Confirmatory Factor Analysis Part 4: Measurement Invariance

- Today's topics:
  - What is measurement invariance?
  - 2 major types of invariance
    - Measurement and Structural
  - Sequence of tests for invariance
    - Metric, Scalar, Residual…. then Structural
  - Wrapping up…

Remember Nested Model Comparisons?

- Can assess whether changing the model (adding or subtracting parameters) impacts the model fit:
  - **Nested models can be compared with χ^2 difference tests**
    - Step 1: Calculate difference of χ^2_{old} and χ^2_{new}
    - Step 2: Calculate difference in df_{old} and df_{new}
    - Compare χ^2_{diff} on df = df_{diff} to critical values table
    - Add 1 parameter? $\chi^2_{diff} > 3.84$, add 2: $\chi^2_{diff} > 5.99$…
  - If **adding** a parameter, the model can either stay the same OR get **better**, as indexed by the χ^2
  - If **removing** a parameter, the model can either stay the same OR get **worse**, as indexed by the χ^2
  - If testing parameters that can't be negative (like variances = 0?), then should use $p < .10$ instead of $p < .05$ (or mixture χ^2 tables)

What is 'Measurement Invariance'?

- Also known as 'factorial invariance' and 'measurement equivalence'
- Concerns the extent to which are the psychometric properties of the observed indicators are transportable (generalizable) across groups or over time
  - In other words, that we are measuring *the same construct in the same way* in different groups or over time
  - In other words, observed scores should depend *only on latent construct scores*, and not on group membership or occasion
  - In other words, that observed differences between groups reflect TRUE differences in the amount of the construct they have
- Relevant concern in many applied settings
  - e.g., across cultures, language, age, modality

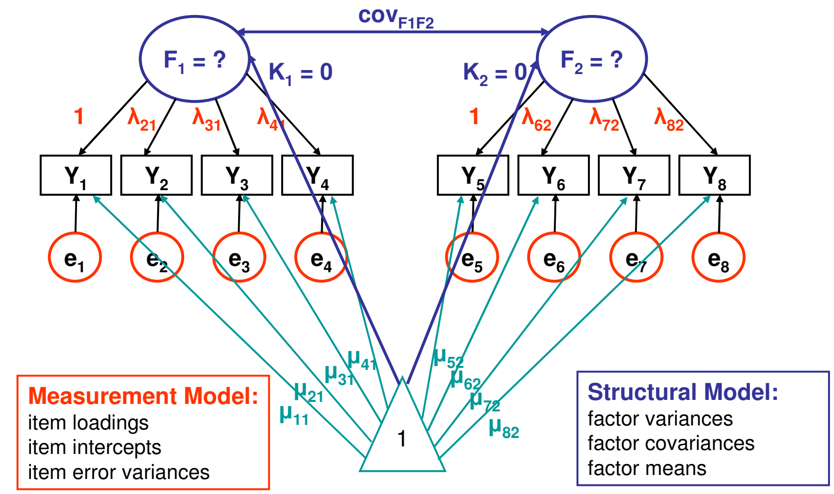
2 Major Types of Factorial Invariance

- **Measurement Invariance** concerns how the items measure the latent construct across groups or over time
  - Measurement model invariance: same factor loadings, same item intercepts, (possibly) same residual variances
  - **Measurement model invariance is a precursor to ANY group comparison (whether it is tested/acknowledged or not)**
- Measurement invariance is often assumed, not tested
  - Even a t-test assumes measurement invariance
  - Modeling change over time assumes measurement invariance
  - People tend to accept this assumption unless you try to use a factor model… then they usually insist on testing invariance

2 Major Types of Factorial Invariance

- **Structural Invariance** concerns how the latent factors are distributed and related in the separate populations
  - Structural model invariance: same factor variances and covariances (or same higher-order structure) and factor means
- Structural invariance may not hold… that may be reality
  - Assuming measurement invariance holds, structural invariance represents ‘real’ differences in the construct across groups/time
  - Structural non-invariance does **not** indicate a problem with your instrument – group structural differences may be of interest
    - e.g., real growth of factors over time
    - e.g., differentiation and de-differentiation

CFA Baseline Model For Multiple Group Invariance Tests



Levels of Invariance across Groups

- **Step 0: Omnibus test of equality of the overall indicator covariance matrix across groups**
  - Do the matrices differ between groups, on the whole?
  - If not, game over. You are done. You have invariance. Congratulations.
  - Many people disagree with the necessity or usefulness of this test to begin testing invariance… why might that be?
    - People also differ in whether invariance should go from top-down or bottom-up directions… I favor bottom-up for the same reason.
- Let’s proceed with an example with 2 factors, 8 items (4 per factor), and 2 groups…
  - Total possible # parameters = 36

