

# Cutting-Edge Psychometric and Statistical Methods in the Era of Educational Big Data

Chun Wang, University of Washington  
[wang4066@uw.edu](mailto:wang4066@uw.edu)

Annual IES Principal Investigators Meeting  
January 25-27, 2022

*Advancing Equity and Inclusion in the Education Sciences*





# Introduction

## Four Presentations

**Uncovering the Hidden Complexity in Statistical Models**

**A New Sentiment Lexicon for Student Evaluation of Teaching Text Data**

**Model-based Approach to Oral Reading Fluency Assessment**

**VEMIRT: A Variational EM Algorithm-based Shiny App for Multi-dimensional IRT Applications**

## Panelists

**Wesley Bonifay, University of Missouri**

**Zhiyong Zhang, University of Notre Dame**

**Akihito Kamata, Southern Methodist University**

**Chun Wang, University of Washington**

# 4

## VEMIRT: A Variational EM Algorithm-based Shiny App for Multi-dimensional IRT Applications

Chun Wang ([wang4066@uw.edu](mailto:wang4066@uw.edu))



## Acknowledgement

- Dr. Gongjun Xu (University of Michigan)
- Doctoral Students
  - Jiaying Xiao
  - Ruoyi Zhu
  - Chenchen Ma



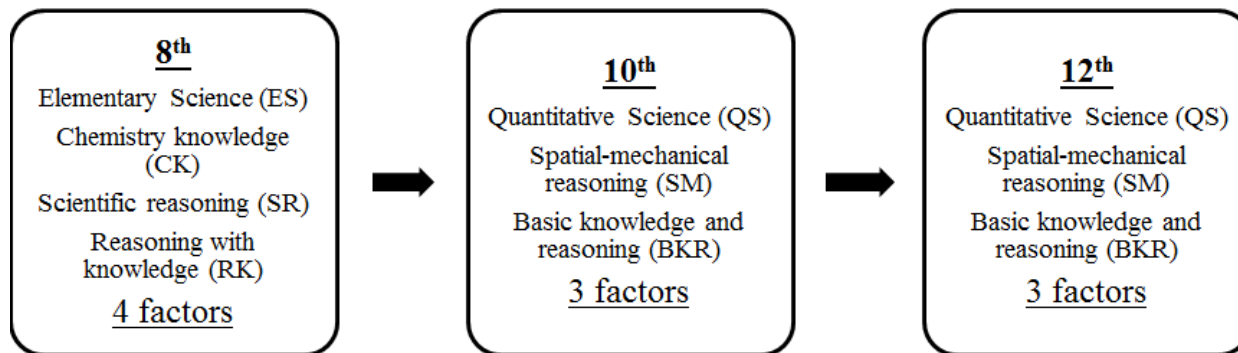
## General Motivation for MIRT: Analysis flowchart





## Aim 1: Multidimensional Constructs

Factor structure of NELS:88 Science achievement data from 8<sup>th</sup> to 12<sup>th</sup> grades



Kuperminzt, Ennis, Hamilton, Talbert, & Snow, 1995, Hamilton,  
Nussbaum, Kupermintz, Kerkhoven, & Snow, 1995



## Gaps in Practice

- Unidimensional IRT models continue to dominate the current practice in building mandatory state assessments.
- Gaussian Variational Expectation Maximization (GVEM) algorithm



## Aim 2: DIF analysis

- Multidimensional constructs
- Sampling weights
- Multiple-group comparison
- GVEM with covariates (GVEM-C)

