Exploratory and Confirmatory Factor Analysis

I. General Concepts

II. Principal Components and Exploratory Factor Analysis

III. Confirmatory Factor Analysis

I. General Concepts

Factor analysis provides information about reliability, item quality, and construct validity

General goal is to understand whether and to what extent items from a scale may reflect an underlying hypothetical construct or constructs, known as *factors*

An analytic method with high sensitivity to identify problematic items and assess the number of factors

I. General Concepts

In general, factor analysis methods decompose (or break down) the covariation among items in a measure into meaningful components

Higher inter-item correlations should reflect greater overlap in what the items measure, and, therefore, higher inter-item correlations reflect higher internal reliability

I. General Concepts

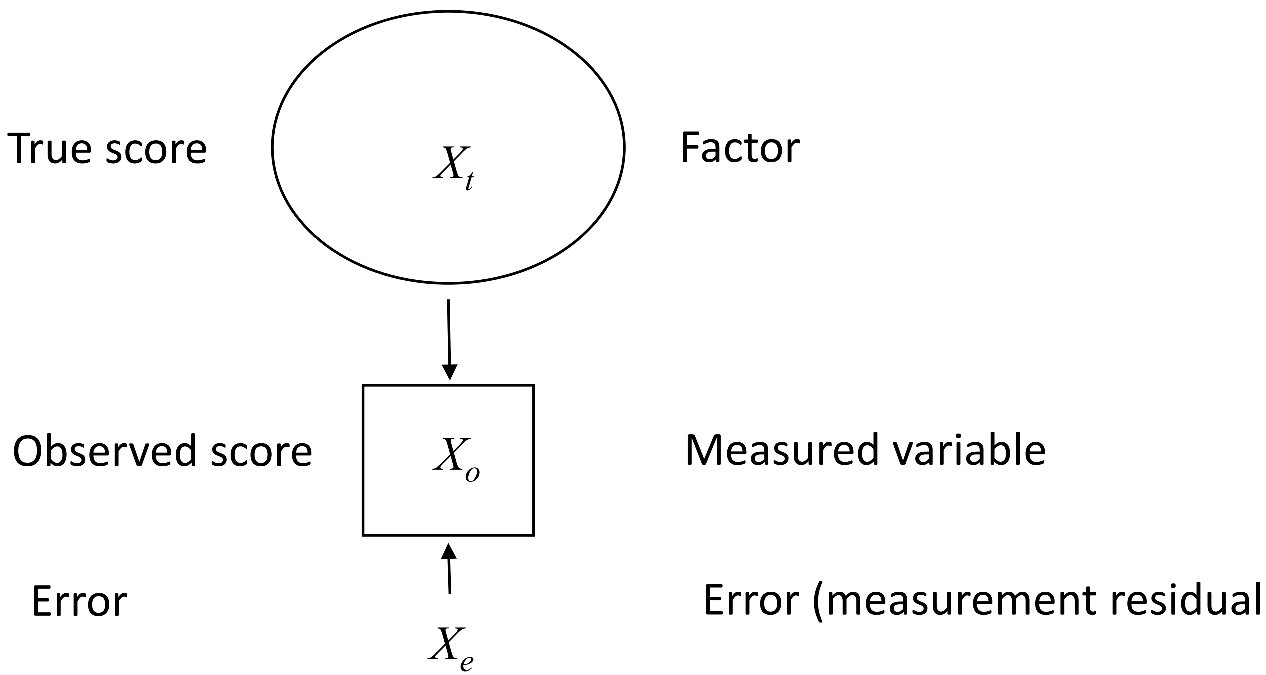
Classical Test Theory (CTT)

