What Is Factor Analysis? A *Simple* Explanation…

Factor analysis is a statistical procedure used to identify a small number of factors that can be used to represent relationships among sets of interrelated variables. For example, COMPUTER USE BY TEACHERS is a broad construct that can have a number of FACTORS (use for testing, use for research, use for presentation development, etc.).

BASIC ASSUMPTION: underlying dimensions – or *factors* – can be used to explain complex events or trends.

GOAL: to identify otherwise not-directly-observable factors on the basis of a set of observable variables.

FOUR STEPS:
1. Compute a correlation matrix for all variables.
2. Determine the number of factors necessary to represent the data and the method of calculating them (*factor extraction*):.
3. Transform the factors to make them interpretable (*rotation*)
4. Compute scores for each factor.

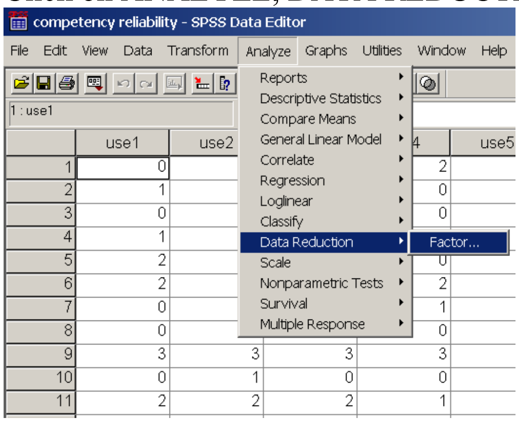
BARTLETT'S TEST OF SPHERICITY is used to test the hypothesis that the correlation matrix is an identity matrix (*all diagonal terms are one and all off-diagonal terms are zero*). You are looking for SIGNIFICANCE (*less than .05*) because you WANT the variables to be correlated. In other words, picture a correlation matrix: all items are perfectly correlated with themselves (one), and have *some level* of correlation with the other items. If they *are not correlated with the other items* then they can't be part of the same factor!

KAISER-MEYER-OLKIN: measure of sampling adequacy is used to compare the magnitudes of the observed correlation coefficients in relation to the magnitudes of the *partial* correlation coefficients. Large KMO values are good because correlations between pairs of variables (i.e., *potential factors*) can be explained by the other variables. If the KMO is below .5, don't do a factor analysis.

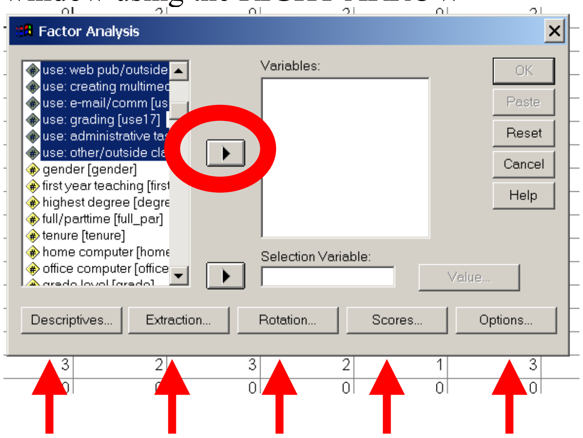
The formula for the KMO is (*the sum of the observed correlation coefficients*) (*the sum of the observed correlation coefficients*) +(*the sum of the partial correlation coefficients between all pairs of variables*). If the sum of the partial correlation coefficients between all pairs of variables is small when compared to the observed correlation coefficients, the KMO measure will be close to ONE. (Remember, a *partial correlation* is a measure of the strength of the relationship between any two variables when the other variables are held constant).

