, Varimax).

Oblique rotation means factors are correlated and this is usually reasonable. **Oblimin** and **Promax** are two oblique options provided in my version of SPSS.

Examine factor loadings to determine factor composition and description; factor loadings are used to name factors.

4. Example 1: Autonomy Support and Student Ratings of Instruction

Student Ratings Data

http://www.bwgriffin.com/gsu/courses/edur9131/2018spr-content/12-factor-analysis/studentratingsdata.sav

Two constructs of interest:

Construct 1: Autonomy Support

24. The instructor was willing to negotiate course requirements with students.
25. Students had some choice in course requirements or activities that would affect their grade.
26. The instructor made changes to course requirements or activities as a result of student comments or concerns.

Construct 2: Student Ratings of Instructor and Course

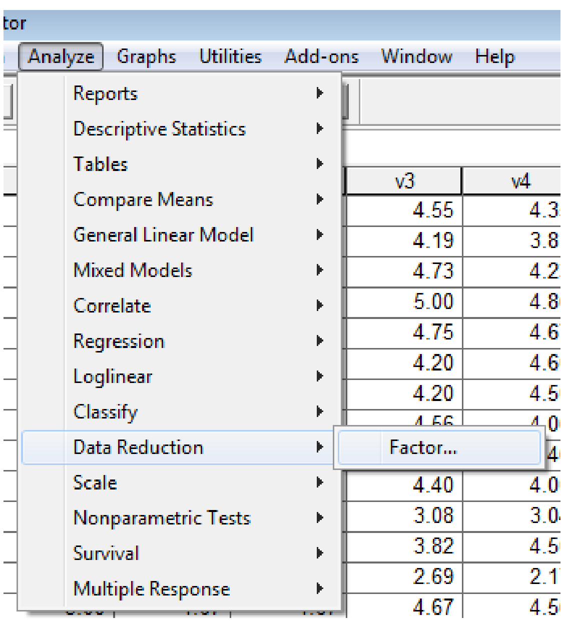
5. The instructor presented the material in a clear and understandable manner.
6. Course materials were well prepared and organized.
8. The instructor made students feel welcome in seeking help/advice in or outside of class.
9. The content of this course is useful, worthwhile, or relevant to you.
10. Methods of evaluating student work were fair and appropriate.
13. The instructor gave students useful/helpful feedback on work.
29. Overall, how would you rate this course?
30. Overall, how would you rate this instructor?

The purpose of this EFA is to assess the internal structure of these two scales (construct validity check). Ideally two distinct factors would emerge, one for autonomy support (only items 24, 25, and 26 load on this factor), and one for student ratings (all other variables should load on this factor). We hope to see weak loadings across factors for variables that were not designed to measure that construct, i.e., hope to find simple structure.

SPSS Factor Analysis Commands

Before proceeding to EFA, first check the correlation matrix among variables to ensure things look appropriate. Always check that items are reversed scored as needed, and that reversed items are included instead of the original un-reversed items.

Analyze -> Data Reduction (or Dimension Reduction) -> Factor

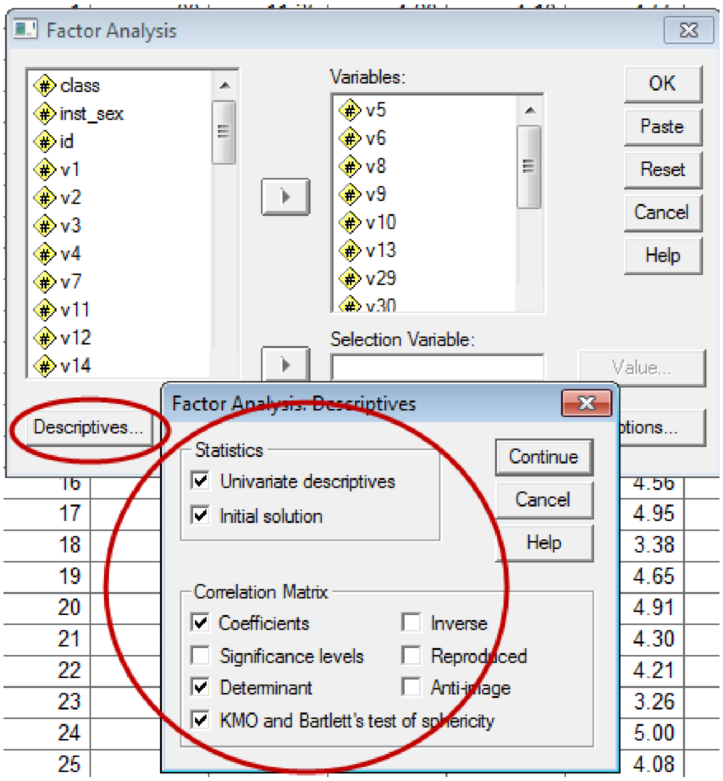


Move items identified above into the Variables box.

Items include the following items: v5, v6, v8, v9, v10, v13, v29, v30, v24, v25, v26

Descriptives -> Check boxes noted below

- Univariate Descriptives
- Initial Solution
- Coefficients
- Determinant
- KMO and Bartlett's test of spherecity



Extraction –>

Method = Principal Axis Factoring

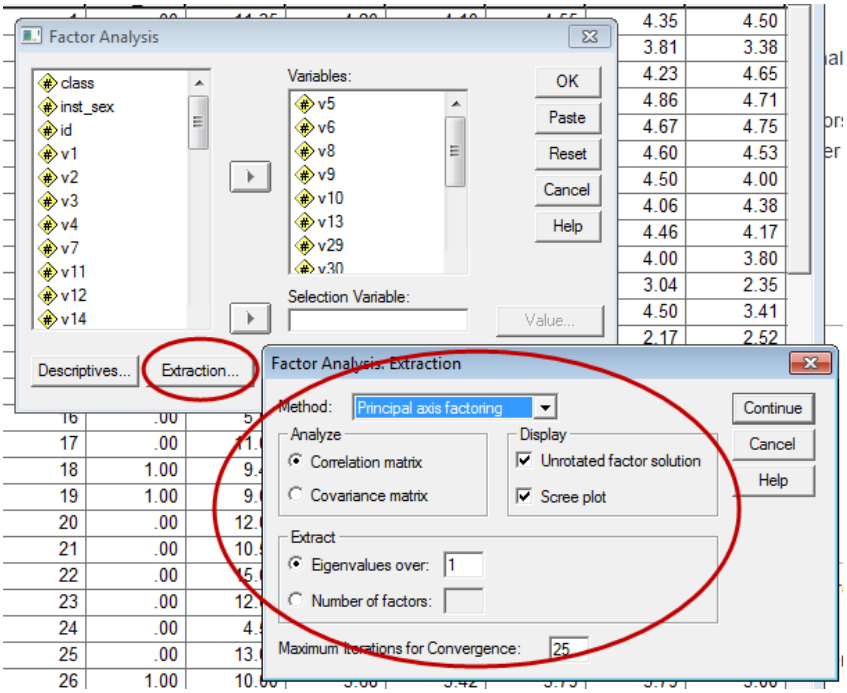
(don't use Principal Component Analysis, that is different from factor analysis)