Observed Score	*=*	*True Score*	*+*	*Error*

$$X_o = X_t + X_e$$

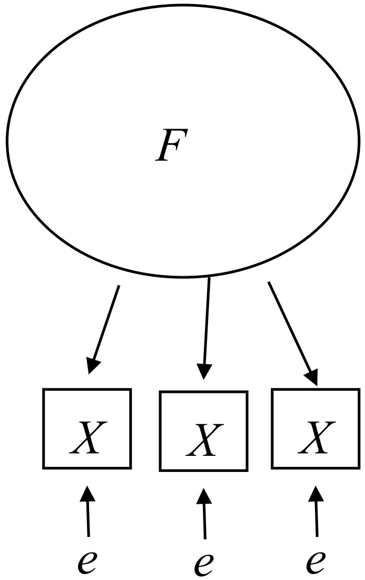
I. General Concepts

Factor model concept is analogous to CTT



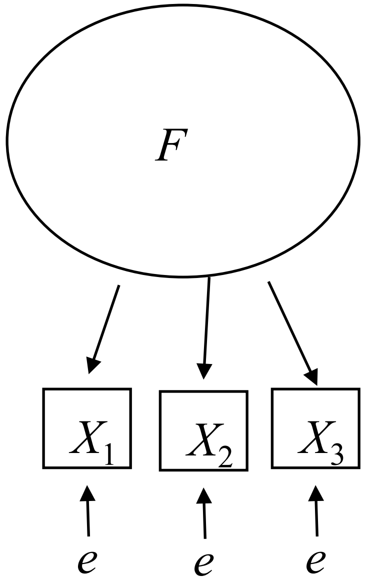
I. General Concepts

In practice, a factor cannot be estimated with one item
Should only be estimated with three or more items
Items with higher correlation with factor contribute more to the measure



I. General Concepts

Items are referred to as *indicators*
Regression slopes between factor and indicators are referred to as *loadings*



I. General Concepts

Patterns of high inter-item correlations among subsets of items suggest more than one factor because the items tend to "cluster" together

Any number of factors might underlie a set of items, up to the total number of items (which would imply no common factor)

- Example: set of six items might assess extroversion and openness

I. General Concepts

Table 4.1 (Hypothetical) Correlation Matrix for a Two-Factor Set of Items

	Talkative	*Assertive*	*Outgoing*	*Creative*	*Imaginative*	*Intellectu*
Talkative	1.00					
Assertive	.66	1.00				
Outgoing	.54	.59	1.00			
Creative	.00	.00	.00	1.00		
Imaginative	.00	.00	.00	.46	1.00	
Intellectual	.00	.00	.00	.57	.72	1.00

Furr, R. M., & Bacharach, V. R. (2013). *Psychometrics: an introduction, second edition*. Sage.

I. General Concepts

We never know the meaning of the factors, however; we can only use theory to decide what they mean and then test their validity

The factors may be related or not related—correlated or orthogonal (uncorrelated)

If those who are extroverted tend to be a little more open, then the factors are correlated (contrary to what is suggested by the table)

II. Principal Components and Exploratory Factor Analysis

Two major types of factor analysis

Exploratory factor analysis (EFA)[1]

Confirmatory factor analysis (CFA)[2]

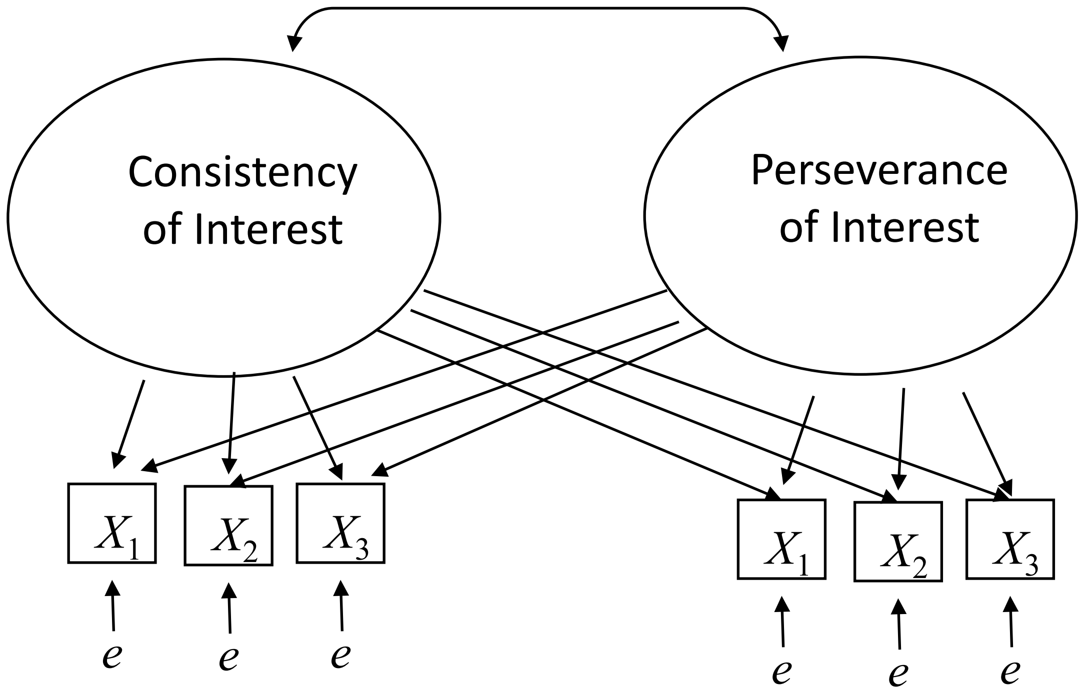
Major difference is that EFA seeks to discover the number of factors and does not specify which items load on which factors