How To Conduct a Factor Analysis in SPSS

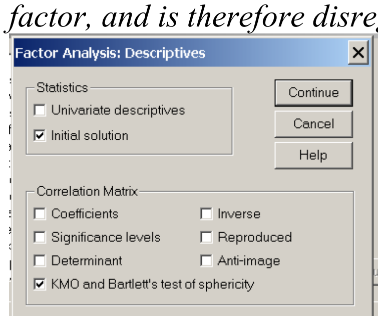
Click on ANALYZE, DATA REDUCTION, FACTOR…



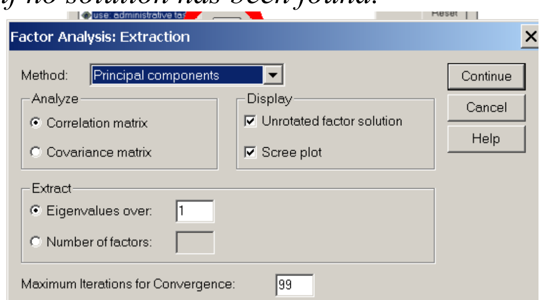
Highlight the items you want to include in the analysis, and move them to the VARIABLES window using the RIGHT ARROW

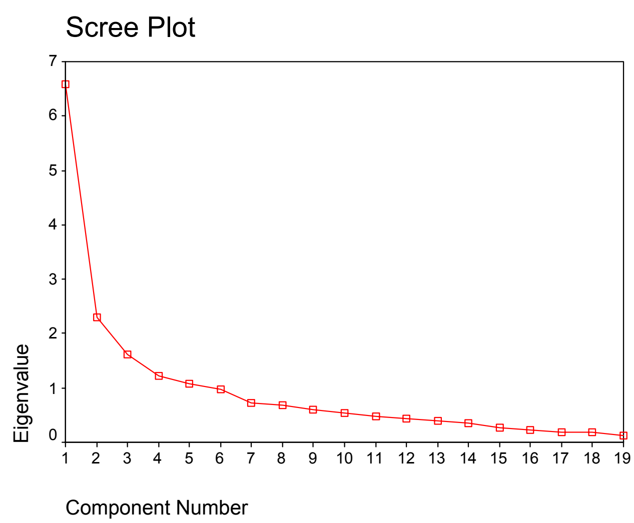


Moving from left to right, select each of the buttons and select the following:
DESCRIPTIVES: Initial Solution (under Statistics) and KMO and Bartlett's test of sphericity under Correlation Matrix, then click CONTINUE. *<u>Initial solution</u> displays initial communalities, eigenvalues, and the percentage of variance explained. "<u>Communality</u>" is the proportion of variance accounted for by the common factors (or 'communality') of a variable. The '<u>Eigenvalue</u>' is the total variance explained by each factor. Any 'factor' that has an Eigenvalue of less than one does not have enough total variance explained to represent a unique factor, and is therefore disregarded.*

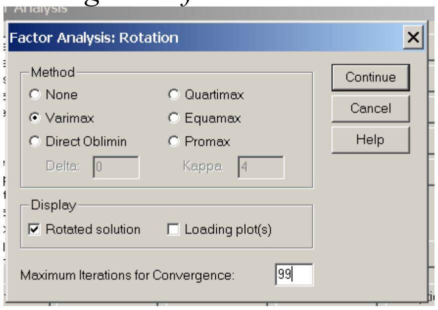


EXTRACTION: Method: Principal Components; Analysis: Correlation Matrix; Display: Unrotated factor solution and Scree Plot; Extract: Eigenvalues over 1, Maximum Iterations for Convergence change to 99, and then click CONTINUE. *Principal Components Analysis is a method of factor extraction where linear combinations of the observed variables are formed. The first 'principal component' is the combination of variables (or items) that accounts for the largest amount of variance in the sample. The second 'principal component' accounts for the next largest amount of variance, and is not correlated with the first component, and so on. When different methods are selected, the percentage of variance explained will be different.*
The 'unrotated factor solution' is the result prior to 'rotating' the solution – rotation is the transformation of the initial matrix into one that can be interpreted.
A 'scree plot' is a graphic that plots the total variance associated with each factor. It is a visual display of how many factors there are in the data. In the scree plot sample below, you can see that, although there are 19 'principal components', only 5 factors have eigenvalues over one, so you will expect to find 5 principal components in the data. 'Maximum Iterations for Convergence' is simply telling the computer how many times to rotate the data before giving up if no solution has been found.



