Validity

Nils Myszkowski, PhD

Outline

Structural validity
Criterion validity
Diagnostic validity
Review

Structural validity
Criterion validity
Diagnostic validity
Review

Structural validity
Criterion validity
Diagnostic validity
Review

Aims of this session

- At the end of this course, you should be able to:
  - Define validity in its multiple forms
  - Understand the importance of test validity
  - Evaluate the validity of a test from empirical results

Structural validity
Criterion validity
Diagnostic validity
Review

Definitions

- Common general definitions of validity notably include these:
  - "Validity is the degree to which all the accumulated evidence supports the intended interpretation of test scores for the proposed purpose." (APA, 1999)
  - Importance of the relationship between what the test results actually mean (how they can be interpreted) and the purpose of the test.

Structural validity
Criterion validity
Diagnostic validity
Review

Definitions

- Common general definitions of validity notably include these:
  - "Validity is an integrated evaluative judgment of the degree to which empirical evidence and the theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores or other modes of assessment" (Messick, 1989)
  - Importance of the use of test scores and the appropriateness of basing them on the scores of a specific test

Structural validity
Criterion validity
Diagnostic validity
Review

Definitions

- Common general definitions of validity notably include these:
  - "Validity is the extent to which a test measures what it purports to measure." (National Association of the Directors of Educational Research, 1995)
  - At the heart of validity is how well measures do in "doing their job" of measuring the construct.

Outline

Structural validity
6
Theoretical structures
Observed structures
Individuals and structures
Criterion validity
Diagnostic validity
Review

Structural validity

Criterion validity

Diagnostic validity

Review

Structural validity

Structural validity 7
Theoretical structures
Observed structures
Individuals and structures
Criterion validity
Diagnostic validity
Review

Factor structure

- Structural validity designates the adequateness of the network of relationships between the observable elements (measured facets) of the measure of a construct.
- In other words, structural validity will be verified by examining the relationships between the measured "parts" of our construct.

Structural validity

Structural validity 8
Theoretical structures
Observed structures
Individuals and structures
Criterion validity
Diagnostic validity
Review

Factor structure

- As a consequence, we are interested in the structure of the data, and in the extent to which it corresponds to the theoretical structure of the construct.
- We will thus compare theoretical expectations with the actual **structure** of the data.

Structural validity

Structural validity
Theoretical structures
Observed structures
Individuals and structures
Criterion validity
Diagnostic validity
Review

Structural validity

- Theoretical structures
- Observed structures
- Individuals and structures

Structural validity
Theoretical structures
Observed structures
Individuals and structures
Criterion validity
Diagnostic validity
Review

Typical structures

- In this part we will review the most typical theoretical test structures, but this list is not exhaustive.
- We will see test structures that are :
  - Unidimensional
  - Multidimensional with independent factors
  - Multidimensional hierarchical
  - Complex structures

Structural validity
Theoretical structures 11
Observed structures
Individuals and structures
Criterion validity
Diagnostic validity
Review

Unidimensional structures

- Unidimensional structures refer to measures of theoretically unidimensional constructs.
- In other terms, such structures are expected of measures of constructs that do not have different facets.
- Ex: The Rosenberg Self-Esteem Scale (RSES), The Perceived Stress Scale (PSS), The Standard Progressive Matrices (SPM)

Structural validity
Theoretical structures

Structural validity
Theoretical structures
Observed structures
Individuals and structures
Criterion validity
Diagnostic validity
Review

Statement	Strongly Agree	Agree	Disagree	Strongly Disagree
1. I feel that I am a person of worth, at least on an equal plan with others.				
2. I feel that I have a number of good qualities.				
3. All in all, I am inclined to feel that I am a failure.				
4. I am able to do things as well as most other people				
5. I feel I do not have much to be proud of.				
6. I take a positive attitude toward myself.				
7. On the whole, I am satisfied with myself.				
8. I wish I could have more respect for myself				
9. I certainly feel useless at times.				
10. At times, I think I am no good at all.				

Figure: The Rosenberg Self-Esteem Scale (Rosenberg, 1989) is a theoretically unidimensional scale.

Structural validity

Theoretical structures

Structural validity
Theoretical structures
13
Observed structures
Individuals and structures
Criterion validity
Diagnostic validity
Review

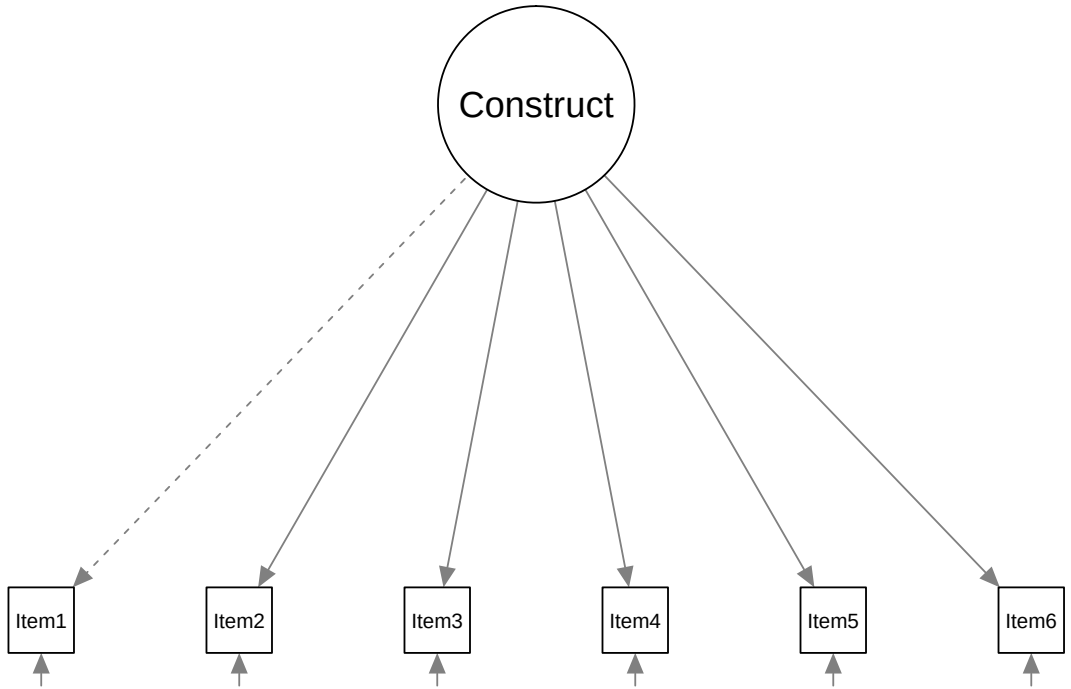


Figure: A path diagram representation of a unidimensional structure

Structural validity
Theoretical structures
Observed structures
Individuals and structures
Criterion validity
Diagnostic validity
Review

Multidimensional structures with independent factors

- Multidimensional structures with independent factors refer to measures of theoretically multidimensional constructs, for which each trait or facet is independent from the others.
- Ex: The Big Five Inventory (BFI)

Structural validity

Theoretical structures

Structural validity
Theoretical structures
Observed structures
Individuals and structures
Criterion validity
Diagnostic validity
Review

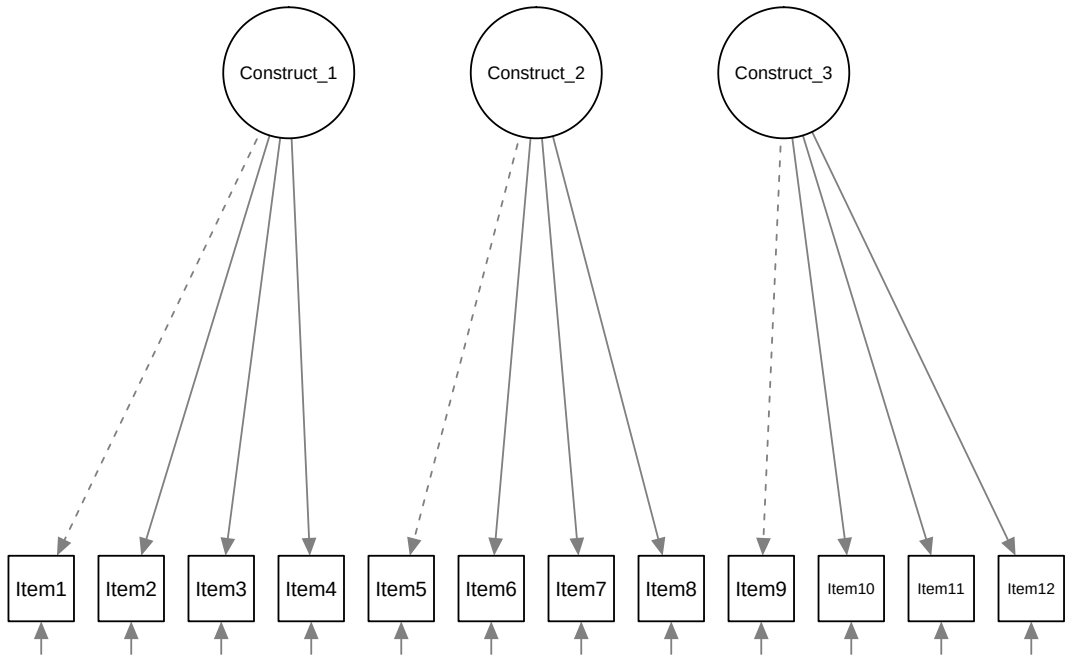


Figure: A path diagram representation of a multidimensional structure with independent factors

Structural validity
Theoretical structures

Structural validity
Theoretical structures
Observed structures
Individuals and structures
Criterion validity
Diagnostic validity
Review

Multidimensional structures with correlated factors

- Some instruments are conceptualized has having multidimensional structures where the different constructs measured are correlated, but not explained by a general factor.
- Ex: The Primary Mental Abilities test (PMA), Domain-Specific creativity measures

Structural validity

Theoretical structures

Structural validity
Theoretical structures
Observed structures
Individuals and structures
Criterion validity
Diagnostic validity
Review

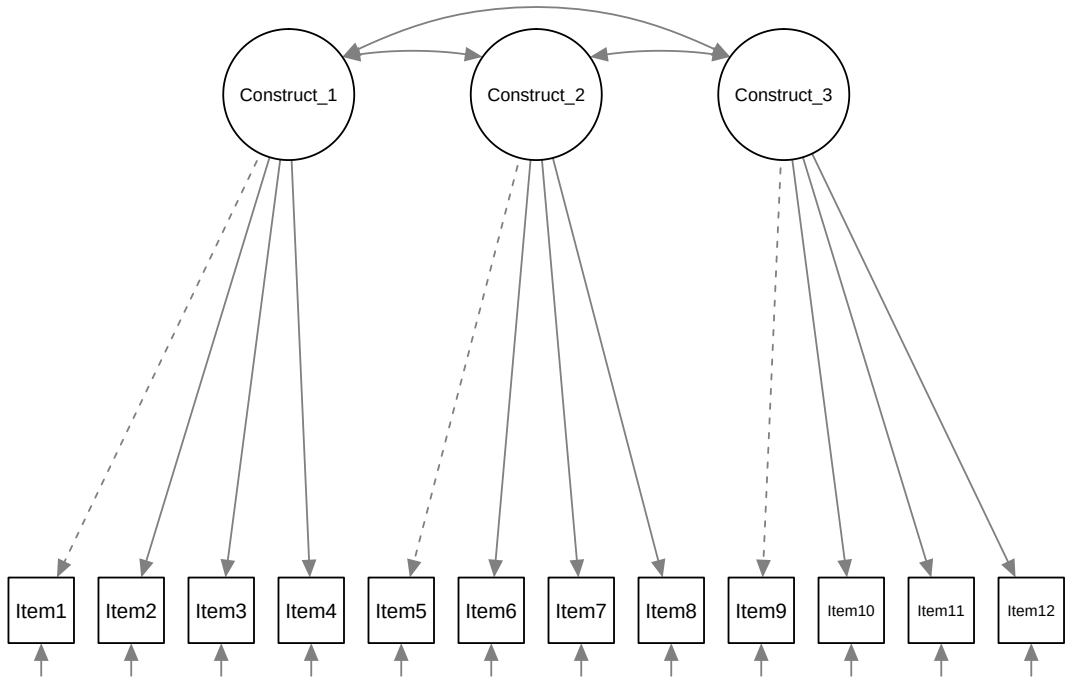


Figure: A path diagram representation of a multidimensional structure with correlated factors

Structural validity
Theoretical structures
Observed structures
Individuals and structures
Criterion validity
Diagnostic validity
Review

Multidimensional hierarchical structures

- Multidimensional hierarchical structures refer to measures of theoretically multidimensional constructs, for which each trait or facet is explained by a general (sometimes called "G", or second order) factor.
- Ex: The Revised Self-Monitoring Scale (RSMS)

Structural validity

Theoretical structures

Structural validity
Theoretical structures
19
Observed structures
Individuals and structures
Criterion validity
Diagnostic validity
Review

