

Constructing Common Measures for Integrative Data Analysis: Model Performance

*Patrick Curran, Nisha Gottfredson, Veronica Cole,
Daniel Bauer & Andrea Hussong*



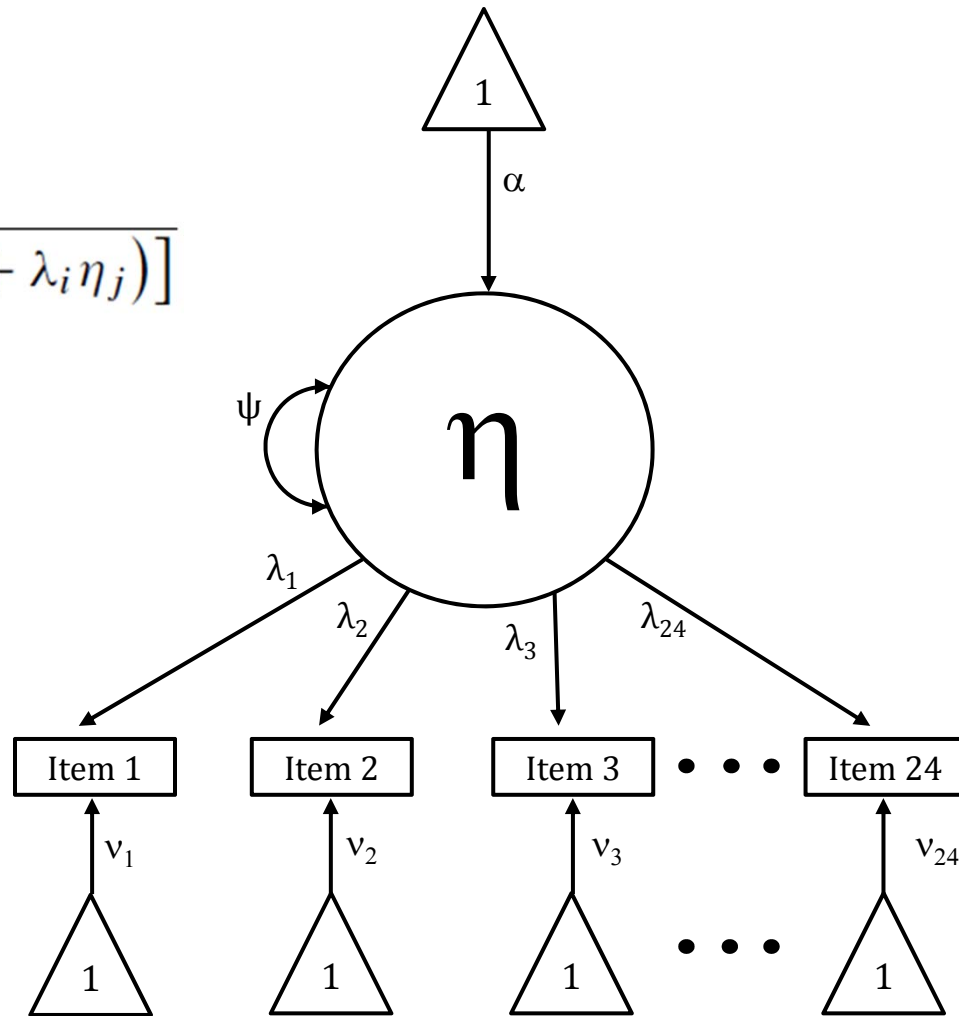
THE UNIVERSITY
of NORTH CAROLINA
at CHAPEL HILL

Integrative Data Analysis

- The fitting of models to data that have been pooled over two or more studies
- Many advantages / many challenges
- **Core goal:** *obtain scores based on different items drawn from different studies anchored to same metric for subsequent modeling*
- We have used MNLFA to accomplish this, but we have not yet studied with controlled design

