Measuring Multidimensional Growth
A Higher-Order IRT Perspective

Steven W. Nydick
Pearson VUE

Chun Wang
University of Minnesota

Xinhui Xiong
CTB/McGraw-Hill

April 6, 2014

Growth Models

Growth models have typically fallen into two camps:

❶ Longitudinal item response theory models

- Positive: Directly models traits via ordinal variables
- Negative: Difficulty accounting for growth structures

❷ Latent growth curve models

- Positive: Separate the measurement/growth parts
- Negative: Do not easily consider ordinal outcomes

Growth in Higher-Order IRT

First propose higher-order IRT model:

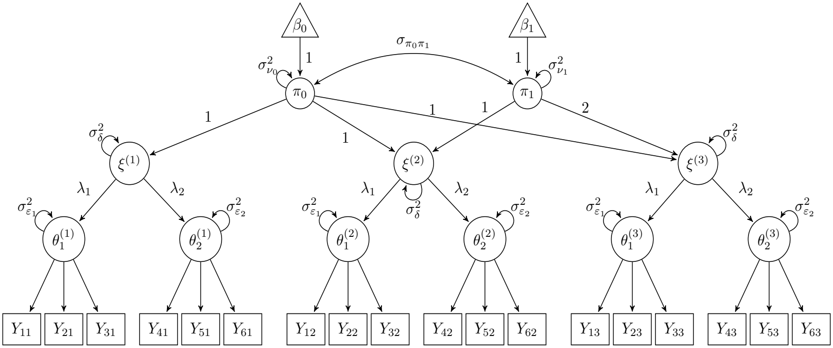
$$\mathbf{\Gamma}_i = \begin{bmatrix} \xi_i^{(1)} & \theta_{i1}^{(1)} & \theta_{i2}^{(1)} & \cdots & \theta_{iK}^{(1)} \\ \xi_i^{(2)} & \theta_{i1}^{(2)} & \theta_{i2}^{(2)} & \cdots & \theta_{iK}^{(2)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \xi_i^{(T)} & \theta_{i1}^{(T)} & \theta_{i2}^{(T)} & \cdots & \theta_{iK}^{(T)} \end{bmatrix}$$

Then link higher-order ability with regression line:

$$\xi_i^{(t)} = \pi_{0i} + \pi_{1i} \times (t-1) + \delta_i^{(t)}.$$

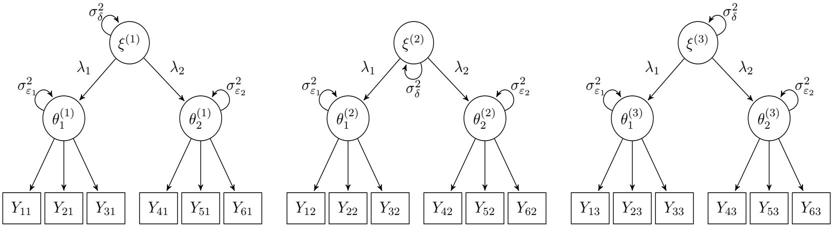
How can we visualize/calibrate this structure?

Longitudinal Path Diagram



Option 1: Build the higher-order structure into the model.

Separate Path Diagrams



Option 2: Separate calibrations, linking, and post-hoc growth.

Purpose of Study

What are the answers to the following questions?

1. Does including higher-order trends into the model sufficiently improve estimation precision?
2. Is there a difference between three and five domain abilities in terms of accuracy?
3. Does bias/MSE change for higher-order/lower-order abilities when increasing time?
4. Can Mplus adequately capture changes in a higher-order IRT model over time?

This study is one step in comparing different longitudinal IRT models, methods of estimating them, and their ability to help detect change and growth of ability over time.

Item Parameter Generation

❶ Models

- C-MIRT
- Items loaded onto only one dimension.
- Same items used across time.

❷ Size of Item Bank

- $J = J_K \times K$
- J_K is the items per dimension.
- K is the number of dimensions.

