INTRODUCTION TO EXPLORATORY FACTOR ANALYSIS

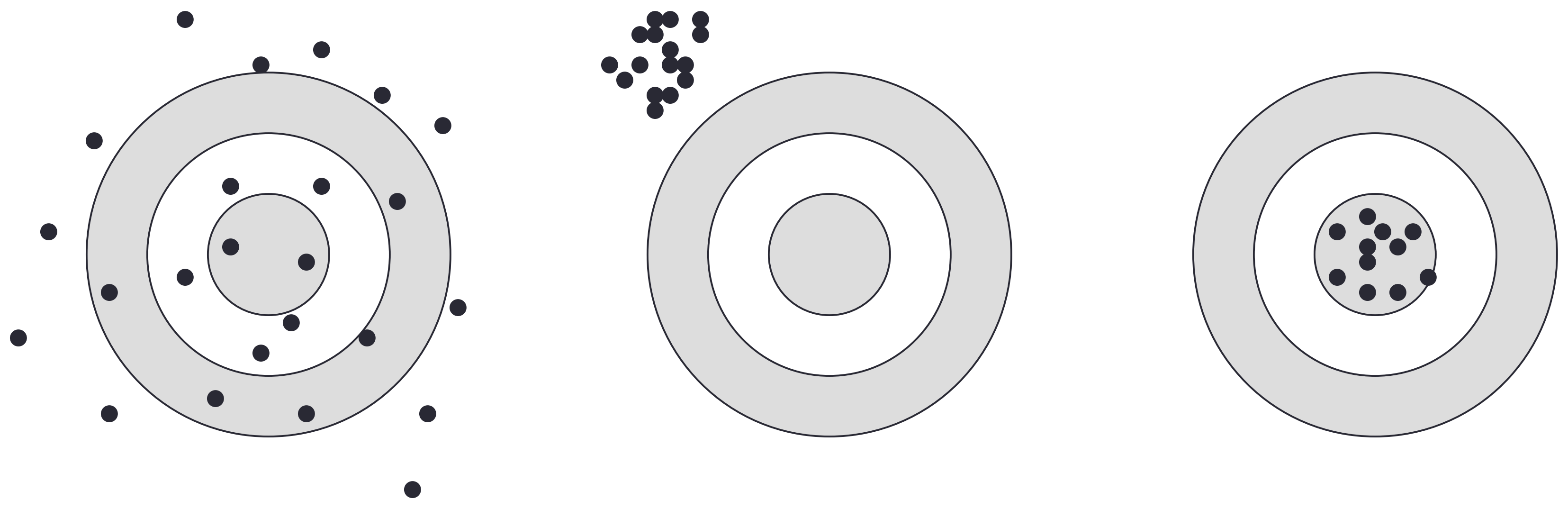
Chapter 3
Principle components analysis
Subjective

Confirmatory factor analysis
Structural equation modeling
These have hypothesis tests

Topics

1. What is Factor Analysis?
2. Questionnaire design and factor analysis.
3. Designing a factor analysis.
4. Example.

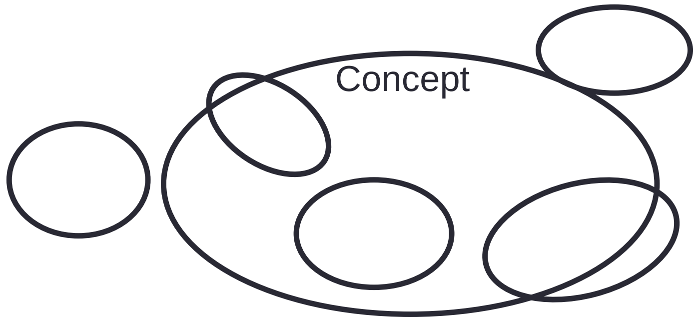
EFA & Validity and Reliability



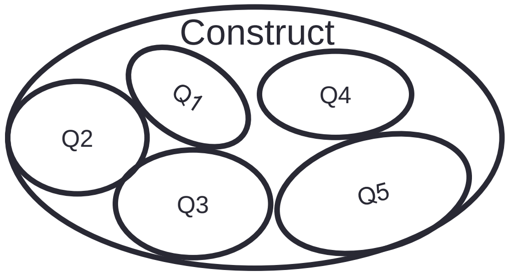
Neither reliable nor valid. | Highly reliable but not valid. | Highly reliable and valid.

- **Reliability:** Repeated measurements of the same thing give consistent results.
- **Validity:** Measuring what you actually mean to measure.

Put simply, in our way of thinking, concepts are precursors to constructs in making sense of organizational worlds…



A construct is formulated so it can be measured; its primary purpose is to delineate a domain of attributes that can be operationalized and preferably quantified as variables.



Forming Groups

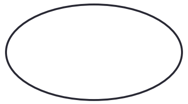
R-type forms groups of related variables.
We focus here.

Variables

Respondents	1	2	3	4	5
1	7	5	6	4	3
2	4	5	2	2	2
3	5	4	2	5	5
4	2	4	2	4	4
5	4	5	5	1	1
6	6	2	3	6	5
7	3	4	4	5	2
8	5	6	7	2	4
9	4	6	3	6	5

Q-type forms groups of related respondents, such as cluster analysis does. This is a rare technique, so we choose to **ignore it.**

