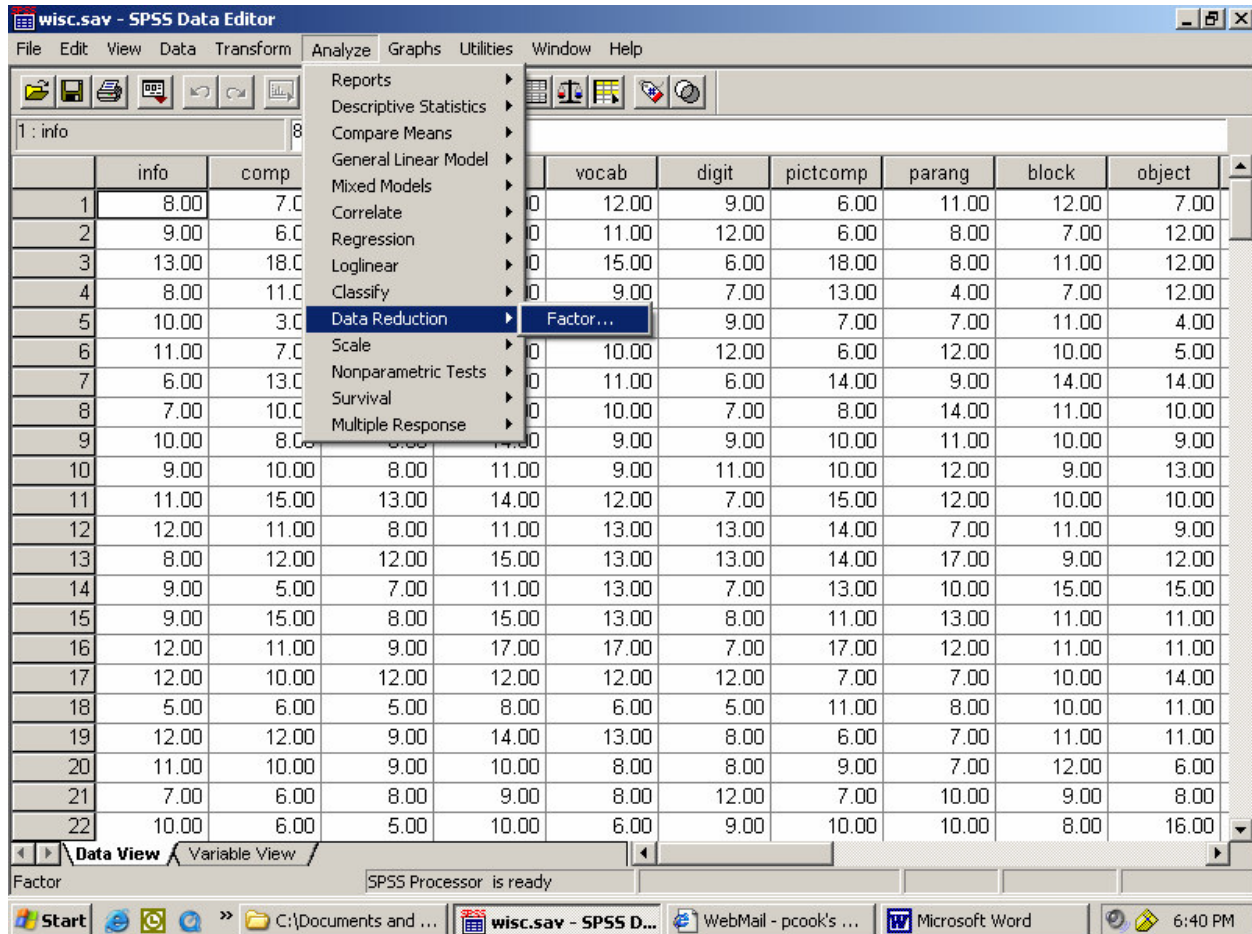


## Factor Analysis in SPSS

To conduct a **Factor Analysis**, start from the “Analyze” menu. This procedure is intended to *reduce the complexity in a set of data*, so we choose “Data Reduction” from the menu. And the choice in this category is “Factor,” for factor analysis.



