Factor Analysis

“Data reduction”

- In SPSS, factor analysis is listed as “data reduction”. The researcher strives to avoid two major obstacles:
  - Using a single score from a multi-item measure in which there is great heterogeneity; or
  - Using several factor scores that are highly correlated or unreliable.
- The goal is to obtain the “right” number of factor scores (and it might be one).

The important stuff that you need to know about FA

- The basic issue is the degree of correlation among a set of items. We expect to find “clumps” of items sometimes, and these are called factors.
- If <u>all</u> of the items correlate highly with each other, then you have a simple outcome: a single factor. A Cronbach’s alpha tells you how well these correlate with each other. If it’s high (say .80 or .90), then it’s probably one factor.
- However, we often find that we don’t have a single homogenous factor; instead, we have several clumpings or factors.

Assumptions

- You have a bunch of items, and you don’t know how many factors you have.
- This is the process of how you would determine the correct number of factors.
- Incidentally, like mediation and moderation, there is no single button to push in SPSS to obtain an optimal solution. It’s a process.

Key points we'll cover

- Factors: how many?
- Extraction: how does the programme generate factor loadings on the factors?
- Rotation: how does the programme identify factors in relation to each other?
- Tools: scree plot, Cronbach's alpha, correlations among factors, etc.
- Terms: eigenvalue, communality, etc.

Factors?

- So how do we identify these clumpings?
- Statisticians refer to “extraction” as the way to identify factors.

Differences between FA and PCA

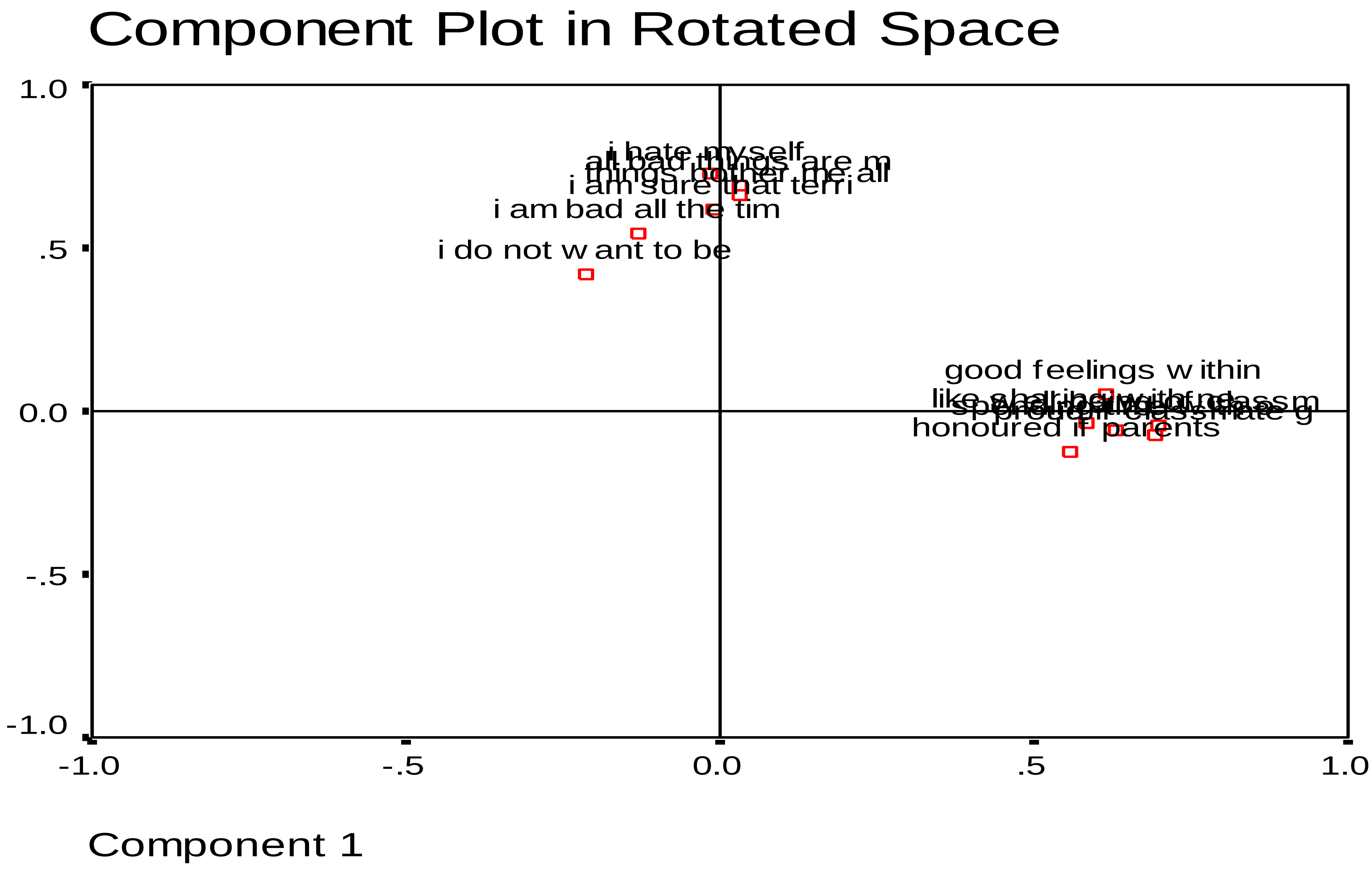
- There are a number of statistical reasons between the two (won’t go into them here, but they are in your textbook).
- The main difference is that FA attempts to eliminate unique and error variance from factors, i.e., obtain latent factors.
- PCA attempts to replicate the associations among items and factors while also including unique and error variance. It is “closer” to the real, unvarnished data.
- Bottom line: Most people use PCA because they want factors that contain error, i.e., are close to observed data.