Notation



Ovals are latent unobserved (latent) variables

Squares are observed variables

Straight arrows indicate causality

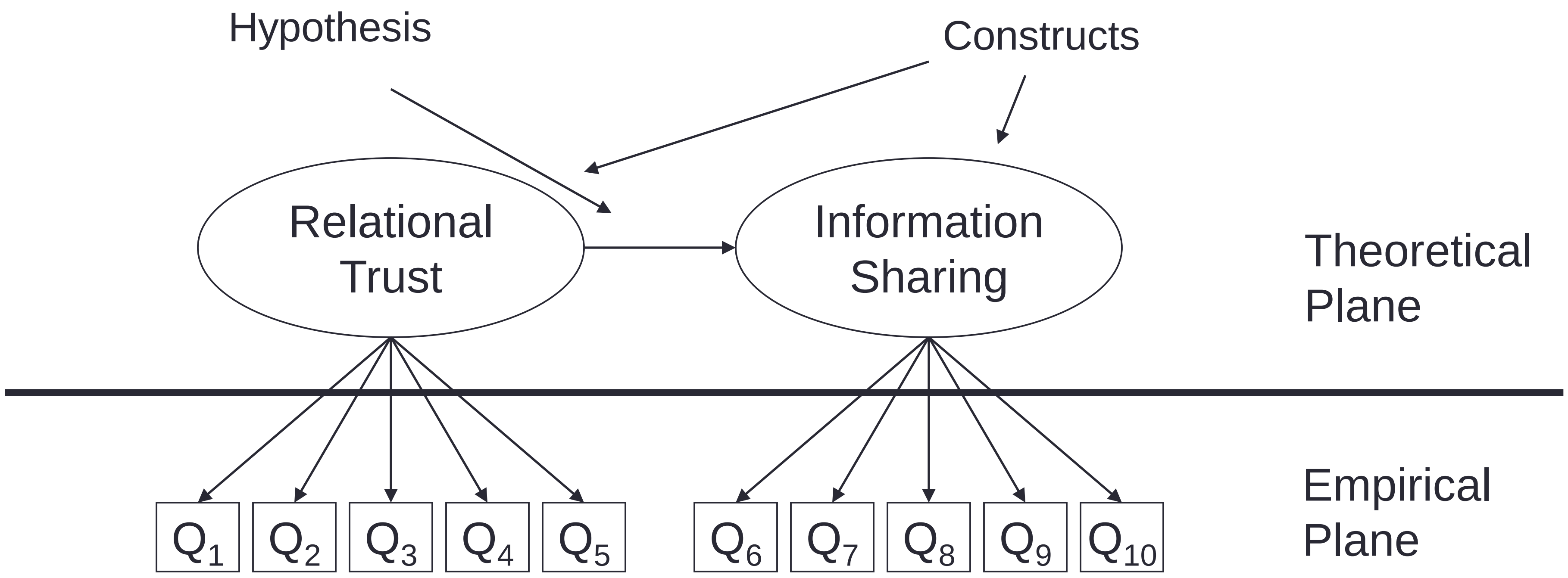
Curved arrows indicate association/correlation

1. What is Factor Analysis?

- It considers the inter-relationship between variables.
- You input a ”bunch” of variables, and it divides them into groups (factors) based on underlying common dimensions.
  - E.g. Questionnaire items like trust (5 indicators) and information sharing (5 indicators).
- You can use it to explore the dimensionality of data, or to confirm beliefs about the structure.
  - Do not confuse this with confirmatory factor analysis (as with LISREL), which tests ***hypotheses*** regarding data dimensions.

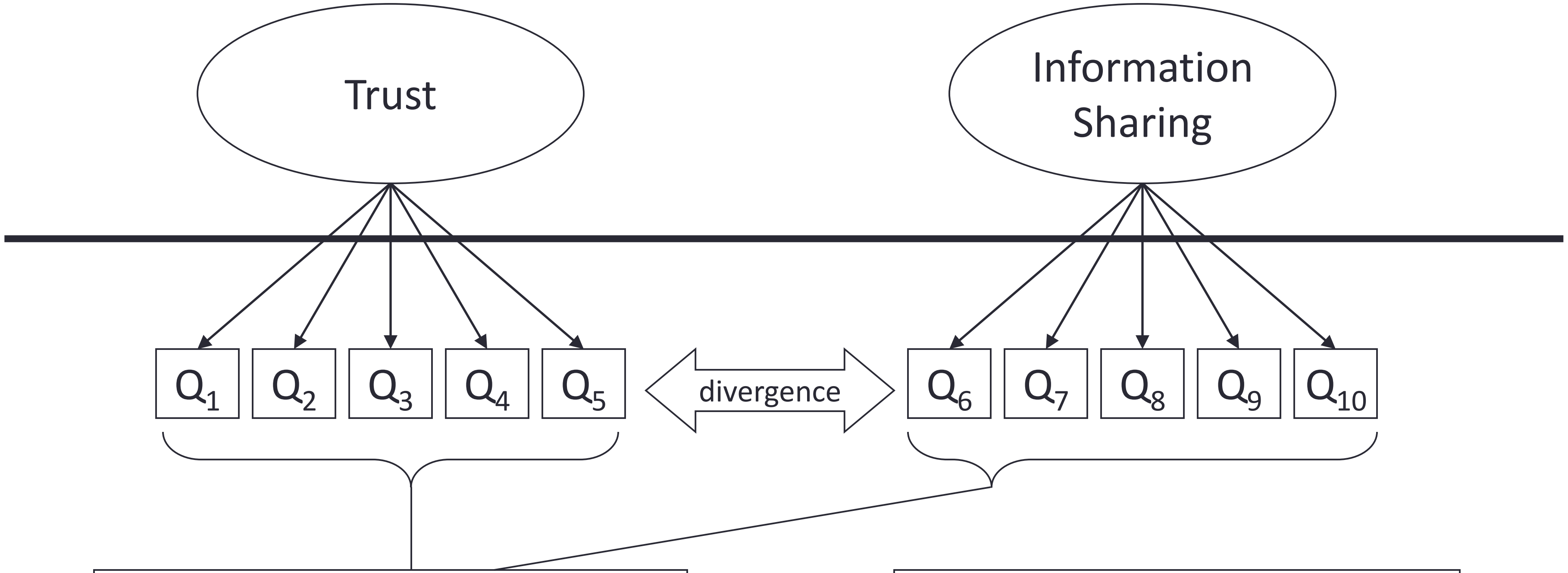
Operationalizing Constructs

Example: Questionnaires are operationalizations of constructs (variables) expressed in hypotheses.



Rule of Thumb: Multi-item measures reduce error.