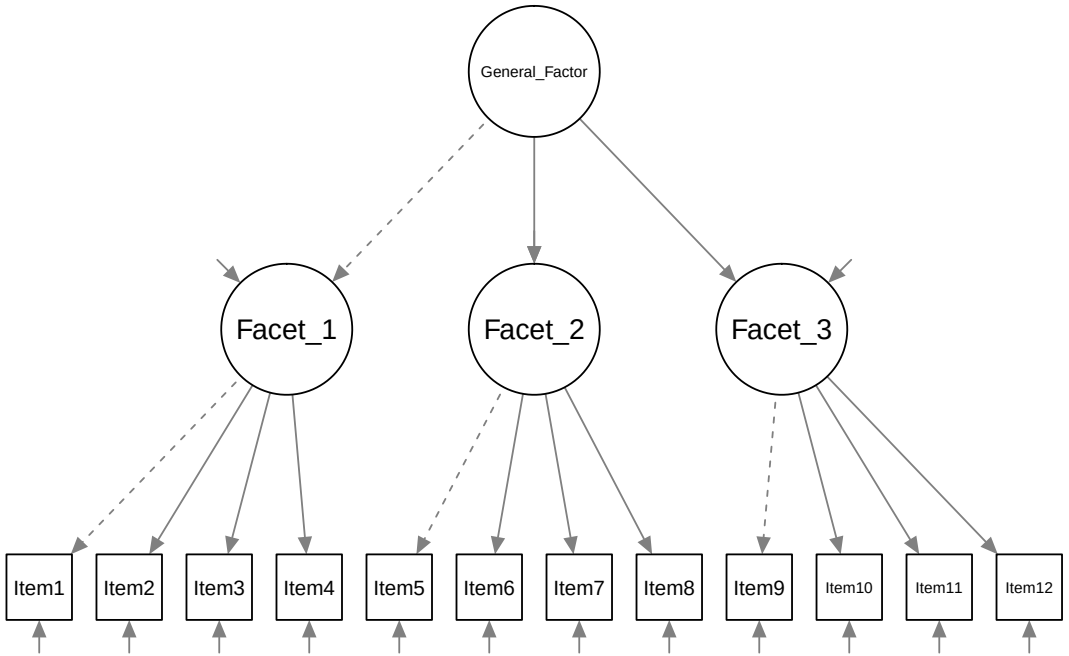


Figure: A path diagram representation of a hierarchical structure

Structural validity
Theoretical structures 20
Observed structures
Individuals and structures
Criterion validity
Diagnostic validity
Review

Multidimensional bifactor structures

- Another approach to multidimensional measures that also integrate a general components is the bifactor model (Holzinger & Swineford, 1937).
- In bifactor models, as opposed to hierarchical models, the general factor influences all items, while specific factors influence some of the items.

Structural validity
Theoretical structures
Observed structures
Individuals and structures
Criterion validity
Diagnostic validity
Review

Multidimensional bifactor structures

- For example, when answering a math problem that requires knowledge in derivativesfor exercise 1 and integrals for exercise 2, math influences solving the questions of both exercise 1 and 2, while specific knowledge of derivatives (independent from math ability) influences solving exercise 1, and knowledge of integrals (independent from math ability) influences solving exercise 2.
- Ex: It has been discussed as an efficient approach in domains where we want to measure something while controlling domain effects (for example, if one wants to control for specificities of subtests).

Structural validity

Theoretical structures

Structural validity
Theoretical structures
22
Observed structures
Individuals and structures
Criterion validity
Diagnostic validity
Review

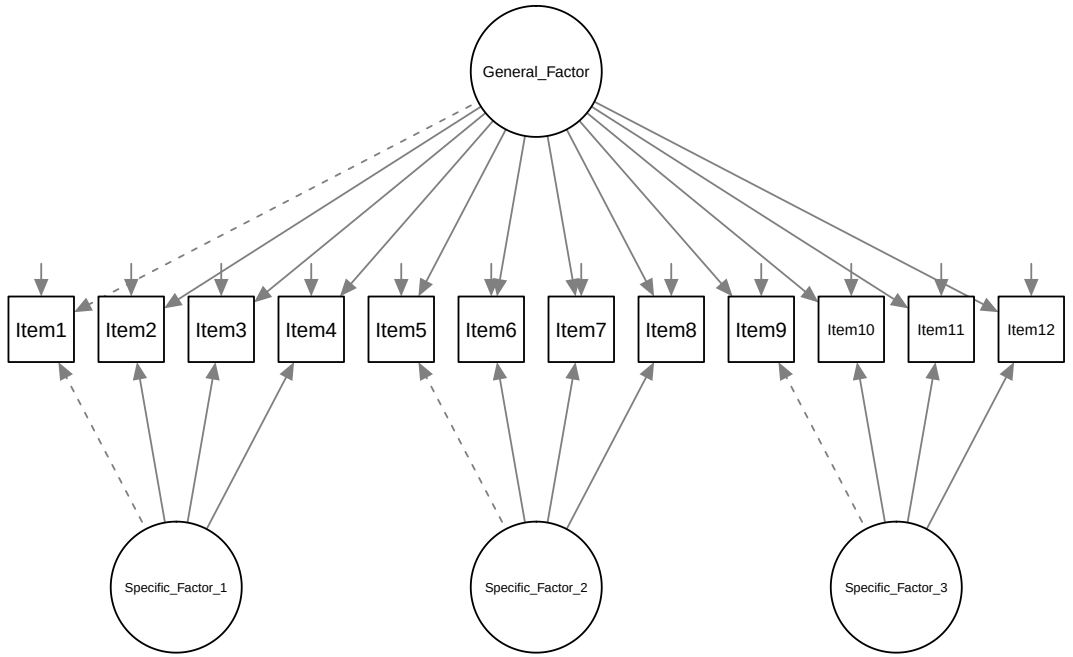


Figure: A path diagram representation of a bifactor structure

Structural validity
Theoretical structures 23
Observed structures
Individuals and structures
Criterion validity
Diagnostic validity
Review

Complex structures

- The structures presented here are typical but do not represent every factor structure possible.
- For example, some structures are a mix of hierarchical structures and independent factor structures.
- Ex: The NEO-PI, in which the five independent traits are measured through facets. In this inventory, each trait has a hierarchical organization but is independent from the others. You can say that you have independent hierarchical traits.

Structural validity

Structural validity
Theoretical structures
Observed structures
Individuals and structures
Criterion validity
Diagnostic validity
Review

Structural validity

- Theoretical structures
- Observed structures
- Individuals and structures

Structural validity
Theoretical structures
Observed structures
Individuals and structures
Criterion validity
Diagnostic validity
Review

Observed structures

- Achieving adequate construct validity implies the observation of an empirical factor structure that is (to some extent) identical to the theoretical factor structure.
- In this section, we will review the techniques that are used to "match" empirical factor structures and theoretical structures:
  - Direct correlation matrix observation
  - Exploratory Factor Analysis (EFA)
  - Confirmatory Factor Analysis (CFA) (through Structural Equation Modelling or Item Response Theory analysis)

Structural validity
Observed structures

Structural validity
Theoretical structures
Observed structures 26
Individuals and structures
Criterion validity
Diagnostic validity
Review

Correlation matrix observation

- This technique implies to simply compare observed correlations to expected correlations between the dimensions that are measured by a test.
- To do so, the correlation matrix is usually reorganized in **clusters**.
- Reorganizing in clusters means simply regrouping the variables that are theoretically (or, in the absence of theory, empirically) correlated.

Structural validity

Observed structures

Structural validity
Theoretical structures
Observed structures 27
Individuals and structures
Criterion validity
Diagnostic validity
Review

Correlation Matrix

		cog1	cog2	cog3	cog4	cog5	aff1	aff2	aff3	aff4	aff5
Correlation	cog1	1.000	.571	.666	.149	.692	.041	.088	.097	.051	.036
	cog2	.571	1.000	.483	.124	.545	.073	.080	.116	.070	-.066
	cog3	.666	.483	1.000	.127	.606	.021	.017	.116	-.006	-.068
	cog4	.149	.124	.127	1.000	.162	-.030	-.045	.005	.026	-.035
	cog5	.692	.545	.606	.162	1.000	.053	.023	.103	.110	-.021
	aff1	.041	.073	.021	-.030	.053	1.000	.578	.632	.587	.242
	aff2	.088	.080	.017	-.045	.023	.578	1.000	.584	.534	.167
	aff3	.097	.116	.116	.005	.103	.632	.584	1.000	.620	.218
	aff4	.051	.070	-.006	.026	.110	.587	.534	.620	1.000	.257
	aff5	.036	-.066	-.068	-.035	-.021	.242	.167	.218	.257	1.000

Figure: Correlation matrix in SPSS

Structural validity

Observed structures

Structural validity
Theoretical structures
Observed structures 28
Individuals and structures
Criterion validity
Diagnostic validity
Review

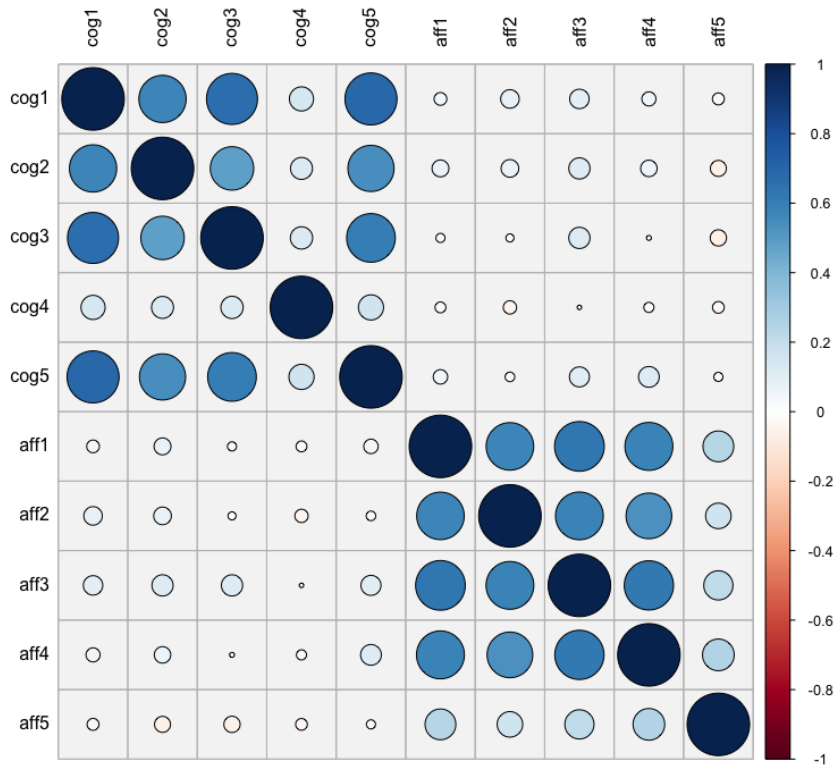


Figure: Correlation Plot (not in SPSS)

Structural validity
Theoretical structures
Observed structures
Individuals and structures
Criterion validity
Diagnostic validity
Review

Exploratory Factor Analysis (EFA)

- Exploratory Factor Analysis (EFA) is a statistical technique which consists in finding a small set of latent factors that explain the variance of the initial (observed) scores.
- The goal is to arrive at a parsimonious representation of the associations between the items.
- In other words, it is a dimensionality reduction technique.
- It extracts a limited set of factors that explain a specific percentage of the variance of the item scores.
- These factors are then "rotated" to be interpreted by the psychometric researcher.

Structural validity
Theoretical structures
Observed structures
Individuals and structures
Criterion validity
Diagnostic validity
Review

Exploratory Factor Analysis (EFA)

- The main methods of factor extraction are:
  - Principal Components Analysis (PC or PCA, invented in 1901 by Pearson): Aims at explaining a maximum of the total variance in the item scores - notably used for data reduction purposes, but should be avoided for factor structure investigation purposes. PCA is often not considered a factor analysis method.
  - Maximum Likelihood Estimation (ML or MLE): Maximizes the likelihood of the observed (co)variances of the items - has many advantages (measures of model fit, significance testing of loadings) but has strong assumptions, notably normality.
  - Principal Axis Factoring (PAF): The first factor accounts for as much shared variance as possible, then the second factor next most variance, and so on - notably used when the assumption of normality is violated.

Structural validity
Theoretical structures
Observed structures
Individuals and structures
Criterion validity
Diagnostic validity
Review

Exploratory Factor Analysis (EFA)

- EFA extracts as many factors as there are items.
- If we keep and use all of them, we will not have reduced dimensionality. So we need to choose to keep only the ones that are the most useful.
- The usefulness of each factor is determined by its **eigenvalue**. Eigenvalues are the total variance explained by the factor (for example, if we have 10 items with a variance of 1, the total variance to explain is 10).
- Alternatively, we can also use the percentage of variance explained by each factor (which is nothing but a factor's eigenvalue divided by the total variance to explain).
- We typically present factors from highest eigenvalue to the lowest.
- A **scree plot** is a graphical representation of the eigenvalues per factor.