Rotation?

- You have several choices of whether you want to rotate your factors. This refers to how you choose to orient items in 2-dimensional space. Example on next page.
- Basic choice is between oblique and orthogonal.
- <u>Oblique</u>: called "oblimin" in SPSS, this refers to correlated factors.
- <u>Orthogonal</u>: called "varimax", "quartimax", and "equamax"; these try to minimize correlation among obtained factors.
- Most people opt for varimax. Differences among these approaches are not great.

Unrotated solution

Two dimensions: Nicely rotated by varimax



Number of factors?

- How does one know how many factors are lurking in the data?
- Goal: maximum number of uncorrelated factors with high internal reliability. (Some people might argue for the “minimum” number but that depends on what you’re trying to do.)

Examples

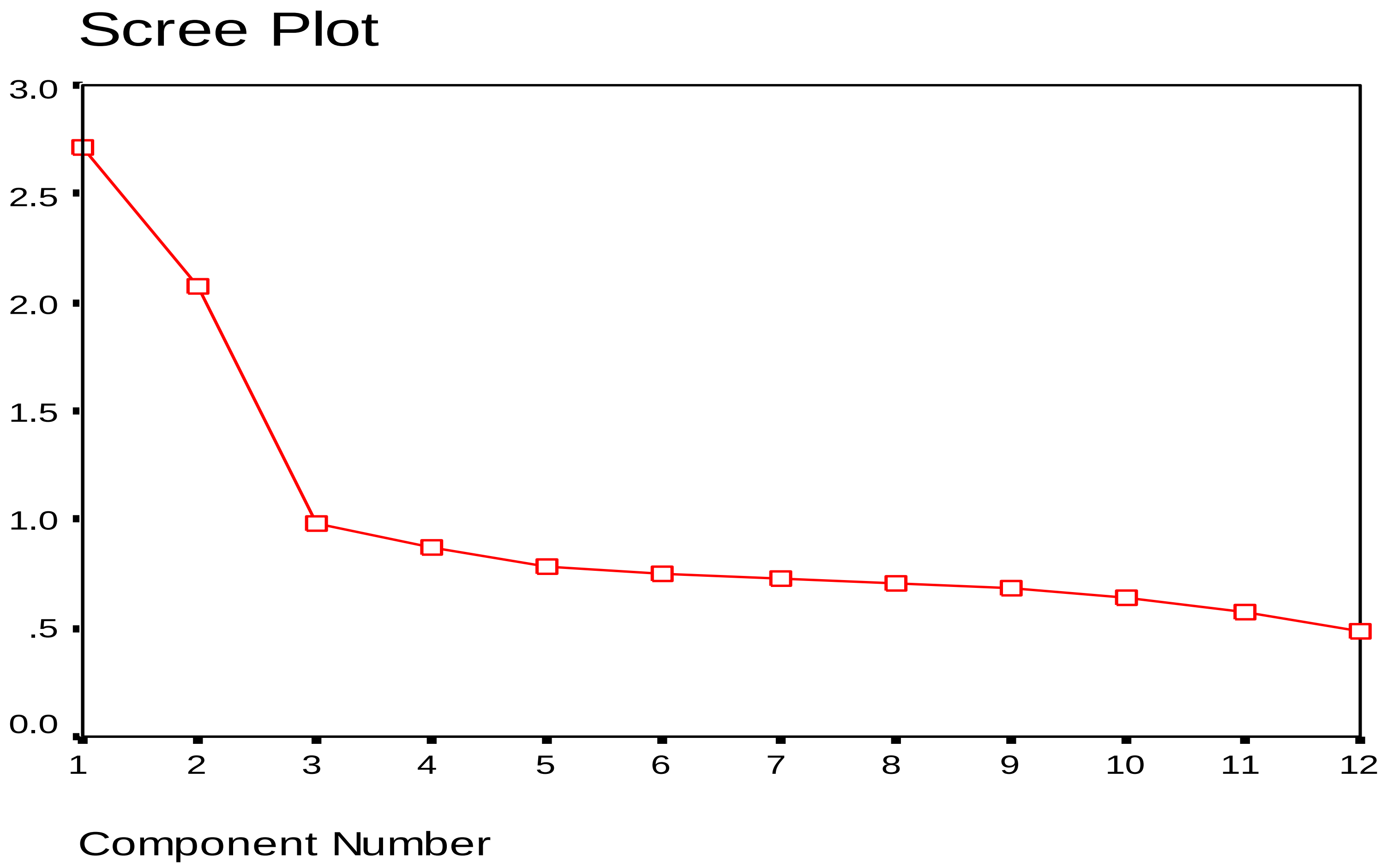
- One factor: α = .88
- Two factors: αs = .85 & .80, $\underline{r}$ = .26
- Three factors: αs = .84, .80, and .75, $\underline{r}$s = .24, .17, and .39.
- Four factors: αs = .84, .79, .74, and .60, $\underline{r}$s = .25, .20, .11, .72, .69, and .68.

- What is the optimal number? Is it always clear cut? Not always.

So how do we get SPSS to give us the optimal number of factors?

- There isn't a button for this.
- It's a process.
- In the first run, don't specify number of factors.
- Look at the scree plot.
- Choose the maximum number of factors that seem advisable from the scree plot. The "rule" is that the "elbow" signifies the optimal number of factors.
- Choose a liberal number of factors and run it again.

A classic elbow: They're not all so clean



Keep going

- Take a look to see whether you have a reasonable number of items for each factor.
- See if there are a lot of double-loading items (indication of more factors).
- Now, do Cronbach’s alphas and correlations on the obtained factors.
- Go to the next fewest factors model, and so forth.
- Choose the best solution based upon high alphas and low intercorrelations (as in our example).

Four factor solution: Too many

Rotated Component Matrix[a]

	Component			
	1	2	3	4
All bad things are my fault	.745			
I hate myself	.716			
I am bad all the time	.658			-.320
Things bother me all the time	.564			.391
I am sure that terrible	.530			.322
proud if classmate gets award		.798		
honoured if parents receive a reward		.736		
well-being of classmates important		.557		.344
I do not want to be with people			-.810	
spending time with others is pleasureable		.362	.580	
good feelings within my group				.731
like sharing with neighbours			.386	.450

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

a. Rotation converged in 8 iterations.