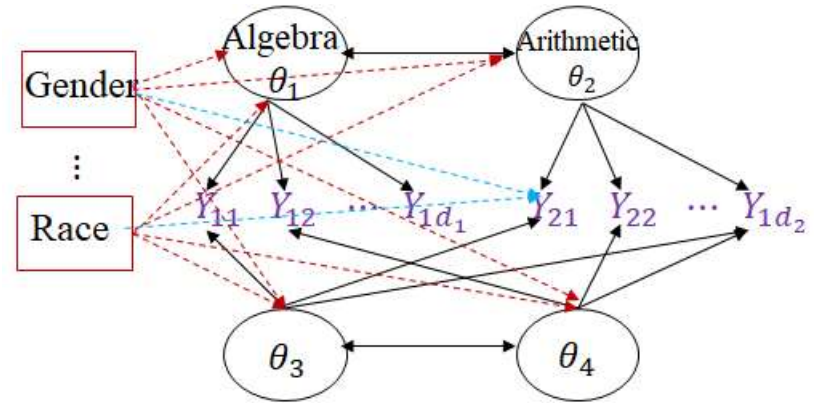
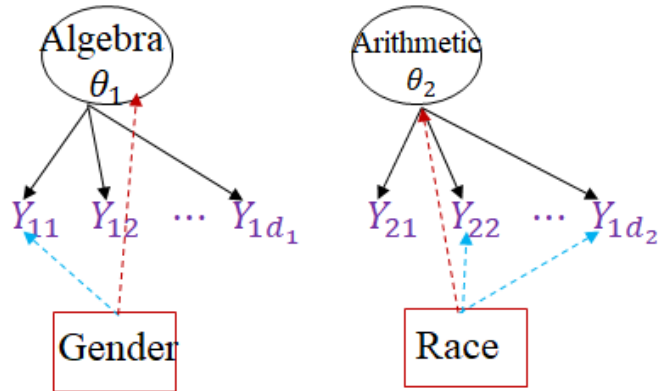




## Two Aims Together

Separate  
calibration

↓  
DIF analysis





## VEMIRT Modules

### Models

- Multidimensional 1PL, 2PL, 3PL (within-item, between-item, multidimensional, bi-factor, hierarchical factor structures)

### Exploratory analysis

- Parallel analysis
- Exploratory item factory analysis using either rotation or regularization methods

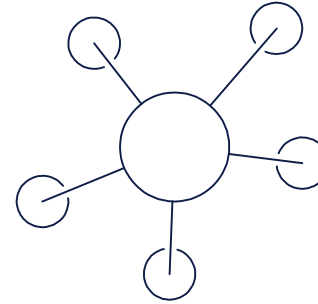
### Confirmatory analysis

- Estimate item parameters (and their SE)
- Estimate person's  $\theta$

### DIF

- Flag DIF items and estimate DIF size

# VEMIRT App 1.0





## VEMIRT 1.0 Functionalities

- Step 1. Identify the number of factors via parallel analysis

[https://uwpmetrics.shinyapps.io/parallel\\_analysis/](https://uwpmetrics.shinyapps.io/parallel_analysis/)

- Step 2. Exploratory and confirmatory M2PL models
  - Exploratory: GVEM+Post-hoc rotation; GVEM+Lasso; GVEM+Adaptive Lasso
  - Confirmatory: GVEM

<https://uwpmetrics.shinyapps.io/m2pl/>

- Step 2 (Alternative). Exploratory and confirmatory M3PL models

- Exploratory: Stochastic GVEM+Post-hoc rotation; SGVEM+Lasso; SGVEM+Adaptive Lasso
- Confirmatory: SGVEM

<https://uwpmetrics.shinyapps.io/m3pl/>

- DIF analysis

<https://uwpmetrics.shinyapps.io/RegDIF/>



## Some Algorithmic Details

- GVEM
  - Super fast for M2PL
  - Stochastic version for M3PL (sGVEM) (Hoffman, Blei, Wang, & Paisley, 2013)
    - At the  $(t+1)$ th iteration, choose a random subset of sample  $S_{t+1}$
    - $(1 - \varepsilon_{t+1})\hat{Q}_t + \varepsilon_{t+1}\hat{Q}_{t+1}$
    - $\varepsilon_{t+1} = (t + 1 + \tau)^{-r}$



## Some Algorithmic Details Cont.

- GVEM+C

- $$P(u_{ij} = 1 | \boldsymbol{\theta}_i, \mathbf{X}_i) = \frac{1}{1 + e^{-[a_j^T \boldsymbol{\theta}_{i(g)} + d_j + (\mathbf{x}_i \boldsymbol{\gamma}_j) \boldsymbol{\theta}_{i(g)} + \mathbf{x}_i \boldsymbol{\beta}_j]}}$$

