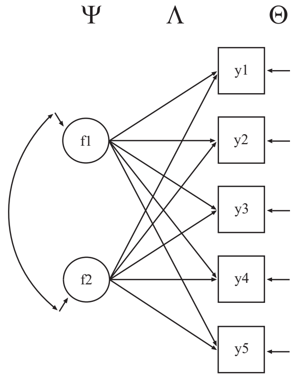
1 EFA background

EFA Model with Two Factors (M = 2)



- The goal of EFA is to find the smallest number of factors that explain the correlations among the observed variables
- EFA specifies only the number of factors - unlike confirmatory factor analysis (CFA), EFA has no hypothesis about zero loadings
- The EFA model is not identified but has m^2 indeterminacies
- The EFA model is made identified by applying a "rotation" that eliminates the m^2 indeterminacies and gives an interpretable model

Bengt Muthén | Exploratory Factor Analysis | 3/ 139

Slide 3 shows an EFA model for 6 factor indicators measuring 2 factors. The single-headed arrows represent linear regressions when indicators are continuous and probit or logit regressions for categorical indicators.

The goal of EFA is to find the smallest number of factors that explain the correlations among the observed variables.

EFA specifies only the number of factors, no left-out arrows from the factors to the indicators as in CFA. Characteristic of EFA, both factors influence all 6 indicators.

As stated in the third bullet, the EFA model is not identified but has $m^2 = 4$ indeterminacies. The correlations among the 6 indicators is expressed by 3 parameters in the factor covariance matrix Ψ - the covariance and the two variances of the factors - as well as $2 \times 6 = 12$ parameters in the factor loading matrix Λ, for a total of 15 parameters. Although there are also 15 sample correlations, the model is not identified, that is, there is no unique solution for the parameter estimates but there are indeterminacies. For example, a change in the factor variance can be compensated by a change in the loadings resulting in the same variances and covariances for the indicators.

As stated in the last bullet, the indeterminacies can be removed by EFA rotation. This solves the problem of indeterminacies and presents an opportunity to explore a solution that is clearly interpretable.

As a start, an unrotated model is created where the factor variances are standardized to 1 and the factor covariance is set at 0 (referred to as orthogonal rotation; see EFA theory section). This leaves 1 remaining indeterminacy which can be settled by fixing a loading to zero. Based on this unrotated solution, the rotation then chooses a simple and more interpretable set of factor loading values where simple means that an indicator should have one or two large loadings in a row of the loading matrix to help define the factor but also small loadings for other factors. The rotation can also relax the zero factor correlation (referred to as an oblique solution) to further simplify the loading pattern. Rotation for two factors is illustrated graphically for two examples in the EFA theory section.

Stages of Measurement Development Using Factor Analysis

Stage	Specification	Technique
Early	# of factors only	EFA, ESEM PSEM with GEOMIN priors
Middle	Pre-specified zero loadings Pre-specified key loadings	Target rotation Cross-loadings constrained using PSEM ALF priors
Late	# of factors, key loadings, and zero cross-loadings	CFA Modification indices

Turning to slide 4, it is useful to distinguish between different stages of measurement development using factor analysis: Early, Middle, and Late.

At an early stage, relatively little is known about how well factor indicators are able to capture hypothesized factors. You may be ready to specify only how many factors the indicators measure or perhaps only a range of number of factors. At this stage, traditional EFA is the technique of choice. Automatic rotation procedures search for a simple, interpretable factor loading matrix. Indicators shown to be poor measures of factors are either reformulated or deleted. For a well-defined EFA model, the techniques of ESEM and PSEM make it possible to move to more general analysis settings.

At the middle stage, experimentation with choices of factor indicators makes it possible to specify not only the number of factors but also which indicators do not load on certain factors, leading to the technique of target rotation, that is, a rotation that is guided by prior knowledge rather than automatic criteria of simplicity. Alternatively, it is possible to specify which indicators are key to measuring which factors, leading to the technique of studying cross loadings using PSEM ALF priors.

At the late stage, it may be possible to specify the number of factors, key loadings, as well as zero cross loadings, leading to a model with more restrictions than EFA. This leads to the technique of CFA (confirmatory factor analysis; see Jöreskog's 1969 Psychometrika article) which is geared towards testing specific factor structures. Model misfit is captured by modification indices.

Two comments on these three stages are of interest.

First, a common problem of factor analysis is that the late stage approach of CFA is often the first approach used, skipping the two earlier stages. This typically happens in the context of testing a hypothesis that has been well established in the literature - except not for the indicators or population at hand. As a consequence, poor fit is obtained and a long search for better fit via modification indices is carried out. As argued by Michael Brown (2001) quoted in the Asparouhov & Muthén (2009) ESEM article, EFA is in this case a better alternative where rotation is a more direct approach to finding misfitting cross loadings than using modification indices.

Second, in many applications, CFA is not a possible or desirable late stage analysis. The CFA model may not fit well despite modifications. Due to the nature of what is measured, it is often not reasonable to expect that relevant factor indicators measure only one factor but have loadings on other factors, albeit with smaller values. The factor loading pattern is then more complex and better suited to an EFA model. In this situation, staying with an ill-fitting CFA model tends to inflate the factor correlations. In the past, an ill-fitting CFA was a hindrance to further analyses in settings such as structural equation modeling, multiple groups, and multiple timepoints. With the ESEM and PSEM advances shown in this presentation, however, it is possible in all these settings to replace the CFA measurement model with an EFA counterpart. These arguments are further discussed in Asparouhov & Muthén (2009) as well as in Muthén & Asparouhov (2012).

Technical Terms Used in the Talk

- Rotation
  - A way to identify the model and get interpretable factor loadings
    - GEOMIN and many others, orthogonal or oblique (uncorrelated vs correlated factors), target, bi-factor (see EFA Theory section)
- ESEM: Exploratory structural equation modeling
  - An approach to do use EFA instead of CFA measurement models in an SEM context
- PSEM: Penalized Structural Equation Modeling
  - A rotation approach and specification of small effects using priors
- Priors
  - A way to restrict the size of a parameter or sets of parameters
  - Common with Bayes estimation but here used with PSEM for maximum-likelihood and WLSMV estimation
    - GEOMIN priors for loadings with rotation, ALF (alignment loss function) for small effects
- AESEM
  - ESEM alignment of measurements across groups or timepoints

Slide 5 gives a summary of terms that will be used throughout the presentation.

Bullet 1: We've already touched on rotation as a way to make the EFA model identified and to obtain interpretable factor loadings. GEOMIN is the Mplus default rotation and is discussed in the 2009 Asparouhov-Muthén ESEM article. Both orthogonal and oblique rotations are available as listed in the Mplus User's Guide. Graphical representation of rotation examples are shown in the EFA Theory section of the presentation.

Bullet 2: ESEM - exploratory SEM - will be featured prominently in this presentation and its main features are described in the 2009 Asparouhov-Muthén article with several more recent additions. The main advantage is to be able to do SEM with an EFA measurement model instead of CFA.

Bullet 3: PSEM - penalized SEM - has two main uses. One is to do EFA rotations with GEOMIN also in non-standard models such as second-order factor analysis. The more general use is with effects hypothesized to be small using ALF priors.

Bullet 4: Using priors is a way to restrict the size of parameters. PSEM priors are used with maximum-likelihood or - with categorical indicators - WLSMV estimation. GEOMIN priors are used for rotating loading matrices. ALF (Alignment Loss Function) priors are used for small effects.

Bullet 5: AESEM stands for alignment using ESEM. Alignment refers to measurement invariance assuming only configural invariance across groups or timepoints.