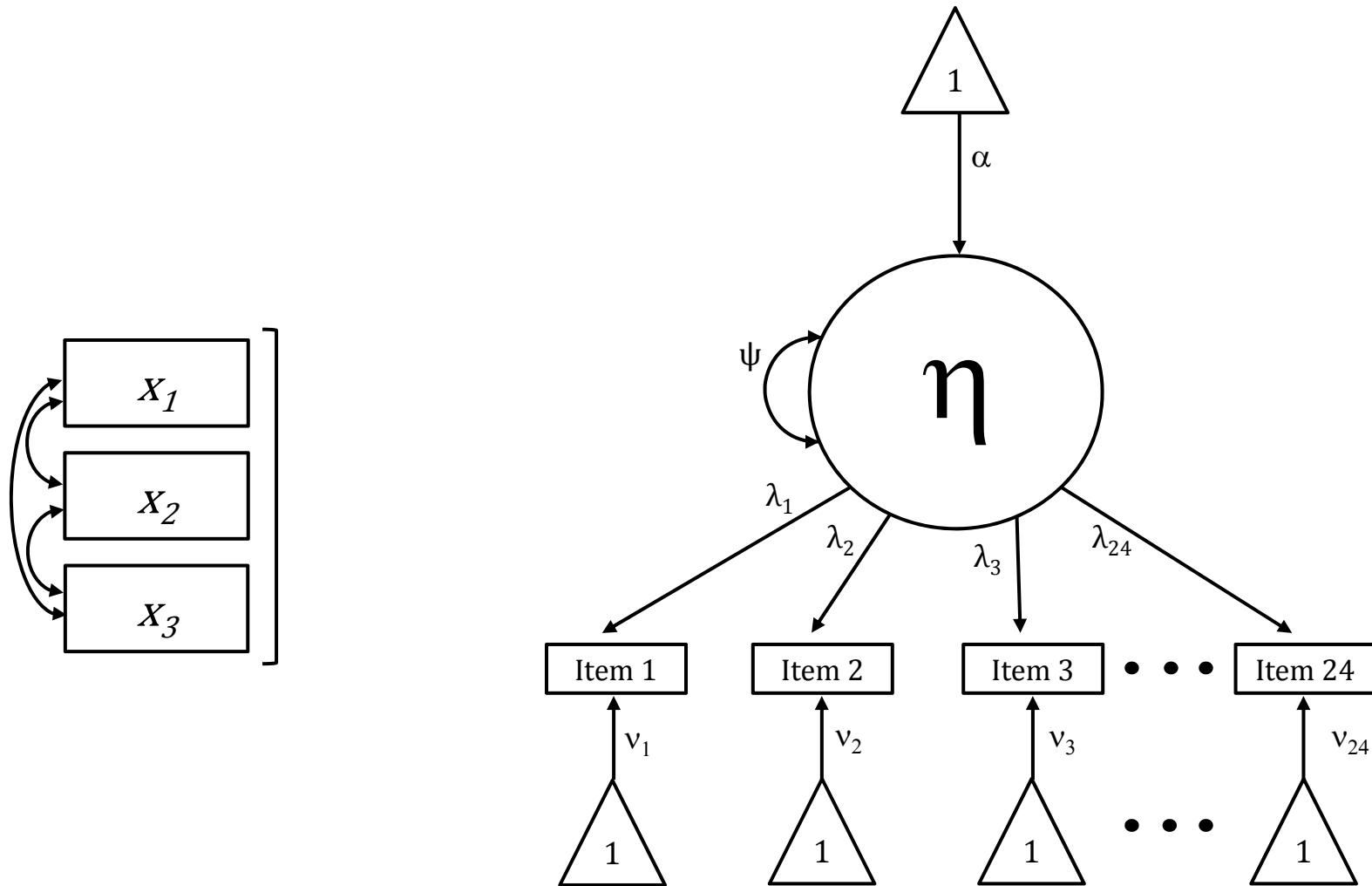
Single Group 2PL-IRT / Binary CFA

$$\mu_{ij} = \frac{1}{1 + \exp[-(v_i + \lambda_i \eta_j)]}$$

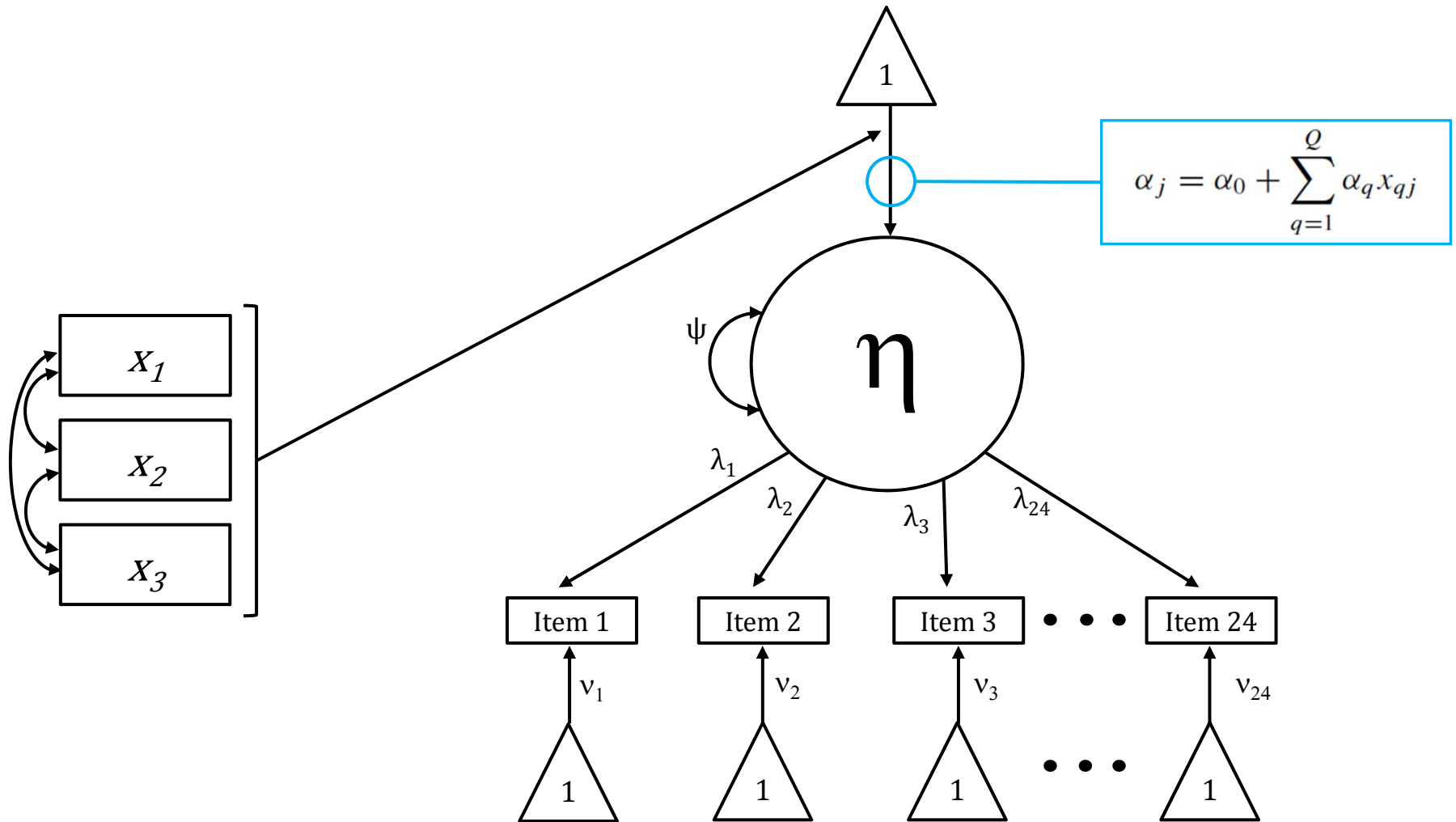


... but assumes invariance over group