Levels of Invariance across Groups

- **Step 1: Test “configural” invariance**
  - Do the groups have the **same factor structure**, broadly construed?
  - Same number of factors, same pattern of free/0 loadings → same conceptual definition of constructs being measured
  - Test factor structure within each group separately, pray they are ‘close enough’ (if not, game over, pretty much)
  - Then estimate a combined model in which all model parameters are allowed to differ across groups
    - This will be the baseline model for further comparisons
    - Model χ^2 and df will be additive across groups
      - Keep in mind that different sample sizes across groups will result in differential weighting of the obtained χ^2 across groups

Configural Invariance Model: Same Factor Structure; All Parameters Separate

Group 1 (last subscript = 1):

- **DF = 36 – 25 = 11**
- $Y_{11} = \mu_{111} + 1F_{11} + e_{11}$
- $Y_{21} = \mu_{211} + \lambda_{211}F_{11} + e_{21}$
- $Y_{31} = \mu_{311} + \lambda_{311}F_{11} + e_{31}$
- $Y_{41} = \mu_{411} + \lambda_{411}F_{11} + e_{41}$
- $Y_{51} = \mu_{521} + 1F_{21} + e_{51}$
- $Y_{61} = \mu_{621} + \lambda_{621}F_{21} + e_{61}$
- $Y_{71} = \mu_{721} + \lambda_{721}F_{21} + e_{71}$
- $Y_{81} = \mu_{821} + \lambda_{821}F_{21} + e_{81}$
- Factor 1 and Factor 2 have own variances, and covariance, but mean fixed to 0

Group 2 (last subscript = 2):

- **DF = 36 – 25 = 11**
- $Y_{12} = \mu_{112} + 1F_{12} + e_{12}$
- $Y_{22} = \mu_{212} + \lambda_{212}F_{12} + e_{22}$
- $Y_{32} = \mu_{312} + \lambda_{312}F_{12} + e_{32}$
- $Y_{42} = \mu_{412} + \lambda_{412}F_{12} + e_{42}$
- $Y_{52} = \mu_{522} + 1F_{22} + e_{52}$
- $Y_{62} = \mu_{622} + \lambda_{621}F_{22} + e_{62}$
- $Y_{72} = \mu_{722} + \lambda_{721}F_{22} + e_{72}$
- $Y_{82} = \mu_{822} + \lambda_{821}F_{22} + e_{82}$
- Factor 1 and Factor 2 have own variances, and covariance, but mean fixed to 0

Levels of Invariance across Groups

- **Step 2: Test "metric" invariance**
  - Also called "**weak** factorial invariance"
  - Do the groups have the **same factor loadings**?
    - Each congeneric item is still allowed to have a different loading
    - Loadings for same item are constrained to equality across groups
  - **Marker items** (that are fixed=1 for identification) are assumed invariant – because they are already fixed, they cannot be tested
    - For this reason, I suggest moving to an alternative specification: **Estimate all factor loadings, but fix the factor variance(s) to 1 in the reference group only** (still free them in the alternative group)
    - This allows us to evaluate ALL loadings and still identify the model

Metric Invariance Model: Same Factor Loadings (saves +6 df here)

Group 1 (last subscript = 1):

- $Y_{11} = \mu_{111} + \boldsymbol{\lambda_{11}F_{11}} + e_{11}$
- $Y_{21} = \mu_{211} + \boldsymbol{\lambda_{21}F_{11}} + e_{21}$
- $Y_{31} = \mu_{311} + \boldsymbol{\lambda_{31}F_{11}} + e_{31}$
- $Y_{41} = \mu_{411} + \boldsymbol{\lambda_{41}F_{11}} + e_{41}$
- $Y_{51} = \mu_{521} + \boldsymbol{\lambda_{52}F_{21}} + e_{51}$
- $Y_{61} = \mu_{621} + \boldsymbol{\lambda_{62}F_{21}} + e_{61}$
- $Y_{71} = \mu_{721} + \boldsymbol{\lambda_{72}F_{21}} + e_{71}$
- $Y_{81} = \mu_{821} + \boldsymbol{\lambda_{82}F_{21}} + e_{81}$
- **Factor 1 and Factor 2 have variance fixed to 1**, free covariance, but mean fixed to 0