Structural validity

Observed structures

Structural validity
Theoretical structures
Observed structures
Individuals and structures
Criterion validity
Diagnostic validity
Review

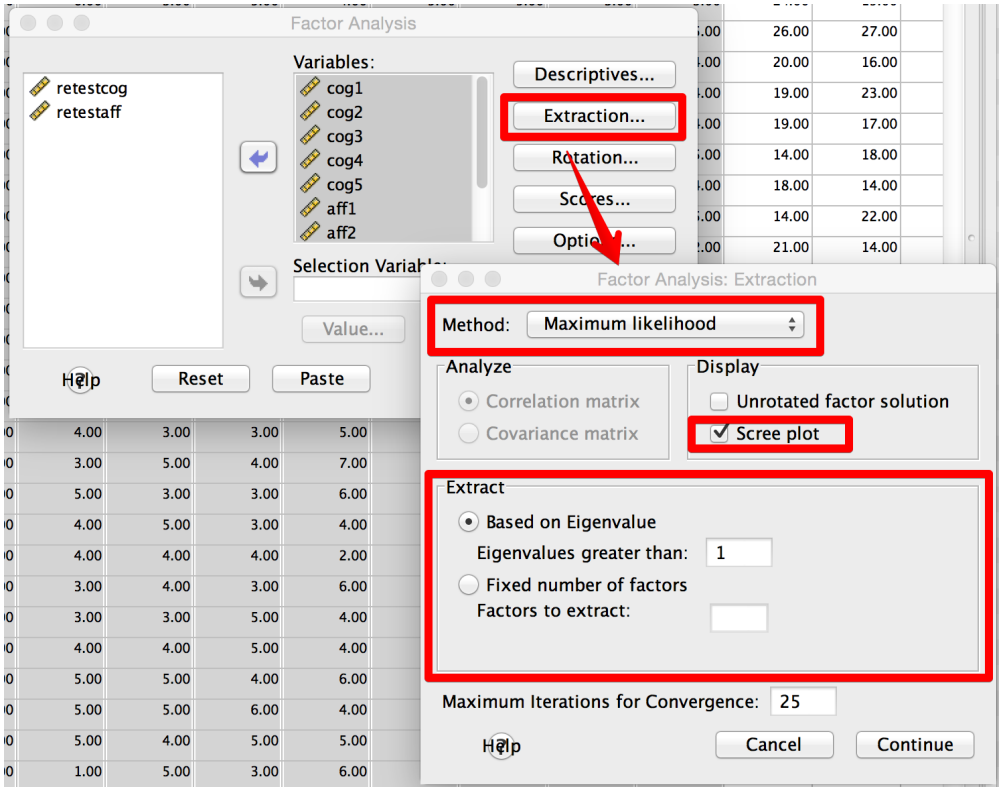


Figure: Choosing extraction method

Structural validity

Observed structures

Structural validity
Theoretical structures
Observed structures 33
Individuals and structures
Criterion validity
Diagnostic validity
Review

Total Variance Explained

Factor	Initial Eigenvalues			Rotation Sums of Squared Loadings[a]
	Total	% of Variance	Cumulative %	Total
1	3.096	30.960	30.960	2.457
2	2.625	26.245	57.205	2.461
3	.962	9.615	66.821	
4	.911	9.107	75.928	
5	.533	5.332	81.260	
6	.486	4.860	86.119	
7	.428	4.276	90.395	
8	.380	3.799	94.194	
9	.315	3.145	97.339	
10	.266	2.661	100.000	

Extraction Method: Maximum Likelihood.

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

Figure: Eigenvalues

Structural validity
Observed structures

Structural validity
Theoretical structures
Observed structures
Individuals and structures
Criterion validity
Diagnostic validity
Review

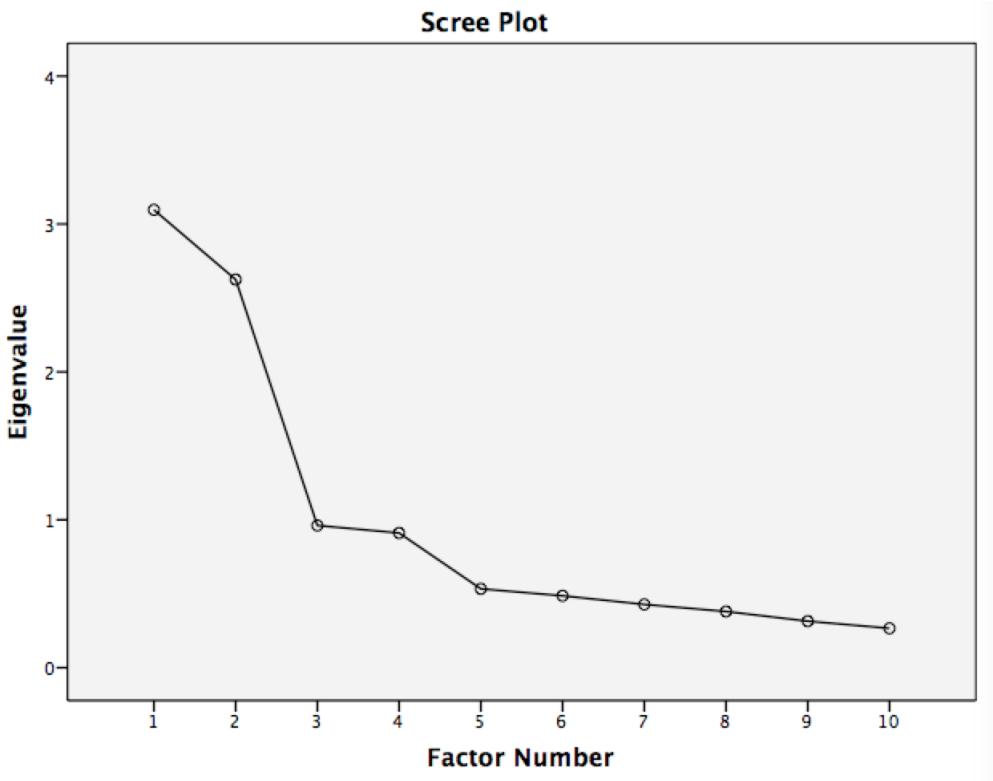


Figure: Scree plot in SPSS

Structural validity
Theoretical structures
Observed structures
Individuals and structures
Criterion validity
Diagnostic validity
Review

Exploratory Factor Analysis (EFA)

- The main criteria to "cut" between the retained and unretained factors are:
  - Kaiser's (1960) K1 : We keep factors with eigenvalues of at least 1. Avoid : Depends a lot on total variance of the items, often leads to "overfactoring", keeping too many factors.
  - Cattell's (1966) Scree test : A substantial eigenvalue drop (with an "elbow") is observed in most scree plots, we keep the factors on the left of that drop. Often used but quite subjective: There are automatized reproducible methods (Acceleration Factor af, Optimal Coordinates oc) that outperform the K1 criterion.

Structural validity
Theoretical structures
Observed structures
Individuals and structures
Criterion validity
Diagnostic validity
Review

Exploratory Factor Analysis (EFA)

- The main criteria to "cut" between the retained and unretained factors are:
  - Horn's (1965) Parallel Analysis (PA) : It conceptually consists in rejecting spurious factors. We compute eigenvalues for a randomly generated correlation matrix (a lot of times), then keep factors with eigenvalues above those found on average in the random simulations. Most recommended, though not in SPSS (directly).
  - Interpretability : Do the items that load on the same factor share some conceptual meaning ? (see next part: Interpretation)

Structural validity
Observed structures

Structural validity
Theoretical structures
Observed structures 37
Individuals and structures
Criterion validity
Diagnostic validity
Review

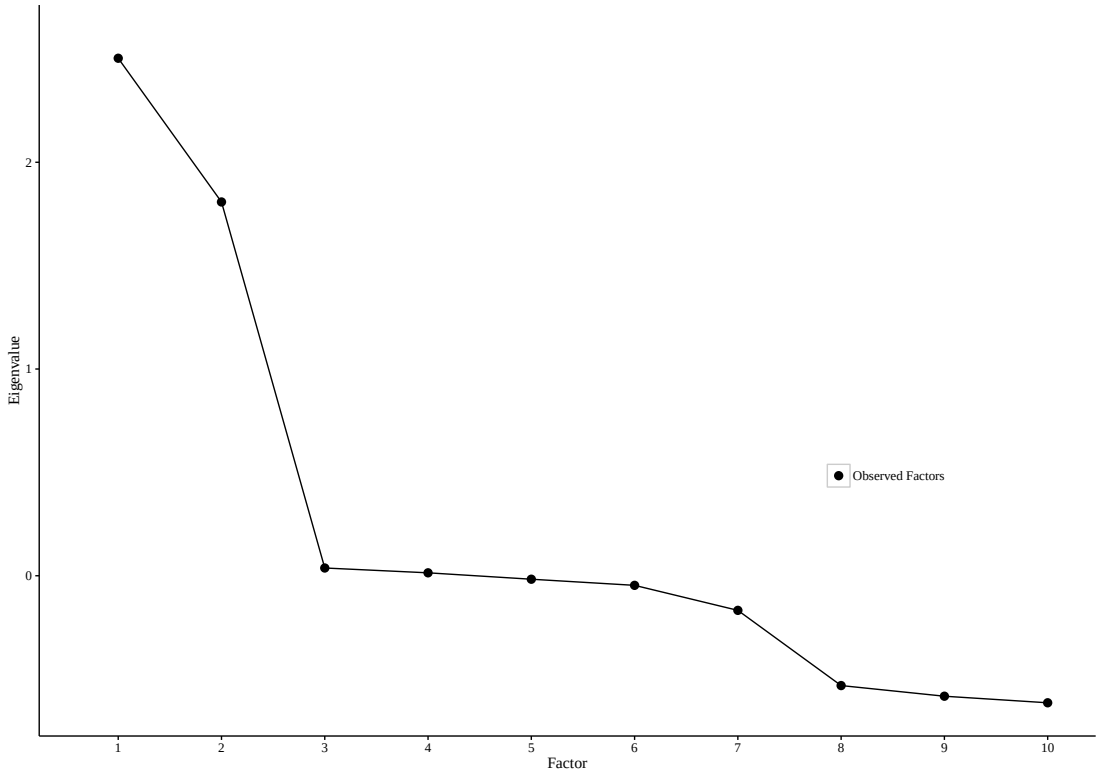


Figure: Scree plot

Structural validity
Observed structures

Structural validity
Theoretical structures
Observed structures
Individuals and structures
Criterion validity
Diagnostic validity
Review

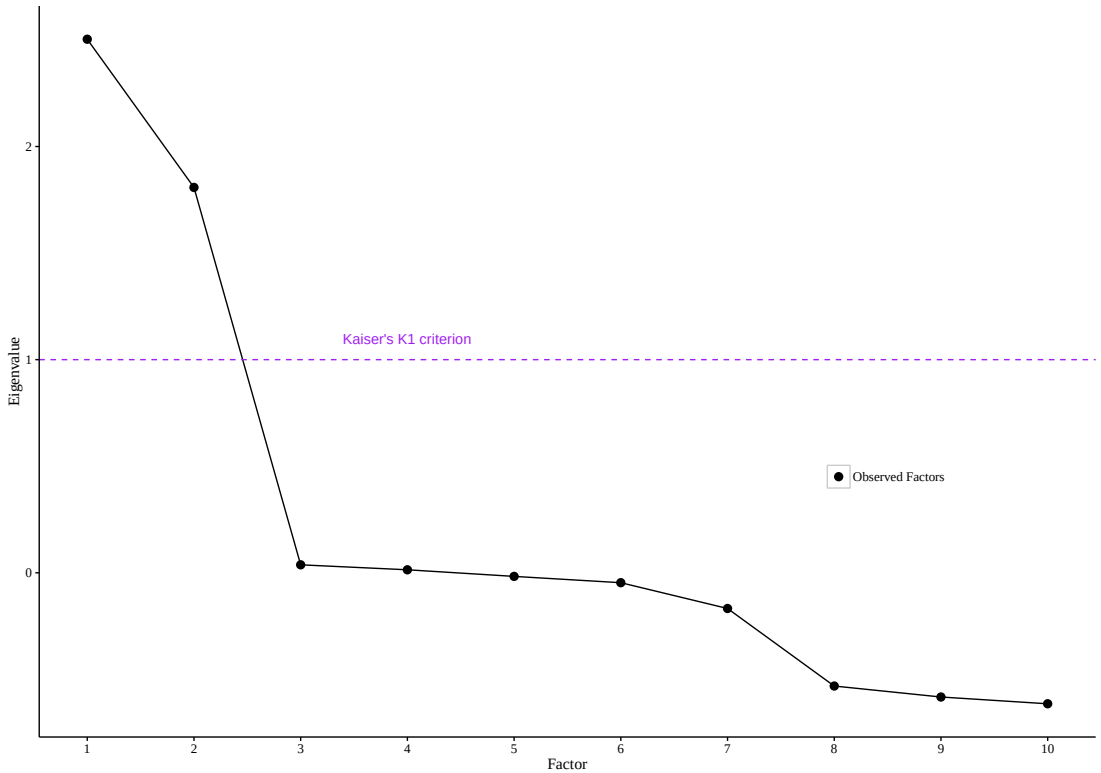


Figure: Scree plot

Structural validity
Observed structures

Structural validity
Theoretical structures
Observed structures
Individuals and structures
Criterion validity
Diagnostic validity
Review