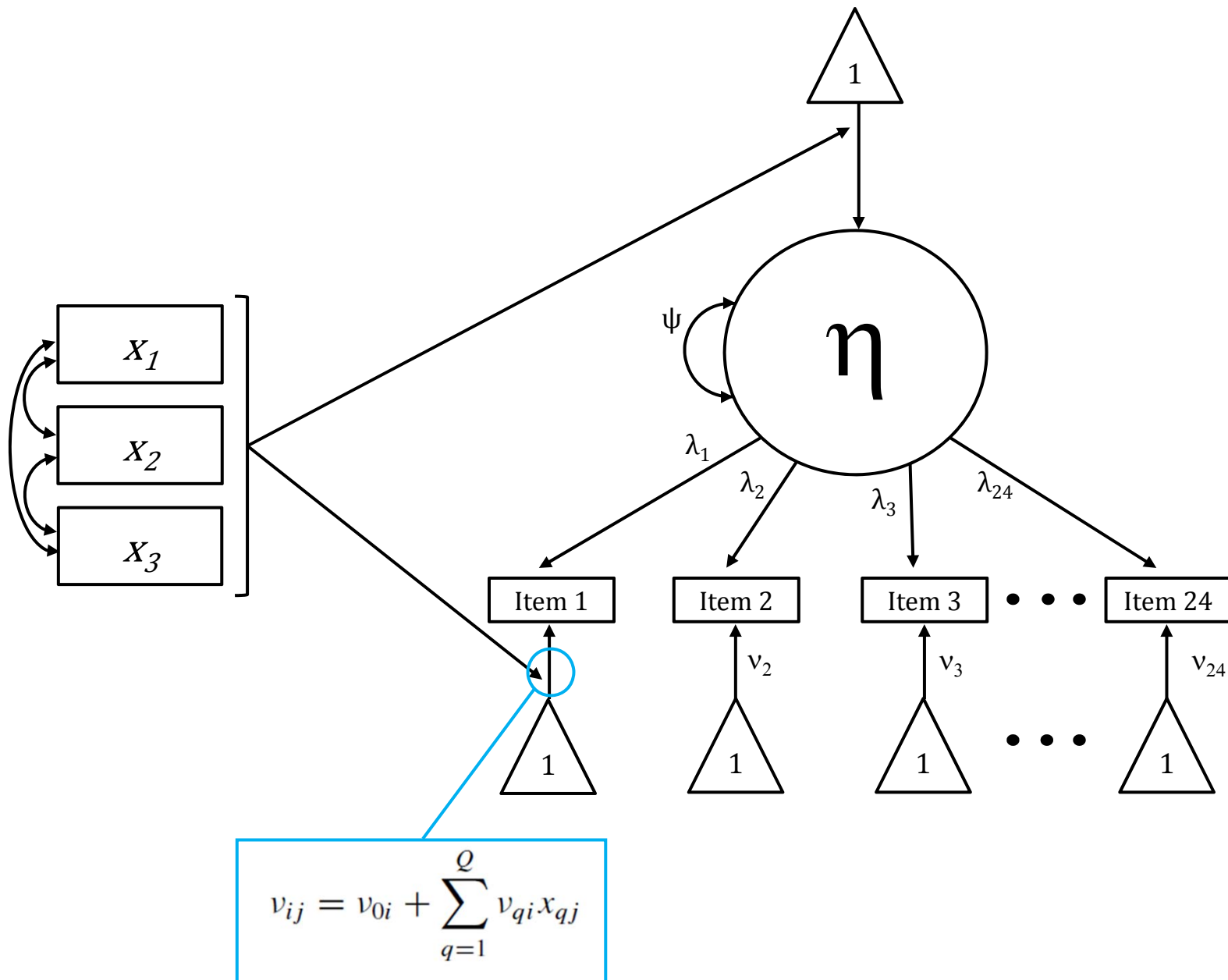
Moderated Nonlinear Factor Analysis



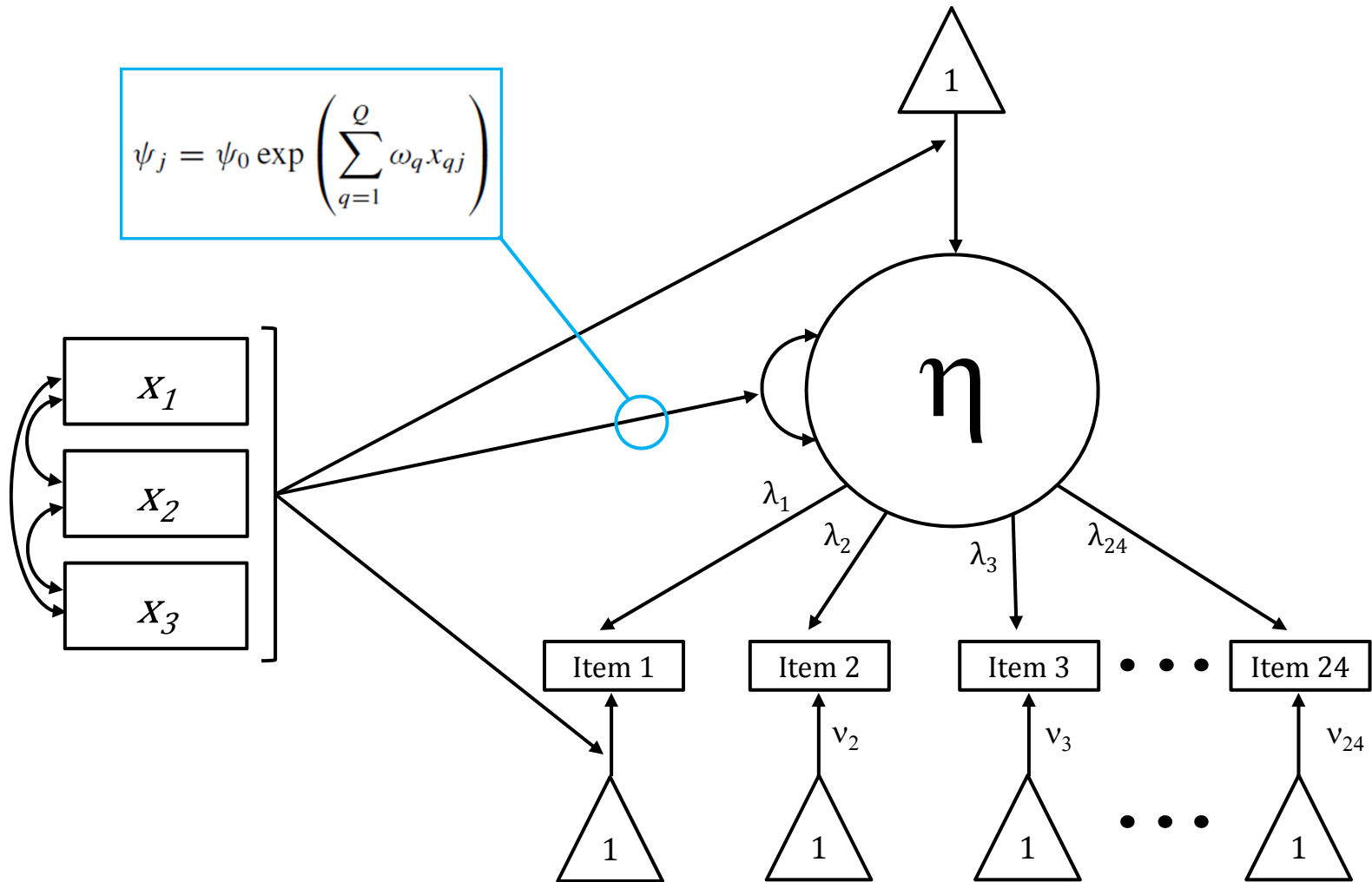
Factor Mean



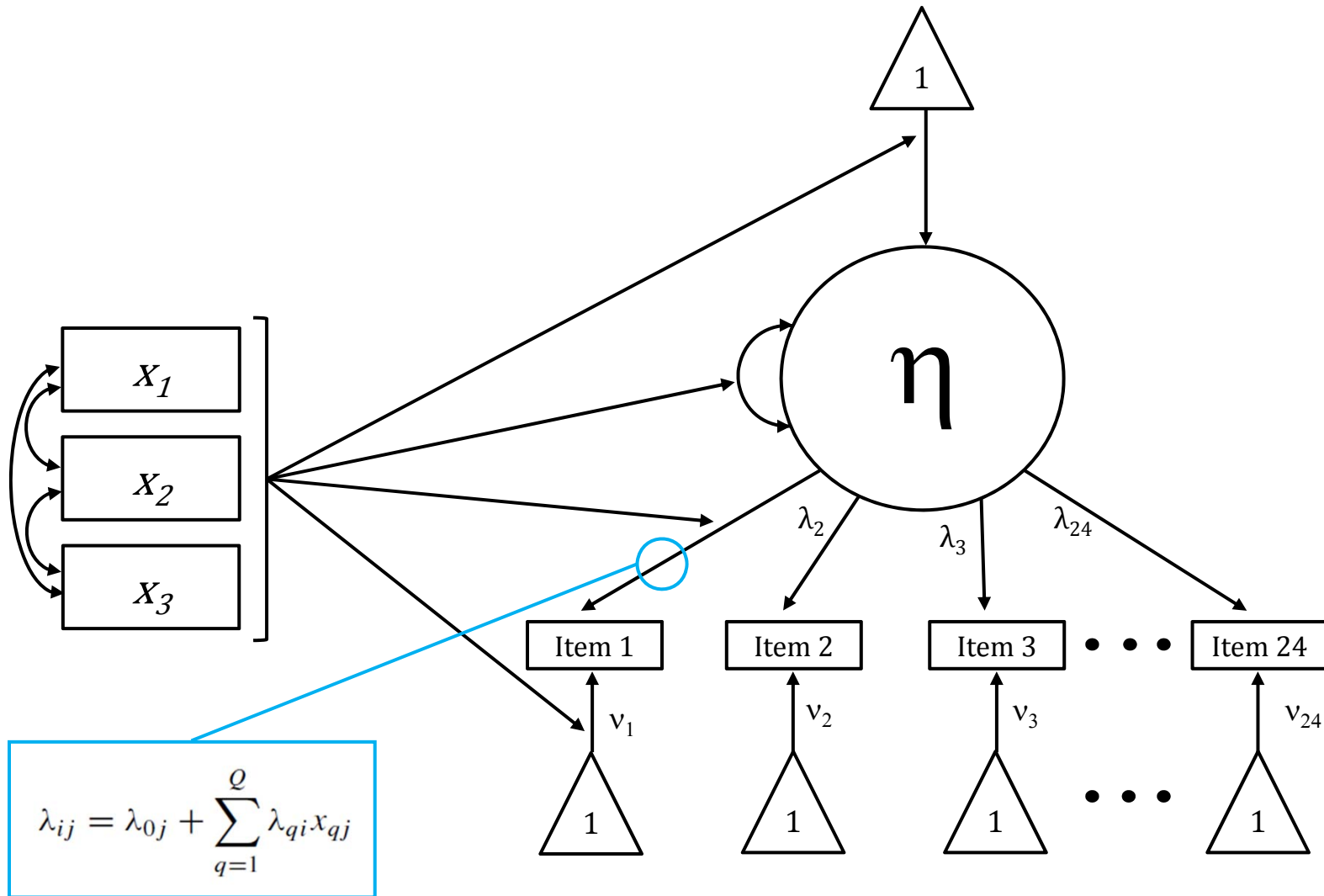
Item Intercept