Group 2 (last subscript = 2):

- $Y_{12} = \mu_{112} + \boldsymbol{\lambda_{21}F_{12}} + e_{12}$
- $Y_{22} = \mu_{212} + \boldsymbol{\lambda_{21}F_{12}} + e_{22}$
- $Y_{32} = \mu_{312} + \boldsymbol{\lambda_{31}F_{12}} + e_{32}$
- $Y_{42} = \mu_{412} + \boldsymbol{\lambda_{41}F_{12}} + e_{42}$
- $Y_{52} = \mu_{522} + \boldsymbol{\lambda_{52}F_{22}} + e_{52}$
- $Y_{62} = \mu_{622} + \boldsymbol{\lambda_{62}F_{22}} + e_{62}$
- $Y_{72} = \mu_{722} + \boldsymbol{\lambda_{72}F_{22}} + e_{72}$
- $Y_{82} = \mu_{822} + \boldsymbol{\lambda_{82}F_{22}} + e_{82}$
- Factor 1 and Factor 2 have free variances and covariance, but mean fixed to 0 (for now)

Metric Invariance

- Compare fit of metric invariance to configural invariance model:
  - Does the model fit not get worse (***X^2 not go up significantly*)?**
    - We are taking parameters away, so it can only get worse…
    - Don't forget to fix variance=1 in reference group only (free in other group)! Otherwise you are imposing a structural constraint too by accident!
  - Either way, inspect the modification indices to see if there are any items whose loadings want to differ between groups
    - Retest the model as needed after releasing one loading at a time
- Do you have at least partial* metric invariance?
  - Congrats! Your construct is measured in the same way across groups
  - If not, it doesn't make sense to evaluate how relationships involving the factor differ across groups (because the factor itself differs)

* No real consensus on how much is "partial", but at least 2!

Levels of Invariance across Groups

- **Step 3: Test "scalar" ("strong") invariance**
  - Do the groups have the **same item intercepts**?
    - Each congeneric item is allowed to have a different intercept
    - Intercepts for same item are constrained to equality across groups
    - Scalar invariance model says factor mean differences cause the item mean differences (*but the item intercepts should still be the same*)
  - If you use marker intercepts (that are fixed=0 for identification), they are assumed invariant – because they are already fixed
    - So we will estimate all intercepts, **but constrain the factor mean(s) to 0 in the reference group** so we can evaluate all intercepts
  - Some folks might say that scalar invariance is not really necessary unless you plan on comparing mean differences…
    - Scalar invariance doesn't always get tested as a result
    - Pry better to error on the side of caution and examine it anyway

Scalar Invariance Model:
Same Indicator Intercepts (saves +6 df here)

<u>Group 1 (last subscript = 1):</u>

- $Y_{11} = \boldsymbol{\mu_{11}} + \lambda_{11}F_{11} + e_{11}$
- $Y_{21} = \boldsymbol{\mu_{21}} + \lambda_{21}F_{11} + e_{21}$
- $Y_{31} = \boldsymbol{\mu_{31}} + \lambda_{31}F_{11} + e_{31}$
- $Y_{41} = \boldsymbol{\mu_{41}} + \lambda_{41}F_{11} + e_{41}$
- $Y_{51} = \boldsymbol{\mu_{51}} + \lambda_{52}F_{21} + e_{51}$
- $Y_{61} = \boldsymbol{\mu_{62}} + \lambda_{62}F_{21} + e_{61}$
- $Y_{71} = \boldsymbol{\mu_{72}} + \lambda_{72}F_{21} + e_{71}$
- $Y_{81} = \boldsymbol{\mu_{82}} + \lambda_{82}F_{21} + e_{81}$

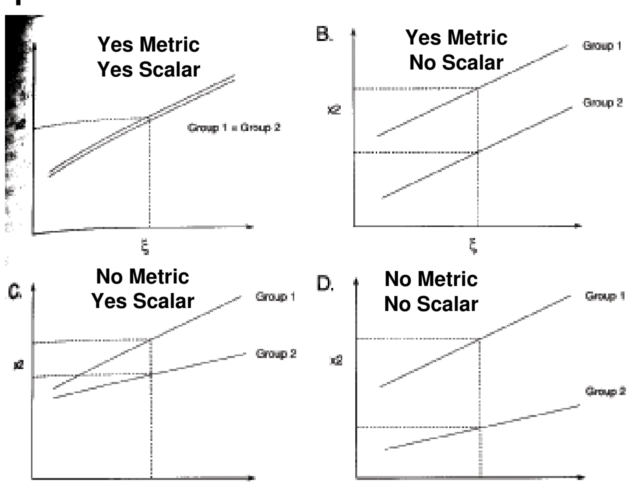
- Factor 1 and Factor 2 have variance fixed to 1, free covariance, but mean fixed to 0 (for group 1)