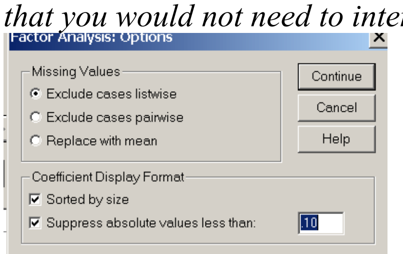


ROTATION: METHOD: select Varimax, DISPLAY: select Rotated solution, Maximum Iterations for Convergence change to 99, and then click CONTINUE. *The purpose of rotation is to simplify the structure of the analysis, so that each factor will have nonzero loadings for only some of the variables without affecting the communalities and the percent of variance explained. The most common method is* <u>*Varimax*</u>*, which minimizes the number of variables that have high loadings on a factor.*



SCORES: There is no need to select this option. It creates one new variable for each factor in the final solution.

OPTIONS: Missing Values: select Exclude cases listwise, Coefficient Display Format: select Sorted by size and Suppress absolute values less than .10, and then click CONTINUE. *The missing values option allows you to specify how missing values within individual items are handled. Missing values listwise are cases that have missing values for any of the variables named will be omitted from the analysis. Selecting 'sorted by size' makes the output easier to read and interpret. 'Suppress absolute values less than .10' will eliminate values from the output that you would not need to interpret anyway, making the output of some analyses easier to read.*



An Example

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.800
Bartlett's Test of Sphericity	Approx. Chi-Square	978.718
	df	171
	Sig.	.000

KAISER-MEYER-OLKIN: measure of sampling adequacy is used to compare the magnitudes of the observed correlation coefficients in relation to the magnitudes of the *partial* correlation coefficients. Large KMO values are good because correlations between pairs of variables (i.e., *potential factors*) can be explained by the other variables. If the KMO is below .5, don't do a factor analysis. This is the type of result you want! A .8 is excellent (you're hoping for a .8 or higher in order to continue…)

BARTLETT'S TEST OF SPHERICITY is used to test the hypothesis that the correlation matrix is an identity matrix (*all diagonal terms are one and all off-diagonal terms are zero*). You are looking for SIGNIFICANCE (*less than .05*) because you WANT the variables to be correlated. In other words, picture a correlation matrix: all items are perfectly correlated with themselves (one), and have *some level* of correlation with the other items. If they *are not correlated with the other items* then they can't be part of the same factor!
Like other tests of significance, you are looking for a value of .05 or less. This result, < .001 is good, and is an indication you can continue with the Factor Analysis.

Communalities

	Initial	Extraction
use: presentations	1.000	.661
use: desktop publishing	1.000	.598
use: drill and practice	1.000	.694
use: testing/eval	1.000	.653
use: tutorials	1.000	.421
use: e-mail/comm	1.000	.719
use: web publishing	1.000	.714
use: web browsing	1.000	.705
use: other/classroom	1.000	.682
use: create presentations	1.000	.604
use: homewk assignments	1.000	.595
use: create tests/handouts	1.000	.662
use: browsing outside class	1.000	.812
use: web pub/outside class	1.000	.749
use: creating multimedia	1.000	.588
use: e-mail/comm	1.000	.792
use: grading	1.000	.674
use: administrative tasks	1.000	.662
use: other/outside classroom	1.000	.825

Extraction Method: Principal Component Analysis.

"Communality" is the proportion of variance accounted for by the common factors (or 'communality') of a variable. Communalities range from 0 to 1. Zero means that the common factors DON'T explain any variance; one means that the common factors explain ALL the variance. Since you are looking for relatively high numbers here, this is a good result.

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	6.582	34.642	34.642	6.582	34.642	34.642	3.337	17.563	17.563
2	2.290	12.052	46.694	2.290	12.052	46.694	2.898	15.255	32.819
3	1.621	8.530	55.224	1.621	8.530	55.224	2.598	13.674	46.493
4	1.232	6.485	61.709	1.232	6.485	61.709	2.150	11.317	57.810
5	1.085	5.710	67.419	1.085	5.710	67.419	1.826	9.609	67.419
6	.966	5.085	72.504						
7	.734	3.862	76.366						
8	.685	3.608	79.973						
9	.604	3.178	83.151						
10	.531	2.797	85.948						
11	.468	2.464	88.412						
12	.431	2.270	90.682						
13	.404	2.125	92.807						
14	.362	1.907	94.713						
15	.276	1.452	96.166						
16	.236	1.240	97.406						
17	.192	1.011	98.417						
18	.180	.949	99.366						
19	.120	.634	100.000						

Extraction Method: Principal Component Analysis.