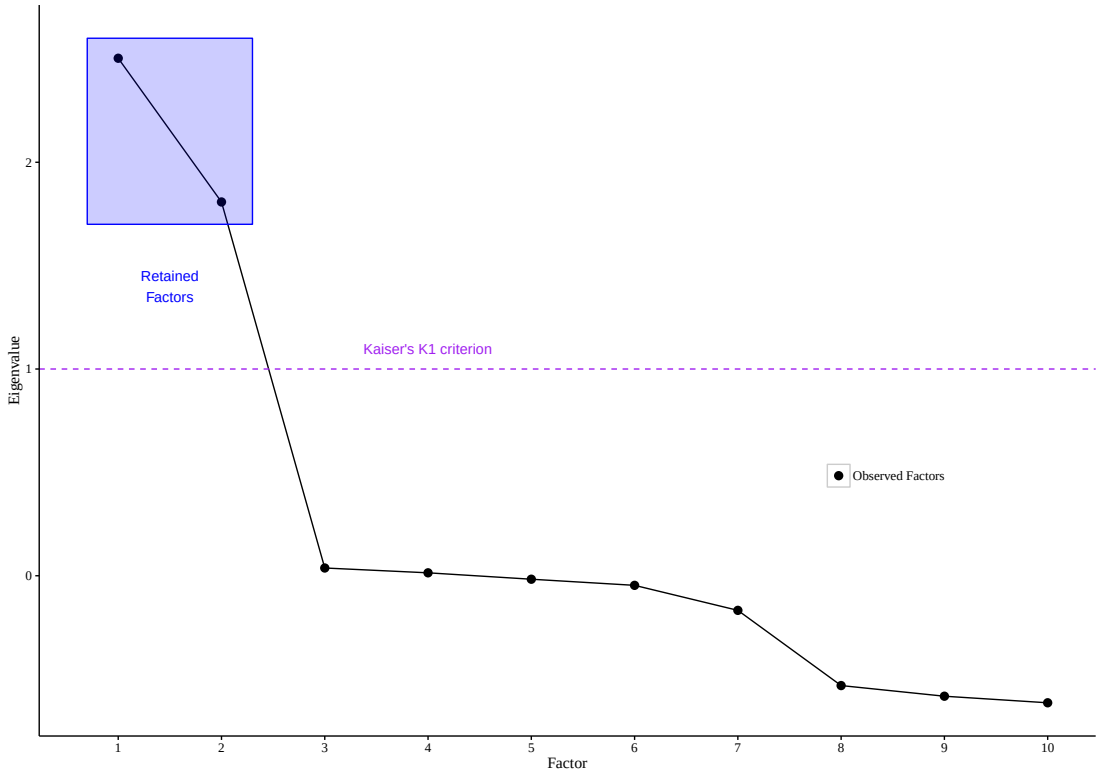


Figure: Scree plot

Structural validity
Observed structures

Structural validity
Theoretical structures
Observed structures
Individuals and structures
Criterion validity
Diagnostic validity
Review

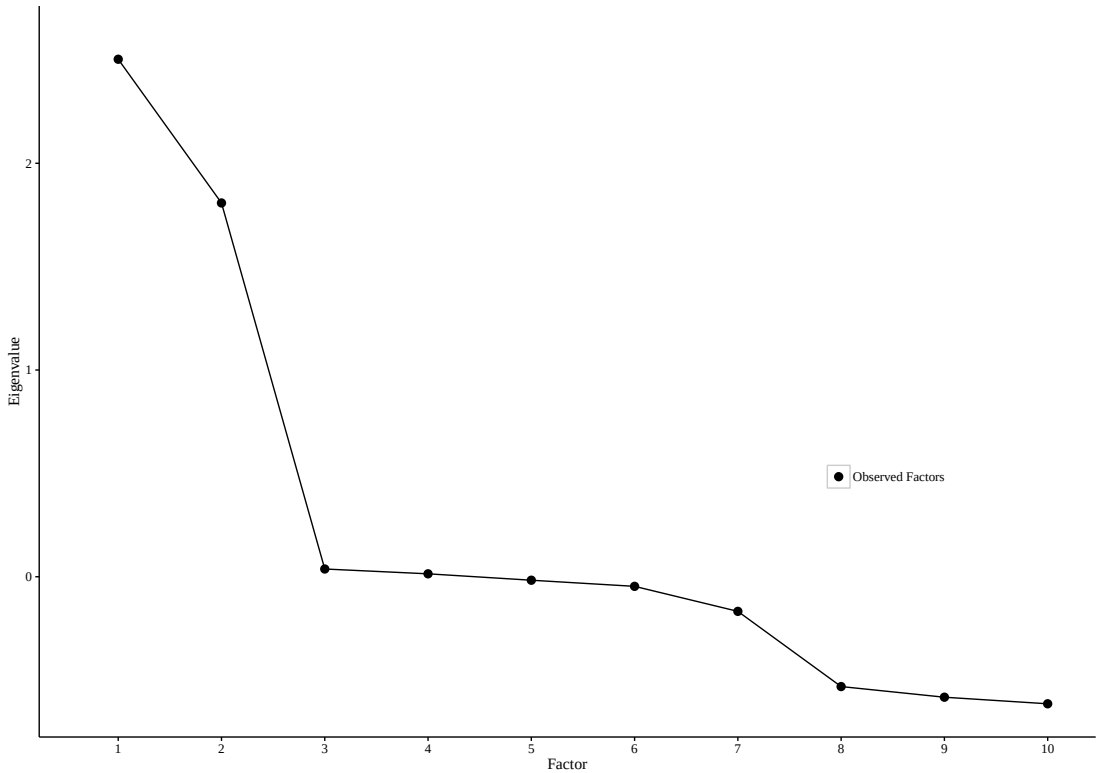


Figure: Scree plot

Structural validity

Observed structures

Structural validity
Theoretical structures
Observed structures
Individuals and structures
Criterion validity
Diagnostic validity
Review

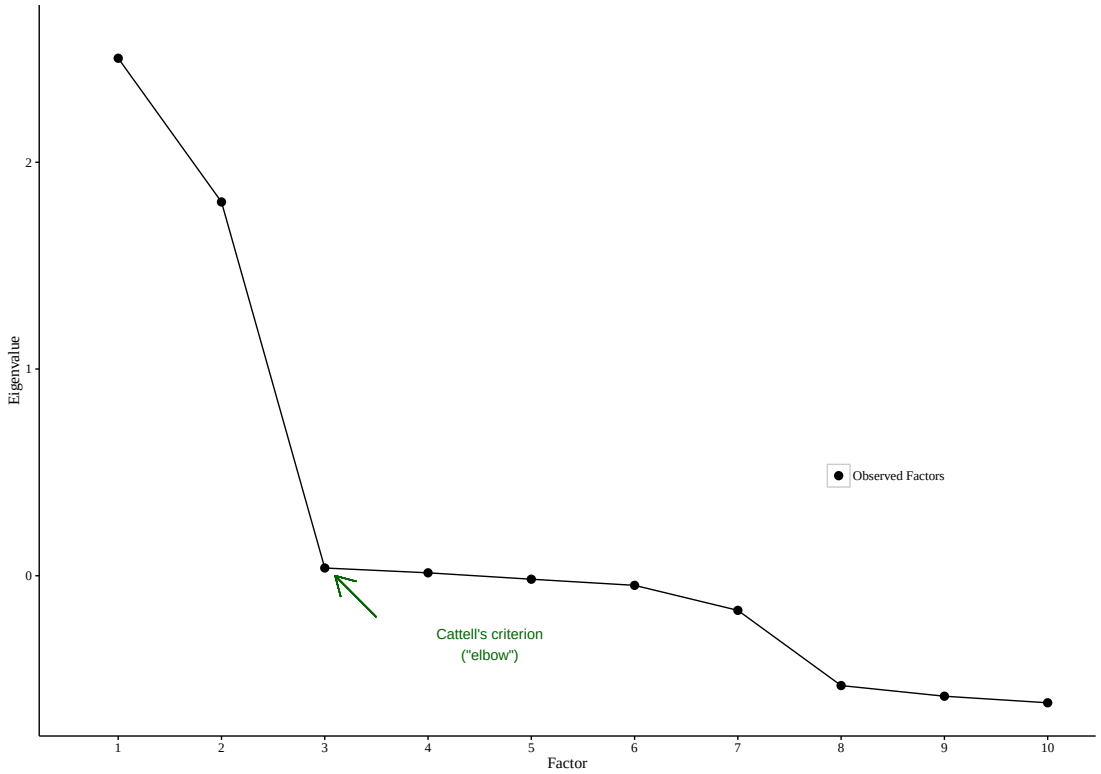


Figure: Scree plot

Structural validity

Observed structures

Structural validity
Theoretical structures
Observed structures
Individuals and structures
Criterion validity
Diagnostic validity
Review

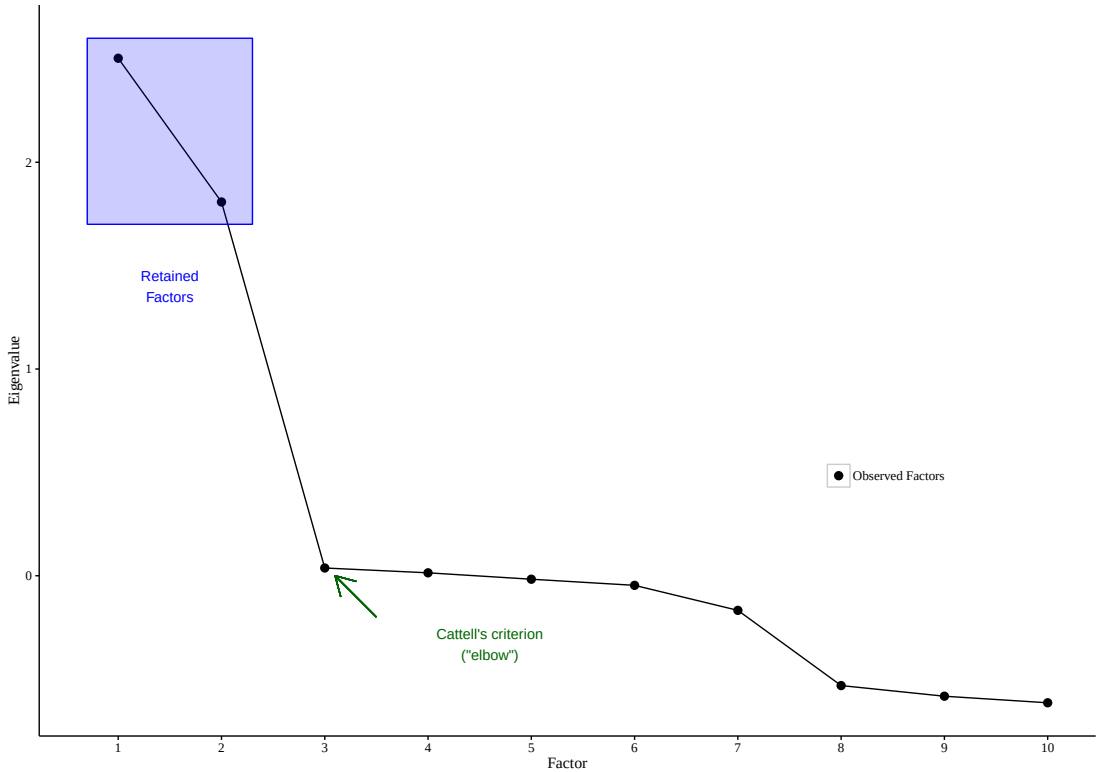


Figure: Scree plot

Structural validity
Observed structures

Structural validity
Theoretical structures
Observed structures
Individuals and structures
Criterion validity
Diagnostic validity
Review

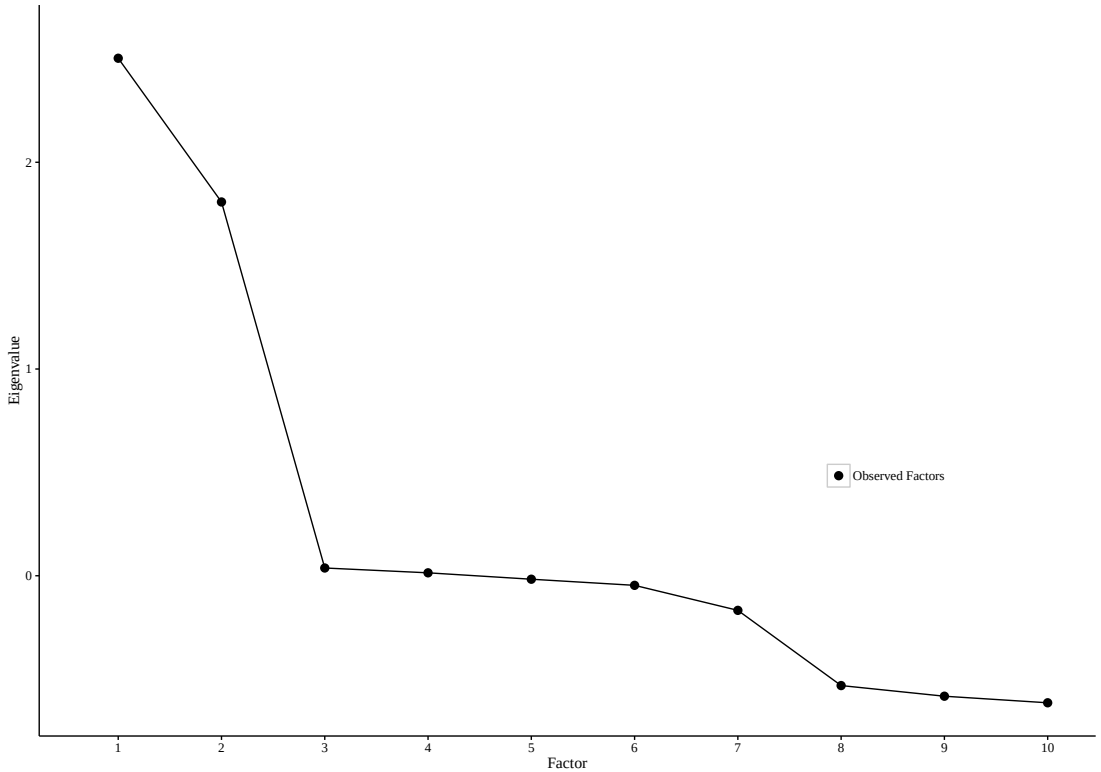


Figure: Scree plot

Structural validity
Observed structures

Structural validity
Theoretical structures
Observed structures
Individuals and structures
Criterion validity
Diagnostic validity
Review