[1] Charles Spearman is credited with developing factor analysis in 1904 and then it was extended by Louis Leon Thurstone in 1931, 1935, and 1947

[2] Confirmatory factor analysis was developed by a number of individuals, including D.N. Lawley (1940), Karl Jöreskog (1967), and Robert Jennrich and Stephen Robinson (1969)

II. Principal Components and Exploratory Factor Analysis

In EFA, loadings are obtained for all items related to all factors



II. Principal Components and Exploratory Factor Analysis

The researcher may discover there is one factor underlying the items or many factors

Items may be eliminated by the researcher if they do not load highly

Researchers choose items that load highly on one factor and low on other factors to achieve *simple structure*

Composite scale scores often created based on the factor analysis to be used in further research

II. Principal Components and Exploratory Factor Analysis

EFA is available in most general statistical software, such as SPSS, R, SAS

Involves several steps and decision points

- Deciding on the number of factors
- Extraction
- Rotation

II. Principal Components and Exploratory Factor Analysis

An initial analysis called *principal components analysis* (PCA) is first conducted to help determine the number of factors that underlie the set of items

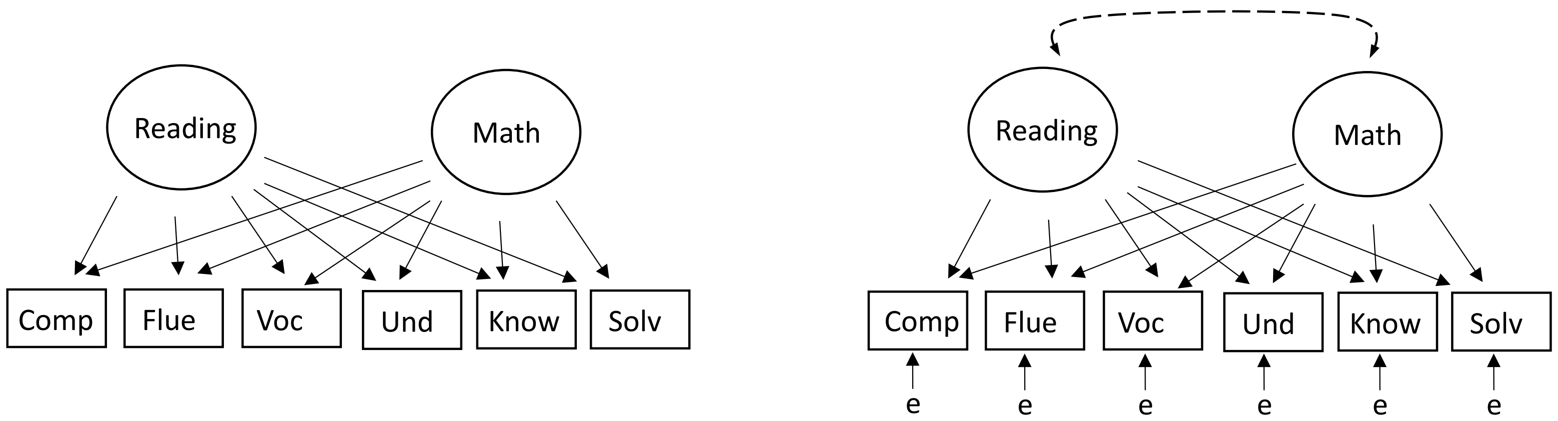
PCA is the default EFA method in most software or the first stage in other exploratory factor analysis methods to select the number of factors

PCA is not considered a “true factor analysis method,” because measurement error is not estimated (Snook & Gorsuch, 1989)

II. Principal Components and Exploratory Factor Analysis

PCA and EFA differ:

- PCA assumes no measurement error
- EFA does not require orthogonal factors



II. Principal Components and Exploratory Factor Analysis

PCA gives *eigenvalues* for the number of components (factors) equal to the number of items

If 12 items, there will be 12 eigenvalues

Each component is a potential "cluster" of highly inter-correlated items

Eigenvalues represent the amount of variance accounted for by each component, but they are not in a standardized metric

Larger eigenvalues indicate a more important (and more likely real) components or factor, with some merely reflecting unimportant factors or random variation

II. Principal Components and Exploratory Factor Analysis

The values sum to the number of items,[1] so if 12 items, then there will be 12 eigenvalues that sum to 12

The proportion or percentage of (co)variance accounted for by each factor can be calculated by dividing by the number of items

[1] This is true if a correlation matrix is used. If a covariance matrix is used then the eigenvalues sum to the total variance of the items

II. Principal Components and Exploratory Factor Analysis

Total Variance Explained

Factor	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings [a]
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total
1	2.195	36.578	36.578	1.836	30.599	30.599	1.836
2	2.173	36.222	72.800	1.808	30.131	60.730	1.808
3	.563	9.382	82.183				
4	.472	7.867	90.050				
5	.333	5.554	95.604				
6	.264	4.396	100.000				

Extraction Method: Principal Axis Factoring.