- $(\mathbf{X}_i)_{1 \times P}$ : all grouping information related to DIF

- $(\boldsymbol{\beta}_j)_{P \times 1}$ : Uniform DIF

- $(\boldsymbol{\gamma}_j)_{P \times K}$ : non-uniform DIF

- Regularization (Bauer et al., 2020, Belzak & Bauer, 2020; Magis et al., 2015, Tutz & Schauberger, 2015)



## Current Facts

- It may take <2.5 min to load
- 5-dimensional M2PL, EFA or CFA, takes around 1 min to run (8G RAM)
  - 75 items, 2000 sample size
- 3-dimensional M3PL, 45 items, sample size 2000
  - EFA: around 5.5 min to complete
- Regularization could be slow (may take 30 min to complete)
  - We had an adaptive mechanism of choosing tuning parameter
    - A warning message will be provided if after two rounds of selecting a range of tuning parameters



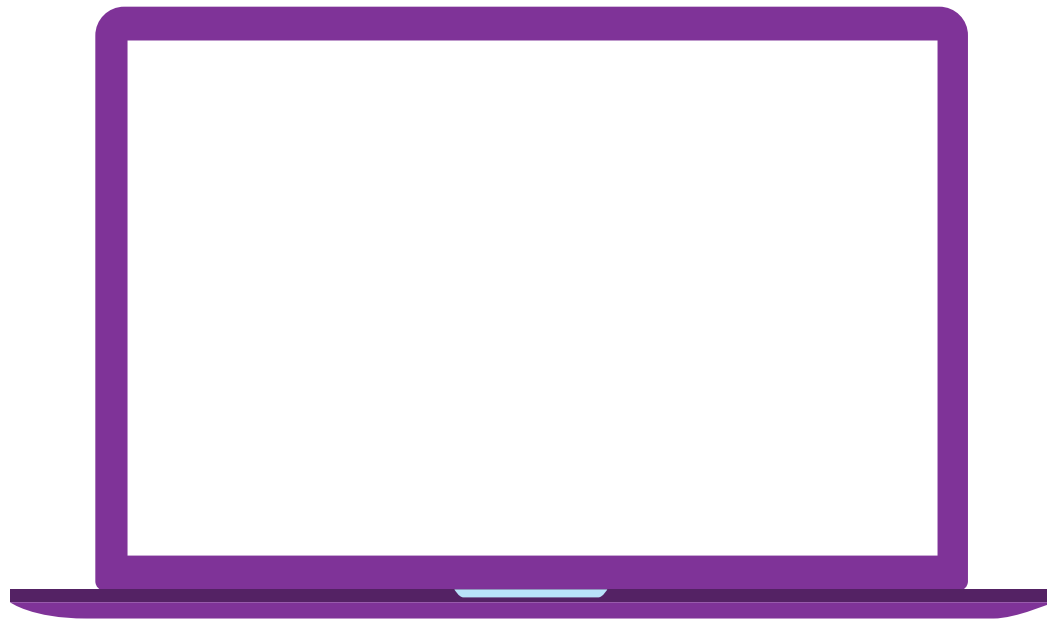
## References

- Wang, C., Zhu, R., & Xu, G. (2021+). Using lasso and adaptive lasso to identify DIF in multidimensional 2PL models. *Multivariate Behavioral Research*.  
 □ <https://github.com/wang4066/MIRT-RegDIF>
- Cho, A., Wang, C., Zhang, X., & Xu, G. (2021). Gaussian Variational Estimation for Multidimensional Item Response Theory. *British Journal of Mathematical and Statistical Psychology*.  
<https://doi.org/10.1111/bmsp.12219>



**Desktop project**

VEMIRT brief demo



THANKS!

**Any questions?**

# 1

## Uncovering the Hidden Complexity in Statistical Models

Wesley Bonifay

# 2

## A New Sentiment Lexicon for Student Evaluation of Teaching Text Data

Zhiyong Zhang

# 3

## **Model-based Approach to Oral Reading Fluency Assessment**

Akihito Kamata