Scale Invariance Testing Differential Item Functioning (DIF): The use of MIMIC Models

By R. L. Brown, Ph.D.

Differential Item Functioning

Differential item functioning (DIF) occurs when people from different groups (e.g gender or ethnicity) with the same underlying latent trait score have a different probability of responding to an item in a particular way.

Group differences in item responses (or on latent variables) do not reflect DIF per se (e.g females score higher than males on a particular item or scale).

DIF is only present if people from different groups with the same underlying ability (or trait level) have a different probability of response. Meaning, the thermometer is different for different groups.

DIF – Measurement Non-Invariance

If the probability of item response is the same (among different sub-groups with the same underlying ability) measurement invariance is assumed.

If the probability of response is different (among different sub-groups with the same underlying ability) than measurement non-invariance (variant) is assumed.

Differential Item Functioning

- For construct validity items of a scale ideally should have little or no DIF
- Items should function in the same way across subgroups of respondents who have the same underlying ability (or level on the latent trait)
- Presence of DIF may compromise comparison across subgroups – give misleading results
- Confound interpretation of observed variables

The Challenge for Developing Culturally Sensitive Instruments

- Populations may give culturally different responses to questions.

- The result is that one group may have higher scores than another group, not because they have higher levels of a trait but because of differences in their cultural beliefs.

- This is known as Differential Item Functioning (DIF) or item bias.

Classic DIF example from the literature

- Azocar, Arean, Miranda, & Munoz (2001) found on the Beck's Depression Inventory:
  - Regardless of the level of depression, Hispanics are more likely to endorse "I feel like crying" than non-Hispanics.
  - Latino culture has practices and symbolisms that portray crying as an acceptable behavior reflecting suffering.

Impact: Differential Item Functioning

- DIF items are a serious threat to the validity of the scale to measure the trait levels of members from different populations or groups.
- Scales containing such items may have reduced validity for between-group comparisons, because their scores may be indicative of a variety of attributes other than those the scale is intended to measure.

DIF Item Conclusions

- Quantitative Methods should co-exist with both qualitative and cognitive methods to build and revise instruments.

- While quantitative methods may detect DIF, it takes review by experts or cognitive interviewing with respondents to determine why an item is exhibiting DIF.

- What do you do with the DIF item?
  - Rewrite the item.
  - Remove the item.

What do you do with the DIF item?

- Remove the item.
  - 1) If you have a large item pool and the item can be replaced with a item measuring similar threshold / discrimination parameters
  - 2) Dropping items might adversely affect the content validity of the instrument.
  - 3) May end up with an instrument that is not comparable to other research using that instrument

Look for causes of DIF

- Rewrite item.
- What do all the DIF items have in common e.g.
  - Are they all negatively or positively worded
  - Are they all at end of study
  - Readability etc
- How do they differ from the invariant items?

MIMIC Model

(Multiple Indicator Multiple Causes)

- First, establish a CFA measurement model.
- Second, add covariate(s) to the model to examine their direct effects on the latent factors and indicators (DIF).

Mimic Model Stages of identifying potential DIF

- Run CFA model without covariates
- Include MIMIC model (add covariate but no direct item effects)
- Add paths from covariate to indicator constrained to 0 - i.e. assuming there is no direct effect (Y1 on ITEM@0)
- Check modification indices
- Add direct path from covariate to indicator for indicator with highest modification indices - rerun model
- Repeat steps 4 & 5 until there are no further significant modification indices , evaluate model fit and significance of the direct effects