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

Furr, R. M., & Bacharach, V. R. (2013). *Psychometrics: an introduction, second edition*. Sage.

II. Principal Components and Exploratory Factor Analysis

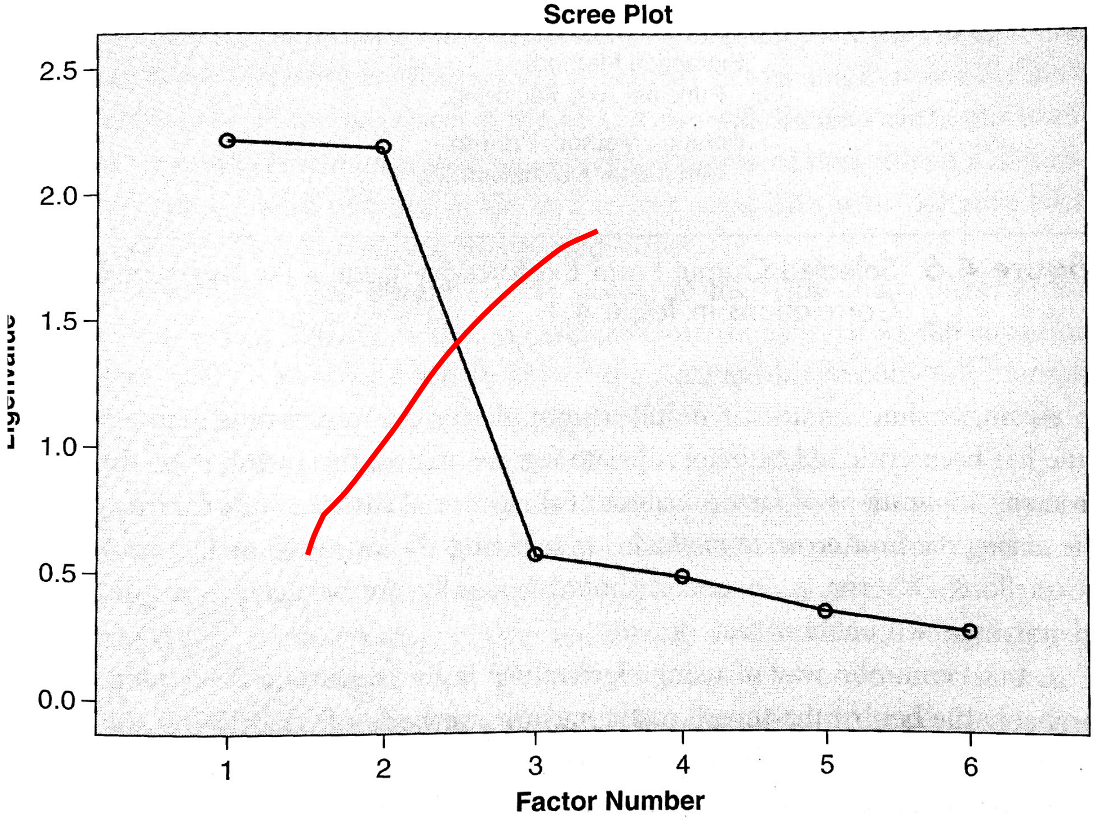
There are several possible rules which may be used for choosing the number of factors based on eigenvalues

The usual rule of greater than 1.0 (the Kaiser-Guttman rule) does not seem to work the best (Preacher & MacCallum, 2003)

Most use the scree plot and a *subjective scree test* by identifying the biggest drop in eigenvalues

The scree test or a more objective version (Cattell–Nelson–Gorsuch test) seems to work well for identifying the correct number of factors (Cattell & Vogelmann, 1977)

II. Principal Components and Exploratory Factor Analysis



Furr, R. M., & Bacharach, V. R. (2013). *Psychometrics: an introduction, second edition*. Sage.