Analyze = Correlation Matrix

Display = Unrotated factor solution

Display = Scree Plot

Eigenvalue over = 1

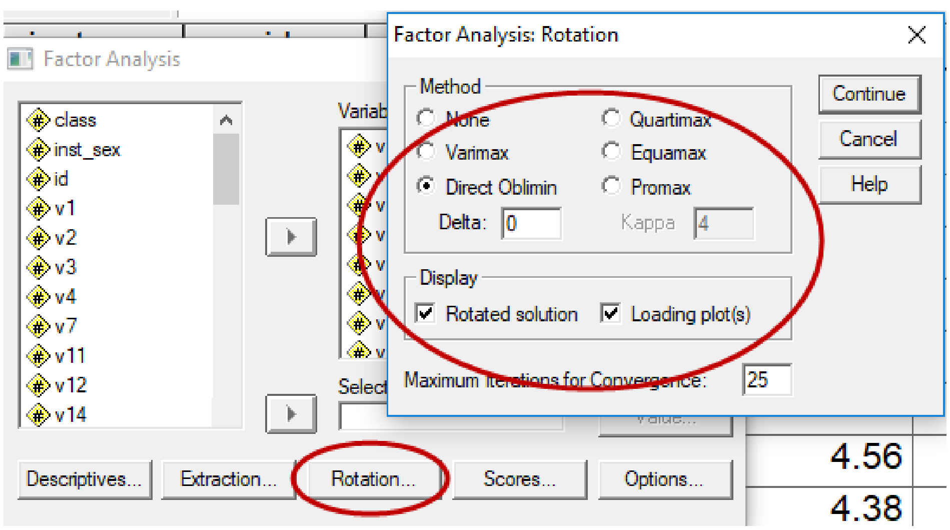


Rotation ->

Method = Direct Oblimin (0) – this is an oblique rotated solution

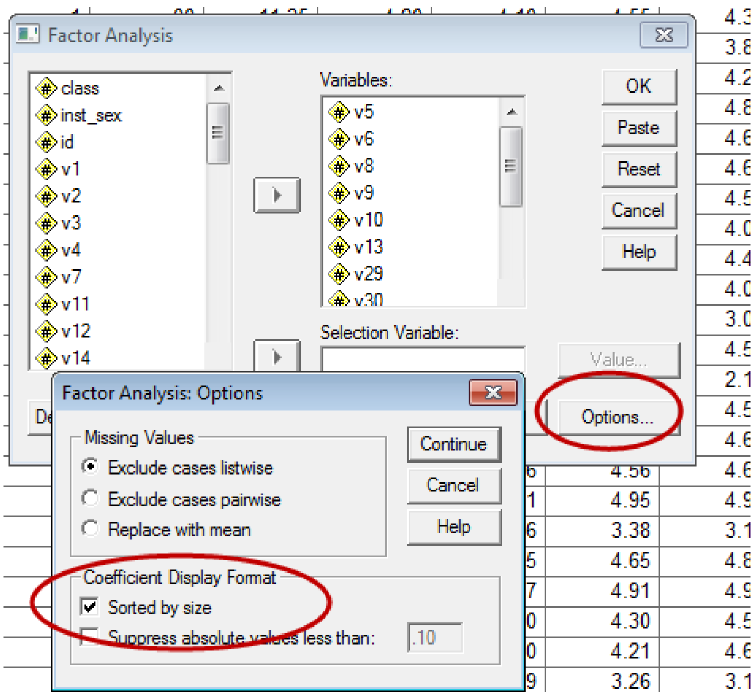
Display = Rotated Solution

Display = Loading Plots



Options ->

Coefficient Display Format = Sorted by size (this will sort loadings by size, easier to see results)



SPSS Factor Analysis Results

What do the results tell us? Do we have two factors? Do the items load on the factors as hoped? Do the results support the internal structure (hence construct validity) of these two scales?

Note – add determinant; it should not be 0.00 otherwise there will be difficulty with computations in EFA.

(a) KMO and Bartlett Tests

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.874
Bartlett's Test of Sphericity	Approx. Chi-Square	745.687
	df	55
	Sig.	.000

The KMO test assesses whether the pattern of correlations in the correlation matrix suggest natural groupings or whether groupings of items appear weak. Recall the correlation matrix at the outset of this presentation – the correlations there formed two groupings highlighted in green and blue. That correlation matrix would work well for factor analysis. KMO should be closer to 1.00. KMO value interpretation:

- below .5 don't attempt EFA,
- .5 and .6 are awful,
- .6 and .7 are acceptable but not good;
- .7 and .8 are good;
- .8 to .9 very good, and above
- .9+ super-duper good.

KMO should be >.6, but ideally .8 and above

Bartlett's test assesses whether the correlation is an identify matrix – this means the correlations between all items is 0.00 (no correlation). If there is no correlation, then EFA is not possible. We want Bartlett's test to be significant because rejecting the null means there are correlations among variables and EFA is suitable.

Bartlett should be significant at .05 level (i.e., Sig < .05)

Both tests suggest these data are appropriate for EFA.

The correlation matrix below illustrates what Bartlett's test assesses as the null hypothesis. If the correlation matrix looked like this, EFA would not be possible because there are no correlations among the variables.

Table 2: Identify Matrix – No correlations Among Items

	1	2	3	4	5	6
Item 1	1.00	.00	.00	.00	.00	.00
Item 2	.00	1.00	.00	.00	.00	.00
Item 3	.00	.00	1.00	.00	.00	.00
Item 4	.00	.00	.00	1.00	.00	.00
Item 5	.00	.00	.00	.00	1.00	.00
Item 6	.00	.00	.00	.00	.00	1.00

(b) Communalities (symbol h^2)

The table below shows communalities. The **Initial** column are communality estimates before factor extraction and therefore not of interest for interpretation purposes. These are determined by the R^2 obtained in regression where one variable is modeled by all others (e.g., V5 treated as a DV and all others – V6 through V30 – are the IVs predicting V5).

The **Extraction** column shows communities for each variable after extraction of the two factors that were retained (see below). These numbers can be interpreted as R^2 values in regression – the proportion of variance in each variable explained or predicted by the extracted factors. For example, the communality for V5 is .88 which means the two extracted factors predict about 88% of the variance in V5.

Our hope is that communalities after extraction are high for each variable. If the communality is low, this means the factors extracted are unable to predict variation in that variable, so it probably does not fit the measurement model examined, i.e., it does not help us measure any factors.

Technically h^2 is the sum of the squared factor loadings for the variable.

Communalities

	Initial	Extraction
v5	.926	.880
v6	.941	.945
v8	.896	.859
v9	.893	.799
v10	.861	.749
v13	.929	.913
v24	.766	.761
v25	.815	.795
v26	.845	.823
v29	.941	.931
v30	.954	.950

Extraction Method: Principal Axis Factoring.

(c) Variance Explained

The **Factor** column is the number of possible factors which always equals the number of variables included in the EFA. Not all 11 factors in this example will be retained.

Total Variance Explained

	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation
Factor	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total
1	7.415	67.405	67.405	7.289	66.264	66.264	7.017
2	2.306	20.960	88.365	2.117	19.243	85.507	3.655
3	.418	3.801	92.165				
4	.290	2.638	94.803				
5	.165	1.502	96.305				
6	.124	1.123	97.428				
7	.098	.891	98.319				
8	.066	.603	98.921				
9	.051	.459	99.381				
10	.039	.358	99.739				
11	.029	.261	100.000				

Extraction Method: Principal Axis Factoring.

a. When factors are correlated, sums of squared loadings cannot be added to obtain a total variance.

The **Eigenvalue** columns include **Total**, **% of Variance**, and **Cumulative %**. Since the correlation matrix was analyzed, each factor has a variance of 1.00, and since there are 11 factors, the **Total** column, if summed, would equal 11. The total column shows that two eigenvalues exceeded 1.00, Factor 1 = 7.415 and Factor 2 = 2.306. This means factor 1's variance was 7.415 and factor 2's variance was 2.306. Of the **total variance** possible, 11 in this case since there are 11 variables, Factor 1 accounted for 7.415 / 11 = 0.674 or 67.4% of the total variance in factors. Factor 2 accounted for 2.306 / 11 = .209 or 20.9% of the total variance.

Together, Factors 1 and 2 accounted for 85.507% of the **common variance** in factors after extraction of the two factors. Notice that in the columns labeled **Extraction of Sums of Squared Loadings** there are only two rows – this set of columns presents only information for the number of factors extracted.

The **Rotation Total** column is the total common variance for the retained factors.

(d) Determining Number of Factors to Extract

Identified above were several approaches to determining the number of factors to extract. Each will be considered below.