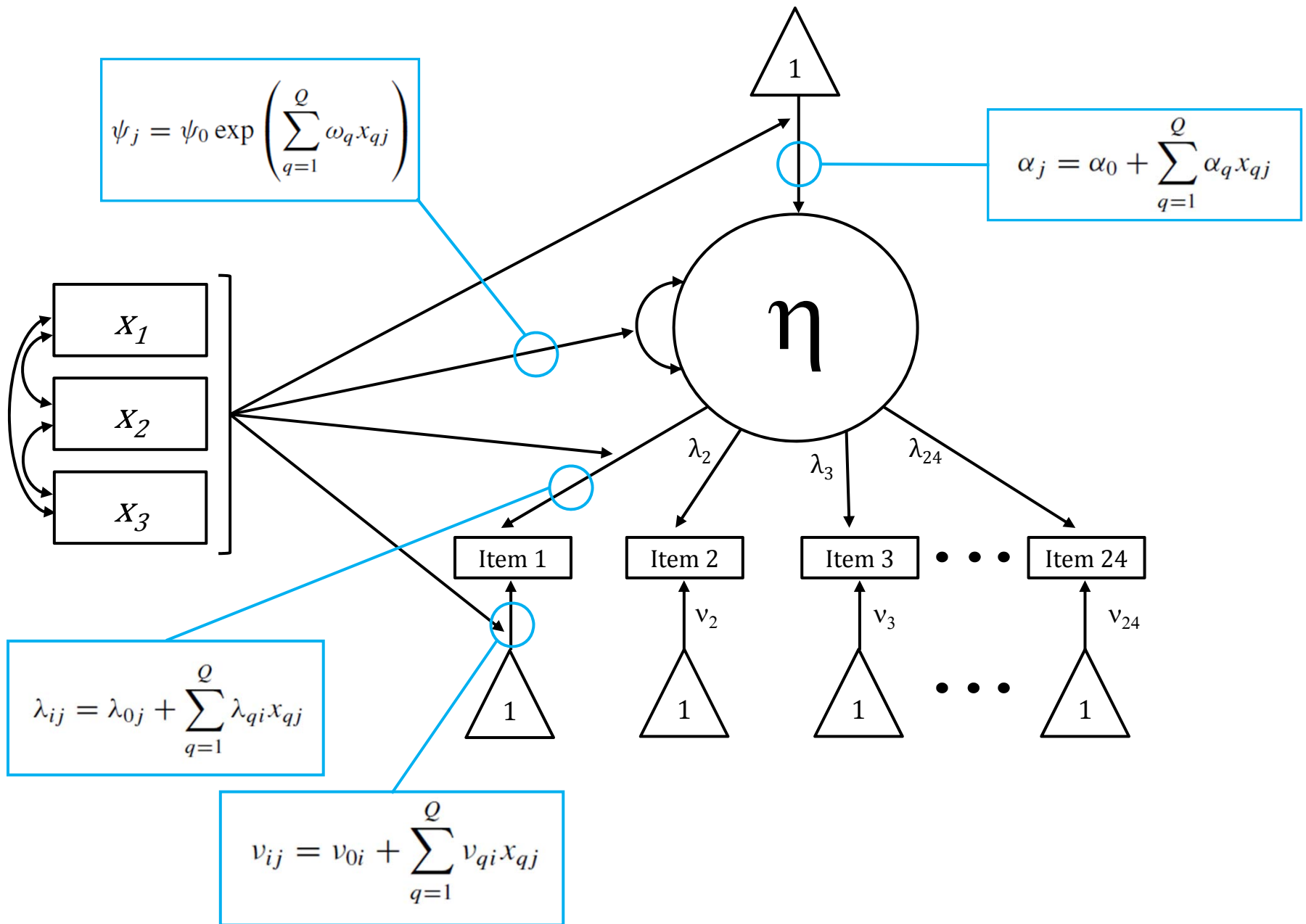
Factor Variance

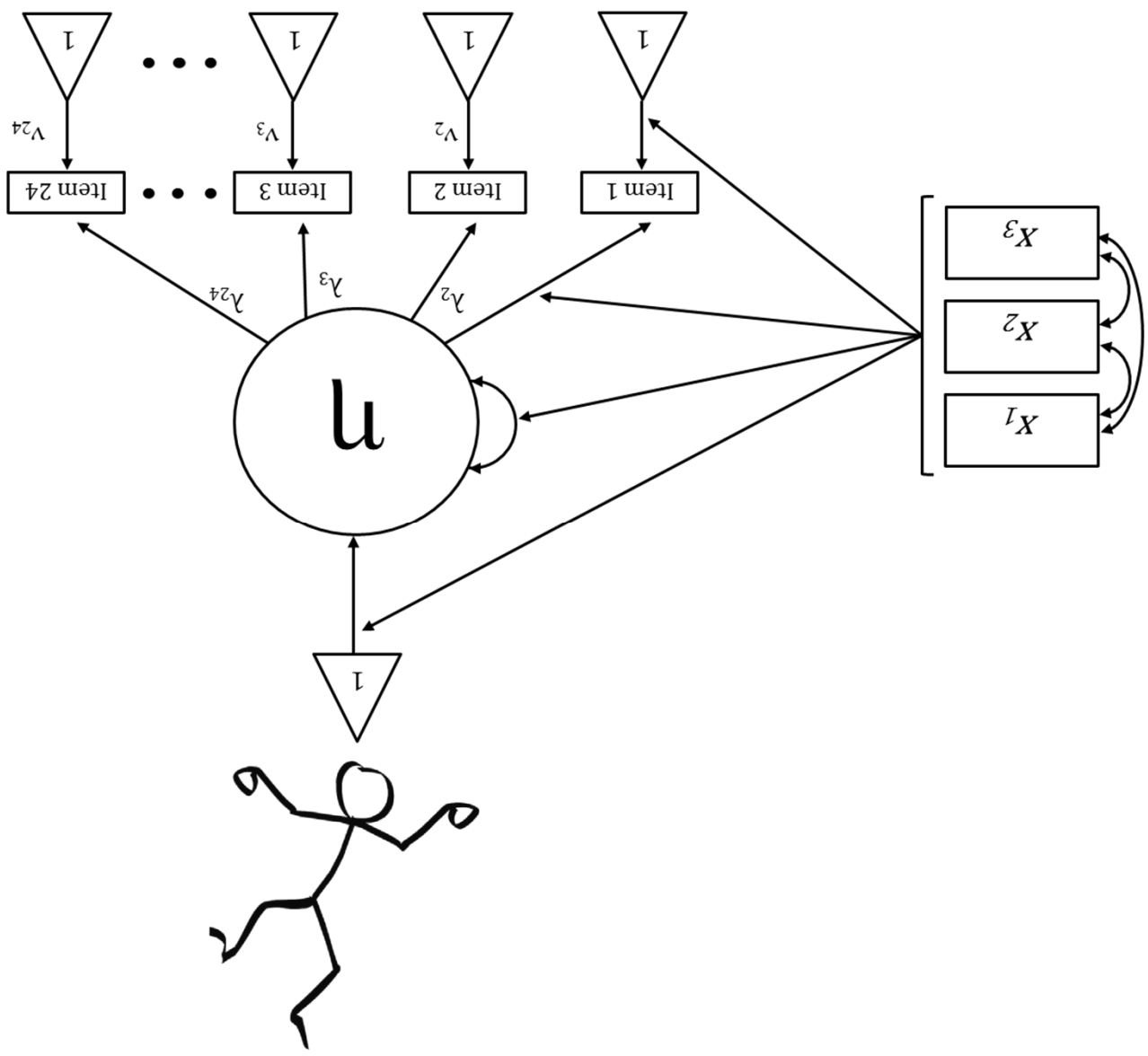


Item Loading

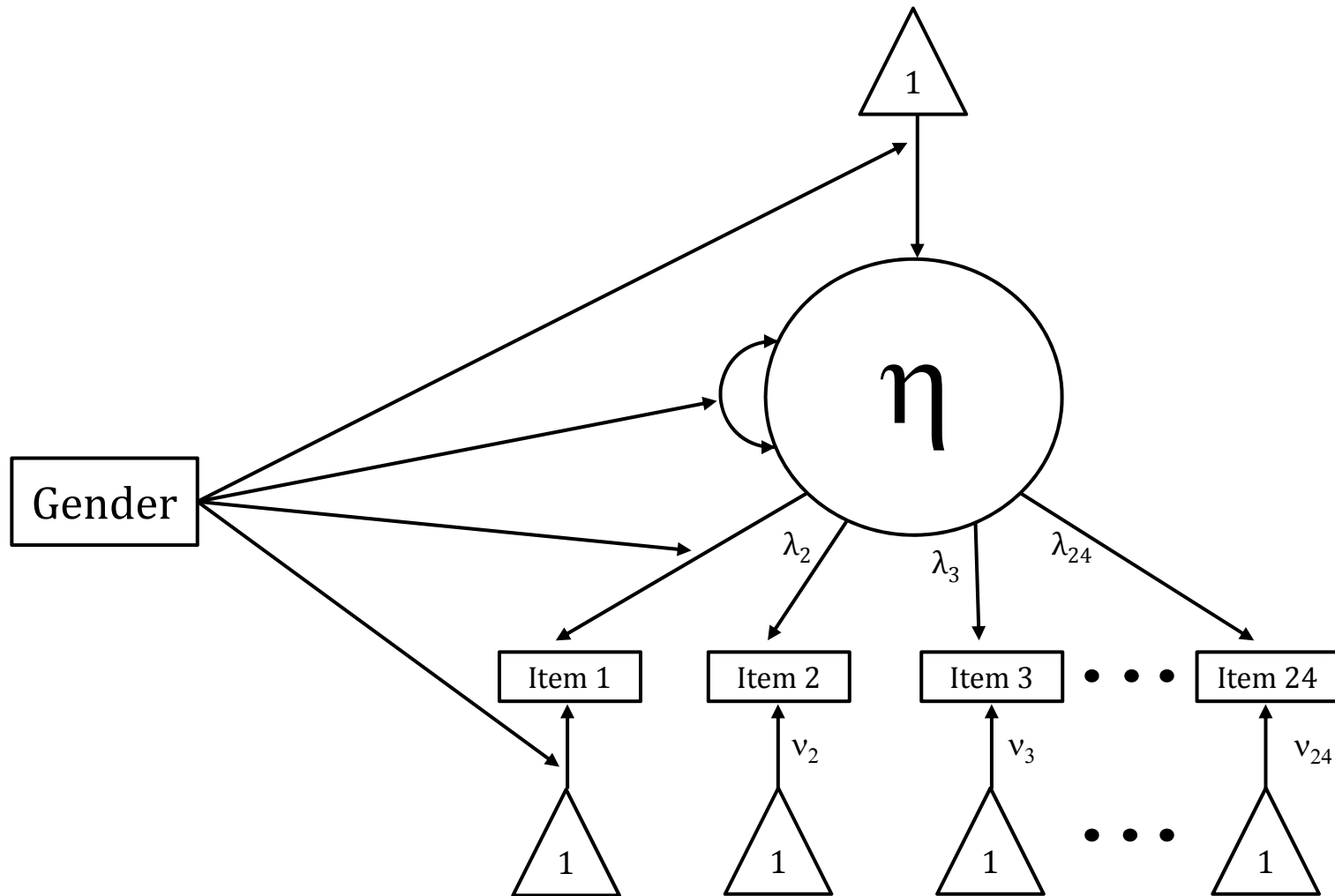


Full MNLFA

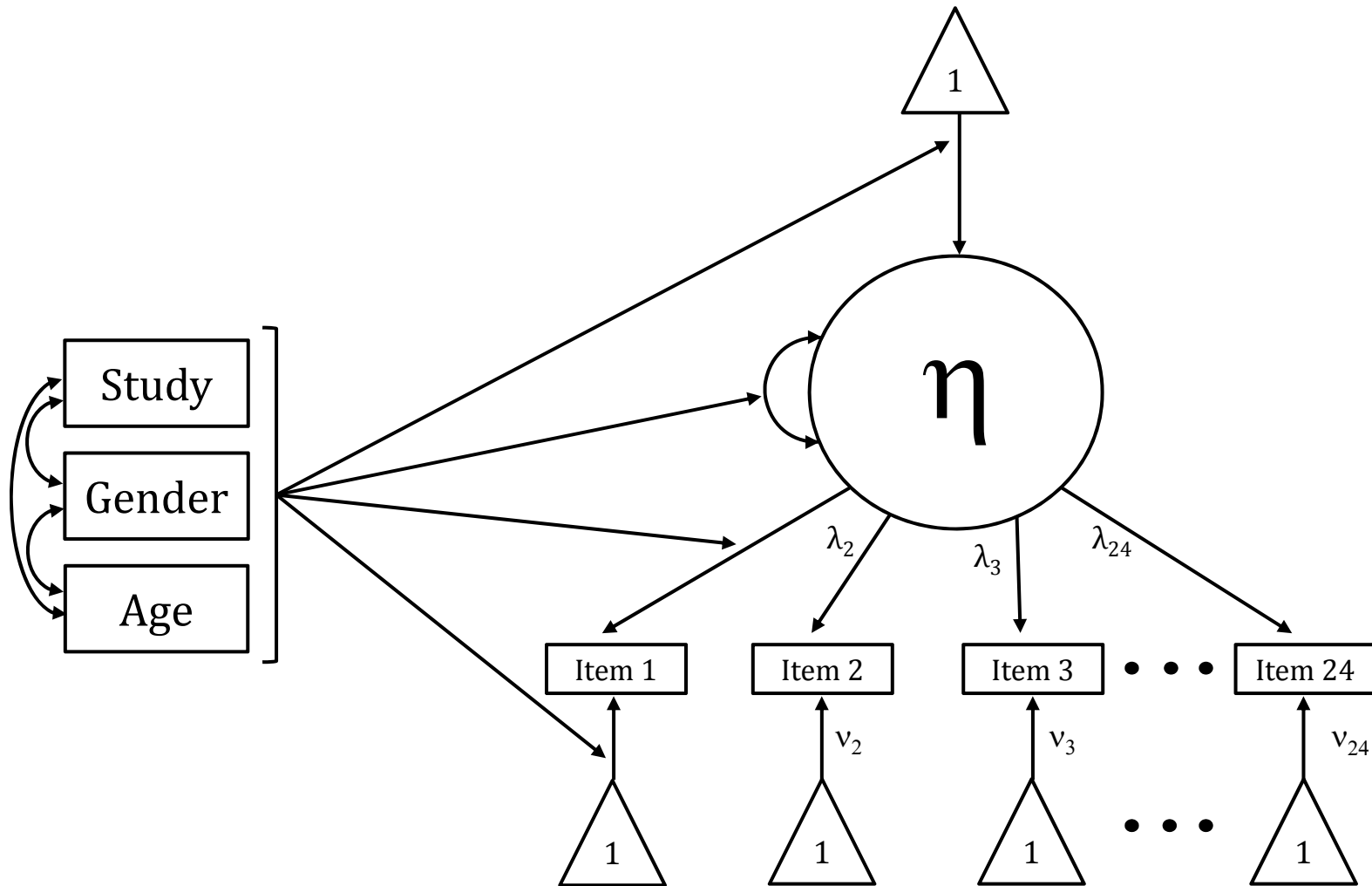




Standard Two-group CFA / IRT



Our Population MNLFA Model



Motivating Research Questions

- Are MNLFA parameters adequately recovered across varying design conditions
 - bias & efficiency
- Are valid and reliable factor scores obtained across varying design conditions
 - score fidelity and predictive validity
- ***Can MNLFA be profitably used in common applied research settings?***