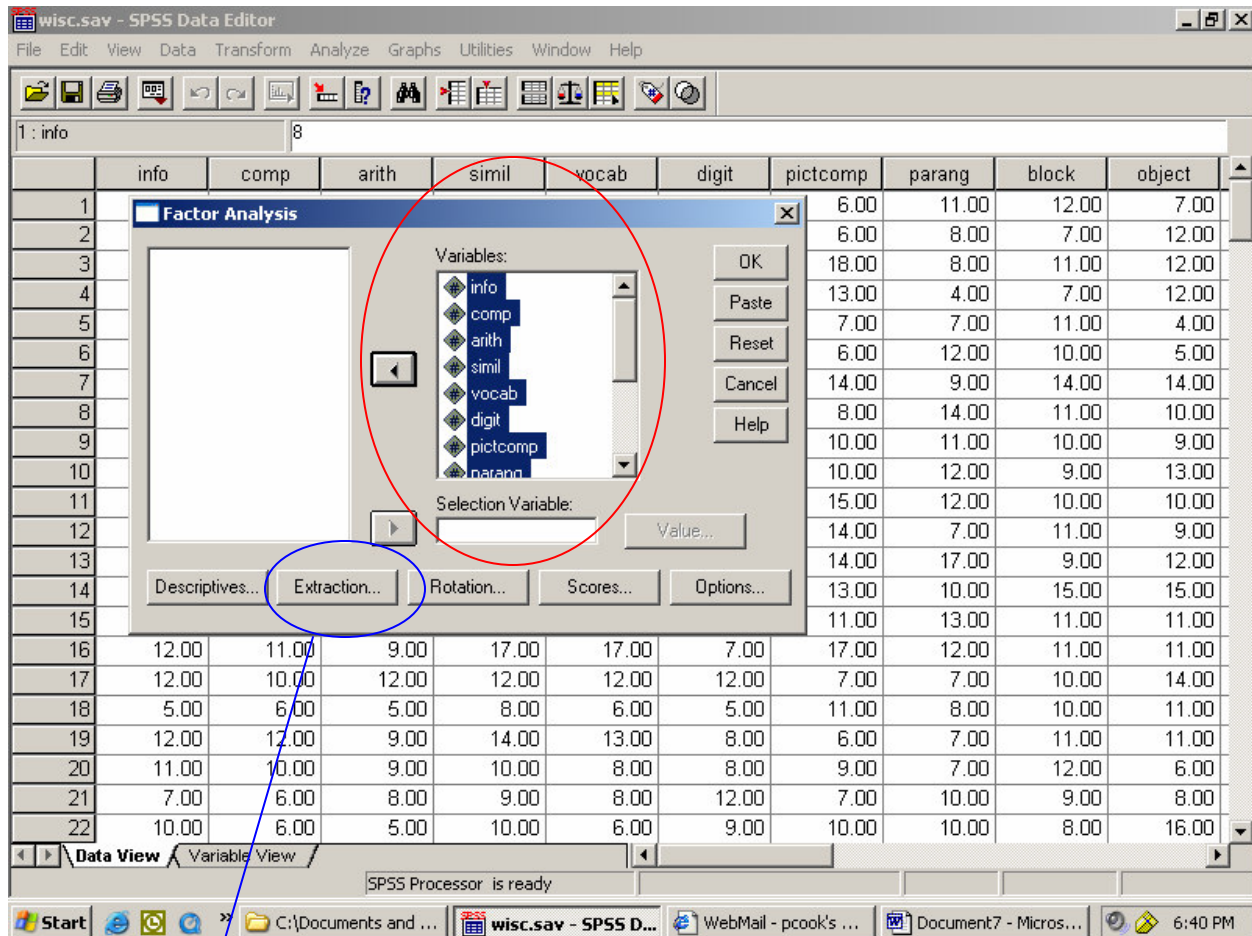
The screenshot shows the SPSS Data Editor window with the 'Analyze' menu open. The 'Data Reduction' option is selected, and the 'Factor...' sub-option is highlighted. The data table below shows scores for 22 cases across 8 variables: info, comp, vocab, digit, pictcomp, parang, block, and object.

