Measurement in Heterogeneous Populations: Beyond Multiple Groups

Daniel J. Bauer, Patrick J. Curran, Andrea M. Hussong

Measurement Invariance

- We are often interested in analyzing data from heterogeneous populations, e.g., individuals that differ in age, gender, ethnicity, or culture
- Measurement Invariance is required for scores to be comparable across different populations
  - Observed items/indicators must have the same relationship to the underlying characteristic to be measured (latent factor) for all individuals.
  - Items that do not meet this requirement are said to display Differential Item Functioning (DIF)

Measurement Invariance Over Groups

- Typically, measurement invariance / DIF is evaluated by testing whether item parameters are equal over groups.
  - Equal intercepts and factor loadings in factor analysis
  - Equal difficulty and discrimination parameters in IRT models
- Usually, invariance / DIF is tested within a multiple groups factor analysis or multiple groups IRT
- These models assume that population heterogeneity can be summarized via a single discrete variable
  - The majority of applications compare measures across two groups

Beyond Multiple Groups

- Measurement invariance is a broader concept, however, that does not apply only across levels of a single discrete grouping variable
- In some cases, it may be of interest to examine measurement invariance as a function of multiple discrete variables simultaneously, or across levels of a continuous variable.
- In these cases, the usual multiple groups factor model or IRT model do not serve us well.

Motivating Example

- Wish to pool data from two studies of alcohol use.
  - AFDP (Chassin) – Community-based sample, Arizona, Ages 10-22 in Waves 1-4, N=454
  - AHBP (Sher) – College sample, Missouri, Ages 17-22 in Waves 1-4, N=484
- Measurement properties may differ across studies
  - differences in the wording and presentation of items
  - demographic differences between the sampled populations
- Measurement properties may differ across age
  - Use of alcohol by a child represents somewhat deviant behavior whereas use of alcohol by a young adult is more normative.
- Need to simultaneously evaluate invariance with respect to study (discrete) and age (continuous)

A General Modeling Approach

- Bauer & Hussong (2009) presented a Moderated Nonlinear Factor Analysis (MNLFA) model that includes:
  - Generalized modeling features to accommodate alternative response scales across items (e.g., mixed response scales)
  - Simultaneous moderation of item parameters by discrete and/or continuous variables (e.g., study, age) for DIF detection

Moderated Nonlinear Factor Analysis

- The MNLFA model can be written as

$$g_i(\mu_{ij}) = \nu_{ij} + \lambda_{ij}\eta_j \qquad \eta_j \sim N(\alpha_j, \psi_j)$$

where

i is the item/indicator

j is the person

g is the link function

and μ_{ij} is the expected value of the conditional distribution for the item/indicator

- The use of different link functions permits indicators of mixed scale types (e.g., binary, continuous, count, etc).

Moderation of Parameters

- Differences in latent variable distribution:

$$\alpha_j = \alpha_0 + \sum_{q=1}^{Q} \alpha_q x_{qj} \qquad \psi_j = \psi_0 \exp\left(\sum_{q=1}^{Q} \omega_q x_{qj}\right)$$

- Differential item functioning (non-invariance):

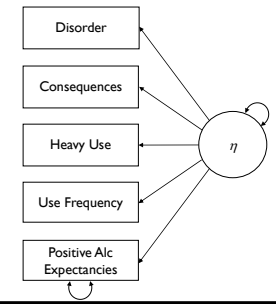
$$\nu_{ij} = \nu_{0i} + \sum_{q=1}^{Q} \nu_{qi} x_{qj} \qquad \lambda_{ij} = \lambda_{0i} + \sum_{q=1}^{Q} \lambda_{qi} x_{qj}$$

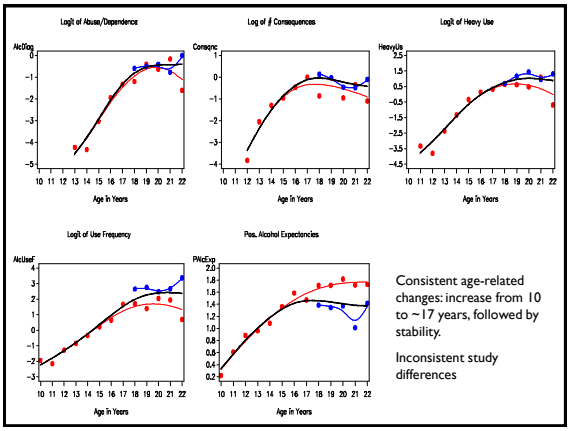
(If conditional distributions include variance or threshold parameters, these parameters can also be expressed as a linear or log-linear function of observed covariates)

Factor Model for Alcohol Involvement

$$g_i(\mu_{ij}) = \nu_{ij} + \lambda_{ij}\eta_j$$

Conditional Distribution	Link Function
Bernoulli	Logit
Poisson	Log
Multinomial	Cumulative Logit
Multinomial	Cumulative Logit
Censored Normal	Identity





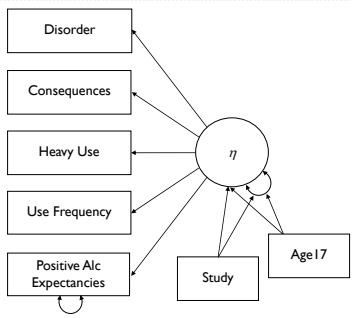
Consistent age-related changes: increase from 10 to ~17 years, followed by stability.

Inconsistent study differences

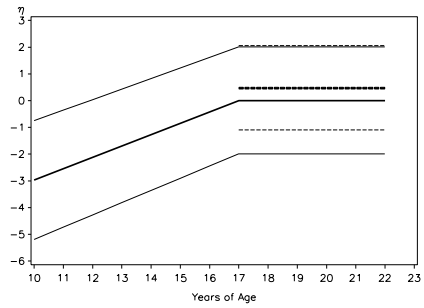
Age and Study Differences in Factor Scores

Piecewise Linear Model for Factor Mean

Log-Linear Model for Factor Variance



Age and Study Differences in Factor Scores

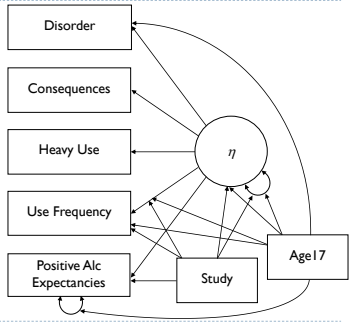


DIF Detection

Evaluated DIF in stepwise fashion

Linear model for intercepts and loadings

Log-linear model for expectancies residual variance



Scoring

- As in conventional IRT, scoring can be done by computing expected a posteriori (EAP) or modal a posteriori (MAP) values
- EAPs and MAPs are computed using information about population heterogeneity in the distribution of the latent factor as well as potential DIF in the items/indicators
- Scores are comparable across individuals in heterogeneous populations to the extent that invariance is supported in the model

Conclusions

- Multiple groups factor models and IRT models are currently the primary tools used in invariance testing
- These models, however, evaluate invariance only across a single grouping variable
- In some cases, population heterogeneity is not so neatly summarized
- The MNLFA proposed by Bauer & Hussong (2009) permits simultaneous evaluation of measurement invariance across multiple discrete or continuous variables

This research was supported by NIH grant DA015398 (PIs: Hussong & Curran)