<u>Group 2 (last subscript = 2):</u>

- $Y_{12} = \boldsymbol{\mu_{11}} + \lambda_{11}F_{12} + e_{12}$
- $Y_{22} = \boldsymbol{\mu_{21}} + \lambda_{21}F_{12} + e_{22}$
- $Y_{32} = \boldsymbol{\mu_{31}} + \lambda_{31}F_{12} + e_{32}$
- $Y_{42} = \boldsymbol{\mu_{41}} + \lambda_{41}F_{12} + e_{42}$
- $Y_{52} = \boldsymbol{\mu_{51}} + \lambda_{52}F_{22} + e_{52}$
- $Y_{62} = \boldsymbol{\mu_{62}} + \lambda_{62}F_{22} + e_{62}$
- $Y_{72} = \boldsymbol{\mu_{72}} + \lambda_{72}F_{22} + e_{72}$
- $Y_{82} = \boldsymbol{\mu_{82}} + \lambda_{82}F_{22} + e_{82}$

- Factor 1 and Factor 2 have own variances, and covariance, **but mean is now free (for group 2) and represents factor mean diff**

Implications of Non-Invariance



Latent Factor

Latent Factor

Scalar Invariance

- Only test those intercepts for which metric invariance holds
  - Different slopes can create different intercepts as an artifact
- Compare fit of scalar invariance to metric invariance model:
  - Does the model fit not get worse (***X^2 not go up significantly***)**?**
  - Either way, inspect the modification indices to see if there are any items whose intercepts want to differ between groups
    - Retest the model as needed after releasing one intercept at a time
- Do you have at least partial* scalar invariance?
  - Your construct accounts for the item mean differences across groups
  - If not, it doesn't make sense to evaluate mean differences in the factor across groups (because other things create item mean differences)

* No real consensus on what is "partial", but at least 2!

Levels of Invariance across Groups

- **Step 4: Test "residual variance" invariance**
  - Also called "**strict** factorial invariance"
  - Do the groups have the **same item residual variances**?
    - Each congeneric item is still allowed to have a different residual variance
    - Residual variances for the same item are constrained to equality across groups
    - Testing residual variances is the last step in assessing measurement invariance
      - People disagree as to whether or not this is necessary
      - Note: Equal residual variances are commonly mis-interpreted to mean "equal reliabilities" – this is ONLY the case if the factor variances are the same across groups, too
        - » We test that one next…

Residual Variance Invariance Model: Same Error Variances (saves +8 df here)

Group 1 (last subscript = 1):

- $Y_{11} = \mu_{11} + \lambda_{11}F_{11} + \mathbf{e_1}$
- $Y_{21} = \mu_{21} + \lambda_{21}F_{11} + \mathbf{e_2}$
- $Y_{31} = \mu_{31} + \lambda_{31}F_{11} + \mathbf{e_3}$
- $Y_{41} = \mu_{41} + \lambda_{41}F_{11} + \mathbf{e_4}$
- $Y_{51} = \mu_{52} + \lambda_{52}F_{21} + \mathbf{e_5}$
- $Y_{61} = \mu_{62} + \lambda_{62}F_{21} + \mathbf{e_6}$
- $Y_{71} = \mu_{72} + \lambda_{72}F_{21} + \mathbf{e_7}$
- $Y_{81} = \mu_{82} + \lambda_{82}F_{21} + \mathbf{e_8}$

- Factor 1 and Factor 2 have variance fixed to 1, free covariance, but mean fixed to 0 (for group 1)

Group 2 (last subscript = 2):

- $Y_{12} = \mu_{11} + \lambda_{11}F_{12} + \mathbf{e_1}$
- $Y_{22} = \mu_{21} + \lambda_{21}F_{12} + \mathbf{e_2}$
- $Y_{32} = \mu_{31} + \lambda_{31}F_{12} + \mathbf{e_3}$
- $Y_{42} = \mu_{41} + \lambda_{41}F_{12} + \mathbf{e_4}$
- $Y_{52} = \mu_{52} + \lambda_{52}F_{22} + \mathbf{e_5}$
- $Y_{62} = \mu_{62} + \lambda_{62}F_{22} + \mathbf{e_6}$
- $Y_{72} = \mu_{72} + \lambda_{72}F_{22} + \mathbf{e_7}$
- $Y_{82} = \mu_{82} + \lambda_{82}F_{22} + \mathbf{e_8}$