	info	comp	vocab	digit	pictcomp	parang	block	object
1	8.00	7.00	12.00	9.00	6.00	11.00	12.00	7.00
2	9.00	6.00	11.00	12.00	6.00	8.00	7.00	12.00
3	13.00	18.00	15.00	6.00	18.00	8.00	11.00	12.00
4	8.00	11.00	9.00	7.00	13.00	4.00	7.00	12.00
5	10.00	3.00	10.00	12.00	6.00	12.00	10.00	5.00
6	11.00	7.00	11.00	6.00	14.00	9.00	14.00	14.00
7	6.00	13.00	10.00	7.00	8.00	14.00	11.00	10.00
8	7.00	10.00	9.00	9.00	10.00	11.00	10.00	9.00
9	10.00	8.00	9.00	10.00	11.00	10.00	9.00	13.00
10	9.00	10.00	8.00	11.00	9.00	11.00	10.00	10.00
11	11.00	15.00	13.00	14.00	12.00	7.00	15.00	12.00
12	12.00	11.00	8.00	11.00	13.00	13.00	14.00	7.00
13	8.00	12.00	12.00	15.00	13.00	13.00	14.00	17.00
14	9.00	5.00	7.00	11.00	13.00	7.00	13.00	10.00
15	9.00	15.00	8.00	15.00	13.00	8.00	11.00	13.00
16	12.00	11.00	9.00	17.00	17.00	7.00	17.00	12.00
17	12.00	10.00	12.00	12.00	12.00	12.00	7.00	7.00
18	5.00	6.00	5.00	8.00	6.00	5.00	11.00	8.00
19	12.00	12.00	9.00	14.00	13.00	8.00	6.00	7.00
20	11.00	10.00	9.00	10.00	8.00	8.00	9.00	7.00
21	7.00	6.00	8.00	9.00	8.00	12.00	7.00	10.00
22	10.00	6.00	5.00	10.00	6.00	9.00	10.00	8.00
							8.00	16.00