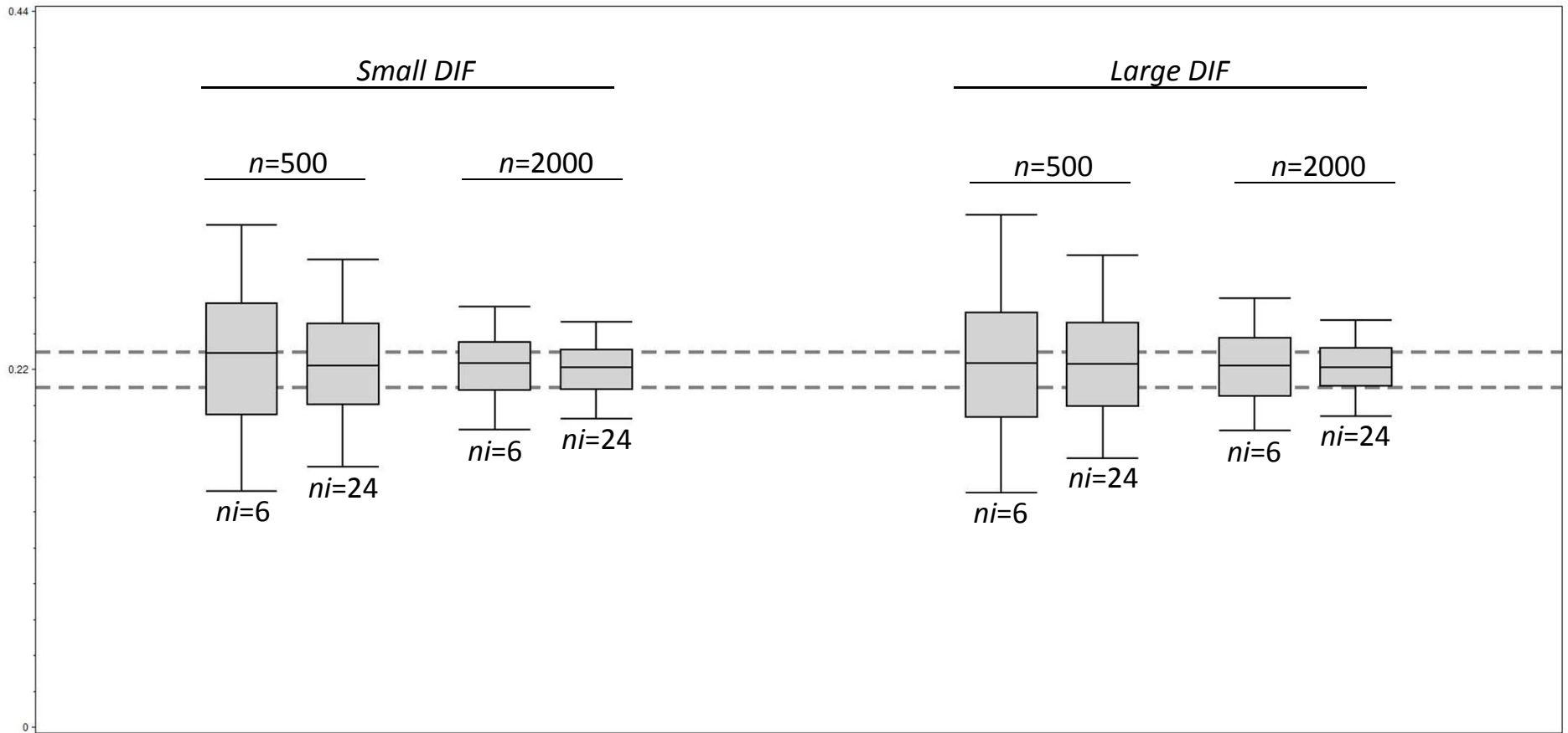
Limited Design for Today

- Held constant:
 - 2 studies, 1 latent factor, multiple binary items
 - 3 covariates: gender, study, age
 - medium impact ES for latent mean and variance
 - 33% of total items have DIF
- Varied:
 - total sample size: 500 vs. 2000
 - total items: 6 vs. 24
 - magnitude of DIF: small vs. large

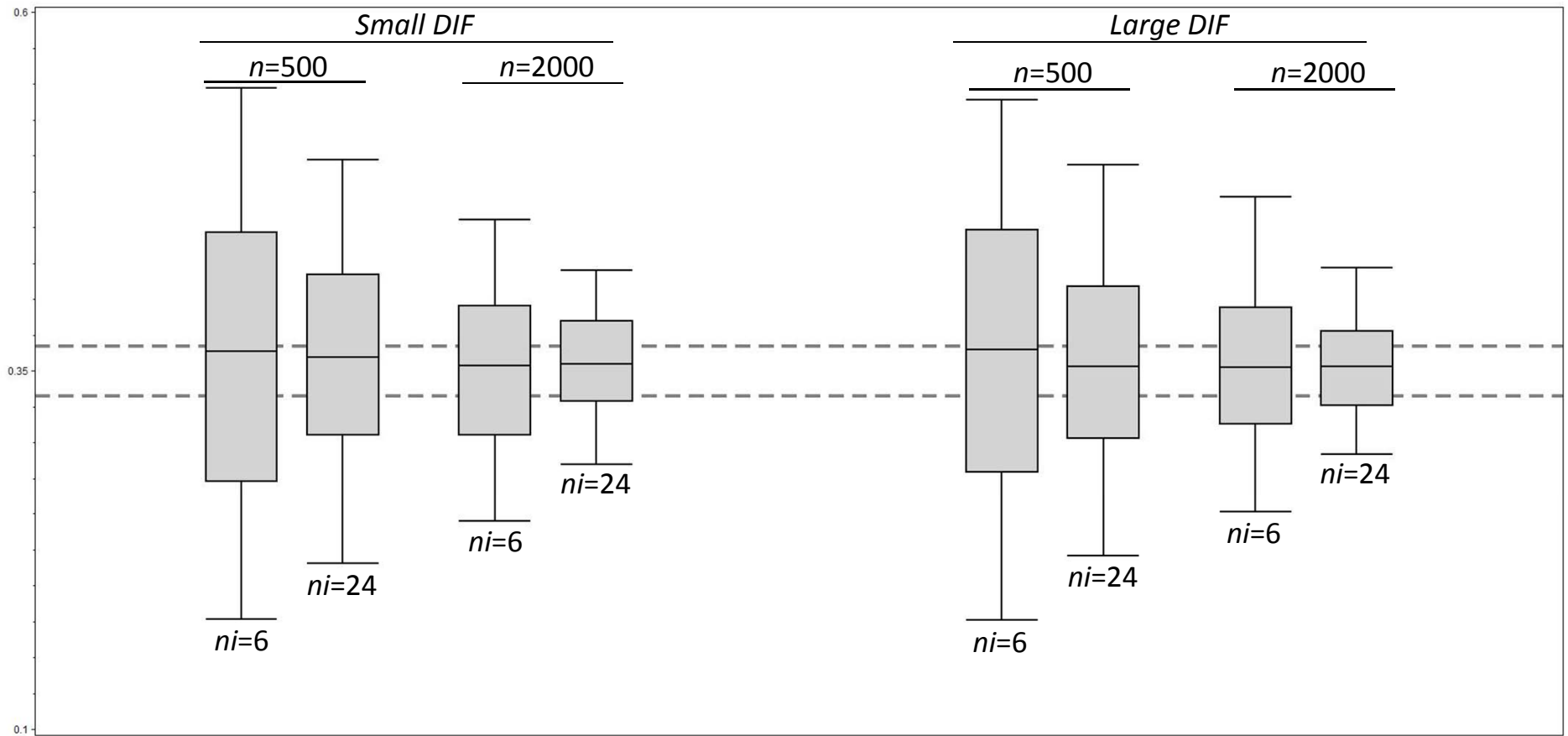
Core Steps in Generation & Analysis

- Eight total cells of design
 - sample size (2) by number of items (2) by DIF (2)
- Data generation in SAS from model equations
 - 500 replications per cell
- Models fit in *Mplus* under logit link and ML estimation with nonlinear constraints
 - all models converged
 - parameter estimates retained (for me)
 - factor score estimates retained (for Dan)