CFA MIMIC Model
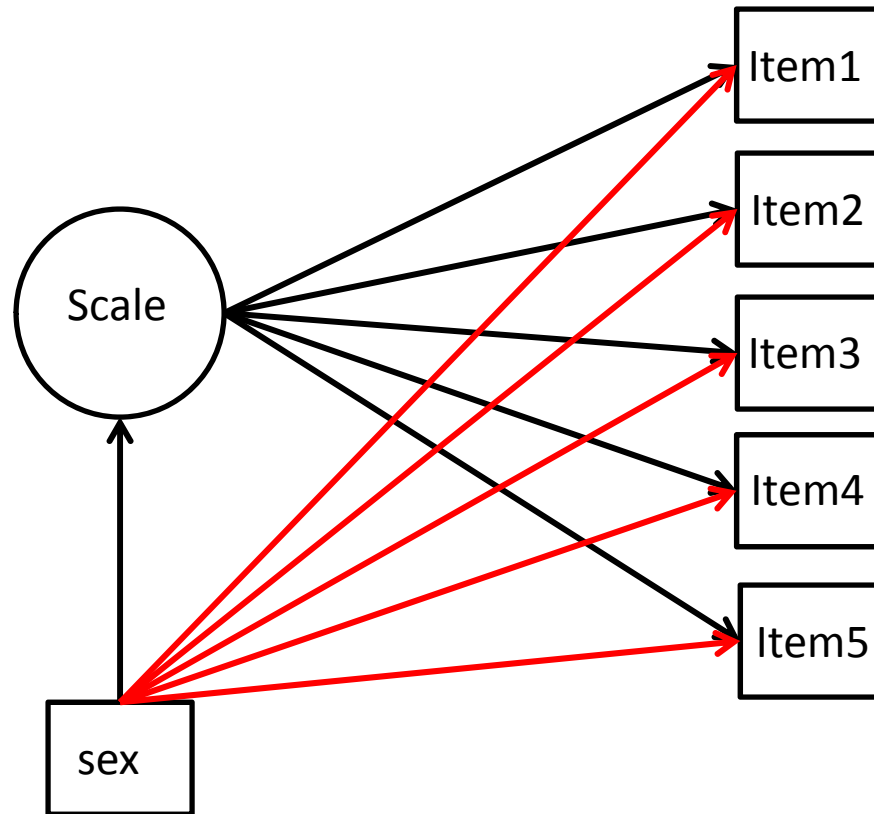
Mplus Example

USEVARIABLES are item1 item2 item3 item4 item5 sex;
CATEGORICAL are item1 item2 item3 item4 item5 ;

   Missing are all .;

 ANALYSIS:
     ESTIMATOR IS wlsmv;
     ITERATIONS = 1000;
     CONVERGENCE = 0.00005;

  MODEL:
     CONDUCT by item1 item2 item3 item4 item5;
     CONDUCT on sex;
     item1 – item5 on sex;

OUTPUT: SAMPSTAT STANDARDIZED RES MOD(10) ;

Check Modification Indices

	M.I.	E.P.C.	Std E.P.C.	StdYX E.P.C.
ON Statements				
ITEM1 ON SEX	82.578	-0.354	-0.354	-0.176
ITEM2 ON SEX	23.839	0.143	0.143	0.071

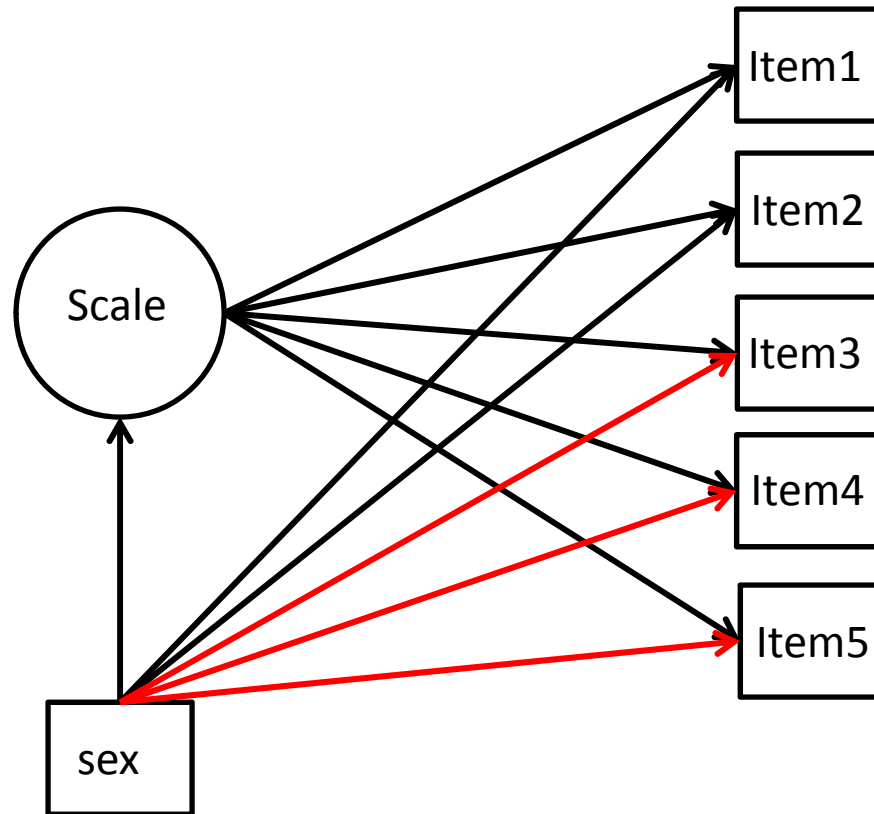Include item with largest MI as a direct effect in model

ITEM1 on SEX;
ITEM2- ITEM5 ON SEX@0;

Recheck mod indices and repeat if necessary

M.I. = Modification Indices are the amount chi-square will drop if this parameter if freed in the model.

E.P.C. = Expected parameter change when freed

CFA MIMIC Model

Final MIMIC Results

Initial Mimic Model (no direct effects)

	Estimate	S.E.	Est./S.E.	P-Value	Std
Scale ON SEX	-0.126	0.022	-5.789	0.000	-0.169

Add 2 item direct effects

	Estimate	S.E.	Est./S.E.	P-Value	Std
Scale ON SEX	-0.113	0.022	-5.203	0.000	-0.152
ITEM1 ON SEX	-0.336	0.044	-7.597	0.000	-0.336
ITEM2 ON SEX	0.112	0.032	3.481	0.000	0.112

DIF (Differential Item Functioning)