Three factor solution

Rotated Component Matrix[a]

	Component		
	1	2	3
proud if classmate gets award	.720		
well-being of classmates important	.680		
honoured if parents receive a reward	.639		
good feelings within my group	.625		
like sharing with neighbours	.509		-.337
I hate myself		.719	
All bad things are my fault		.708	
Things bother me all the time		.642	
I am sure that terrible		.598	
I am bad all the time		.589	
I do not want to be with people			.805
spending time with others is pleasureable	.505		-.523

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

a. Rotation converged in 5 iterations.

Two factor solution: Looks like the best

Rotated Component Matrix[a]

	Component	
	1	2
well-being of classmates important	.698	
proud if classmate gets award	.694	
spending time with others is pleasureable	.628	
good feelings within my group	.613	
like sharing with neighbours	.581	
honoured if parents receive a reward	.559	
I hate myself		.725
All bad things are my fault		.691
Things bother me all the time		.658
I am sure that terrible		.618
I am bad all the time		.546
I do not want to be with people		.423

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

a. Rotation converged in 3 iterations.

Cronbach's alphas and correlation

- Internal reliability:
  - Collectivism: .70 (barely acceptable); and
  - Depression: .73 (slightly above acceptable).
- Correlation:
  - $r(2003) = -.16$ (good example of two factors not correlated)
- Conclusion: We have two sub-factors with acceptable internal reliability that are not significantly correlated.
- Use: Would use these two sub-factors separately, and would not combine.

A common question: Size of sample?

- Depends upon two things:
  - How many items you have (about 10 subjects per item); and
  - How sensitive you want to be.
- A colleague proposed a 44-item coping measures. Ouch! Would need about 400 subjects. I had him cut back a bit.
- Your textbook says that 5 subjects per item is okay. Well, perhaps with larger samples. I wouldn’t want to do a PCA on 25 subjects for a questionnaire of 5 items. At the lower end, use the 10:1 ratio.

Terms, jargon, and obscure labels

- What do all of the terms mean?
  - Eigenvalue: a “latent root”, a clumping of items that may or may not meet other criteria for being called a factor.
  - Communality: amount of variance an item shares with all other items (want this high)
  - Factor loadings: correlation between an item and the factor (higher the better)
  - Bartlett’s test of sphericity: degree of intercorrelation among items (similar to Cronbach’s alpha)

Uses of PCA (factor analysis)