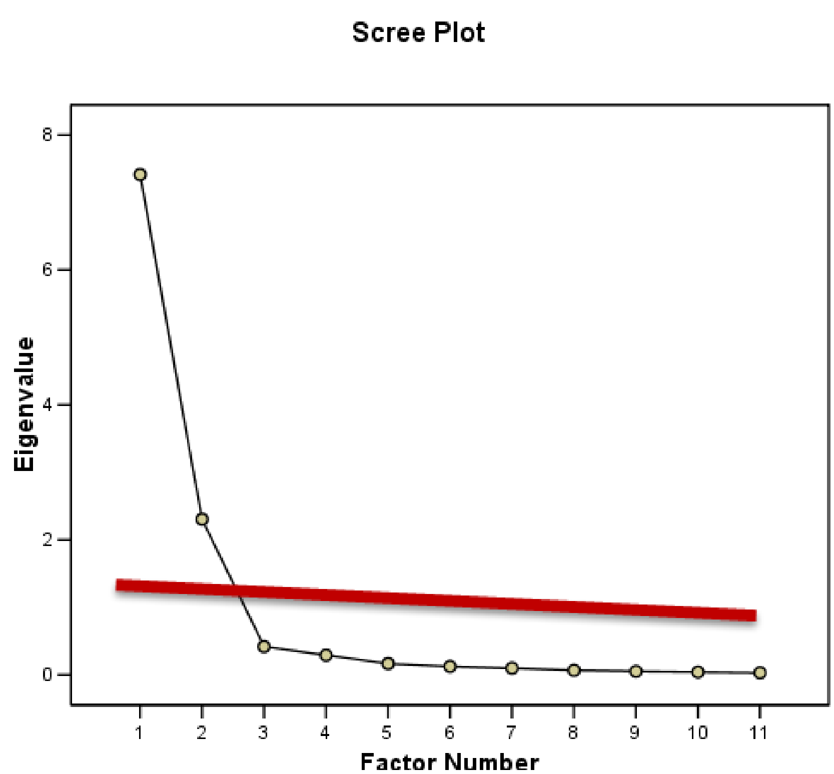
d1. Theoretical Model

Two scales were included with these data, Autonomy Support and Student Ratings, so there should be two factors identified.

d2. Scree Plot

This plot shows eigenvalues by number of factors. The idea is that clear factors will form a vertical line, and non-factors will form a horizontal line. Where these join forms an elbow or right bend, and that is the area used to make a cut between which factors to retain and which to drop.

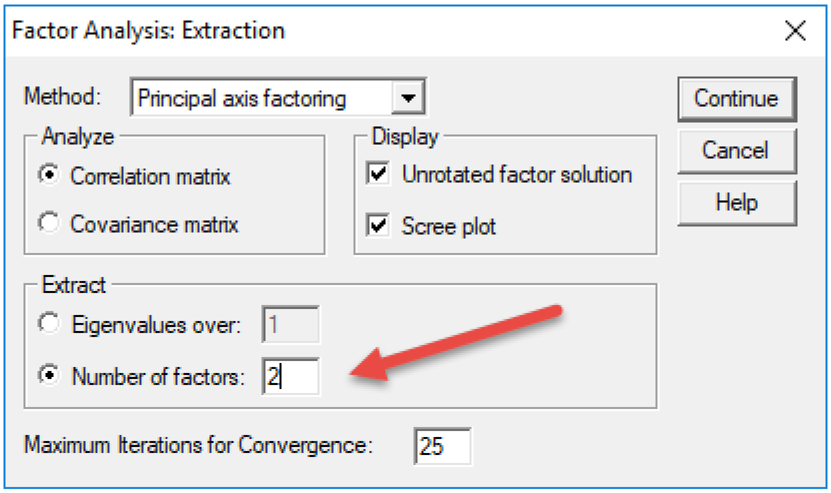
In the graph below a red line has been added separating the vertical and horizontal pattenrs – factors above the red line should be retained. In this case, two factors are identified. Origin of scree plot idea comes from rubble located at bottom of mountains.



d3. Eigenvalues Size

Another option is to retain factors with eigenvalues that exceed a pre-specified level, which is often defined as 1.00. Using this criterion, two factors had eigenvalues greater than 1.00, so two factors should be retained.

This is the default in SPSS and the reason two factors were extracted in this first run of EFA. If we believe the number of factors extracted by eigenvalue size is incorrect, we can easily specify the number of factors to retain in the the **Factor Analysis: Extracton** screen by placing that number in the box below indicated by the red arrow.



d4. Percentage Explained

Another approach factor count determination is to retain the number of factors that account for 70% or 80% of the total factor variance. In this example the two factors accounted for 88.365% of the factor variance, so this suggests two factors should be retained.

Total Variance Explained

Factor	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total
1	7.415	67.405	67.405	7.289	66.264	66.264	7.017
2	2.306	20.960	88.365	2.117	19.243	85.507	3.655
3	.418	3.801	92.165				
4	.290	2.638	94.803				
5	.165	1.502	96.305				
6	.124	1.123	97.428				
7	.098	.891	98.319				
8	.066	.603	98.921				
9	.051	.459	99.381				
10	.039	.358	99.739				
11	.029	.261	100.000				

Extraction Method: Principal Axis Factoring.

a. When factors are correlated, sums of squared loadings cannot be added to obtain a total variance.

d5. Parallel Analysis

Parallel analysis estimates the size of factor eigenvalues from a large number randomly generated correlation matrices. The logic: randomly generated correlations will lead to purely random eigenvalues, so compare the eigenvalues obtained from real data against those generated from random data; if the real eigenvalues are larger than their random counterparts, then those must be real factors; if eigenvalues from real data are less than eigenvalues from random data, then those must be random factors embedded in the real data.

Parallel analysis eigenvalues can be obtained from this site; results are shown below.

https://analytics.gonzaga.edu/parallelengine/

Parallel Analysis

Number of Variables in Your Dataset to be Factor Analyzed (Please change)

11

Sample Size of Your Dataset (Please change)

47

Type of Analysis

Factors

Number of Random Correlation Matrices to Generate (default of 100 currently set)

1000

Percentile of Eigenvalues (default of 95th percentile currently set)

95

Seed

1000

About this Application

Patil et al. (2008) presented a

Using this Application

Based on parameters provided by the researcher, this engine calculates eigenvalues from randomly generated correlation matrices. These can be then compared with eigenvalues extracted from the researcher's dataset. The number of factors to retain will be the number of eigenvalues (generated from the researcher's dataset) that are larger than the corresponding random eigenvalues (Horn 1965).

The default (and recommended) values for number of random correlation matrices and percentile of eigenvalues are 100 and 95 respectively (see Cota et al. 1993; Glorfeld 1995; Turner 1998; Velicer et al. 2000). Based on the nature of their particular dataset, researchers, can override these default options. Higher (lower) values of number of correlation matrices generated increase (decrease) computation time but provide more (fewer) data points in the distribution of different eigenvalues. The percentile determines the desired eigenvalue from this distribution, which is then used for comparison purposes. Lower values of the percentile tend to lead to over extraction (extraction of more factors than necessary).

Component or Factor	Mean Eigenvalue	Real Data Eigenvalues	Percentile Eigenvalue (Random Eigenvalues)
1	1.150419	7.415	1.496861
2	0.826487	2.306	1.094097
3	0.604206	0.418	0.803992
4	0.417641	0.290	0.586459
5	0.257536		0.396175
6	0.122790		0.242392
7	0.003827		0.104622
8	-0.106547		-0.025044
9	-0.202628		-0.135487
10	-0.292470		-0.234641
11	-0.385586		-0.325141

Note – discuss screenshot above in class.

The parallel analysis shows that 2 factors should be retained.

All approaches considered above – theorectical mode, scree plot, eigenvalue size, percent explained, and parallel analysis – suggest that 2 factors should be retained.

(e) Initial Factor Loadings

The **Factor Matrix** (pattern matrix) table contains the unrotated factor loadings which are the correlations between variables and factors. This table can provide some ideas about which variables load on which factors. Highlighted in red are the highest loadings for each factor. It seems Factor 1 is composed of the Student Ratings items, and Factor 2 the three Autonomy Support items.

These loadings are easy to read and show clearly that Factor 1 represent student ratings and Factor 2 represents autonomy support. In cases like this, factor rotation to simplify interpretation is not needed because the factors are easy to read. Sometimes loadings are not so easy to read, and factor rotation can help clarify the picture.

Factor Matrix[a]

	Factor	
	1	2
v30	.962	-.157
v13	.953	-.077
v29	.928	-.265
v8	.920	.115
v6	.914	-.331
v5	.893	-.287
v10	.853	.149
v9	.820	-.355
v26	.488	.765
v24	.454	.745
v25	.522	.723

Extraction Method: Principal Axis Factoring.

a. 2 factors extracted. 5 iterations required.