Convergent & Discriminant Validity



Factor analysis examines the ***convergence*** of operationally related indicators onto one construct (factor) = **convergent validity.**

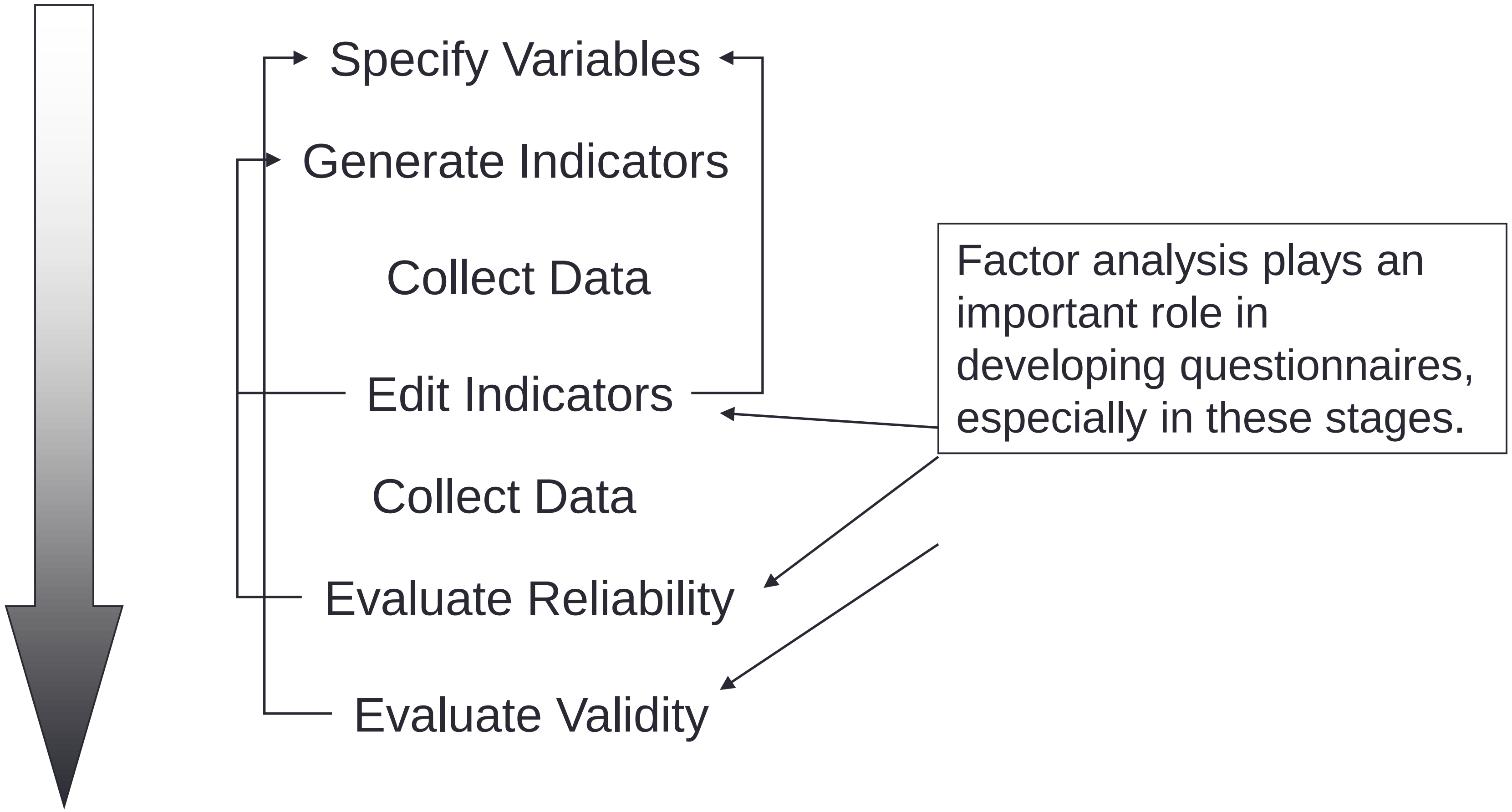
AND

The ***divergence*** of operationally unrelated indicators for different constructs = **discriminant validity.**

2. Questionnaire Design

- In market analysis, questionnaires are the most common way to collect data.
- They function to standardize the communication between the interviewer and respondent, thus aiding in quantifying the data.
- They are directly based on the research question, with the aim of providing the data to answer the question.

Developing Questionnaires



Generating Indicators

- Precisely define what is included and excluded in the domain of the construct.
- How was it measured before?
- Use a garbage-can approach to create a pool of measures:
  - Empirical studies.
  - Exploratory research.
- Evaluate the measures through logic, consultation with experts, and ***factor analysis***.

Definitions

A working alliance is the joining of a patient's reasonable side with a therapist's working or analyzing side (Gelso and Hayes 1998), and consists of: a) a collaborative nature, b) an affective bond, and c) the joint ability to agree on goals and tasks (Martin et al. 2000).

Theoretical Plane

A common facet of the scales is that they measure working alliance as the collaborative nature, affective bond, and joint ability to agree on goals and tasks (Martin et al. 2000).

Empirical Plane

Q1: The advisor and I agree on what will be discussed during the session. (agree)

Q2: After the session the advisor and I have the same understanding of how we will proceed so that I will get the help I need from the bank. (collab)

Q3: I feel respected and accepted by the advisor. (affect)

3. Designing a Factor Analysis

Two issues:

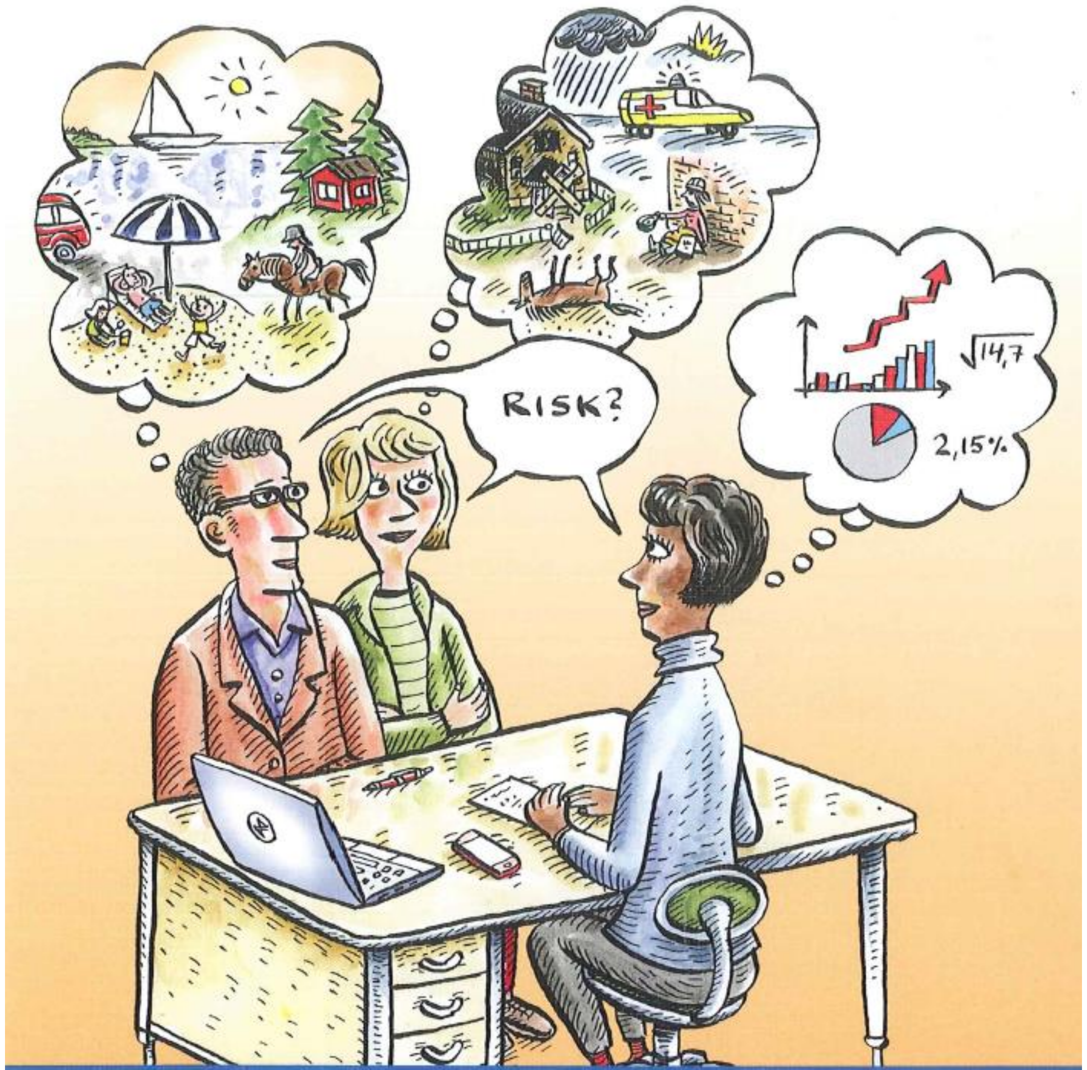
1. Choosing the variables.
   - Metric variables (using dummies is possible).
   - Minimize number of variables.
     - Too many makes it meaningless.
2. Sample size.
   - Minimum 100.
   - 5/1 (10/1) ratio of observations to variables.

Assumptions

- Some multicollinearity is good.
- Check the bivariate correlation matrix.
  - You want a ”substantial” number of significant correlations.
- Check the measure of sampling adequacy (KMO).
  - Minimum .5, but higher is better.
- Check the anti-image correlation matrix.
- Outliers can substantially affect results.
- Consider homogeneity:
  - e.g. Of male/female with regard to the structure.