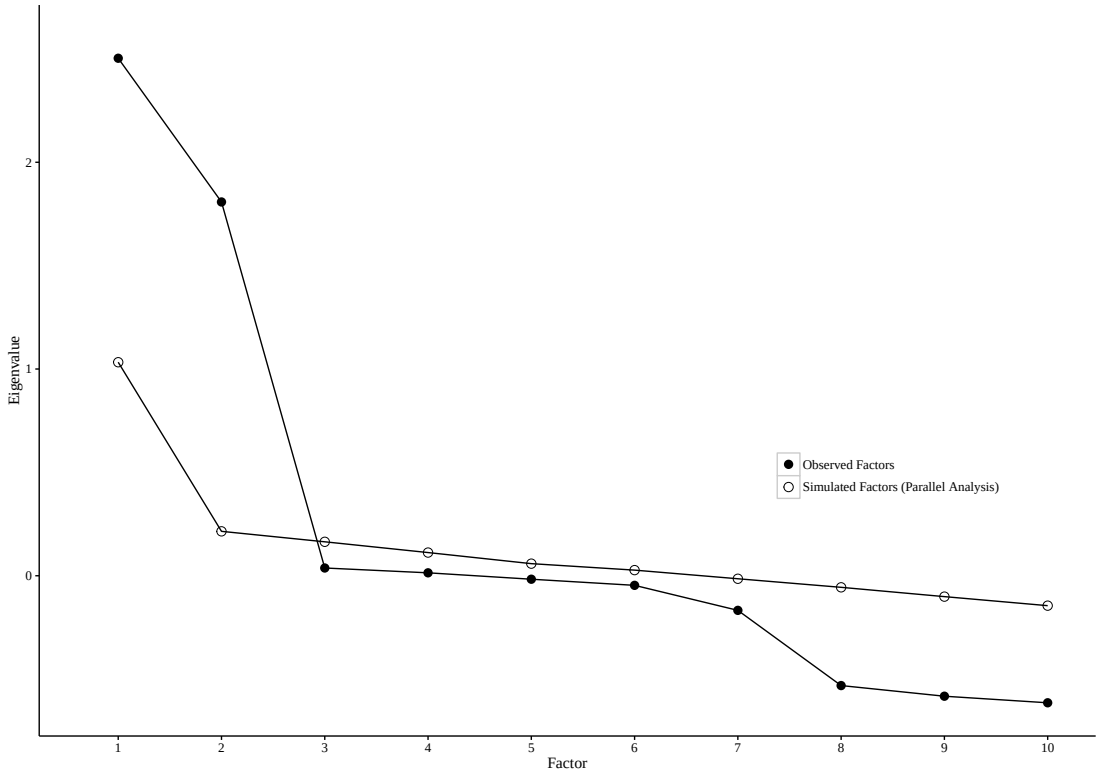


Figure: Scree plot

Structural validity
Observed structures

Structural validity
Theoretical structures
Observed structures
Individuals and structures
Criterion validity
Diagnostic validity
Review

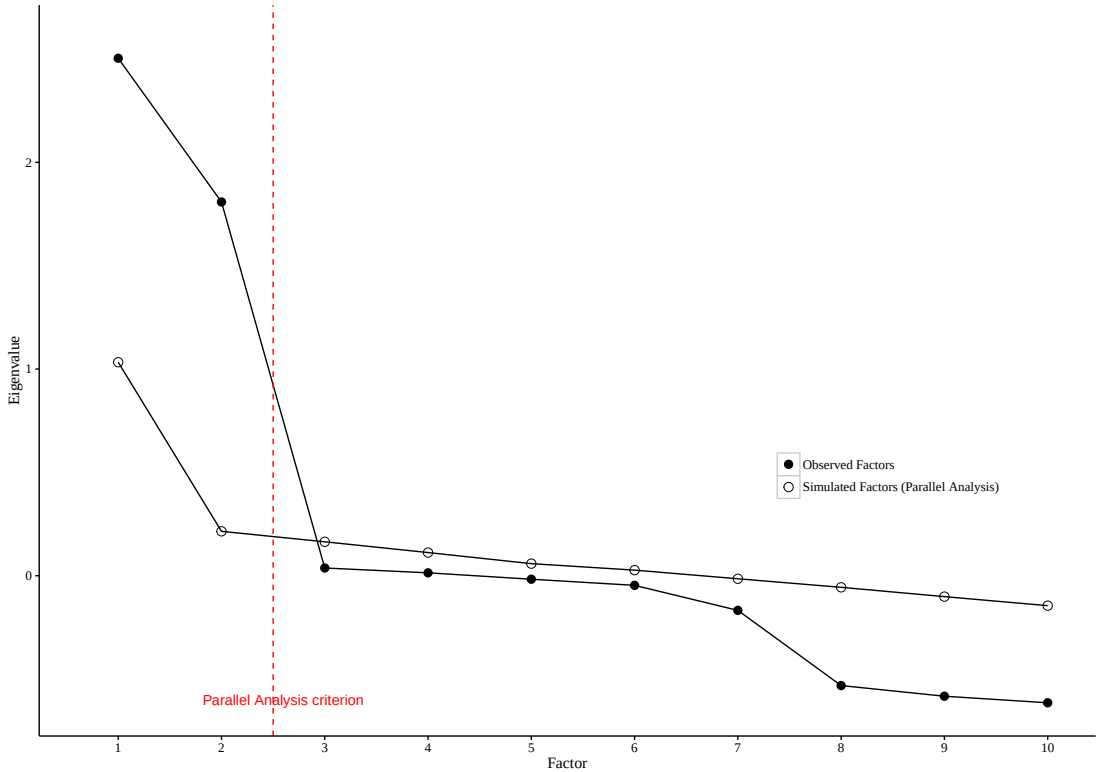


Figure: Scree plot

Structural validity
Observed structures

Structural validity
Theoretical structures
Observed structures
Individuals and structures
Criterion validity
Diagnostic validity
Review

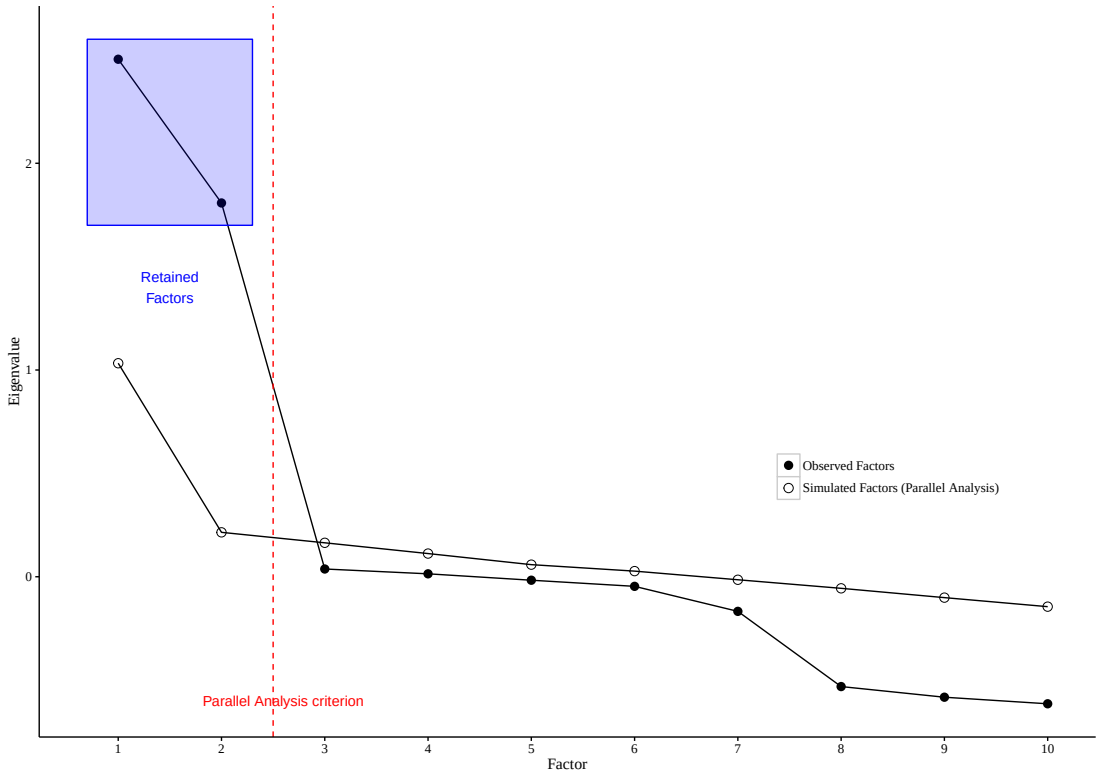


Figure: Scree plot

Structural validity
Observed structures

Structural validity
Theoretical structures
Observed structures
Individuals and structures
Criterion validity
Diagnostic validity
Review

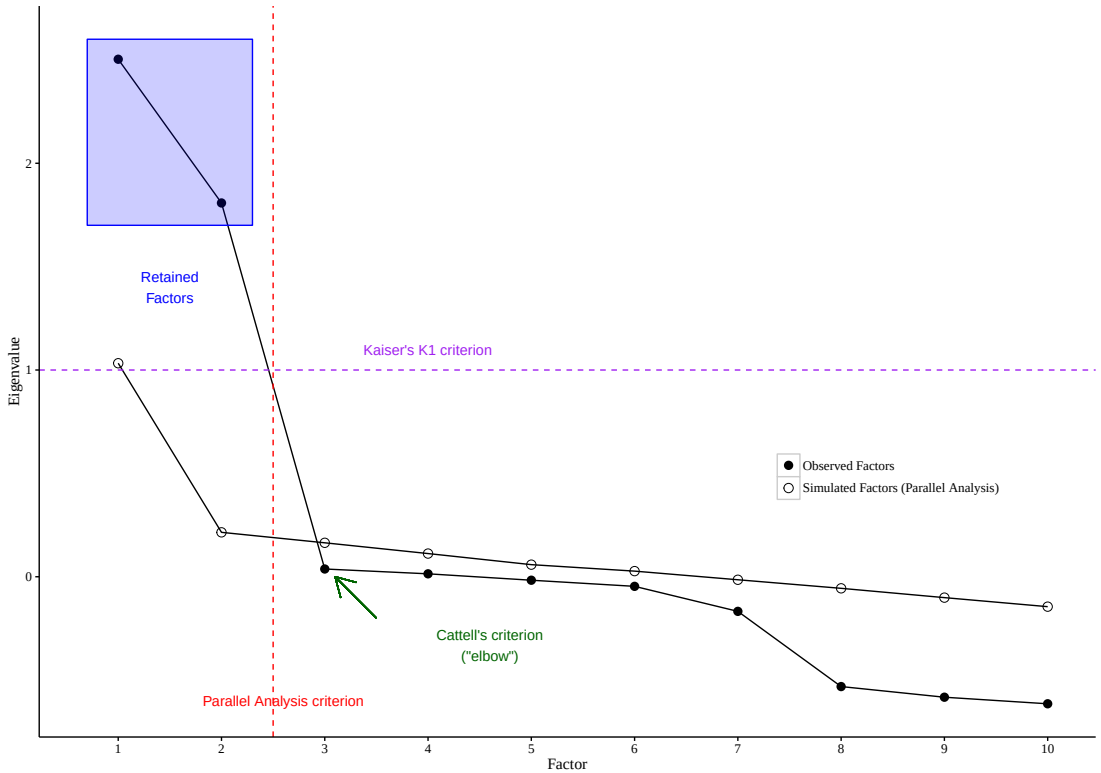


Figure: Scree plot

Structural validity
Theoretical structures
Observed structures
Individuals and structures
Criterion validity
Diagnostic validity
Review

Exploratory Factor Analysis (EFA)

- The next steps is interpreting the retained factors: What do they represent?
- To do so, we look at the relation between the factors retained and the items. These relations are represented by the items' **factor loadings**. We also talk of items that "load" on a factor.
- Loadings are not exactly correlations, but they are interpreted in similar ways, and also range from -1 to $+1$. Stronger loadings are close to -1 or $+1$.
- Loadings are typically considered weak below $.3$.