The communality for each variable can be found by squaring and adding the loadings reported in the **Factor Matrix** table (note – I think this does not work for rotated tables, check this). For example, V30 loadings squared and summed are $.962^2 + -.157^2 = .95$, the same value reported above in the communalities table.

Note: The unrotated **Factor Matrix** is both factor coefficient and factor correlation (i.e., both pattern and structure). For orthogonal rotations, both pattern and structure matrices, described below, are the same.

(e) Factor Rotation and Interpretation

Recall that an oblique rotation was requested (oblimin) which allows factors to correlate. Rotated factor loadings are shown below in both the Pattern and Structure Matrices.

Pattern matrix – Coefficients for linear combinations of variables (factor coefficients used to reproduce variable scores, predict variable scores, like regression coefficients). Most use this matrix for interpretation of factors; usually both pattern and structure provide similar interpretations for factors.

Structure matrix – Correlations between factors and variables after oblique rotation

Rotation – Redistribute variable loadings on each factor in such a way to help produce a simple structure to make interpretation easier. Common rotation options are briefly described below.

Orthogonal Rotations (Factors Uncorrelated)

- Varimax: Minimizes the number of variables that have high loadings on each factor identified.
- Quartimax: Minimizes number of factors with high loadings with variables.
- Equamax: Combination of both varimax and quartimax.
- I do not recommend use of any Orthogonal rotation methods since these may artificially discard useful information about how factors are related.

Oblique Rotations (Factors Correlated)

- Oblimin (or Direct Oblimin): The degree of correlation between factors is controlled by delta. In SPSS the default is 0. Negative values of delta result in weaker factor correlations and positive values result in stronger factor correlations. It is not clear of the potential range for delta, but 0 appears to be a mid-range value in terms of producing factor correlations. The upper value for delta is 0.80; I am uncertain about the lowest value of delta. Values below -4 tend to produce factors that are nearly uncorrelated.
- Promax: Another orthogonal rotation method – often presented as quicker than Oblimin, but that is not a concern for those who use computers to rotate factors.
- I recommend using Oblique rotation methods, and do not have a recommend for which, oblimin or promax, to use. If using oblimin, unless there is reason to change delta, leave it at 0.00.

Pattern Matrix[a]

	Factor	
	1	2
v6	1.006	-.120
v29	.980	-.048
v5	.962	-.079
v30	.948	.072
v9	.937	-.167
v13	.893	.152
v8	.753	.343
v10	.673	.361
v26	-.008	.910
v24	-.027	.881
v25	.045	.875

Extraction Method: Principal Axis Factoring.
Rotation Method: Oblimin with Kaiser Normalization.

a. Rotation converged in 4 iterations.

Structure Matrix

	Factor	
	1	2
v30	.973	.393
v6	.965	.221
v29	.964	.284
v13	.945	.455
v5	.935	.247
v9	.880	.150
v8	.869	.598
v10	.796	.590
v26	.300	.907
v25	.342	.891
v24	.271	.872

Extraction Method: Principal Axis Factoring.
Rotation Method: Oblimin with Kaiser Normalization.

The **Pattern Matrix** presents the **pattern loadings**, or coefficients, linking each factor with each variable. The pattern matrix is often used to interpret factors. Results shown in the pattern matrix demonstrate what is known as **simple structure** – high loadings on one factor and low loadings on the other factor. Item V29, for example, has a high loading on Factor 1 (.980), but a low loading on Factor 2 (-.048), so this indicates V29 is aligned closely with Factor 1 but not with Factor 2.

Interpretation of the pattern matrix is the process of identifying which variables, or scale items, load well and poorly on a factor. Those items with high loadings must be considered when naming a factor. Factor 1 in the current pattern is dominated by the Student Ratings variables, shown below.

Factor 1 = Student Ratings

5. The instructor presented the material in a clear and understandable manner.
6. Course materials were well prepared and organized.
8. The instructor made students feel welcome in seeking help/advice in or outside of class.
9. The content of this course is useful, worthwhile, or relevant to you.
10. Methods of evaluating student work were fair and appropriate.
13. The instructor gave students useful/helpful feedback on work.
29. Overall, how would you rate this course?
30. Overall, how would you rate this instructor?

Factor 2 has a very simple structure – Autonomy Support items, see below, load very well on Factor 2, and Student Ratings items show almost no loading on Factor 2.

Factor 2 = Autonomy Support

24. The instructor was willing to negotiate course requirements with students.
25. Students had some choice in course requirements or activities that would affect their grade.
26. The instructor made changes to course requirements or activities as a result of student comments or concerns.

The **Structure Matrix** shows the **structure loadings**, or the correlation between each variable and factor. This table can also be used to interpret factors, and the interpretation results are the same as shown in the Pattern Matrix – items V26, V25, and V24 load best on Factor 2, so Factor 2 is the Autonomy Support factor.

Factor Correlations

Factor Correlation Matrix

Factor	1	2
1	1.000	.339
2	.339	1.000

Extraction Method: Principal Axis Factoring.
Rotation Method: Oblimin with Kaiser Normalization.

The **Factor Correlation Matrix** table shows the correlation between Factor 1 (which I named Student Ratings) and Factor 2 (which I named Autonomy Support). It appears that autonomy support correlates .339 with Student Ratings. This correlation was obtained with the **oblimin delta = 0.00**.

Below are estimates of factor correlations between student ratings and autonomy support for different values of **oblimin delta**.

Oblimin Delta	Correlation between Factors 1 and 2
.8	.765
.5	.550
.2	.392
0	.339
-.2	.306

Oblimin Delta	Correlation between Factors 1 and 2
-.5	.274
-.8	.252
-1	.240
-2	.195
-4	.138

With delta = -500, the correlation was .005. It appears that as delta approaches -∞ the factor correlation approaches 0.00.

As a comparison, I computed mean composite scores for both autonomy and student ratings. The correlation obtained from the composite scores was .367 which is close to the value or .339 provided when delta = 0.00.

Factor Plots

The plot below shows how the items cluster in space for Factors 1 and 2. This clustering demonstrates clear separation thereby confirming the two-factor solution.

Figure x: Factor Plot with Unrotated Factors

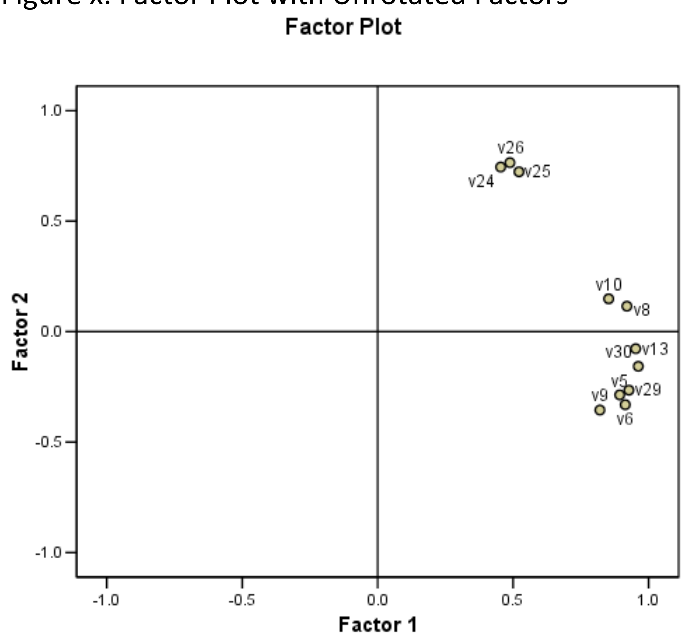
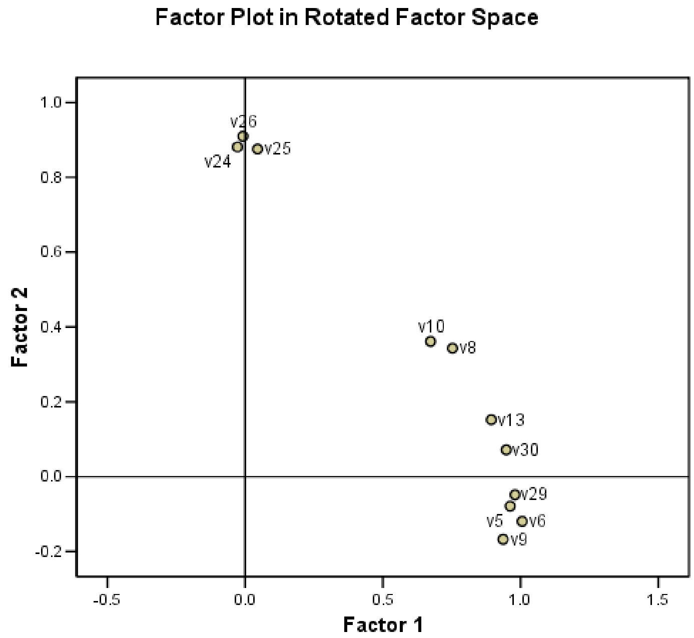


Figure x: Factor Plot with Rotated Factors (Oblimin, delta = 0)



Note - Factor Scores, to be added; discuss possible use, see Kootsra Exploratory Factor Analysis p. 8 for uses.

