

# Simultaneous Evaluation of DIF and its Sources Using Hierarchical Explanatory Models

Skorupski, W. P., Brussow, J. A., & Loughran, J. T.

National Council on Measurement in Education, Washington, DC, April 2016

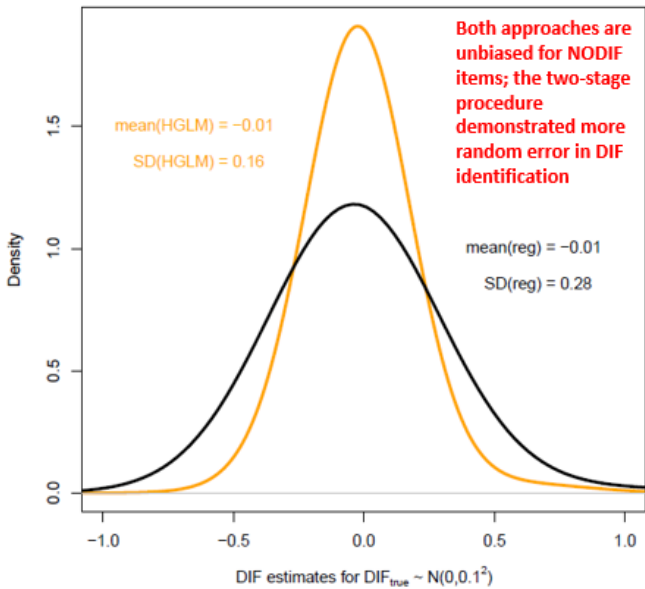
## Purpose

- Evaluate use of item-level features as explanatory variables for understanding DIF
- Understanding item-level features' impact could change the way we write items to avoid DIF

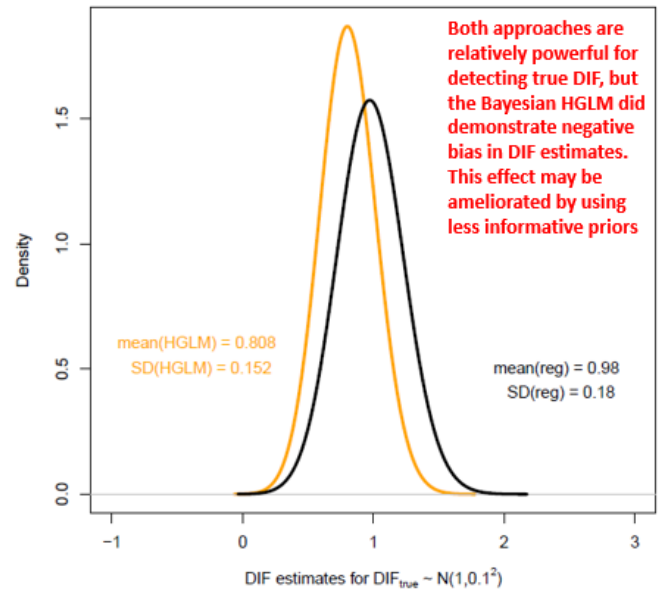
## Methods

- Simulated data from 2-PL IRT model
  - 2,000 examinees taking 60-item test; Focal and reference groups proportions: 1/1, 3/1, 9/1
  - Reference group  $\theta \sim N(0, 1)$ ; Focal group  $\theta \sim N(-0.5, 1)$
  - Difficulty:  $b \sim N(0, 1)$ ; Discrimination:  $a \sim U(0.5, 3.5)$
  - 3 items with uniform DIF:  $D \sim N(1, 0.1^2)$ ; 57 items w/ random, negligible DIF:  $D \sim N(0, 0.1^2)$
- Comparing 500 simulated datasets w/ different relationships between DIF & "ancillary info":  $\rho = 0.4, 0.5, 0.6$ 
  - Logistic regression followed by ordinary least squares regression (two-stage)
  - Fully Bayesian Hierarchical Generalized Linear Model (simultaneous)

Density plots of DIF estimates for items simulated to have negligible DIF.

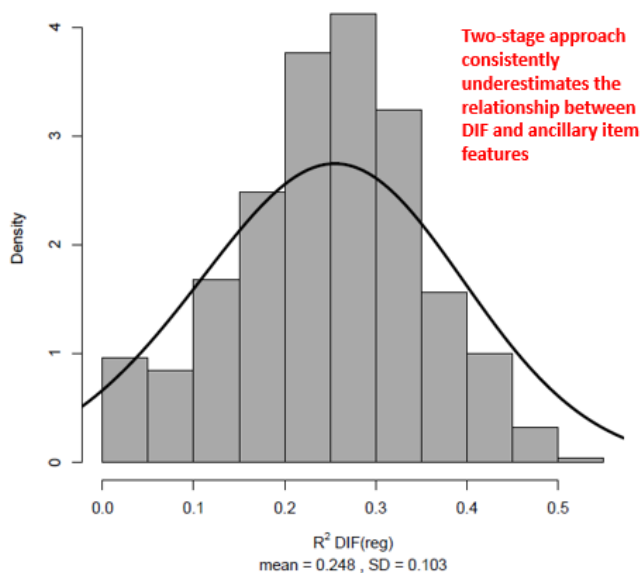


Density plots of DIF estimates for items simulated to have meaningful DIF.



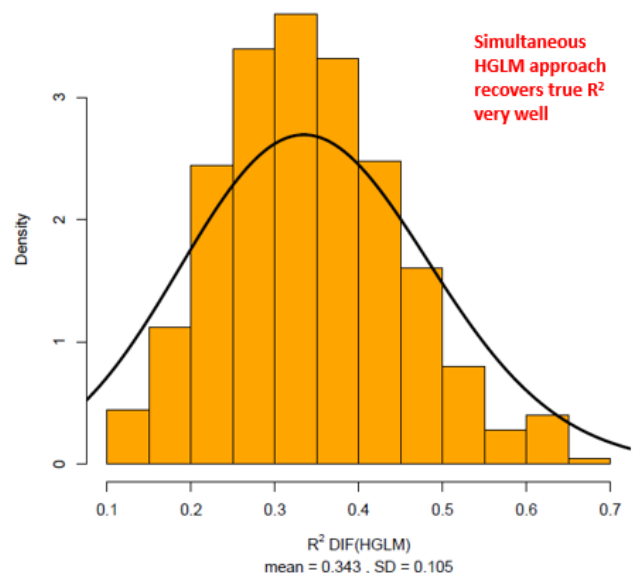
Histogram of  $R^2$  estimates from the linear regression of DIF estimates (two-stage)

True = 0.36 Prop(ref) = 0.9



Histogram of  $R^2$  estimates from the Bayesian HGLM procedure

True = 0.36 Prop(ref) = 0.9



- Bayesian HGLM is a reliable procedure for evaluating & interpreting DIF with respect to interpretable item features
- Bayesian HGLM had lower Type 1 error but also less power than the two-stage procedure; less informative priors may fix this
- Bayesian HGLM was more precise and accurate in estimating DIF & relationship (i.e.,  $R^2$ ) between DIF & item features