This table is the ‘Initial Solution’. The ‘Eigenvalue’ is the total variance explained by each factor. Any ‘factor’ that has an Eigenvalue of less than one does not have enough total variance explained to represent a unique factor, and is therefore disregarded. Note that components 6 down have eigenvalues less than 1.0, so they are eliminated from the rest of the analysis. Note that the Cumulative % is less than 100%. This is because not all of the variance is explained when only some of the factors are retained in the final analysis (6 through 19 were eliminated, although together they represent over 30% of the variance explained. However, any one of the factors account for very little variance).

Component Matrix[a]

	Component				
	1	2	3	4	5
use: browsing outside class	.745	.304	-.244	-.168	-.278
use: presentations	.695	-.242	-.107	-.120	.305
use: e-mail/comm	.687	.190	-.123	-.173	-.489
use: e-mail/comm	.681	-.320		-.181	-.346
use: web browsing	.676	-.224	-.207	-.157	-.362
use: create presentations	.667	.321	-.203		.117
use: creating multimedia	.661	.137	-.220	.110	.270
use: desktop publishing	.650	-.347		-.142	.184
use: web publishing	.616	-.113	-.473	.307	
use: create tests/handouts	.595	.505		-.118	.174
use: web pub/outside class	.576	.103	-.423	.413	.237
use: drill and practice	.559	-.429	.433		
use: tutorials	.549	-.258	.121	-.153	.123
use: homewk assignments	.519	.375	.406	-.115	
use: other/classroom	.480	-.425	.387	.308	-.164
use: testing/eval	.471	-.564	.232		.242
use: administrative tasks	.474	.505	.324	.265	
use: other/outside classroom	.422	.329	.520	.505	-.116
use: grading	.213	.421	.207	-.571	.287

Extraction Method: Principal Component Analysis.

a. 5 components extracted.

This Component Matrix indicates how each item in the analysis correlates with each of the five retained factors. Negative and positive correlations carry the same weight. This chart gives you an idea how the items correlate with the factors, but is not as easy to interpret as the rotated solution.

Rotated Component Matrix[a]

	Component				
	1	2	3	4	5
use: drill and practice	.794		.140	.195	
use: testing/eval	.793	.139			
use: desktop publishing	.641	.329	.233		.142
use: other/classroom	.639		.160	.391	-.309
use: presentations	.584	.479	.183		.227
use: tutorials	.566	.172	.199		.174
use: web pub/outside class		.839		.154	
use: web publishing	.208	.747	.283		-.182
use: creating multimedia	.225	.661	.162	.153	.227
use: create presentations		.591	.307	.238	.314
use: e-mail/comm		.203	.817	.229	.156
use: browsing outside class		.413	.720	.191	.291
use: web browsing	.366	.245	.713		
use: e-mail/comm	.535		.650		
use: other/outside classroom	.125	.102		.893	
use: administrative tasks		.197	.164	.737	.229
use: grading					.817
use: create tests/handouts		.357	.210	.388	.580
use: homewk assignments	.206		.158	.509	.513

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 7 iterations.

The Pattern Matrix for oblique rotations reports the factor loadings for each variable on the components or factors after rotation. The rotated solution gives you a clear indication how each item correlates with each factor. Note that there are missing values – this is because we asked the computer to 'suppress values less than .10' to make the chart easier to read. If an item correlates less than .1, it doesn't belong with that factor. In this example, we could have suppressed values of .3 or even higher, and still matched each item with it's corresponding factor. Review this chart and group all of the items with its appropriate factor. Once you have done that, see what the various items have in common, and see if you can name the CONSTRUCT these items represent. (Each question referred to ways teachers can use computers in and outside the classroom).

Component Transformation Matrix

Component	1	2	3	4	5
1	.535	.533	.516	.323	.241
2	-.701	.149	.041	.455	.527
3	.395	-.613	-.247	.608	.192
4	-.096	.385	-.319	.535	-.674
5	.239	.410	-.755	-.179	.416

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

The factor transformation matrix describes the specific rotation applied to your factor solution. It does not need to be reported or interpreted.