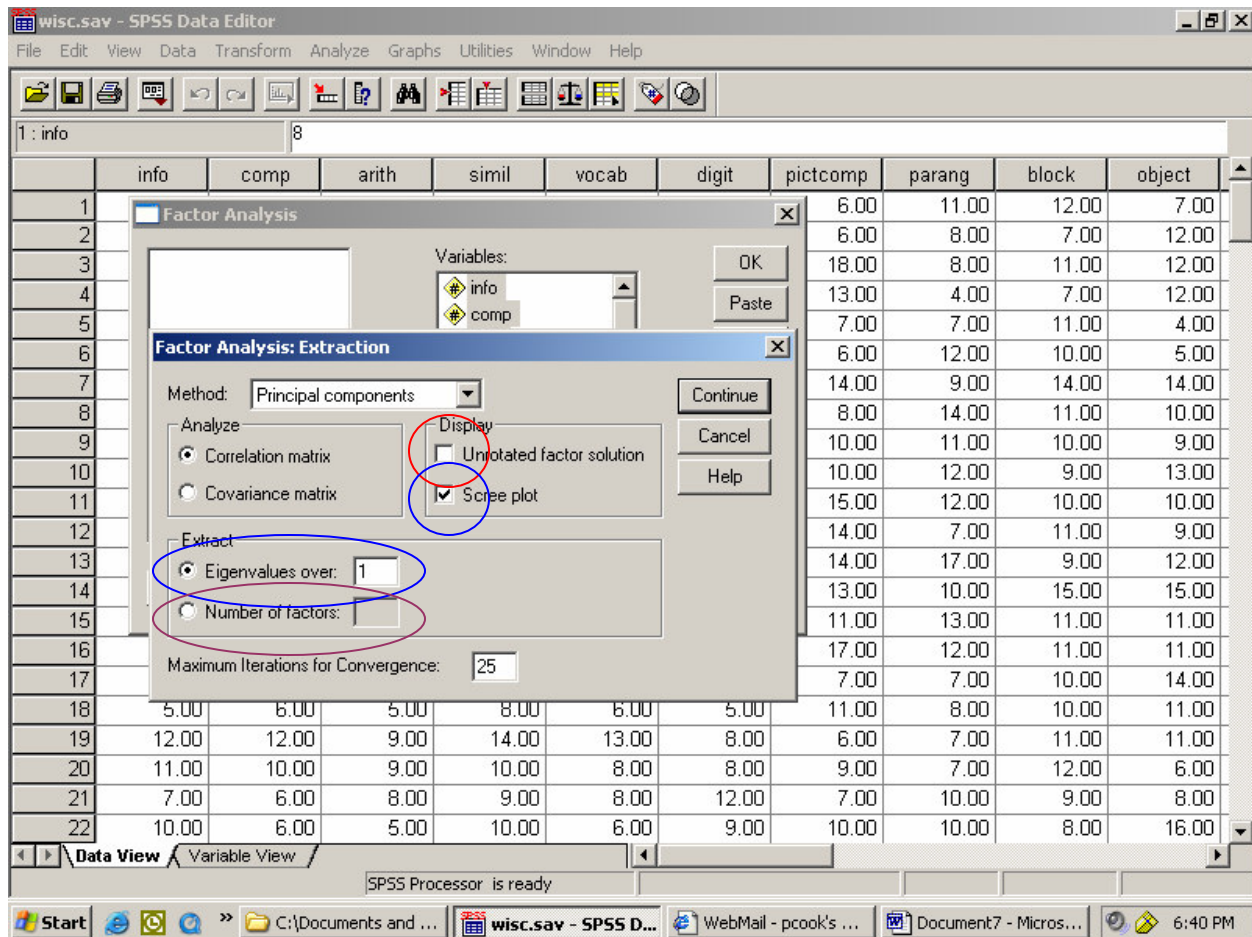
This dataset gives children’s scores on subtests of the Wechsler Intelligence Scale for Children (WISC-III). The Wechsler scales are scored to give you a “verbal” and a “performance” IQ. The question is whether we can reproduce the verbal vs. nonverbal distinction, with the appropriate subtests grouping into each category, using factor analysis.

This dialog box will appear:

Factor analysis has no IVs and DVs, so everything you want to get factors for just goes into the list labeled “variables.” In this case, it’s all the variables. In some datasets, there is also a dummy “subject number” variable included. Be sure that you *don’t* include subject number as one of the variables for your factor analysis!



To do a factor analysis, we need to select an “extraction method” and a “rotation method.” Hit the “Extraction” button to specify your extraction method.