II. Principal Components and Exploratory Factor Analysis

Next steps in an EFA after deciding on the number of factors is to choose a *method of extraction*

The extraction method is the statistical algorithm used to estimate loadings

There are several to choose from, of which *principal factors* (principal axis factoring) or *maximum likelihood* seem to perform the best (Fabrigar et al., 1999)

II. Principal Components and Exploratory Factor Analysis

And *factor rotation*

Factor rotation is a mathematical scaling process for the loadings that also specifies whether the factors are correlated (oblique) or uncorrelated (orthogonal)

Usually no harm in allowing factors to correlate

If the factor correlation is zero, then the same as orthogonal

Orthogonal rotation makes a strong assumption that the factors are uncorrelated, which probably is not likely in most applications

II. Principal Components and Exploratory Factor Analysis

TABLE 2

Summary of Final Results from Exploratory Factor Analysis—Stage Two

Item	Quality	Emotional	Price	Social
has consistent quality	0.82	0.28	0.21	
is well made	0.79	0.31	0.26	0.23
has an acceptable standard of quality	0.76	0.38	0.20	
has poor workmanship (*)	0.76	0.25	0.26	
would *not* last a long time (*)	0.76	0.20		
would perform consistently	0.70	0.31	0.22	
is one that I would enjoy	0.37	0.80		0.28
would make me want to use it	0.32	0.77		0.26
is one that I would feel relaxed about using	0.37	0.76	0.21	
would make me feel good	0.32	0.74	0.21	0.36
would give me pleasure	0.35	0.71		0.33
is reasonably priced			0.90	
offers value for money	0.30		0.82	
is a good product for the price	0.33	0.35	0.76	
would be economical	0.25		0.72	
would help me to feel acceptable				0.83
would improve the way I am perceived				0.83
would make a good impression on other people	0.26	0.29		0.74
would give its owner social approval		0.26		0.60
Eigen value	9.53	2.22	1.47	1.00

Sweeney, J. C., & Soutar, G. N. (2001). Consumer perceived value: The development of a multiple item scale. *Journal of retailing*, *77*(2), 203-220.

III. Confirmatory Factor Analysis

Confirmatory factor analysis (CFA) starts with a hypothesis about how many factors there are and which items load on which factors

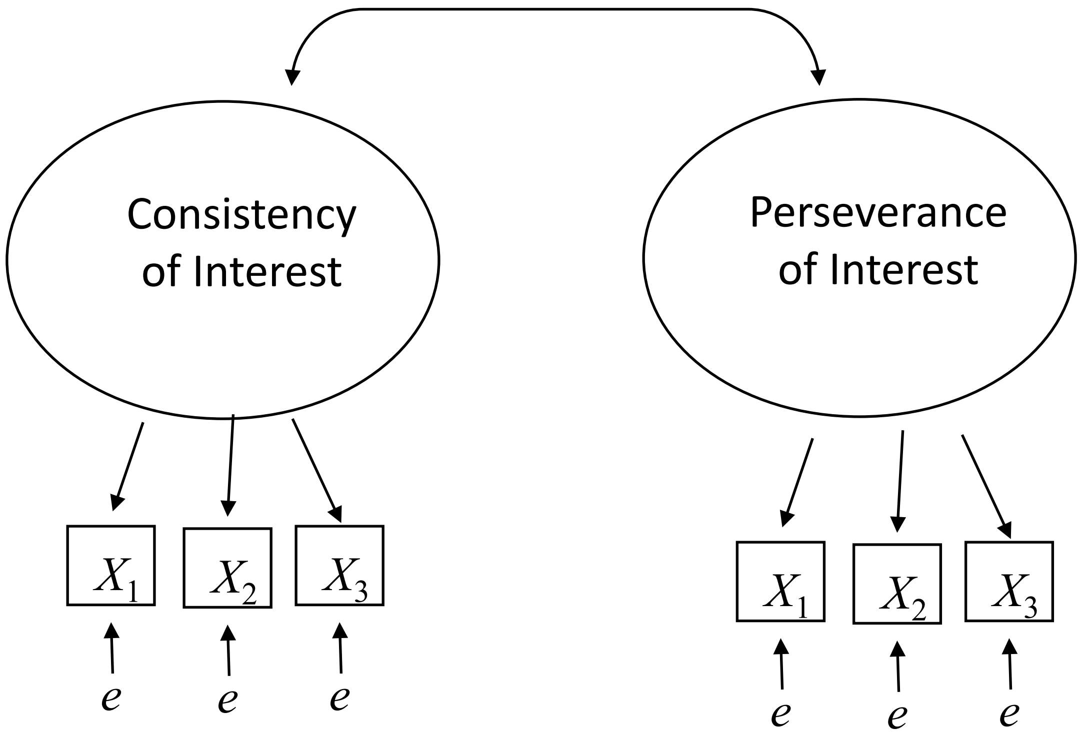
Factor loadings and factor correlations are obtained as in EFA

EFA, in contrast, does not specify a measurement model initially and usually seeks to discover the measurement model

In EFA, all items load on all factors

In CFA, most researchers start with a model in which items load on only one factor (simple structure)

III. Confirmatory Factor Analysis



Duckworth, A. L., & Quinn, P. D. (2009). Development and validation of the Short Grit Scale (GRIT–S). *Journal of personality assessment*, *91*(2), 166-174.