- Factor 1 and Factor 2 have own variances, and covariance, but mean is free (for group 2) and represents factor mean diff

Residual Invariance

- Only test those residual variances for which metric and scalar invariance hold
- Compare fit of residual invariance to scalar invariance model:
  - Does the model fit not get worse (***X^2 not go up significantly***)**?**
  - Either way, inspect the modification indices to see if there are any items whose residual variances want to differ between groups
    - Retest the model as needed after releasing one residual variance at a time
- Do you have at least partial* residual invariance?
  - Your groups have the same amount of "not the factor" in each item
  - If not??? Ongoing debate about the necessity of this

* No real consensus on what is "partial", but at least 2!

Next, Structural Invariance

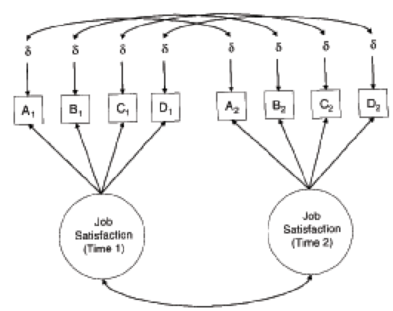
- Are the **factor variances** the same across groups? (+2 df)
  - Fix the factor variance in the alternative group to 1 (as in the ref group)
  - Did model fit get worse? If so, the groups differ in their factor variances
- Is the **factor covariance** the same across groups? (+1 df)
  - Fix the factor covariances equal across groups, did model fit get worse?
  - Factor correlation will only be the same across groups if the factor variances are the same, too
- Are the **factor means** the same across groups? (+2 df)
  - Fix the factor mean in the alternative group to 0 (as in the ref group)
  - Did model fit get worse? If so, the groups differ in their factor means
- **It is not problematic if structural invariance doesn't hold.**
  - Given measurement invariance, this is a **substantive issue** about differences in the latent trait amounts and relations (and that's ok).

Setting up the Invariance Model

- Testing invariance across independent groups? You need a multiple-group approach:
  - Estimate factor model for each group at once, but only the variables per group are related within each subgroup model
  - An alternative approach, MIMIC models in which the grouping variable is entered as a predictor, do not allow testing of equality of factor loadings or factor variances
- Testing invariance across time?
  - Put all the observed indicators into the SAME MODEL
    - Correlate errors from same indicators across time
    - Model gets big and complicated quickly
  - Multiple group approach is not appropriate because observations from same person are not independent

Longitudinal Invariance Example

Residual covariances among same items at different times are usually included by default



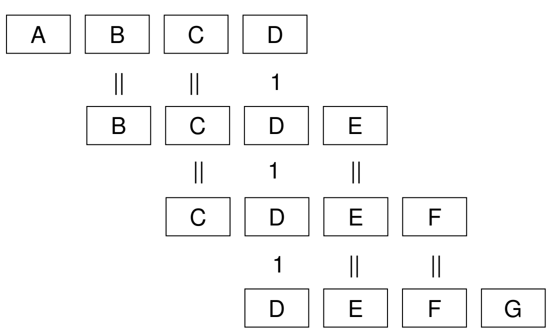
Same series of constraints as in multiple group approach:

Measurement Invariance (should hold if you want to compare change in factors):
Factor loadings (metric, weak)
Item intercepts (scalar, strong)
Residual variances (error, strict)

Structural Invariance (usually not expected to hold):
Factor variances
Factor covariance
Factor means

Partial Invariance over Time

- Partial invariance can be used to 'hold together' a latent construct whose observed measures change over time



Wrapping Up…

- The process of testing factorial invariance has two distinct parts:
  - **Measurement invariance**: Is your construct being measured in the same way? Let's hope so!
    - Better hope for at least "partial" invariance… otherwise, game over.
  - **Structural invariance**: Do your groups differ in their distribution and/or means of the construct? Let's find out!
    - Structural differences are real and interpretable differences *given measurement invariance of the constructs*
- Measurement invariance is always assumed in any statistical analysis…
  - But can be tested explicitly in a latent trait modeling framework