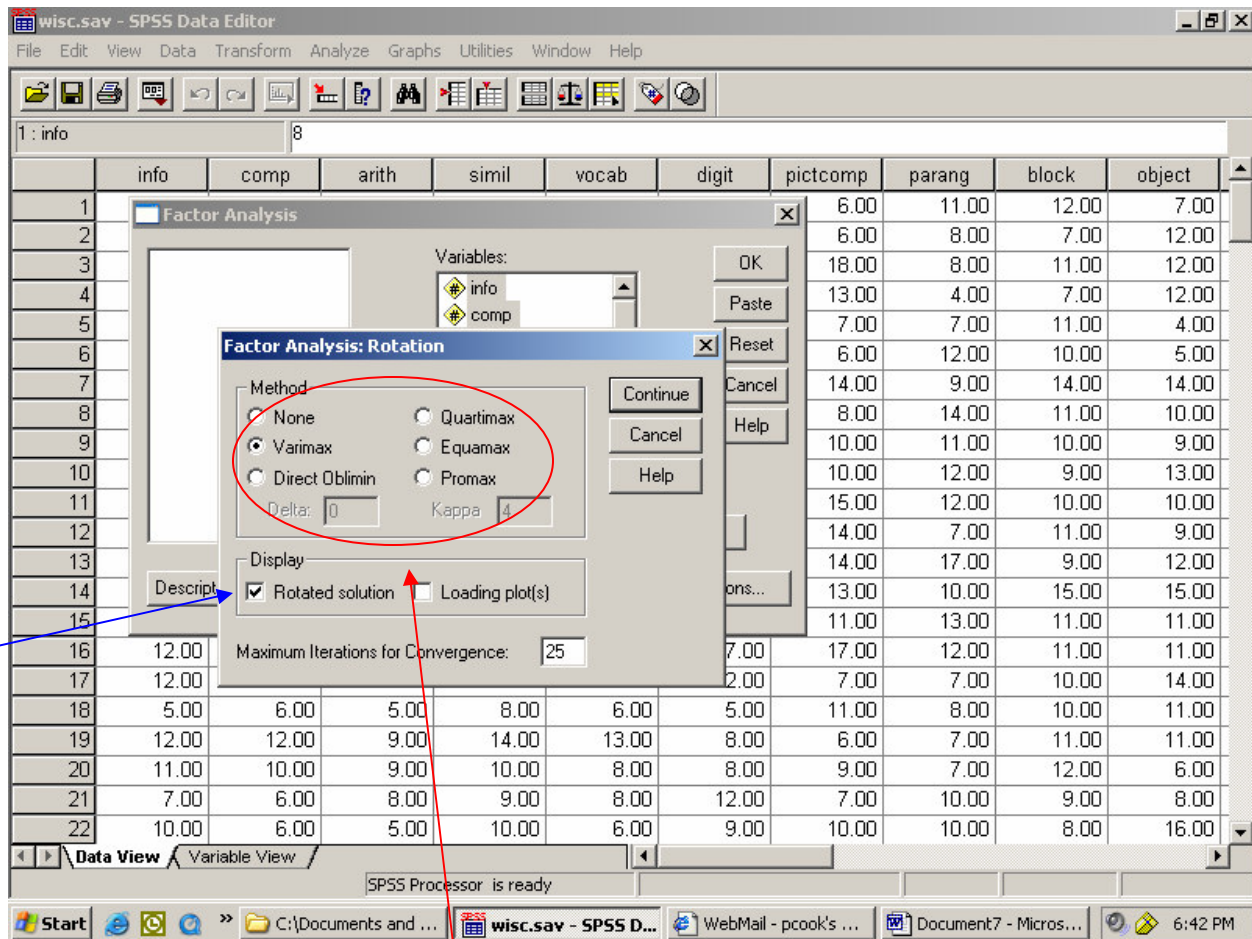


In this dialog box, you can make a number of selections. First, I want you to *un-check* the box labeled “Unrotated factor solution.” This is a default setting for your printout, but it just gives you information that you don’t need, and that may distract you from the real answers. So, always go into the Extraction sub-dialog and un-check this box.

Second, check the box for a “scree plot.” This will give you a scree diagram (as discussed in the lecture notes), which is one way to decide how many factors to extract.

Third, look at the section labeled “Extract.” As you can see, the default setting is for SPSS to use the Kaiser stopping criterion (i.e., all factors with eigenvalues greater than 1) to decide how many factors to extract. You can set a more conservative stopping criterion by requiring each factor to have a higher eigenvalue. Or, if you already know exactly how many factors you think there will be, you can set the extraction method to a specific “Number of factors,” and then put the number into this box.