III. Confirmatory Factor Analysis

A test is computed to investigate how well the hypothesized *factor structure* fits with the data

The fit test seeks to find a non-significant result, indicating good fit to the data

III. Confirmatory Factor Analysis

The model fit is derived from comparing the correlations (technically, the covariances) among the items to the correlations expected by the model being tested

Mathematically, certain models imply certain correlations, e.g.,

- if one-factor model, items should be highly correlated, items that do not correlate highly will lead to a poor fit for a one factor model
- If model specifies that two factors are uncorrelated, then the model will not fit well if items from one factor tend to be correlated with items from another factor

III. Confirmatory Factor Analysis

You may hear about many fit indices, so here are some common examples:

Chi-square, χ^2 lower values indicate better fit

The chi-square is supposed to be nonsignificant to indicate good fit of the model to the data. Large sample sizes makes this a relatively rare occurrence, so other indexes of fit used

RMSEA, lower values indicate better fit (< .06)

SRMR, lower values indicate better fit (< .08)

Comparative Fit Index, higher value indicate better fit (>.95)

Tucker-Lewis Index, higher value indicate better fit (>.95)

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1-55.

III. Confirmatory Factor Analysis

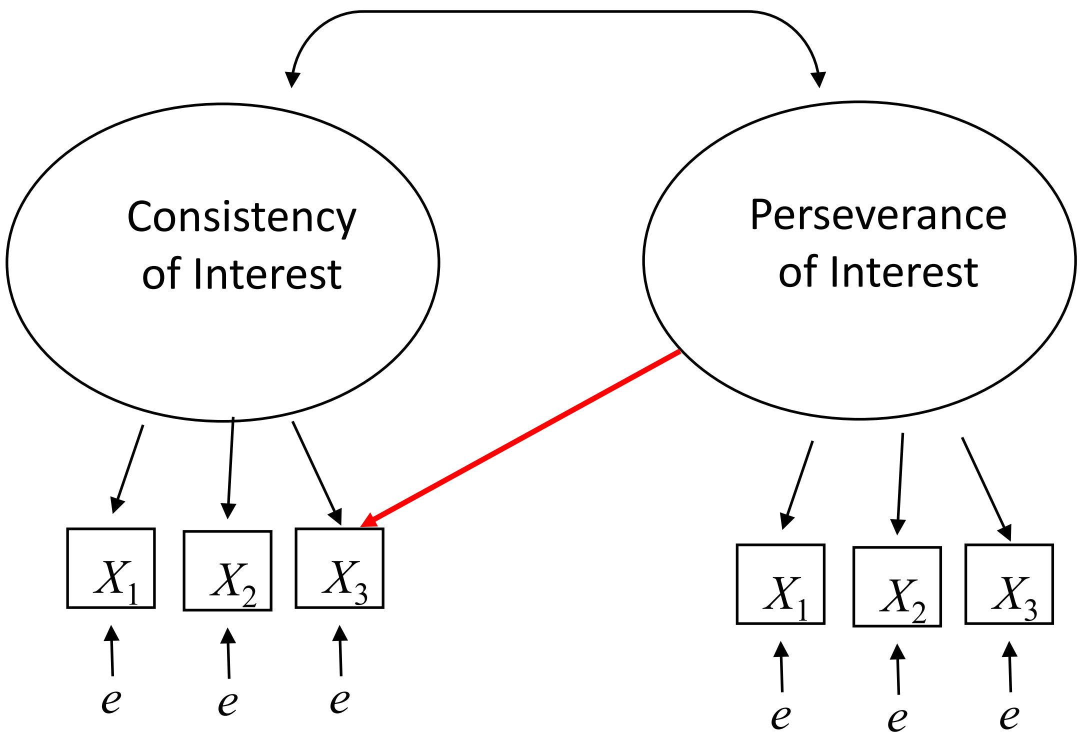
If the model does not fit well, it can be altered and retested

Many possible loading structures can be specified by the user and any item can load on multiple factors if desired

The more changes made, the more the researcher may be capitalizing on chance, running the risk of Type I errors

Many changes constitutes an exploratory approach

III. Confirmatory Factor Analysis



Duckworth, A. L., & Quinn, P. D. (2009). Development and validation of the Short Grit Scale (GRIT–S). *Journal of personality assessment*, *91*(2), 166-174.