Note – add goodness of fit expanded discussion

Briefly, goodness of fit of these two factor EFA?

- KMO = .874 (very good)
- Percent variable predicted = 85.5%
- Communalities = range from low of .749 to high of .95
- Factor pattern = clear factors
- Reproduced correlation matrix = note, add discussion of deviance, how many residual >.05?

Update Example 5 – Not good example, sample size too small to assess three factors. Collect more data.

~~5. Example 2: Employment Thoughts Data~~

~~Some of you completed questionnaire twice for this course. The items were selected from Menon (2001) and were designed to measure three employment related constructs. Responses to each item scaled from Strongly Disagree (1) to Strongly Agree (6).~~

~~Perceived Control~~

~~Q1: I can influence the way work is done in my department~~
~~Q2: I can influence decisions taken in my department~~
~~Q3: I have the authority to make decisions at work~~

~~Goal Internalization~~

~~Q4: I am inspired by what we are trying to achieve as an organization~~
~~Q5: I am inspired by the goals of the organization~~
~~Q6: I am enthusiastic about working toward the organization's objectives~~

~~Perceived Competence~~

~~Q7: I have the capabilities required to do my job well~~
~~Q8: I have the skills and abilities to do my job well~~
~~Q9: I have the competence to work effectively~~

~~SPSS data file link (can be found in Reliability section on course web page):~~

~~http://www.bwgriffin.com/gsu/courses/edur9131/2018spr-content/06-reliability/06-EDUR9131-EmploymentThoughts-Merged.sav~~

5. Example 2: Academic Control

Scores come from the cyber-harassment dataset used in the previous two meetings. Sample size is 500+ and data collected from undergraduate students at Georgia Southern. One scale they completed was Academic Control. Items are shown below.

SPSS data file link:
http://www.bwgriffin.com/gsu/courses/edur9131/stat-data/cyberharass2.sav

Perform a factor analysis to determine whether these eight items appear to form one factor as designed.

Academic Control

40 to 47 = Academic Control alpha = .85	Strongly Disagree	Disagree	Mix of Disagree and Agree	Agree	Strongly Agree
40. My grades are basically determined by things beyond my control and there is little I can do to change.	1	2	3	4	5
41. I see myself as largely responsible for my performance throughout my college career.	1	2	3	4	5
42. No matter what I do, I can't seem to do well in my courses.	1	2	3	4	5
43. There is little I can do about my performance in college/university.	1	2	3	4	5
44. The more effort I put into my courses, the better I do in them.	1	2	3	4	5
45. How well I do in my courses is often the "luck of the draw."	1	2	3	4	5
46. I have a great deal of control over my academic performance in my courses.	1	2	3	4	5
47. When I do poorly in a course, it's usually because I haven't given it my best effort.	1	2	3	4	5

6. Example 3: Toxic Disinhibition

Same data as above. Items for this scale are shown below. Udris (2014) originally developed items 16 to 19, then revised the scale (Udris, 2017) with only items 20 to 22.

Toxic Disinhibition

16 17 18 19 = Toxic Disinhibition alpha = .68	Strongly Disagree	Disagree	Mix of Disagree and Agree	Agree	Strongly Agree
16. I don't mind writing insulting things about others online, because it's anonymous.	1	2	3	4	5
17. It is easy to write insulting things online because there are no repercussions.	1	2 16 to 19 Udris 2014 original items.	3	4	5
18. There are no rules online therefore you can do whatever you want.	1	2	3	4	5
19. Writing insulting things online is not bullying.	1	2	3	4	5
20 21 22 = Toxic Disinhibition alpha = .92					
20. On the Internet it is easier to annoy or disturb someone I don't like.	1	2 20 to 22 Revised scale Udris 2016	3	4	5
21. On the Internet it is easier to blame or criticize someone without fear of revenge or repercussions.	1	2	3	4	5
22. On the Internet it is easier to ridicule or make fun of someone.	1	2	3	4	5
16 through 22 = Toxic Disinhibiton alpha = .85					

Do these seven items form one, or more than one, factor?

7. Example 4: Doctoral Student Efficacy and Anxiety toward the Dissertation Proceess

Efficacy and anxiety toward the dissertation process. Odd items measure efficacy and even items measure anxiety.

Data 2: http://www.bwgriffin.com/gsu/courses/edur9131/temp/alphadata.sav
Alpha Data
Questions: http://www.bwgriffin.com/gsu/courses/edur9131/activities/Assignment_6_internal_consistency_data.pdf

8. Example 5: Parenting Stress and Coping in Difficult Parenting Situations

Szymańska A, Dobrenko KA. (2017) The ways parents cope with stress in difficult parenting situations: the structural equation modeling approach. PeerJ 5:e3384

https://peerj.com/articles/3384/

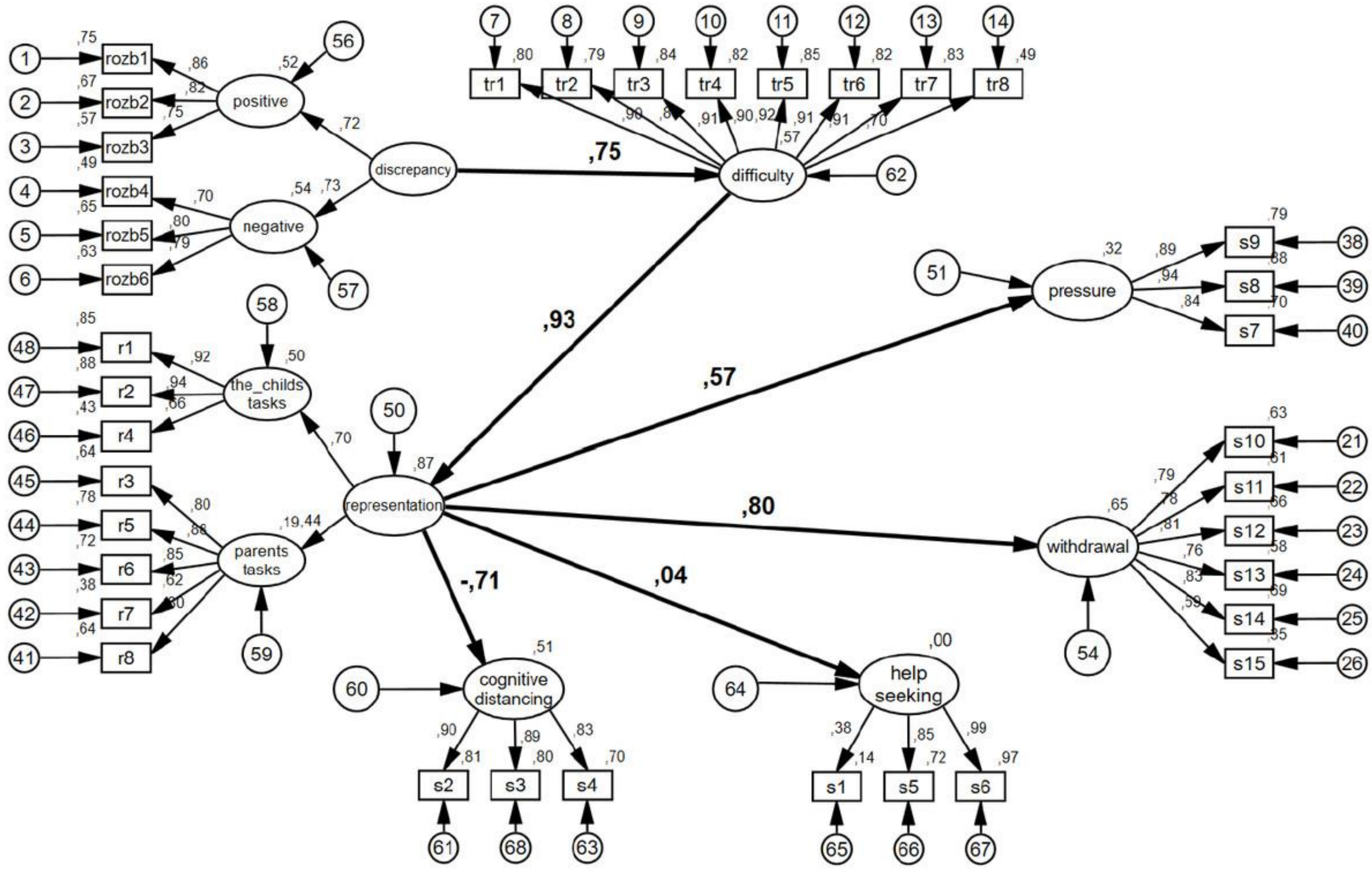
Szymańska and Dobrenko (2017) present the following figure showing relations among a number of constructs. They also present their data in an SPSS file, which is linked below.