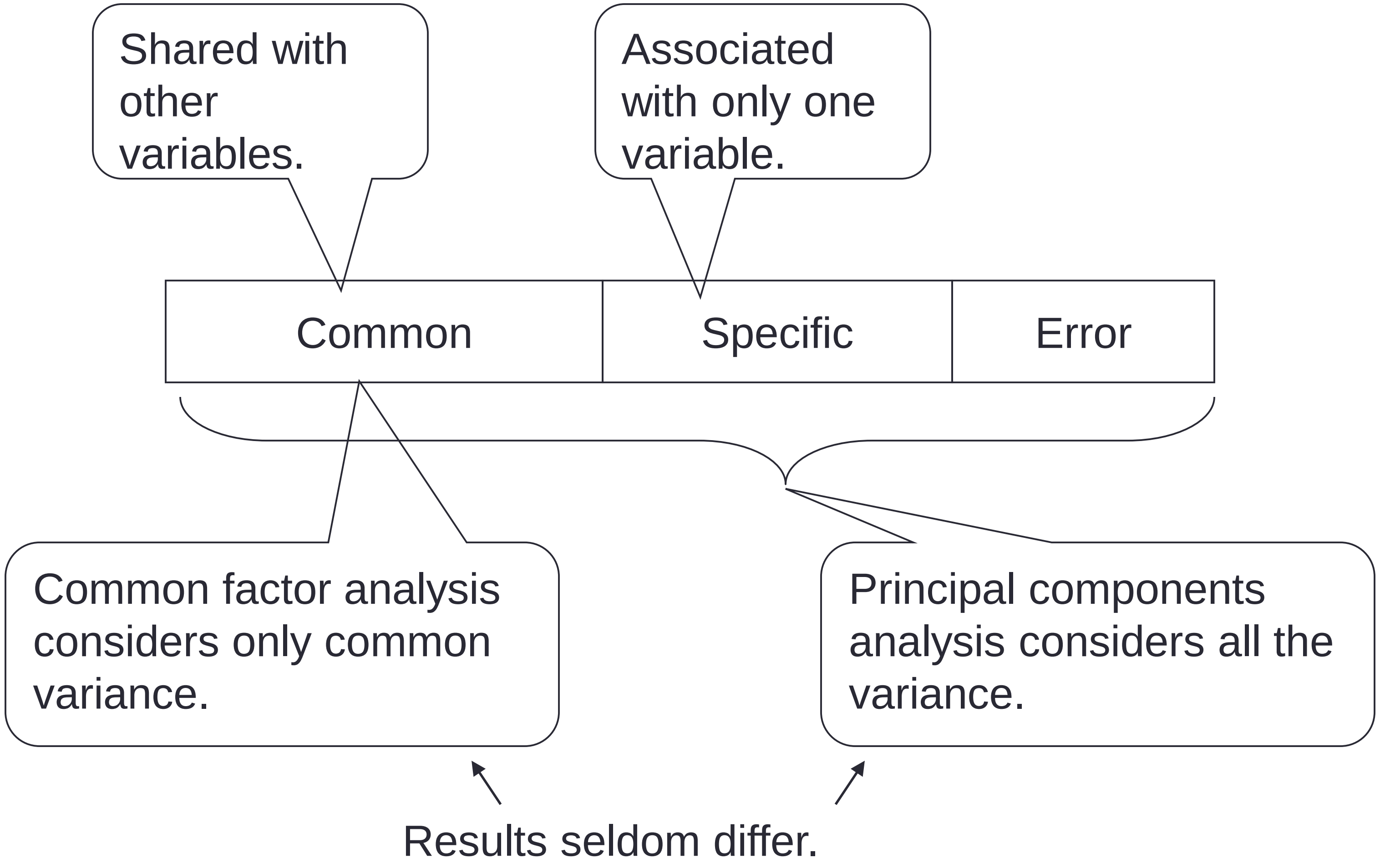
Conceptual Issue

- The sample should be homogeneous with regard to the underlying factor structure.
- E.g. You collect data on service encounters between the provider and customer. Do they share a common understanding (factor structure), or do they differ?
- If they differ they should not be combined into the same structure!



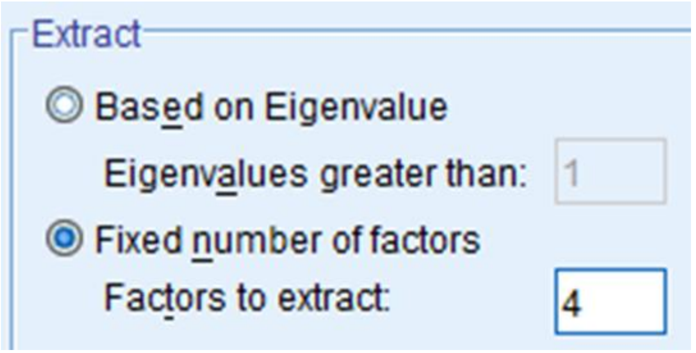
Financial advisory services

Variance


Shared with other variables.
Associated with only one variable.
Common
Specific
Error
Common factor analysis considers only common variance.
Principal components analysis considers all the variance.
Results seldom differ.

Choosing the Number of Factors

- Use the default in the program (SPSS), which is based on the eigen values.
  - The default cutoff is 1, but it can be useful to adjust it to ”force” the number of factors you want.
  - You do this based on theoretical logic and analysis of the eigen values.
  - E.g. If the next highest eigen value is .94, it is close enough to 1 to consider forcing the factor. If it is .54 then you are pushing pretty hard.



Interpreting the Factors

- Interpreting an unrotated factor matrix is difficult, if not impossible.
- Two basic types of rotation to choose from:
    1. Orthogonal: Axis maintained at 90 degrees.
    2. Oblique: Rotating the axis to best fit the data, which is not necessarily 90 degrees.
        - More flexible, but more controversial.
- Rule of Thumb: Try different methods.
    - The results are usually identical.

Factor thresholds

The larger the factor size, the greater its importance for interpreting the factor.

The square of the loading is its contribution to the explained variance of the factor.

Factor Loading	Sample Size
.30	350
.35	250
.40	200
.45	150
.50	120
.55	100
.60	85
.65	70
.70	60
.75	50

Practical Significance

± .3 - .4 is minimal.

± .5 is good.

> ± .7 is excellent.

Which Matrix to Look At?

Assuming that more than one factor is formed:

- Orthogonal (90% axis): Look at the ***rotated factor matrix.***
- Oblique: Look at the ***pattern matrix.***

- Consider each variable and which factor it loads highest on.