III. Confirmatory Factor Analysis

The differences between EFA and CFA are often overstated

Despite their names, both can be used in an exploratory manner

- CFA models can be modified if the model does not fit well
- EFA is sometimes used by researchers even though they have a well-developed idea about the factor structure and wants to confirm it

Both methods are based on discovering number of underlying factors for a set of items and estimating how strongly they relate to the factors

III. Confirmatory Factor Analysis

Both methods of factor analysis are sensitive psychometric analysis that provide information about reliability, item quality, and validity

Scale may be modified by eliminating items or changing the structure of the measure

Either method may be used as a preliminary step to evaluate a measure or set of subscales that will be computed and used in later research

III. Confirmatory Factor Analysis

Specialized software usually required (e.g., Amos, Mplus, LISREL, EQS, the R package lavaan)

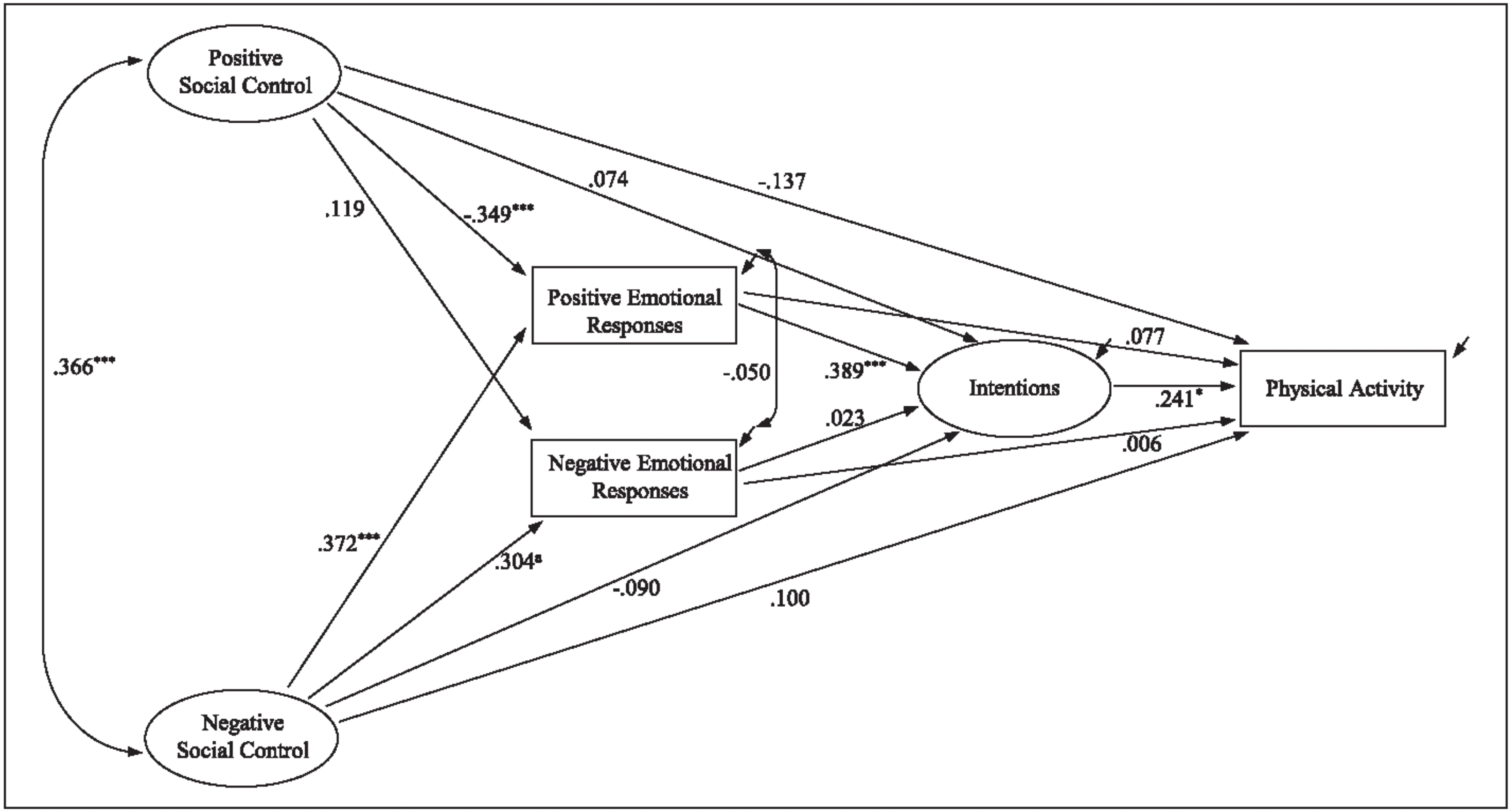
EFA procedures usually available in general statistical software packages like SPSS, SAS, Stata etc.