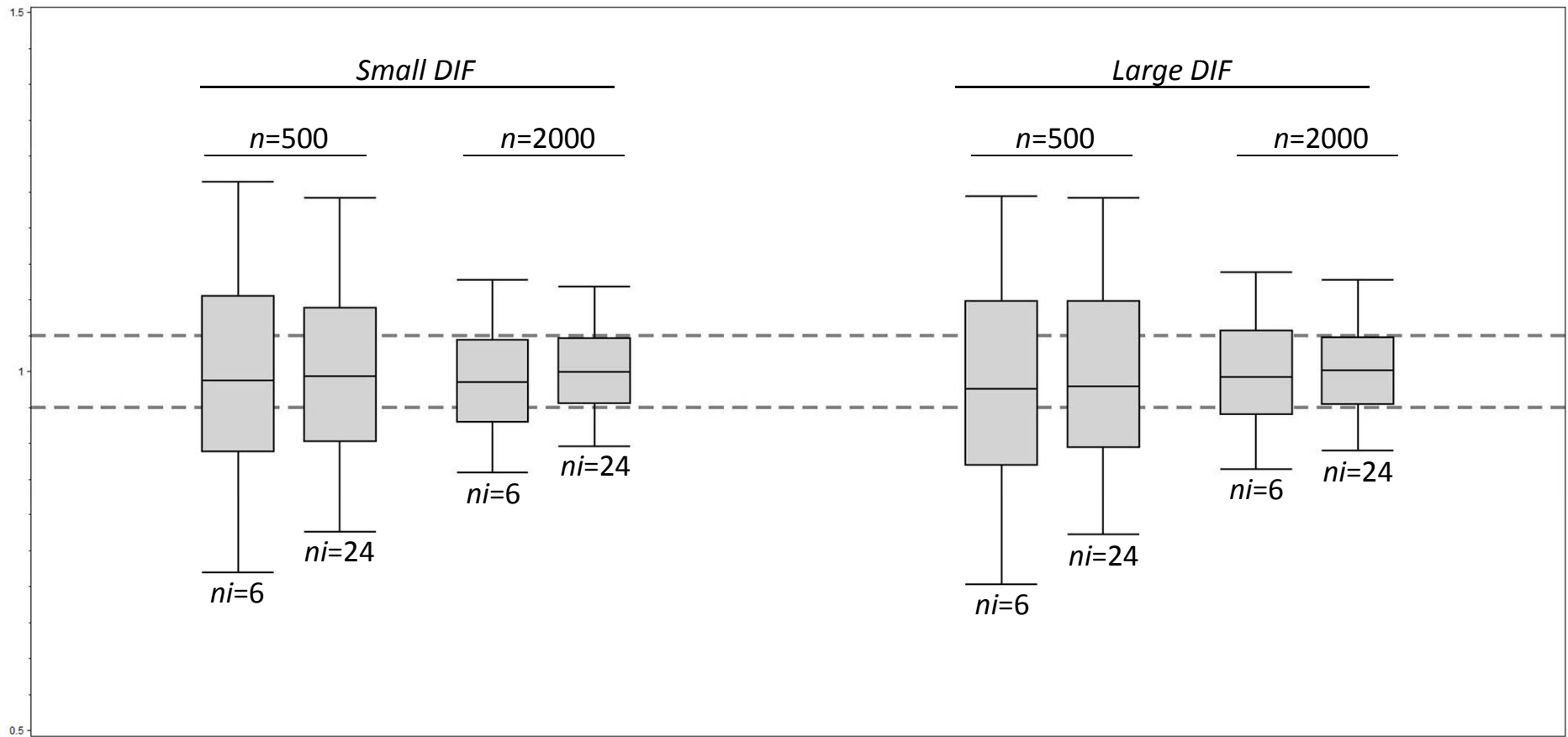
Age Effect on Factor Mean



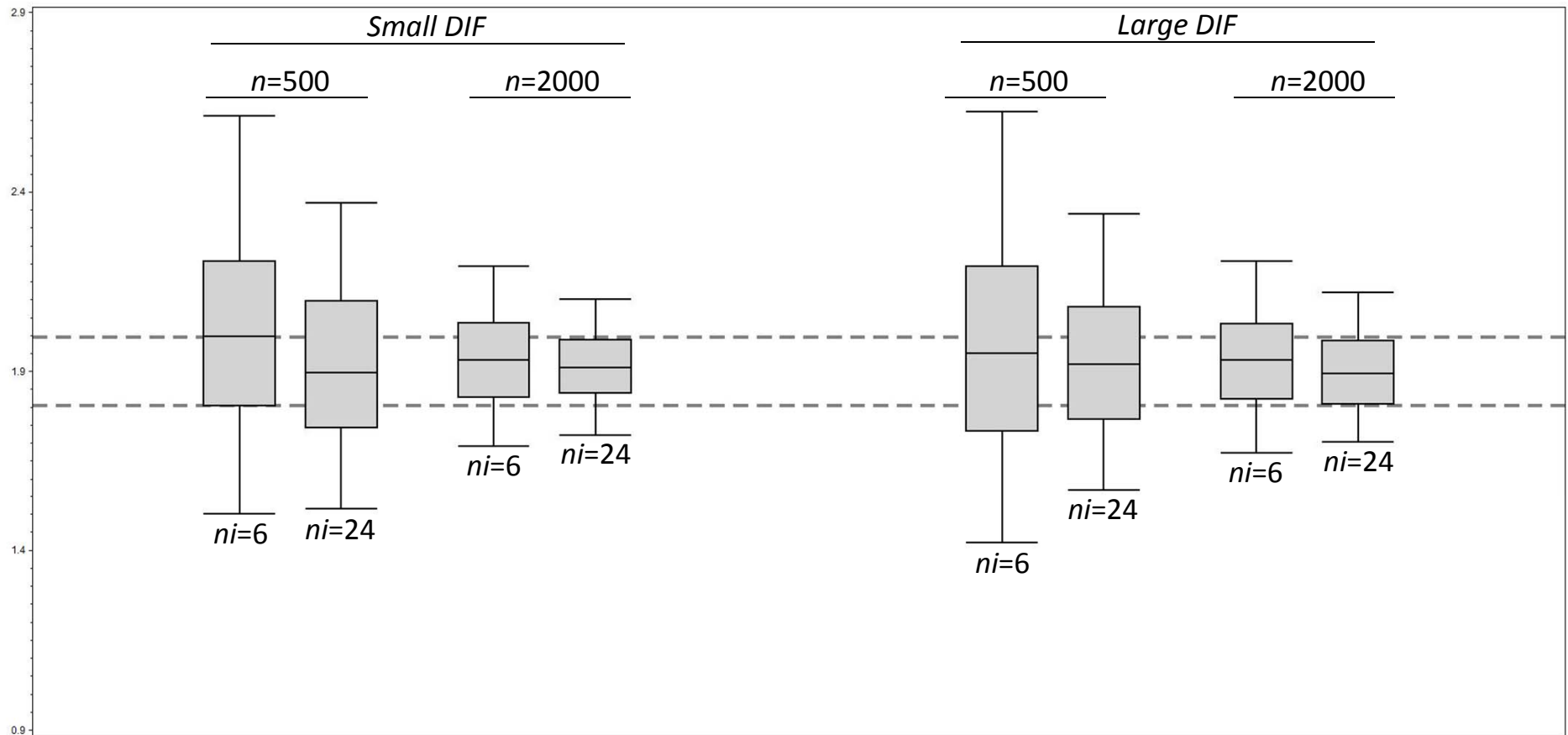
Age Effect on Factor Variance



Invariant Factor Loading



Baseline DIF Factor Loading



Invariant Item Intercept: Same

Baseline DIF Item Intercept: Same

Age Effect on Factor Loading: Same

Age Effect on Item Intercept: Same

Anything on Anything: Same

Conclusions

- For properly specified models:
 - parameter estimates are minimally biased
 - greater available information decreases variability
 - larger DIF modestly increases variability
- Findings are exactly as predicted
 - confirms theory: we weren't always so sure....
- Next steps:
 - consider more conditions, including misspecification
 - evaluate model building strategies
 - examine factor scores estimates...stay tuned for Dan

Thank you!



This project was supported by Award Number R01DA034636 (Bauer, PI) from the National Institute on Drug Abuse. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute on Drug Abuse or the National Institutes of Health.