https://dfzljdn9uc3pi.cloudfront.net/2017/3384/1/base_for_review_stress.sav

I have also saved these data to the course web site, linked below.

http://www.bwgriffin.com/gsu/courses/edur9131/2018spr-content/12-factor-analysis/12-2017-Szymanska-data.sav

While EFA cannot be used to assess the structural relations among constructs in the model below, it can be used to assess whether the measurement model – the factors and their loadings – are similar to those shown in the figure.

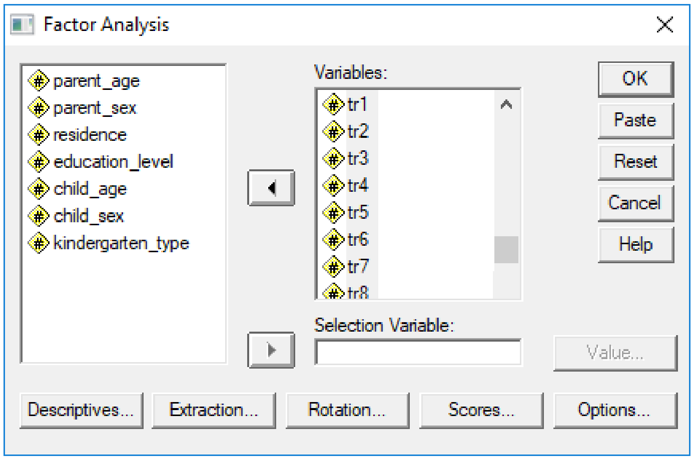


The variables used to measure each construct are identified above.

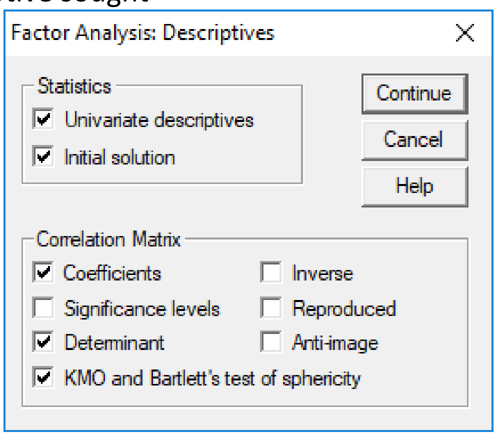
Discrepancy = rozb1 to rozb6
Representation = r1 to r8
Cognitive Distancing = S2 s3 s4
Help Seeking = S1 s5 s6
Difficulty = tr1 to tr8
Pressure = s7 s8 s9
Withdrawal = s10 to s15

EFA with all variables entered. According to their model, there should be 7 overall factors, or possible 9 if Representation and Discrepancy both divide into 2 sub-factors as shown in the figure.

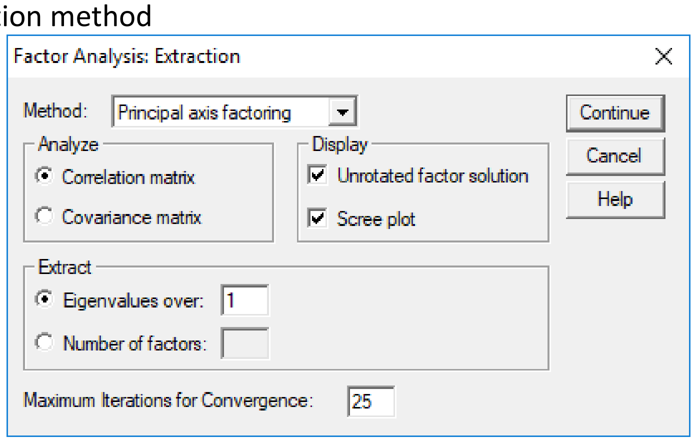
Data entered in EFA



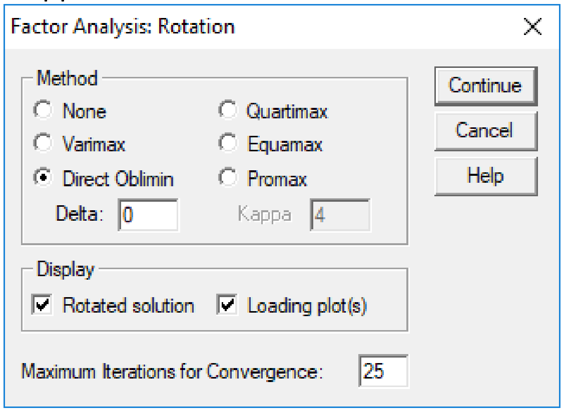
Descriptive sought



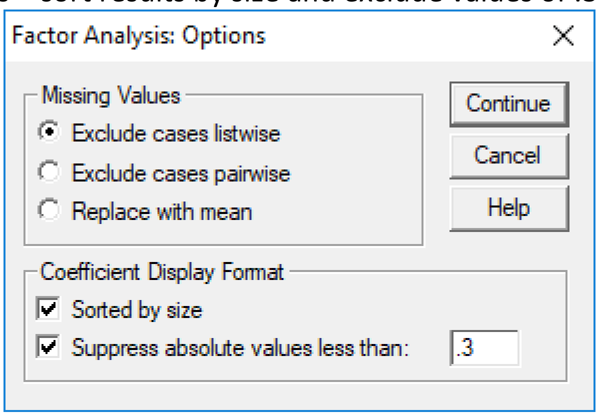
Extraction method



Rotation approach



Options – sort results by size and exclude values of .30 or less in absolute value



SPSS Results

Descriptive show n = 258

Descriptive Statistics

	Mean	Std. Deviation	Analysis N
rozb1	20.9109	37.44469	258
rozb2	19.1667	34.70015	258
rozb3	19.3062	32.52086	258
rozb4	50.5426	49.83022	258

KMO and Bartlett – both are very good

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.922
Bartlett's Test of Sphericity	Approx. Chi-Square	8297.527
	df	666
	Sig.	.000

Variance Explained – total of 37 variables entered, 8 factors with eigenvalues greater than 8, so SPSS extracts 8 factors by default