III. Confirmatory Factor Analysis

CFA is part of a larger analysis framework, called *structural equation modeling* (SEM), which combines CFA with path analysis (regression slopes)

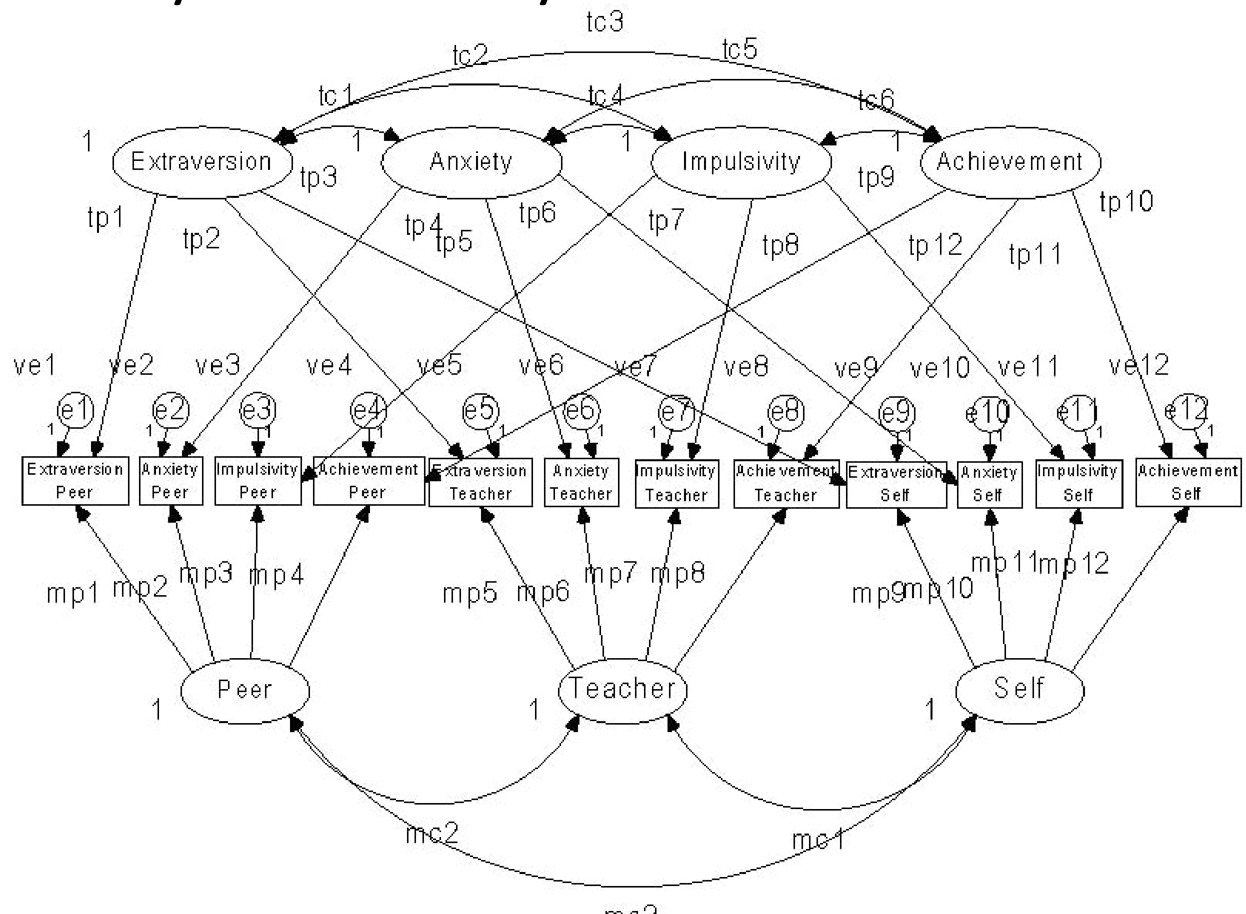
SEM can use factors (or “latent variables”) in regression analysis to predict other variables or be predicted by other variables, with the advantage of estimating and eliminating measurement error from correlation and regression estimates

III. Confirmatory Factor Analysis



Newsom, J.T., Shaw, B.A., August, K.J., & Strath, S.J. (2016). Physical activity–related social control and social support in older adults: Cognitive and emotional pathways to physical activity. *Journal of Health Psychology*, 1-16. Published online July 29, 2016: DOI: 10.1177/1359105316656768

III. Confirmatory Factor Analysis



Bentler, P. M., & Lee, S. Y. (1979). A statistical development of three-mode factor analysis. British Journal of Mathematical and Statistical Psychology, 32(1), 87-104.