Structural validity
Theoretical structures
Observed structures
Individuals and structures
Criterion validity
Diagnostic validity
Review

Exploratory Factor Analysis (EFA)

- In most cases the factors cannot be interpreted from the unrotated factor loadings: The factors need be "rotated" to facilitate interpretation.
- A rotation is a way to observe the factors and their relations with one another to make them easily interpretable. It is a bit similar to observing axes in a 3D space: They are "easier" to understand from a specific point of view.
- Note : Rotation is useless when only one factor is extracted.

Structural validity
Theoretical structures
Observed structures
Individuals and structures
Criterion validity
Diagnostic validity
Review

Exploratory Factor Analysis (EFA)

- There are two main types of rotation : Orthogonal and Oblique.
  - **Orthogonal** rotations, like the typically used **Varimax** rotation, ensure that the different factors extracted are independent (not correlated). They help interpretability when the theoretical structure is unidimensional or comprises multiple independent factors, as they "force" the factors to differentiate one another.
  - **Oblique** rotations, like the typically used **Direct Oblimin** rotation, allow the factors to be correlated. They are useful when the theoretical structure is hierarchical. The **Promax** rotation is a helpful alternative to Oblimin on large datasets (less computationnally heavy).

Structural validity

Observed structures

Structural validity
Theoretical structures
Observed structures 51
Individuals and structures
Criterion validity
Diagnostic validity
Review

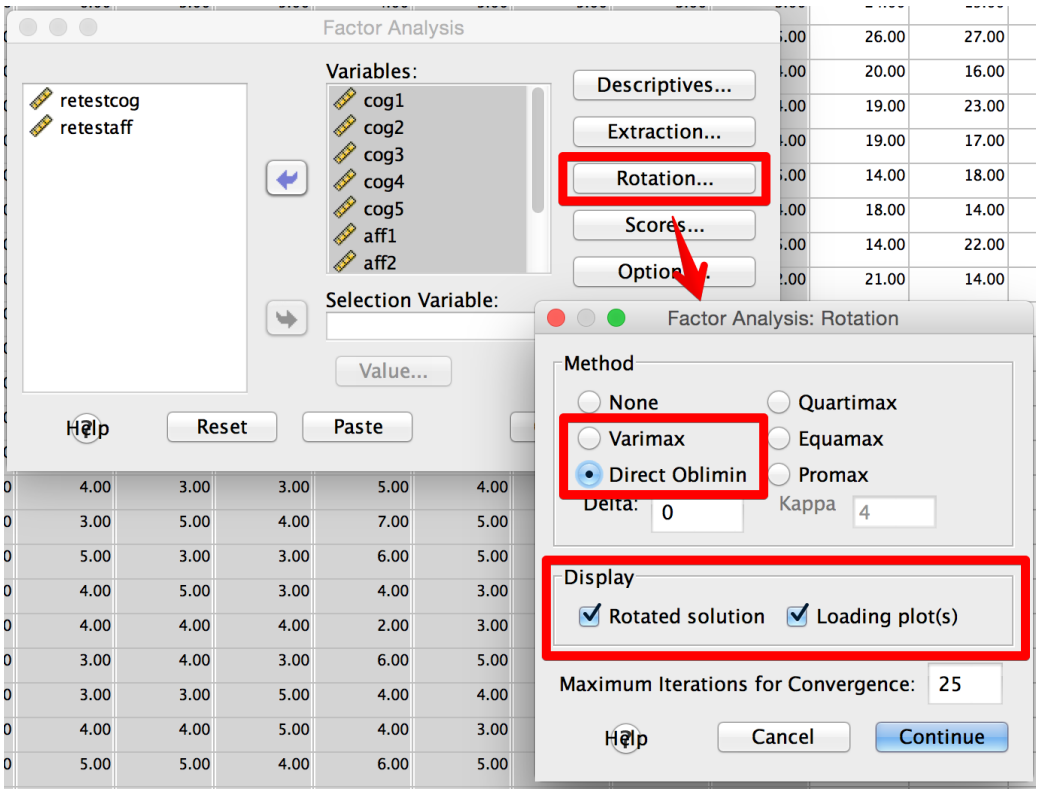


Figure: Choosing a rotation in SPSS

Structural validity

Observed structures

Structural validity
Theoretical structures
Observed structures 52
Individuals and structures
Criterion validity
Diagnostic validity
Review

Pattern Matrix[a]

	Factor	
	1	2
cog1	.864	.025
cog2	.657	.058
cog3	.762	-.006
cog4	.183	-.027
cog5	.801	.035
aff1	-.001	.787
aff2	.018	.718
aff3	.074	.805
aff4	.019	.756
aff5	-.039	.292

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

a. Rotation converged in 3 iterations.

Figure: Factor loadings ("pattern matrix")

Structural validity
Observed structures

Structural validity
Theoretical structures
Observed structures 53
Individuals and structures
Criterion validity
Diagnostic validity
Review

Exploratory Factor Analysis (EFA)

- Items can also be plotted on a "Loadings Plot", with the first and second factor as the x and y axes.
- It's an easy way of interpreting factors, but it obviously is limited to cases where only few factors are to be interpreted.

Structural validity
Observed structures

Structural validity
Theoretical structures
Observed structures
Individuals and structures
Criterion validity
Diagnostic validity
Review

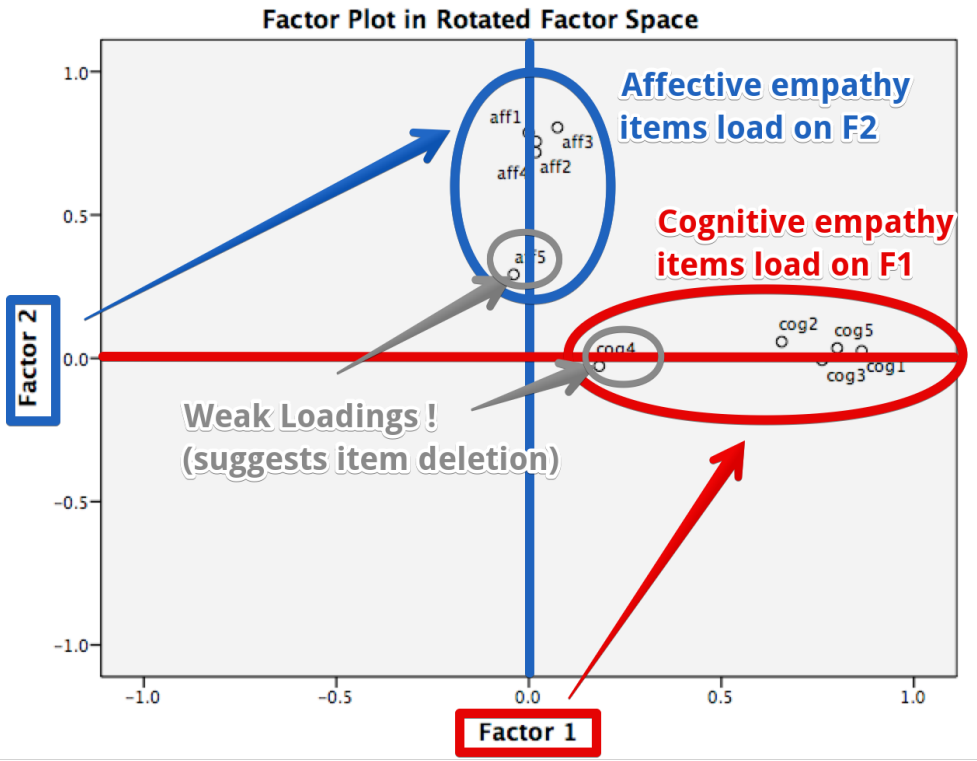


Figure: Factor loadings on a loadings plot

Structural validity

Observed structures

Structural validity
Theoretical structures
Observed structures 55
Individuals and structures
Criterion validity
Diagnostic validity
Review

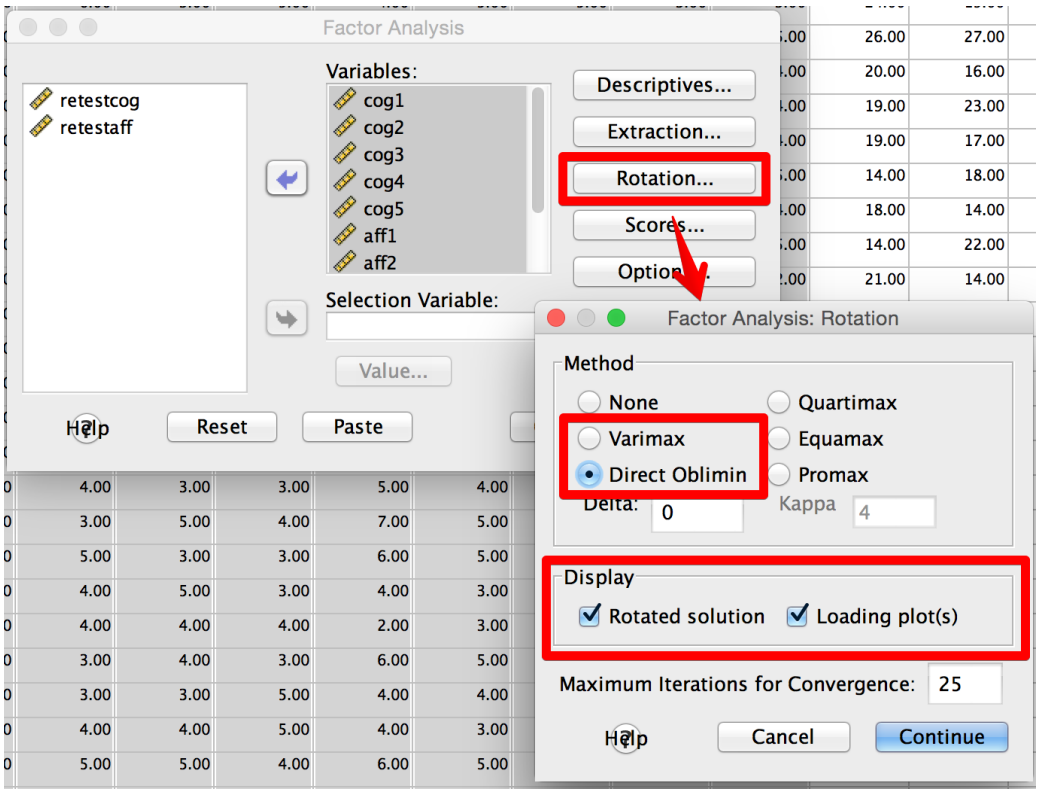


Figure: Factor loadings on a loadings plot

Structural validity
Theoretical structures
Observed structures
Individuals and structures
Criterion validity
Diagnostic validity
Review

Confirmatory Factor Analysis (CFA)

- Confirmatory Factor Analysis (CFA) consists in creating (using Structural Equation Modeling) a model with factor structure that is identical to the theoretical model, and then testing the fit of this model to the empirical data.
- It notably results in indices of model fit (which quantify how much the model built is able to reproduce the covariance matrix), and in factor loadings (like in EFA).
- Because of its flexibility with complex models, and because, contrary to EFA, it is an actual confirmatory method, CFA is nowadays more frequently used than EFA in psychometrical evaluation.

Structural validity
Theoretical structures
Observed structures
Individuals and structures
Criterion validity
Diagnostic validity
Review

Confirmatory Factor Analysis (CFA)

- As explained, CFA is achieved with specialized software, among which:
  - R
  - IBM Amos
  - SPSS
  - SAS
  - MPlus

Structural validity
Theoretical structures
Observed structures 58
Individuals and structures
Criterion validity
Diagnostic validity
Review

Confirmatory Factor Analysis (CFA)

- SEM is a framework that allows the simultaneous estimation of regressions between observed and latent variables.
- A typical representation of a SEM consists in a path diagram, with observed variables as rectangles, latent variables as ovals, and regression paths as arrows pointing to the outcome variable. Two sided arrows represent covariances, and self pointed arrows represent residuals.

Structural validity

Observed structures

Structural validity
Theoretical structures
Observed structures
59
Individuals and structures
Criterion validity
Diagnostic validity
Review

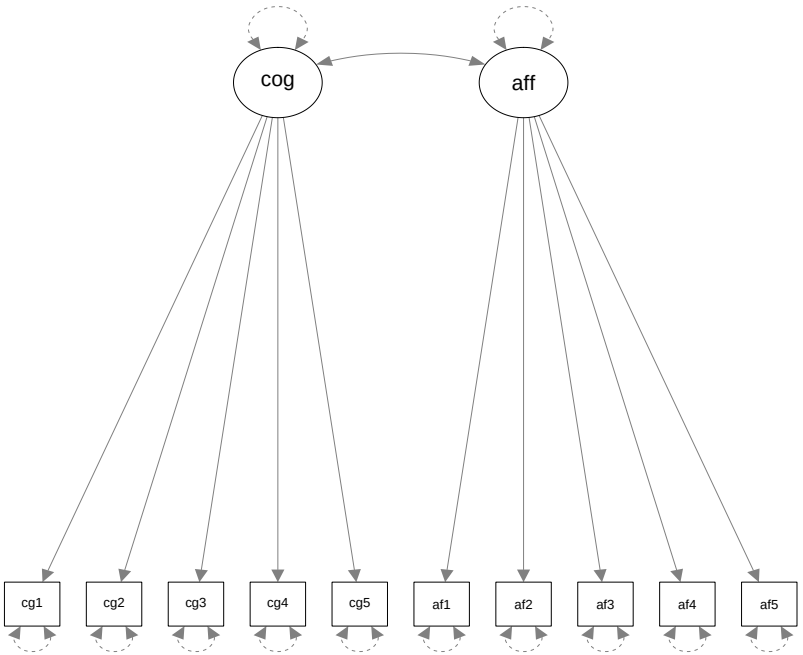


Figure: CFA Empathy Model

Structural validity

Observed structures

Structural validity
Theoretical structures
Observed structures 60
Individuals and structures
Criterion validity
Diagnostic validity
Review

Confirmatory Factor Analysis (CFA)

- The model is specified using either syntax or a path diagram
- The model is then fitted to the data using an estimation method (typically, Maximum Likelihood for assumed normal data)
- SEM software output indices of model fit, which indicate directly the extent to which the data and the hypothetical measurement model "agree".