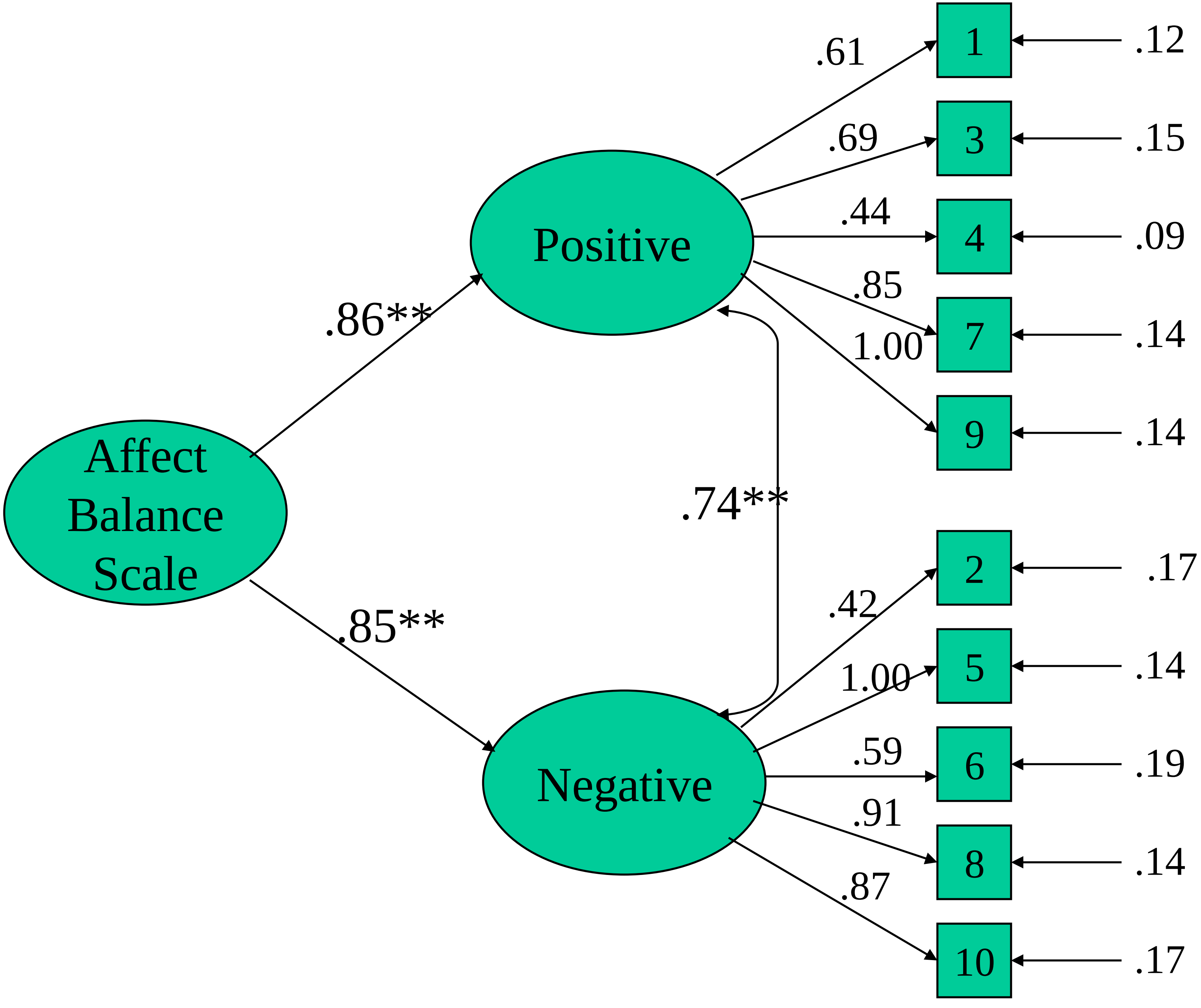
- The assumption I used for the previous example was that I had written 12 items for a new scale. Doing an exploratory factor analysis (EFA) is perfectly fine for that.
- But many people (most people!) use EFA to verify factor structure on a previously FA'ed measure. For example, you want to use a measure, and the initial paper said that there are two factors. What do you do? Most people do an EFA on their data to try to come up with two factors. What's wrong with this?

EFA vs. CFA

- The other main technique is called confirmatory factor analysis (CFA), which should be used when you’re trying to confirm whether a previously obtained factor structure applies to your dataset.
- Where is CFA in SPSS? You won’t find it. Unless you go all the way down the list and choose AMOS, which is a structural equation modeling programme.
- CFA is done in SEM (structural equation modeling). I use LISREL, but AMOS does the same things.

Is CFA hard?

- Well, it is initially, but after you learn it, then it seems simple.
- Basic technique is to tell the program which items load on which factors. Then the program will tell you whether your data are a good fit to the model. Your output is a set of model fit indices that tell you how well the model fits the data.
- Next page gives an example.


Affect Balance Scale
Positive
Negative
.86**
.85**
.74**
.61
.69
.44
.85
1.00
.42
1.00
.59
.91
.87
1
3
4
7
9
2
5
6
8
10
.12
.15
.09
.14
.14
.17
.14
.19
.14
.17

Model fit indices

	Number of Factors for ABS	
Model fit indices	One	Two
Ratio of χ^2/df	2.13	1.91
GFI	.91	.93
RMSEA	.09	.07
CFI	.83	.89
NNFI	.78	.85

Is CFA for everyone?

- Well, maybe not at the present time, but SEM is becoming more widely used and the day is coming soon when undergraduates are expected to know it.
- I can say with authority that more and more of the articles in the journals that you're reading include SEM-type analyses, and if you don't understand the basics you will be left behind.
- It's getting to the point where you cannot publish in a first-tier journal without using some kind of SEM.