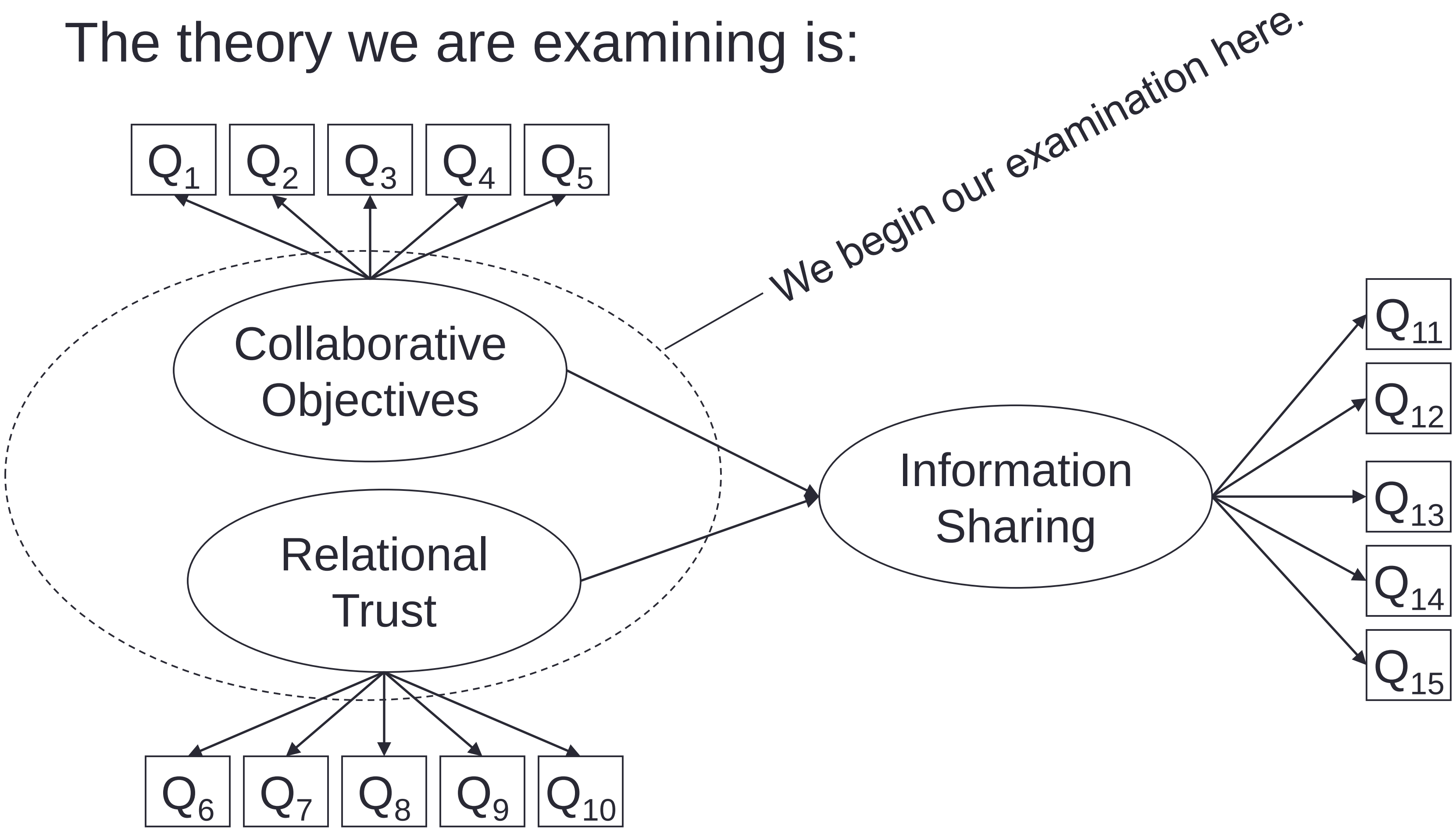
Communalities

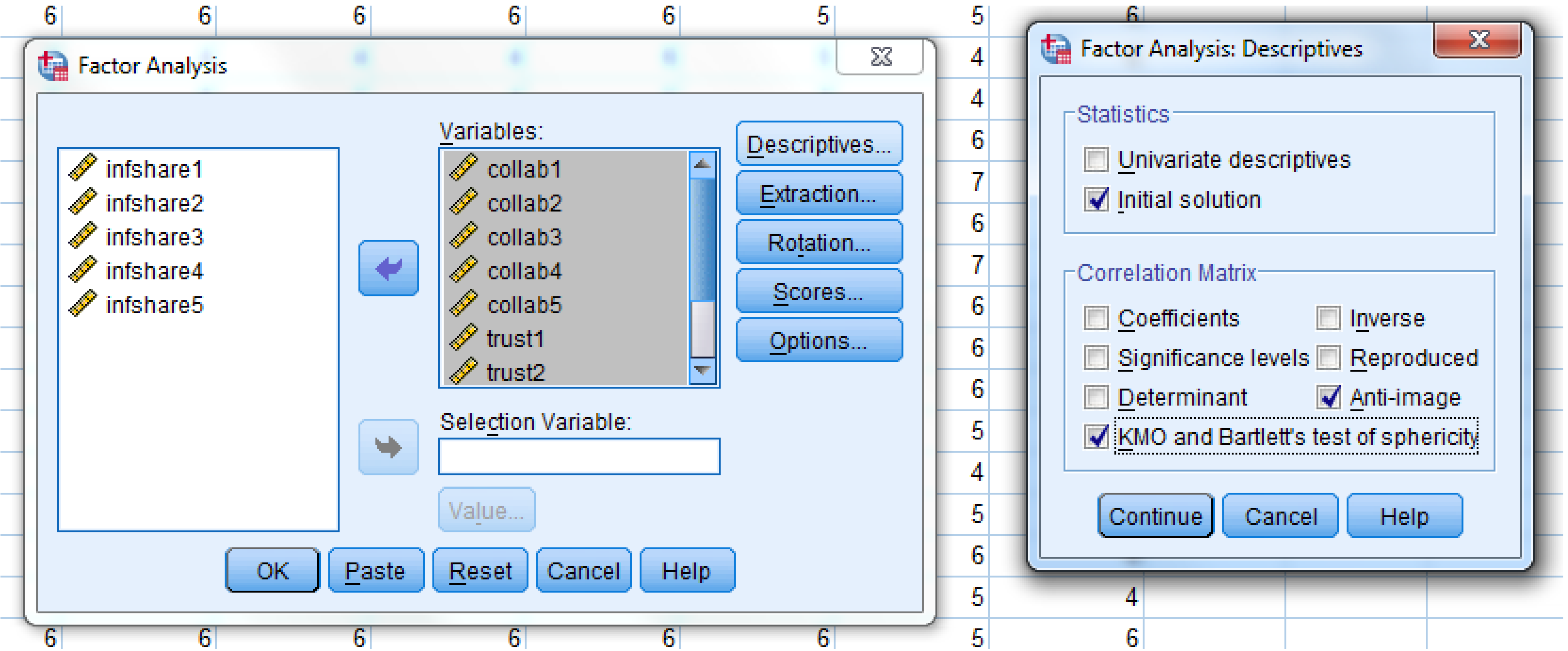
- Communalities represent the explained variance of the factor solution for each variable.
  - Are they too low (No specific value is given, but Hair et al. Mention .5)?
    - Either ignore the variable and interpret the solution,
    - Or, delete the variable and rerun the analysis.

  - Generally, I don’t like including ”useless” variables because they affect the results.