Structural validity
Theoretical structures
Observed structures
Individuals and structures
Criterion validity
Diagnostic validity
Review

Confirmatory Factor Analysis (CFA)

- The extent to which the model fits the data, **model fit**, is quantified by many indices, notably:
  - χ^2: Should be non-significant ($p > .05$) for satisfactory fit (very biased by sample size, highly unrealistic most of the time)
  - CFI, TLI: Should be above $.95$ for satisfactory fit
  - RMSEA, SRMR: Should be below $.08$ for satisfactory fit

Structural validity
Observed structures

Structural validity
Theoretical structures
Observed structures 62
Individuals and structures
Criterion validity
Diagnostic validity
Review


```
lavaan (0.5-22) converged normally after  17 iterations

  Number of observations                           500

  Estimator                                         ML
  Minimum Function Test Statistic               60.954
  Degrees of freedom                                34
  P-value (Chi-square)                           0.003

Model test baseline model:

  Minimum Function Test Statistic             1843.185
  Degrees of freedom                                45
  P-value                                        0.000

User model versus baseline model:

  Comparative Fit Index (CFI)                    0.985
  Tucker-Lewis Index (TLI)                       0.980

Loglikelihood and Information Criteria:

  Loglikelihood user model (H0)              -7459.665
  Loglikelihood unrestricted model (H1)      -7429.187

  Number of free parameters                         21
  Akaike (AIC)                               14961.329
  Bayesian (BIC)                             15049.836
  Sample-size adjusted Bayesian (BIC)        14983.181

Root Mean Square Error of Approximation:

  RMSEA                                          0.040
  90 Percent Confidence Interval          0.023  0.056
  P-value RMSEA <= 0.05                          0.845

Standardized Root Mean Square Residual:

  SRMR                                           0.031
```

Structural validity
Observed structures

Structural validity
Theoretical structures
Observed structures 63
Individuals and structures
Criterion validity
Diagnostic validity
Review


```
Latent Variables:
                   Estimate  Std.Err  z-value  P(>|z|)  Std.lv  Std.all
  cog =~
    cog1              1.232    0.054   22.832    0.000   1.232    0.868
    cog2              0.957    0.061   15.716    0.000   0.957    0.660
    cog3              1.079    0.057   18.877    0.000   1.079    0.758
    cog4              0.182    0.048    3.798    0.000   0.180    0.180
    cog5              0.967    0.047   20.443    0.000   0.967    0.803
  aff =~
    aff1              1.174    0.060   19.576    0.000   1.174    0.785
    aff2              0.968    0.056   17.376    0.000   0.968    0.719
    aff3              1.041    0.051   20.520    0.000   1.041    0.812
    aff4              1.134    0.061   18.612    0.000   1.134    0.756
    aff5              0.257    0.042    6.101    0.000   0.257    0.288

Covariances:
                   Estimate  Std.Err  z-value  P(>|z|)  Std.lv  Std.all
  cog ~~
    aff               0.107    0.051    2.088    0.037   0.107    0.107

Variances:
                   Estimate  Std.Err  z-value  P(>|z|)  Std.lv  Std.all
   .cog1              0.495    0.059    8.454    0.000   0.495    0.246
   .cog2              1.185    0.085   14.011    0.000   1.185    0.564
   .cog3              0.864    0.069   12.564    0.000   0.864    0.426
   .cog4              0.995    0.063   15.736    0.000   0.995    0.968
   .cog5              0.515    0.046   11.300    0.000   0.515    0.355
   .aff1              0.860    0.075   11.390    0.000   0.860    0.384
   .aff2              0.878    0.068   12.914    0.000   0.878    0.484
   .aff3              0.560    0.053   10.477    0.000   0.560    0.341
   .aff4              0.963    0.079   12.154    0.000   0.963    0.428
   .aff5              0.729    0.047   15.574    0.000   0.729    0.917
    cog               1.000                       1.000    1.000
    aff               1.000                       1.000    1.000
```

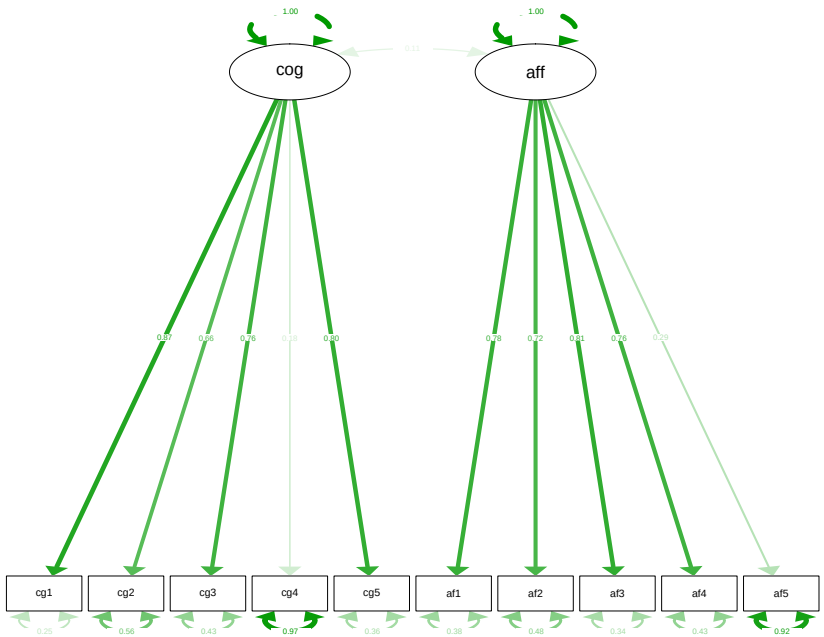
Weak loadings

LOADINGS

CORRELATION BETWEEN LATENTS

Structural validity
Observed structures

Structural validity
Theoretical structures
Observed structures 64
Individuals and structures
Criterion validity
Diagnostic validity
Review



Structural validity
Observed structures

Structural validity
Theoretical structures
Observed structures
Individuals and structures
Criterion validity
Diagnostic validity
Review

Fig. 1. Model A (best fit) with standardized estimates (all significant at $p < 0.001$).

Figure: CFA is often reported in path diagrams, which are simplified to show especially factor laodings, as here (Myszkowski et al., 2017).

Structural validity
Theoretical structures
Observed structures
66
Individuals and structures
Criterion validity
Diagnostic validity
Review

CFA

- When using CFA with mutlidimensional measures, usually a number of models can be considered. For example, independent factors structures, correlated factors structures, general factor models, etc.
- In the psychometric literature, nested models (pairs of models where one is a constrained version of the other) are often compared with Likelihood Ratio Tests (which are presented as χ^2 tests). Because Likelihood Ratio Tests are not available for non-nested models, usually model comparisons are based on either differences (Δ) in traditionnal CFA indices (ΔCFI, $\Delta RMSEA$, etc.) or on Information Criteria (Akaike Information Criterion or AIC, Bayesian Information Criterion or BIC, etc.).

Structural validity
Observed structures

Structural validity
Theoretical structures
Observed structures 67
Individuals and structures
Criterion validity
Diagnostic validity
Review

Table 1
CFA Fit indices of the tested models.

Model Letter	Model	χ^2	df	χ^2/df	CFI	SRMR	RMSEA	AIC
A	5 correlated factors (best model in original version)	277.33	80	3.46	0.938	0.058	0.076	9706.0
B	5 factors with 2 correlated second order factors (with 1 inequality constraint)	305.24	84	3.63	0.930	0.068	0.079	9726.0
C	5 factors with 1 s order factor (with 1 inequality constraint)	341.80	85	4.02	0.919	0.074	0.085	9760.5
D	5 orthogonal factors	920.99	90	10.23	0.738	0.283	0.149	10329.7
E	5 factors with 2 orthogonal second order factors (with 1 inequality constraint)	448.90	85	5.28	0.885	0.184	0.101	9897.6
F	1 factor	997.96	90	11.09	0.713	0.107	0.155	10436.7
G	2 orthogonal factors	822.21	90	9.14	0.769	0.204	0.140	10230.9
H	2 correlated factors	721.39	89	8.11	0.800	0.106	0.130	10132.1

Note. CFI – Comparative Fit Index; SRMR – Standardized Root Mean Square Residual; RMSEA – Root Mean Square Error of Approximation; AIC – Akaike Information Criterion.

Figure: An example of reporting of fit indices.

Structural validity
Theoretical structures
Observed structures
Individuals and structures
Criterion validity
Diagnostic validity
Review

CFA meets IRT

- Note that Confirmatory Factor Analysis is not necessarily done using a Structural Equation Modeling framework.
- It can be performed also in the IRT (Item Response Theory) modeling framework, which provides more appropriate tools for categorical items (for example, success/failure in performance tests).
- In the IRT tradition, researchers often focus on comparing models together and on "item fit", they identify how many and which items do not fit well within the factor structure (hopefully none!).
- However, recently (Maydeu-Olivares & Joe, 2006), the computation of "SEM-like" model fit indices have become available in (most of) IRT modeling (CFI, TLI, RSMR, RMSEA, etc.).

Structural validity

Observed structures

Structural validity
Theoretical structures
Observed structures
Individuals and structures
Criterion validity
Diagnostic validity
Review

Table 2
IRT fit indices of the VAST-R (Sample 2 and full original sample).

Model	Sample	χ^2	*df*	CFI	SRMR	RMSEA	AICc
1PL	Sample 2	430.04	299	0.843	0.100	0.040	6316.1
	Full original sample	557.60	299	0.865	0.088	0.040	12,395.6
2PL	Sample 2	373.90	275	0.929	0.066	0.036	6269.3
	Full original sample	503.69	275	0.930	0.055	0.039	12,304.6
3PL	Sample 2	325.38	250	0.946	0.065	0.033	6329.8
	Full original sample	420.54	250	0.948	0.056	0.035	12,333.3
4PL	Sample 2	316.62	225	0.934	0.065	0.039	6386.8
	Full original sample	380.18	225	0.953	0.055	0.036	12,306.2

Note. CFI – Comparative Fit Index; SRMR – Standardized Root Mean Square Residual; RMSEA – Root Mean Square Error of Approximation; AICc – Akaike Information Criterion (corrected).

Figure: An example of reporting of fit indices, here in IRT (Myszkowski & Storme, 2017).

Structural validity

Structural validity
Theoretical structures
Observed structures
Individuals and structures
70
Criterion validity
Diagnostic validity
Review

Structural validity

Theoretical structures

Observed structures

Individuals and structures

Structural validity
Theoretical structures
Observed structures
Individuals and structures 71
Criterion validity
Diagnostic validity
Review

Individuals and structures

- Keep in mind that factor structures remain general assumptions about how a test works in a population.
- Specific structures can be better applicable to certain individuals or groups of individuals but not to others, depending on the underlying processes. This is is addressed through **measurement invariance** (MI) analyses.
- For example, some individuals may be have a high general mental ability, and this might result in high scores in most subtests. A unidimensional factor structure would then "fits" the scores of such individuals. Oppositely, other individuals may, because of different processes being in use, have higher verbal ability than spatial ability, for example. Multidimensional factor structures would then "fit" the scores of such individuals.

Structural validity
Theoretical structures
Observed structures
Individuals and structures 72
Criterion validity
Diagnostic validity
Review

Individuals and structures

- In the framework of Confirmatory Factor Analysis, measurement invariance is verified by comparing factor structures and their characteristics (e.g., the loadings, etc.) between groups of individuals.

Structural validity
Theoretical structures
Observed structures
Individuals and structures 73
Criterion validity
Diagnostic validity
Review

Measurement invariance

- Note that the issue of measurement invariance is also relevant in longitudinal studies.
- Typically, if a measure has been administered at different time points, it may be relevant to question whether the same structure is observed at the different time points.

Structural validity
Individuals and structures

Structural validity
Theoretical structures
Observed structures
Individuals and structures 74
Criterion validity
Diagnostic validity
Review

Rules

- Because of a potential lack of measurement invariance, some tests may apply specific rules regarding how scores should be interpreted for each respondent.
- For example, discrepancies between subtest or indices scores may lead to a more detailed interpretation, while similarities between them would lead to a simpler interpretation.