Total Variance Explained

Factor	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total
1	14.573	39.385	39.385	14.305	38.662	38.662	10.743
2	3.119	8.430	47.816	2.809	7.593	46.255	5.199
3	2.327	6.289	54.105	2.016	5.449	51.704	1.980
4	2.177	5.882	59.987	1.835	4.959	56.663	8.791
5	1.784	4.821	64.808	1.534	4.147	60.810	5.972
6	1.476	3.988	68.796	1.203	3.252	64.063	4.926
7	1.310	3.540	72.336	1.019	2.755	66.818	6.543
8	1.291	3.490	75.826	.907	2.452	69.270	5.355
9	.958	2.590	78.416				
10	.711	1.922	80.339				
11	.636	1.719	82.057				
12	.555	1.501	83.558				
13	.501	1.353	84.911				
14	.483	1.304	86.216				
15	.452	1.221	87.437				
16	.426	1.150	88.587				
17	.402	1.088	89.675				
18	.366	.990	90.665				
19	.331	.895	91.560				
20	.319	.861	92.421				
21	.297	.802	93.223				
22	.267	.723	93.946				
23	.250	.675	94.621				
24	.232	.626	95.247				
25	.212	.572	95.819				
26	.207	.558	96.377				
27	.192	.519	96.896				
28	.173	.468	97.365				
29	.155	.419	97.783				
30	.147	.398	98.181				
31	.135	.364	98.545				
32	.109	.295	98.840				
33	.106	.287	99.126				
34	.095	.256	99.382				
35	.089	.241	99.624				
36	.084	.226	99.850				
37	.056	.150	100.000				

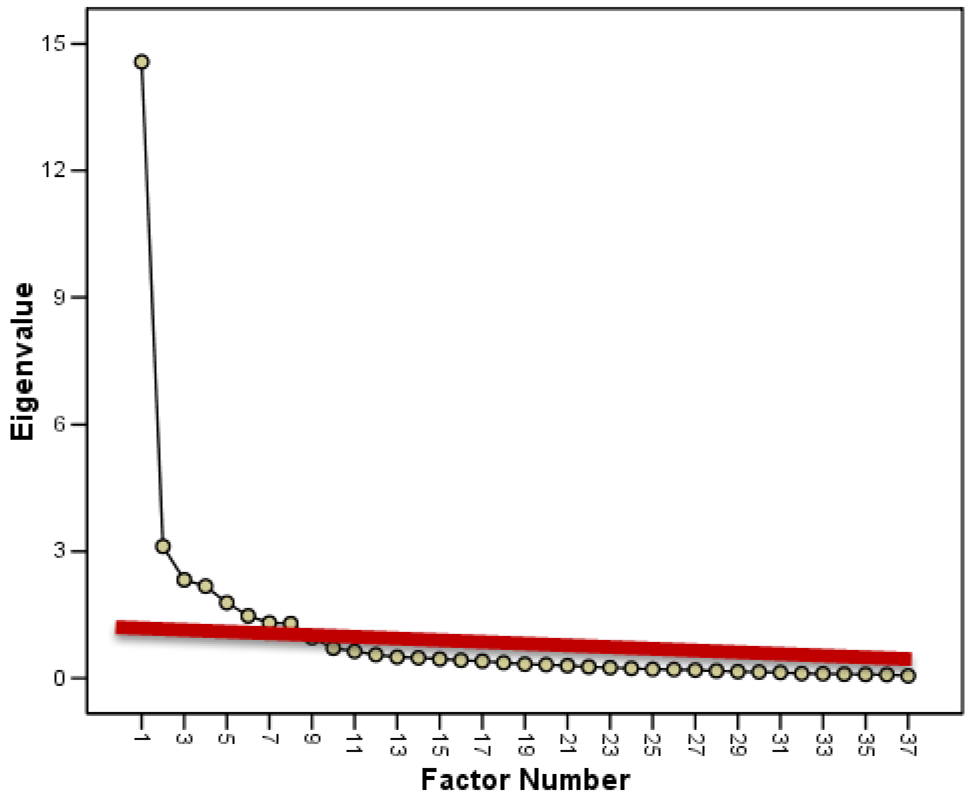
How many factors to retain?

8-1. Theoretical Model

As noted, there are 7 overall factors, and maybe 9 if two factor sub-divide into two factors each.

8-2. Scree Plot

The scree plot is not clear, but maybe 8 according to the line I drew below.



8-3. Eigenvalues Size

There are 8 factors with eigenvalues larger than 1.

8-4. Percentage Explained

Eight factors explain 75.8% of variance, which hits the mark of 70% to 80% variance expalined.

Total Variance Explained

	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation
Factor	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total
1	14.573	39.385	39.385	14.305	38.662	38.662	10.743
2	3.119	8.430	47.816	2.809	7.593	46.255	5.199
3	2.327	6.289	54.105	2.016	5.449	51.704	1.980
4	2.177	5.882	59.987	1.835	4.959	56.663	8.791
5	1.784	4.821	64.808	1.534	4.147	60.810	5.972
6	1.476	3.988	68.796	1.203	3.252	64.063	4.926
7	1.310	3.540	72.336	1.019	2.755	66.818	6.543
8	1.291	3.490	75.826	.907	2.452	69.270	5.355
9	.958	2.590	78.416				

8-5. Parallel Analysis

https://analytics.gonzaga.edu/parallelengine/

Parallel Analysis

Number of Variables in Your Dataset to be Factor Analyzed (Please change)

37

Sample Size of Your Dataset (Please change)

258

Type of Analysis

Factors

Number of Random Correlation Matrices to Generate (default of 100 currently set)

1000

Percentile of Eigenvalues (default of 95th percentile currently set)

95

Seed

1000

About this Application

Patil et al. (2008) presented a web-based parallel analysis engine (Patil et al. 2007) that used SAS. This engine was published at

http://ires.ku.edu/~smishra/parallelengine.htm

Since that application is facing few

Using this Application

Based on parameters provided by the researcher, this engine calculates eigenvalues from randomly generated correlation matrices. These can be then compared with eigenvalues extracted from the researcher's dataset. The number of factors to retain will be the number of eigenvalues (generated from the researcher's dataset) that are larger than the corresponding random eigenvalues (Horn 1965).

The default (and recommended) values for number of random correlation matrices and percentile of eigenvalues are 100 and 95 respectively (see Cota et al. 1993; Glorfeld 1995; Turner 1998; Velicer et al. 2000). Based on the nature of their particular dataset, researchers, can override these default options. Higher (lower) values of number of correlation matrices generated increase (decrease) computation time but provide more (fewer) data points in the distribution of different eigenvalues. The percentile determines the desired eigenvalue from this distribution, which is then used for comparison purposes. Lower values of the percentile tend to lead to over extraction (extraction of more factors than necessary).

Component or Factor	Mean Eigenvalue	Obtained Eigenvalues	Percentile Eigenvalue
1	0.960062	14.57	1.069689
2	0.852197	3.119	0.937988
3	0.773731	2.37	0.841950
4	0.708622	2.17	0.773223
5	0.649628	1.78	0.708783
6	0.595639	1.47	0.650359
7	0.544260	1.31	0.597603
8	0.497105	1.29	0.546581
9	0.451822	.96	0.497962
10	0.409867	.71	0.455636
11	0.368465	.63	0.411374
12	0.329715	.55	0.371410
13	0.291491	.50	0.331271
14	0.255572		0.294080

The parallel analysis shows that more than 13 factors should be retained, so this does not appear to be a useful assesment.

Overall most approaches to assessing factor extraction seem to suggest 8 factors, so we will proceed with 8 factors.

Pattern Matrix – overall the results are very good (see below). In most cases each factor has loadings that are unique to that factor (simple structure) except for Difficulty which is correlated to Representation (child's task). Given the number of items (n = 37) and the number of constructs to measure (7 or 9), this EFA did well recreating the factor structure.

Pattern Matrix[a]

	Factor							
	1	2	3	4	5	6	7	8
tr1	.741							
tr2	.705							
tr3	.688							
tr4	.666							
r2	.660							
r1	.641							
tr5	.600							
r4	.576							
tr7	.553						-.330	
tr6	.547							
r5		.826						
r6		.806						
r8		.760						
r3		.710						
r7		.596						
s6			.923					
s5			.844					
s1			.416					
s12				-.875				
s13				-.806				
s11				-.745				
s10				-.580				
s14				-.541				
s15				-.525				
s8					.912			
s7					.849			
s9					.814			
rozb2						.857		
rozb1						.786		
rozb3						.577		
s3							.840	
s2							.801	
s4							.594	
tr8							-.307	
rozb5								.751
rozb6								.748
rozb4								.650

Difficulty + r1, r2, r4 = Child's task of Representtion

Parent's task of Representation

Help Seeking

Withdrawal

Pressure

Positive Discrepancy

Cognitive Distancing

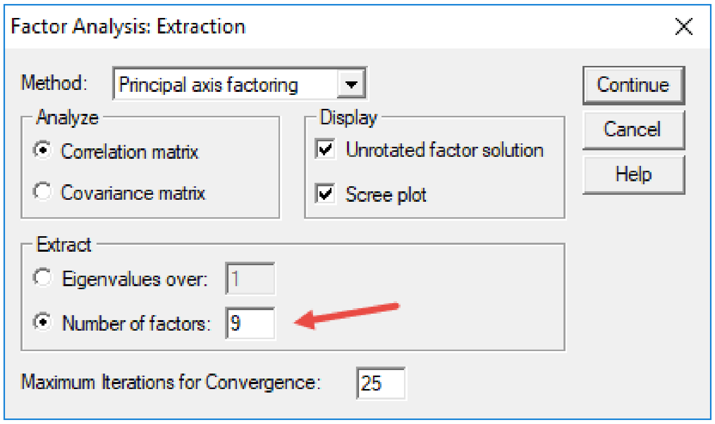
Negative Discrepancy

Extraction Method: Principal Axis Factoring.
Rotation Method: Oblimin with Kaiser Normalization.

a. Rotation converged in 9 iterations.