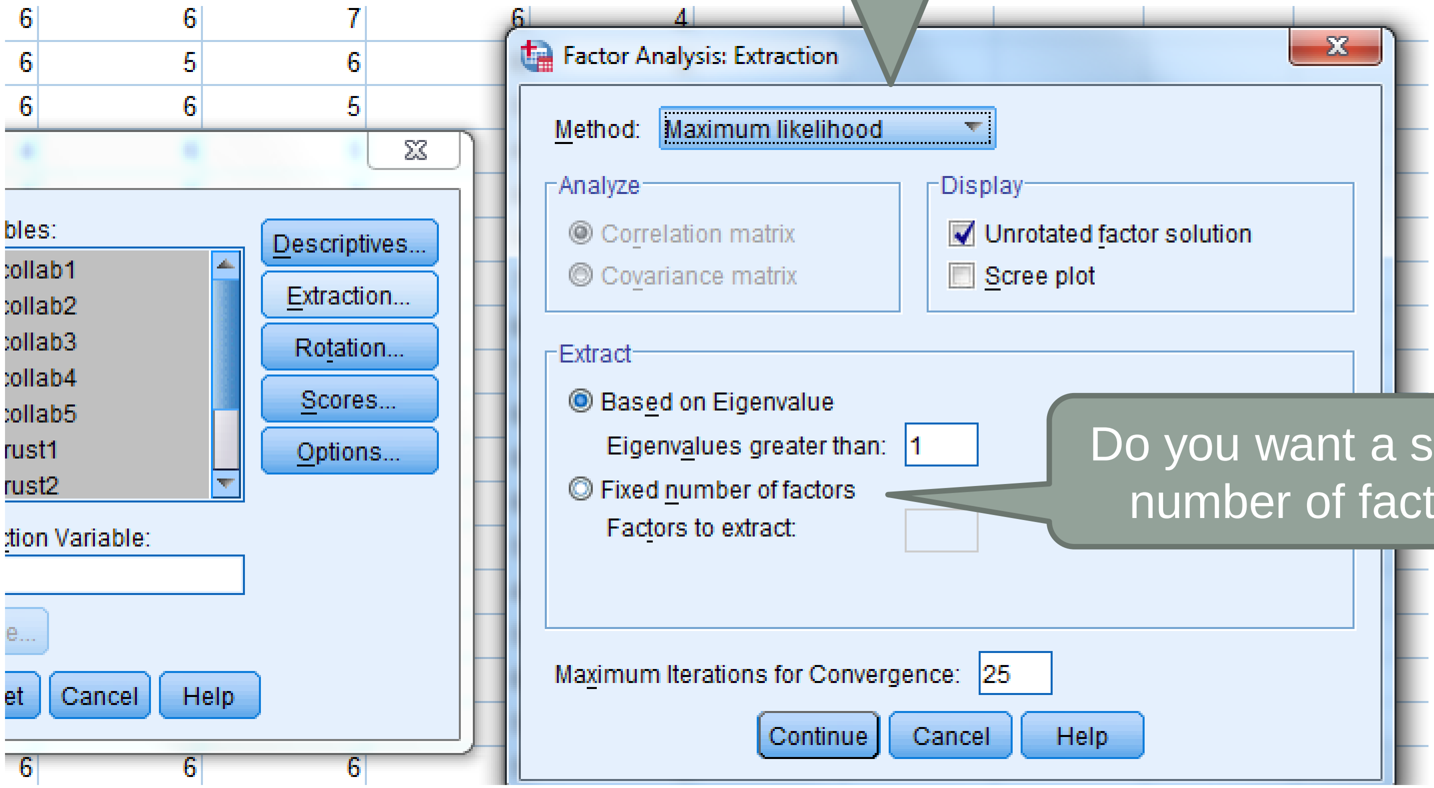
4. An Example

The theory we are examining is:

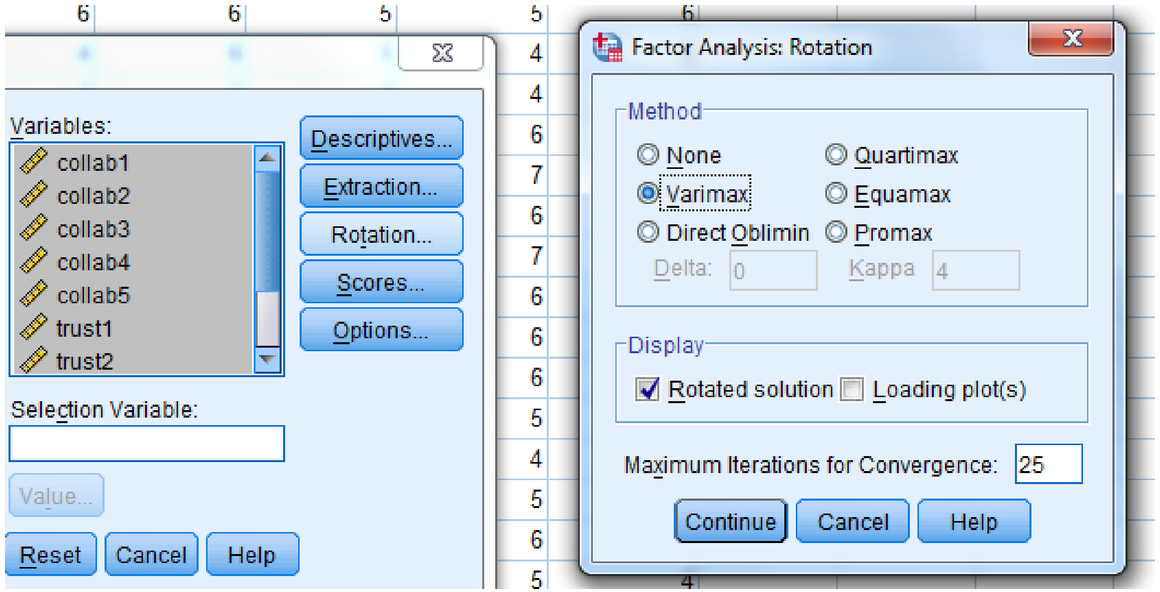


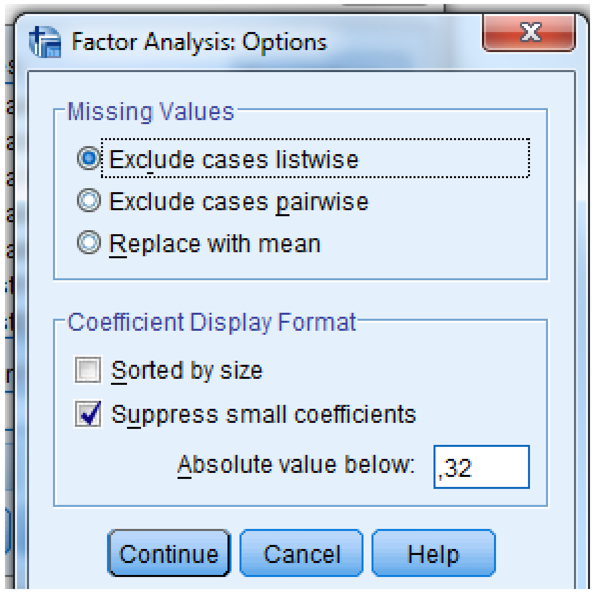

Factor Analysis
infshare1
infshare2
infshare3
infshare4
infshare5
Variables:
collab1
collab2
collab3
collab4
collab5
trust1
trust2
Descriptives...
Extraction...
Rotation...
Scores...
Options...
Selection Variable:
Value...
OK
Paste
Reset
Cancel
Help
Factor Analysis: Descriptives
Statistics
Univariate descriptives
Initial solution
Correlation Matrix
Coefficients
Inverse
Significance levels
Reproduced
Determinant
Anti-image
KMO and Bartlett's test of sphericity
Continue
Cancel
Help

Principal Components or Maximum Likelihood?



Do you want a specific number of factors?





Depends on sample size

Correlation Matrix

Correlations

	CO	CO	CO	CO	CO	RT	RT	RT	RT	RT
CO	1.0									
CO	.779	1.0								
CO	.718	.775	1.0							
CO	.721	.709	.793	1.0						
CO	.746	.735	.706	.769	1.0					
RT	.482	.469	.389	.433	.408	1.0				
RT	.472	.483	.403	.515	.435	.681	1.0			
RT	.501	.443	.373	.509	.420	.661	.747	1.0		
RT	.476	.461	.426	.541	.431	.676	.751	.763	1.0	
RT	.486	.442	.432	.534	.441	.646	.701	.708	.758	1.0

All correlations are significant at the .01 level.

KMO

You want this above .5.

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.920
Bartlett's Test of Sphericity	Approx. Chi-Square	2486.581
	df	45
	Sig.	.000

>.9 Marvelous
.8+ Meritorious
.7+ Middling
.6+ Mediocre
.5+ Miserable
<.5 Unacceptable

Anti-Image Matrix (correlations)

Anti-image Matrices

	CO	CO	CO	CO	CO	RT	RT	RT	RT	RT
CO	.931[a]									
CO	-.345	.906[a]								
CO	-.118	-.383	.886[a]							
CO	-.097	.033	-.459	.897[a]						
CO	-.263	-.218	-.011	-.370	.923[a]					
RT	-.087	-.084	-.024	.132	-.019	.947[a]				
RT	.067	-.125	.061	-.068	-.004	-.218	.931[a]			
RT	-.153	.020	.126	-.087	.034	-.139	-.284	.922[a]		
RT	.049	-.009	-.015	-.111	.036	-.169	-.230	-.293	.920[a]	
RT	-.039	.084	-.052	-.073	-.018	-.153	-.153	-.162	-.320	.941[a]

a. Measures of Sampling Adequacy(MSA)

Look at the diagonal for values under .5. When KMO is low this is a good way to weed out troublesome variables.

Communalities

Communalities

	Initial	Extraction
CO	.708	.733
CO	.731	.757
CO	.733	.768
CO	.756	.764
CO	.690	.724
RT	.574	.597
RT	.680	.732
RT	.688	.745
RT	.721	.787
RT	.652	.697

Extraction Method: Maximum Likelihood.

Rule of Thumb: >.5

Some authors suggest looking at communalities **after** you have decided on the number of factors. They also disagree as to a cut-off. Simply beware of small communalities!

Eigen values & Explained Variance

Total Variance Explained

Factor	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total
1	6.196	61.961	61.961	5.926	59.257	59.257	5.079
2	1.640	16.400	78.361	1.377	13.773	73.030	5.016
3	.429	4.293	82.654				
4	.332	3.323	85.977				
5	.304	3.041	89.018				
6	.284	2.836	91.854				
7	.254	2.541	94.395				
8	.219	2.187	96.582				
9	.194	1.937	98.519				
10	.148	1.481	100.000				

Extraction Method: Maximum Likelihood.

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

Rotated Solution

Suppressed values under .32

Rotated Factor Matrix[a]

	Factor	
	1	2
collab1	,790	,330
collab2	,821	,288
collab3	,850	,215
collab4	,794	,365
collab5	,811	,257
trust1	,278	,721
trust2	,279	,809
trust3	,256	,824
trust4	,274	,844
trust5	,295	,781

Extraction Method: Maximum Likelihood.
Rotation Method: Varimax with Kaiser Normalization.

Rotated Factor Matrix[a]

	Factor	
	1	2
collab1	,790	,330
collab2	,821	
collab3	,850	
collab4	,794	,365
collab5	,811	
trust1		,721
trust2		,809
trust3		,824
trust4		,844
trust5		,781

Extraction Method: Maximum Likelihood.
Rotation Method: Varimax with Kaiser Normalization.

Principal Components (rotated)

Rotated Compon… trix[a]

Two principal components

	Component	
	1	2
collab1	,827	,320
collab2	,853	,278
collab3	,880	,200
collab4	,824	,350
collab5	,855	,238
trust1	,258	,789
trust2	,268	,845
trust3	,248	,854
trust4	,268	,863
trust5	,284	,821

Component scores

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

a. Rotation converged in 3 iterations.

Assessment – we have construct validity

- Sample size – 315 – good.
- Correlations – good.
- KMO - .920 – good.
- Anti-image – all over .8 – good.
- Communalities – lowest near .6 – good.
- Cumulative explained variance – 78% - good.
- Two factors – good.
- All variables load significantly on the correct factor – good.
- No significant cross-loadings – good.