Structural validity
Individuals and structures

Structural validity
Theoretical structures
Observed structures
Individuals and structures 75
Criterion validity
Diagnostic validity
Review

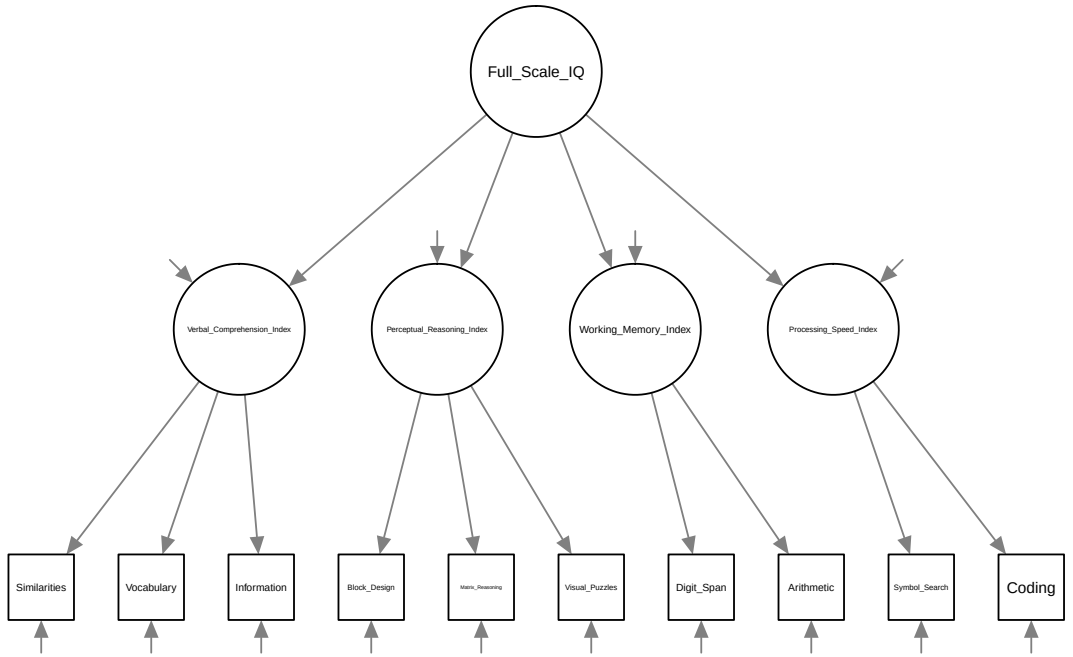


Figure: For individuals among which this structure is an adequate representation, interpreting a full scale IQ makes more sense...

Structural validity
Individuals and structures

Structural validity
Theoretical structures
Observed structures
Individuals and structures 76
Criterion validity
Diagnostic validity
Review

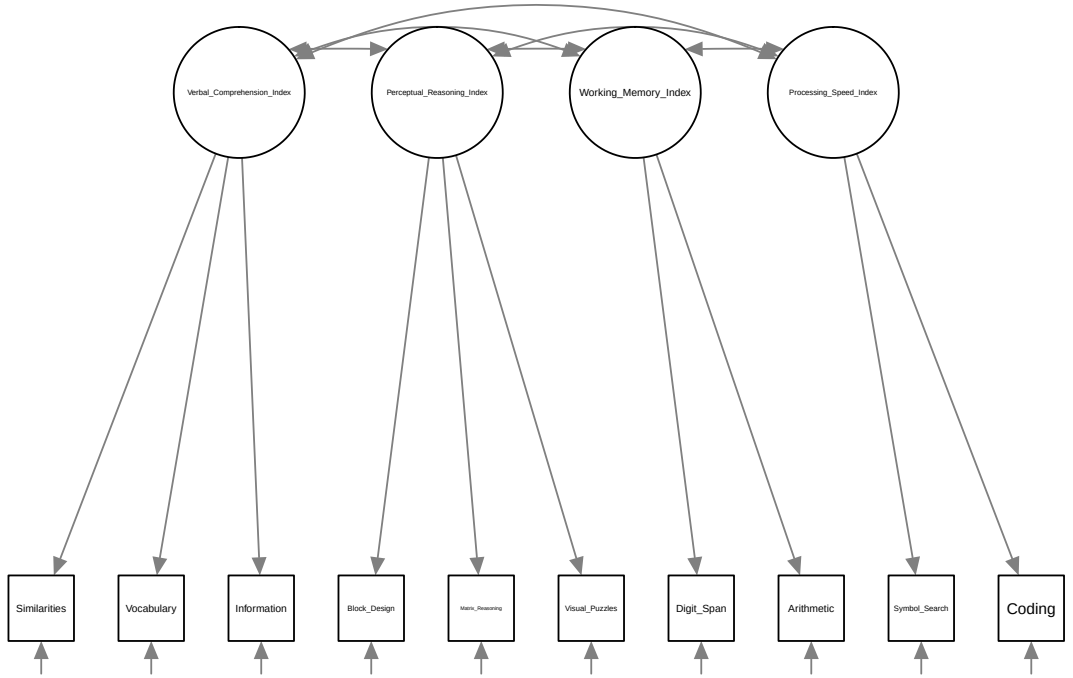


Figure: For individuals among which a weaker second order structure is observed, interpreting a full scale IQ makes less sense...

Structural validity

Individuals and structures

Structural validity
Theoretical structures
Observed structures
Individuals and structures 77
Criterion validity
Diagnostic validity
Review

Scores

- In your practice, watch for which scores should be used and under which conditions.
- Keep in mind that interpreting of giving as a feedback a score that is a sum or a mean implies that you aggregated responses together : Is it relevant to aggregate them? Do they reflect the same psychological attribute in the same way?

Outline

Structural validity
Criterion validity
78
Convergent Validity
Divergent validity
Predictive validity
Diagnostic validity
Review

Structural validity
Criterion validity
Diagnostic validity
Review

Criterion validity

Structural validity
Criterion validity
Convergent Validity
Divergent validity
Predictive validity
Diagnostic validity
Review

Criterion Validity

- Apart from examining the internal structure of a test, psychometricians and practitioners are usually interested in the relationships between the scores obtained with a test and the scores obtained with other measures of theoretically related (or not) constructs.
- Examining these relationships is investigating the **criterion validity** of a test (also called external validity).

Structural validity
Criterion validity
Convergent Validity
Divergent validity
Predictive validity
Diagnostic validity
Review

Criterion validity

- Criterion validity is usually done through examining the relationships between the scores obtained with the test and the scores obtained with other measures (called **external criteria**, or **validity criteria**).
- These relationships are usually examined through bivariate **correlations** between the score of the test and the score of the criterion.

Criterion validity

Structural validity
Criterion validity
Convergent Validity
Divergent validity
Predictive validity
Diagnostic validity
Review

Criterion validity

- Criterion validity has multiple forms, notably:
  - Convergent Validity
  - Divergent Validity
  - Predictive Validity

Criterion validity

Structural validity
Criterion validity
Convergent Validity
Divergent validity
Predictive validity
Diagnostic validity
Review

Criterion validity

- Convergent Validity
- Divergent validity
- Predictive validity

Structural validity
Criterion validity
Convergent Validity
Divergent validity
Predictive validity
Diagnostic validity
Review

Convergent Validity

- Good **convergent validity** is shown when the scores of a measure are (strongly) correlated with the scores of already validated measures of theoretically related constructs.
- For example, a good convergent validity criterion for a spatial reasoning ability test would be an already validated measure of the construct, like the Standard Progressive Matrices.

Structural validity
Criterion validity
Convergent Validity
Divergent validity
Predictive validity
Diagnostic validity
Review

Convergent validity

- Be careful, expected correlations between a test and a convergent validity criteria can be negative in some cases.
- For example, if you use an emotional stability scale as an external criterion for a neuroticism measure, you expect a negative correlation between the two measures.

Criterion validity

Structural validity
Criterion validity
Convergent Validity
Divergent validity
Predictive validity
Diagnostic validity
Review

Criterion validity

Convergent Validity
Divergent validity
Predictive validity

Structural validity
Criterion validity
Convergent Validity
Divergent validity
Predictive validity
Diagnostic validity
Review

Divergent validity

- Good **divergent validity** (also called **discriminant validity**) is shown when the scores of a measure are not (or very weakly) correlated with the scores of already validated measures of theoretically unrelated constructs.
- For example, a good divergent validity criterion for a new logical reasoning test would be an already validated measure of personality, since personality and logical reasoning are distinct constructs constructs.

Criterion validity

Structural validity
Criterion validity
Convergent Validity
Divergent validity
Predictive validity
Diagnostic validity
Review

Criterion validity

- Convergent Validity
- Divergent validity
- Predictive validity

Structural validity
Criterion validity
Convergent Validity
Divergent validity
Predictive validity
Diagnostic validity
Review

Predictive validity

- Tests are not only used to measure, but also sometimes to predict.
- For example, cognitive ability tests can be used to predict future academic success, or future job performance.
- Such relations can be empirically tested.

Structural validity
Criterion validity
Convergent Validity
Divergent validity
Predictive validity
Diagnostic validity
Review

Predictive validity

- Good **predictive validity** is shown when the scores of a measure are (strongly) correlated with the scores of already validated measures of constructs that are theoretically predicted by the test.
- For example, a good predictive validity criterion for a new cognitive ability test for children would be a measure of their academic success 5 years later.

Outline

Structural validity
Criterion validity
Diagnostic validity
Review

Structural validity
Criterion validity
Diagnostic validity
Review

Diagnostic validity

Structural validity
Criterion validity
Diagnostic validity 91
Review

Diagnostic contexts

- Psychological measurement is not always only concerned with observing a quantity, which is generally discussed as its main objective.
- Apart from achieving a quantity, it can be concerned with making classifications of individuals.
- For example, when deciding whether to categorize an individual as clinical or not (“clinical significance”).

Diagnostic validity

Structural validity
Criterion validity
Diagnostic validity 92
Review

Diagnostic validity

- Diagnostic validity is the ability of a test to perform well of accurately categorizing individuals in clinical vs. non-clinical groups.
- In other words, diagnostic validity is the ability of a test to accurately indicate which individuals present a disorder and which do not.

Diagnostic validity

Structural validity
Criterion validity
Diagnostic validity 93
Review

Thresholds

- Because here a test is supposed to lead to a binary outcome (disorder/no disorder), generally a threshold is used as a decision rule (although other procedures can be used, such as latent class analysis).
- When we consider diagnostic validity, we consider the test as the instrument and its threshold. The threshold score therefore plays a role in the diagnostic validity of a test.

Diagnostic validity

Structural validity
Criterion validity
Diagnostic validity 94
Review

Sources of error

- Valid diagnostic tests have high sensitivity and specificity:
  - **Sensitivity** is the ability of the test to identify correctly those who do have the disorder. (correct "hit", true positive)
  - **Specificity** is the ability of the test to identify correctly those who do not have the disorder. (correct rejection, true negative)

Diagnostic validity

Structural validity
Criterion validity
Diagnostic validity 95
Review

Sources of error

- We want to minimize two sources of error:
  - Concluding that an individual has the disorder when the individual does not have the disorder. (False positive, Type I error)
  - Concluding that an individual does not have the disorder when the individual has the disorder. (False negative, Type II error)

Diagnostic validity

Structural validity
Criterion validity
Diagnostic validity
Review

Diagnostic validity

		Reality: Disorder	Reality: No disorder
Test Outcome	**Positive**	True Positive	False Positive (Type I Error)
	Negative	False Negative (Type II Error)	True Negative

$$Sensitivity = \frac{TP}{TP+FN} \qquad Specificity = \frac{TN}{FP+TN}$$

Diagnostic validity

Structural validity
Criterion validity
Diagnostic validity 97
Review

The sensitivity-specificity balance

- There is a trade-off between sensitivity and specificity in each test, although perfecting best tests have both.
- In some situations, sensitivity or specificity may be more or less desirable.
- For example, if we are afraid to fail to detect a very severe disorder, we may want high sensitivity. In other contexts, we could be afraid to detect a disorder in somebody that does not present a disorder (for example if the treatment is very heavy or very expensive and we don't want to give it "for no reason").
- Typically, in diagnostic contexts, a high sensitivity is performed first (to screen for disorders). Then, a high specificity test is used to make sure that we don't "over-detect" the disorder.

Diagnostic validity

Structural validity
Criterion validity
Diagnostic validity
Review

The sensitivity-specificity balance

- In psychological screening for Alzheimer disease, for example, we want to start with high sensitivity tests to screen for a potential presence of a disorder. Doing so, we "make sure we don't forget anybody with such a disease", and we reduce false negatives.
- Then, we use a high specificity test to make sure we don't trigger an expensive and invasive anti-Alzheimer treatment for somebody who does not have Alzheimer disease. Doing so, we reduce false positives.

Outline

Structural validity
Criterion validity
Diagnostic validity
Review

Structural validity
Criterion validity
Diagnostic validity

Review

Review

Structural validity
Criterion validity
Diagnostic validity
Review

What you should remember

- There 3 main forms of validity, that apply to different tests and different contexts in testing:
  - Structural validity
    - Evaluated by examining the internal structure and the relations between the elements of the test
  - Criterion validity
    - Evaluated by examining the relationships between the test and validated measures of other constructs
  - Diagnostic validity
    - Examined by examining the quality of detection of a disorder through the correct detection of present disorders (sensitivity) and the correct non-detection when disorders are absent (specificity).