Now go back to the main dialog box by clicking “Continue.” Once you’re there, click on the button for “Rotation” to see the second sub-dialog:



This dialog allows you to choose a “rotation method” for your factor analysis. As discussed in the class notes, a rotation method gets factors that are as different from each other as possible, and helps you interpret the factors by putting each variable primarily on one of the factors. However, you still need to decide whether you want an “orthogonal” solution (factors are *not* highly correlated with each other), or an “oblique” solution (factors *are* correlated with one another). If you want an oblique solution, the only choice SPSS gives you is “Direct Oblimin.” All of the others are orthogonal solutions—the one that you’ll use most often from these choices is the default value, “Varimax.” Most of the factor analyses you will see in published articles use a Varimax rotation.

While you’re here, make sure that the check-box for a “rotated solution” is on. The rotated solution gives you the factor loadings for each individual variable in your dataset, which are what you use to interpret the meaning of (i.e., make up names for) the different factors.

Hit “Continue” in the sub-dialog, and then “OK” in the main dialog to see the output:

## Factor Analysis

### Communalities

	Initial
INFO	1.000
COMP	1.000
ARITH	1.000
SIMIL	1.000
VOCAB	1.000
DIGIT	1.000
PICTCOMP	1.000
PARANG	1.000
BLOCK	1.000
OBJECT	1.000
CODING	1.000

Extraction Method: Principal Component Analysis.

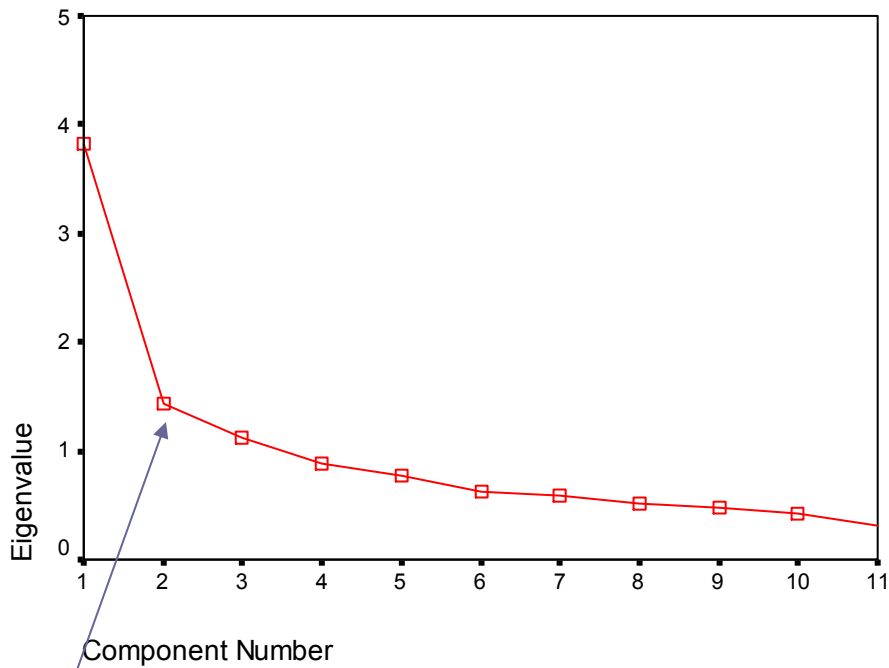
### Total Variance Explained

Component	Initial Eigenvalues			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.829	34.806	34.806	3.023	27.485	27.485
2	1.442	13.109	47.915	2.209	20.084	47.569
3	1.116	10.147	58.062	1.154	10.492	58.062
4	.890	8.092	66.153			
5	.768	6.985	73.138			
6	.633	5.753	78.891			
7	.595	5.412	84.303			
8	.522	4.749	89.051			
9	.471	4.281	93.332			
10	.419	3.806	97.138			
11	.315	2.862	100.000			

Extraction Method: Principal Component Analysis.