Nine-factor Option

Out of curiosity, I re-ran the EFA but specified extraction of 9 factors.



Results are shown below. The EFA almost perfectly reproduced the factor structure expected for the questionnaire – this is a strong indication that the 9-factor extraction is the appropriate solution. Overall their measures of these 9 constructs worked very well to independently assess these 9 constructs. These are excellent results.

Pattern Structure matrix on next page.

Pattern Matrix[a]

	Factor								
	1	2	3	4	5	6	7	8	9
tr3	.854								
tr1	.809								
tr2	.807								
tr4	.718								
tr7	.620								
tr5	.618								
tr6	.601								
tr8	.317								
r6		.827							
r5		.822							
r8		.796							
r3		.692							
r7		.589							
s6			.980						
s5			.894						
s1			.301				.300		
s12				-.862					
s13				-.782					
s11				-.727					
s10				-.577					
s14				-.522					
s15				-.502					
s8					.912				
s7					.858				
s9					.827				
rozb2						.877			
rozb1						.787			
rozb3						.578			
s3							.858		
s2							.806		
s4							.608		
rozb6								.766	
rozb5								.724	
rozb4								.608	
r1									-.809
r2									-.802
r4									-.510

Extraction Method: Principal Axis Factoring.
Rotation Method: Oblimin with Kaiser Normalization.

a. Rotation converged in 9 iterations.

Difficulty

Parent's Tasks of Representation

Help Seeking

Withdrawal

Pressue

Positive Discrepancy

Cognitive Distancing

Negative Discrepancy

Child's Tasks of Representation

9. Reading Factor Analysis Tables

How to select best items using EFA results.

http://www.bwgriffin.com/gsu/courses/edur9131/2018spr-content/12-factor-analysis/12-2004-Tschannen-principal-efficacy.pdf

PSES	Factor 1	Factor 2	Factor 3
Efficacy for management			
Handle the time demands of the job	0.82	0.11	0.11
Handle the paperwork required of the job	0.73	0.14	0.19
Maintain control of your own daily schedule	0.70	0.20	0.22
Prioritize among competing demands of the job	0.63	0.27	0.26
Cope with the stress of the job	0.57	0.21	0.19
Shape the operational policies and procedures that are necessary to manage your school	0.53	0.15	0.30
Efficacy for instructional leadership			
Motivate teachers	0.15	0.81	0.20
Generate enthusiasm for a shared vision for the school	0.15	0.79	0.18
Manage change in your school	0.25	0.67	0.19
Create a positive learning environment in your school	0.17	0.64	0.29
Facilitate student learning in your school	0.22	0.62	0.21
Raise student achievement on standardized tests	0.17	0.45	0.32
Efficacy for moral leadership			
Promote acceptable behavior among students	0.20	0.26	0.78
Promote school spirit among a large majority of the student population	0.18	0.24	0.71
Handle effectively the discipline of students in your school	0.21	0.17	0.59
Promote a positive image of your school with the media	0.21	0.25	0.56
Promote the prevailing values of the community in your school	0.36	0.22	0.51
Promote ethical behavior among school personnel	0.38	0.29	0.43

Notes: $N = 544$; Factor 1: Eigenvalue = 7.4; Cumulative percent of variance explained = 41.12; Factor 2: Eigenvalue = 1.9; Cumulative percent of variance explained = 51.84; and Factor 3: Eigenvalue = 1.4; Cumulative percent of variance explained = 59.64

Table I. Factor loadings for the PSES (Study 3) http://www.bwgriffin.com/gsu/courses/edur9131/2018spr-content/12-factor-analysis/12-2008-Thomas-science-selfefficacy.pdf

Table 4. Factor loadings* of items in the refined version of the SEMLI-S

Original item No (Code in SEMLI-S)	Constructivist Connectivity (CC)	Monitoring, Evaluation & Planning (MEP)	Science Learning Self-efficacy (SE)	Learning Risks Awareness (AW)	Control of Concentration (CO)
5 (CC1)	0.69				
11 (CC2)	0.80				
12 (CC3)	0.77				
22 (CC4)	0.68				
26 (CC5)	0.73				
35 (CC6)	0.53				
39 (CC7)	0.74				
8 (MEP4)		0.67			
21 (MEP1)		0.65			
23 (MEP5)		0.61			
24 (MEP2)		0.76			
32 (MEP3)		0.68			
50 (MEP6)		0.56			
55 (MEP7)		0.60			
57 (MEP8)		0.47			
64 (MEP9)		0.51			
15 (SE1)			0.63		
27 (SE2)			0.65		
33 (SE3)			0.71		
53 (SE5)			0.75		
62 (SE6)			0.63		
69 (SE4)			0.67		
16 (AW1)				0.71	
20 (AW2)				0.73	
36 (AW3)				0.71	
65 (AW4)				0.63	
71 (AW5)				0.60	
41 (CO1)					0.68
61 (CO2)					0.76
68 (CO3)					0.74

*. All loadings smaller than 0.4 have been omitted

Complete wording of items presented in the appendix.

8. Sample Size for EFA

See Costello and Osborne (2005) for discussion of sample size.

to be added

Appendix

Information to Include in Reporting Factor Analyses

Confirmatory Factor Analysis

- Initial proposed model(s)
  - Number and composition of factors
  - Orthogonal versus correlated factors
  - Secondary loadings, correlated error terms
  - Other model constraints (fixed and free parameters)
- Method of estimation
- Goodness of fit
  - Overall fit
  - Relative fit
  - Parsimony
- Any model modification to improve model fit to data
- Factor loadings (lambda) and standard errors
- Communality (or squared correlations of observed variables with the factors)
- Factor correlations and standard errors (significance)

Exploratory Factor Analysis

- Principal component analysis of common factor analysis
- Initial communality estimates (common factor analysis)
- Method of factor extraction
- Criteria for retaining factors
- Eigenvalues, percentage of variance accounted for by the unrotated factors
- Rotation method (and rationale)
- All rotated factor loadings
- Factor intercorrelations (oblique solutions)
- Variance explained by factors after rotation

Received March 1, 1995
Revision received April 19, 1995
Accepted April 21, 1995 ■

References

Coltman, T., Devinney, T.M., Midgley, D.F., & Venaik, S. (2008). Formative versus reflective measurement models: Two applications of formative measurement. Journal of Business Research, 61, 1250-1262,

Costello, Anna B. & Jason Osborne (2005). Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis. Practical Assessment Research & Evaluation, 10(7). Available online: http://pareonline.net/getvn.asp?v=10&n=7

Floyd, F. & Widaman, K.F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. Psychological Assessment, 3, 286-299.

Karen, (2018). "The Fundamental Difference Between Principal Component Analysis and Factor Analysis." Retried 13 April 2018 from: https://www.theanalysisfactor.com/the-fundamental-difference-between-principal-component-analysis-and-factor-analysis/

Udris, R. (2014). Cyberbullying among high school students in Japan: Development and validation of the Online Disinhibition Scale. *Computers in Human Behavior*, *41*, 253-261.

Udris, R. (2017). Psychological and social factors as predictors of online and offline deviant behavior among Japanese adolescents. *Deviant behavior*, *38*(7), 792-809.