This table shows you the actual factors that were extracted. If you look at the section labeled “Rotation Sums of Squared Loadings,” it shows you only those factors that met your cut-off criterion (extraction method). In this case, there were three factors with eigenvalues greater than 1. SPSS always extracts as many factors initially as there are variables in the dataset, but the rest of these didn’t make the grade. The “% of variance” column tells you how much of the total variability (in all of the variables together) can be accounted for by each of these summary scales or factors. Factor 1 accounts for 27.485% of the variability in all 11 variables, and so on.

## Scree Plot



Here's the scree plot that we asked for. You can see that it might lead you to a slightly different conclusion—it looks like the slope of this curve levels out after just two factors, rather than three.

### Component Matrix(a)

a 3 components extracted.

This section would have a table that looks a lot like the following one, but that we don't use for anything. We un-checked a box so that it wouldn't even show up, but the heading is still here.

**Rotated Component Matrix(a)**

	Component		
	1	2	3
INFO	.826	.107	-.017
COMP	.634	.415	-.053
ARITH	.669	.084	.179
SIMIL	.694	.324	-.172
VOCAB	.782	.181	.013
DIGIT	.535	-.067	.428
PICTCOMP	.246	.649	-.280
PARANG	.142	.567	.163
BLOCK	.173	.742	.141
OBJECT	.038	.756	-.026
CODING	-.005	.106	.883

Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization.  
a. Rotation converged in 5 iterations.

Finally, the Rotated Component Matrix shows you the factor loadings for each variable. I went across each row, and highlighted the factor that each variable loaded most strongly on. Based on these factor loadings, I think the factors represent:

--The first 5 subtests loaded strongly on Factor 1, which I'll call "Verbal IQ"

--Picture Completion through Object Assembly all loaded strongly on Factor 2, which I'll call "Performance IQ"

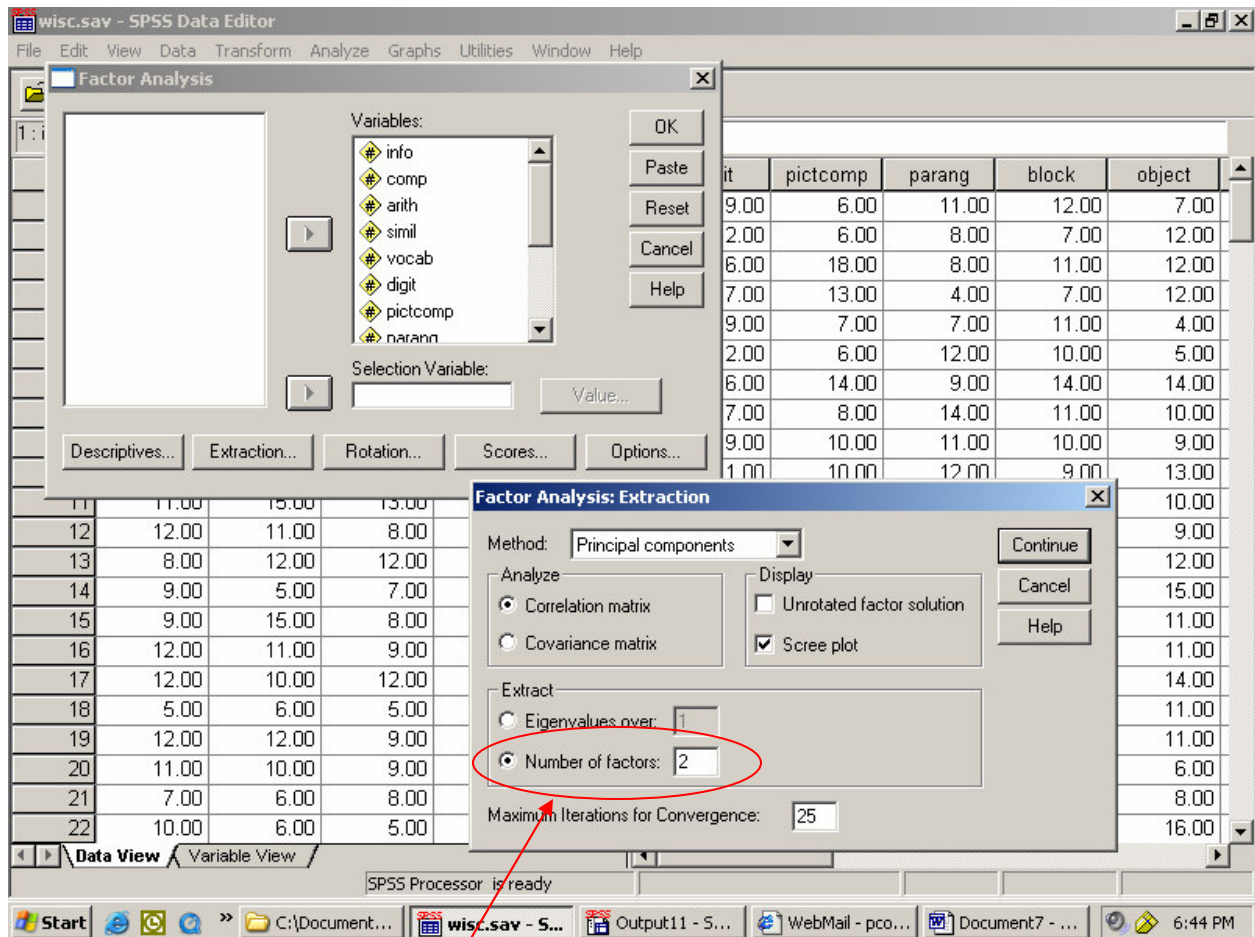
--Coding loaded strongly on Factor 3 (and Digit Span loaded fairly strongly on Factor 3, although it also loaded on Factor 1). Probably Factor 3 is "Freedom from Distraction," because these are concentration-intensive tasks.

**Component Transformation Matrix**

Component	1	2	3
1	.818	.573	.050
2	-.533	.788	-.309
3	-.217	.226	.950

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

Not using this table for anything. It has to do with SPSS performing the orthogonal rotation that we asked for.



Our original interest was in whether we could differentiate verbal from nonverbal tasks. It looks like we got close with a 3-factor solution, but there was still some confusion around Digit Span, and Coding loaded on its own factor. Let's force SPSS's hand, by specifying a set number of factors to extract.

To do this, go back to the main dialog, and then to the "Extraction" sub-dialog. Under "Extract," select "Number of factors" and enter the number two, rather than letting SPSS make this decision empirically based on the eigenvalues.



Here's the revised output, with a two-factor solution:

## Factor Analysis

[table omitted]

**Total Variance Explained**

Component	Initial Eigenvalues			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.829	34.806	34.806	2.727	24.793	24.793
2	1.442	13.109	47.915	2.543	23.122	47.915
3	1.116	10.147	58.062			
4	.890	8.092	66.153			
5	.768	6.985	73.138			
6	.633	5.753	78.891			
7	.595	5.412	84.303			
8	.522	4.749	89.051			
9	.471	4.281	93.332			
10	.419	3.806	97.138			
11	.315	2.862	100.000			

Extraction Method: Principal Component Analysis.

Note that we now have just two factors extracted, and that each one accounts for a larger percentage of the variability.

[same scree plot]

**Rotated Component Matrix(a)**

	Component	
	1	2
INFO	.778	.242
COMP	.550	.516
ARITH	.679	.157
SIMIL	.588	.461
VOCAB	.735	.301
DIGIT	.628	-.059
PICTCOMP	.093	.723
PARANG	.112	.540
BLOCK	.117	.719
OBJECT	-.056	.742
CODING	.201	-.068

Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization.  
a. Rotation converged in 3 iterations.

Here are the revised factor loadings. Coding doesn't load strongly on either of these, but our two factors of "Verbal" and "Performance" IQ have emerged more strongly (larger factor loadings).