Person Parameter Generation: $T = 2$

Higher-Order Parameters

1. Let r be the specified correlation between dimensions.
2. Then $[\xi_i^{(1)}, \xi_i^{(2)}] \sim N\left(\boldsymbol{\mu}_{\boldsymbol{\xi}} = \left[\begin{smallmatrix} 0.0 & 0.3 \end{smallmatrix}\right]^T, \boldsymbol{\Sigma}_{\boldsymbol{\xi}} = \left[\begin{smallmatrix} 1.0 & r \\ r & 1.0 \end{smallmatrix}\right]\right)$

Lower-Order Parameters

1. Let λ be the loading of ξ_t onto $\theta_k^{(t)}$.
2. Then $\epsilon_{ik}^{(t)} \sim N(\mu_\epsilon = 0.0, \sigma_\epsilon^2 = 1 - \lambda^2)$.
3. And $\theta_{ik}^{(t)} = \lambda \xi_i^{(t)} + \epsilon_{ik}^{(t)}$.

Person Parameter Generation: $T = 4$

Higher-Order Parameters

1. Let $\pi_{0i} \sim N(\mu_{\pi_0} = 0.0, \sigma^2_{\pi_0} = 0.5)$.
2. Let $\pi_{1i} \sim N(\mu_{\pi_1} = 0.25, \sigma^2_{\pi_1} = 0.01)$.
3. Let $\delta_i^{(t)} \sim N(\mu_\delta = 0.0, \sigma^2_\delta = 0.1)$.
4. Then $\xi_i^{(t)} = \pi_{0i} + (t - 1) \times \pi_{1i} + \delta_i^{(t)}$.

Lower-Order Parameters

1. Let λ be the loading of $\xi^{(t)}$ onto $\theta_d^{(t)}$.
2. Then $\epsilon_{ik}^{(t)} \sim N(\mu_\epsilon = 0.0, \sigma^2_\epsilon = 1 - \lambda^2)$.
3. And $\theta_{ik}^{(t)} = \lambda\xi_i^{(t)} + \epsilon_{ik}^{(t)}$.

Conditions Tables

Conditions Table: $T = 2$

N	2 (250, 1,000)
J_k	2 (10, 20)
K	2 (3, 5)
λ	2 (0.8, 0.9)
ρ	2 (0.5, 0.75)
Overall	32

Conditions Table: $T = 4$

N	2 (250, 1,000)
J_k	2 (10, 20)
K	2 (3, 5)
λ	2 (0.8, 0.9)
Overall	16

Overall Procedure: $T = 2$

For the $T = 2$ simulation, we did the following:

1. Generated response matrix of all persons to all items across all time using R.
2. Jointly estimated item and persons parameters across all time using Mplus.
3. Separately estimated item and person parameters at each time point using Mplus.
   - Linked parameters at $t = 2$ to parameters at $t = 1$.
4. Repeated each condition 25 times.

Higher-Order Ability: Varying *r*

❶ Combined Calibration:

MSE

r	$\xi^{(1)}$	$\xi^{(2)}$
.50	0.24	0.27
.75	0.21	0.24

Correlation

r	$\xi^{(1)}$	$\xi^{(2)}$
.50	.88	.88
.75	.89	.89

❷ Separate Calibration:

MSE

r	$\xi^{(1)}$	$\xi^{(2)}$
.50	0.25	0.27
.75	0.25	0.27

Correlation

r	$\xi^{(1)}$	$\xi^{(2)}$
.50	.87	.87
.75	.87	.87

Domain Ability: Varying *r*

❶ Combined Calibration: MSE

r	$\theta_1^{(1)}$	$\theta_2^{(1)}$	$\theta_3^{(1)}$	$\theta_1^{(2)}$	$\theta_2^{(2)}$	$\theta_3^{(2)}$
.50	0.40	0.39	0.40	0.42	0.41	0.42
.75	0.38	0.40	0.38	0.40	0.40	0.39

❷ Separate Calibration: MSE

r	$\theta_1^{(1)}$	$\theta_2^{(1)}$	$\theta_3^{(1)}$	$\theta_1^{(2)}$	$\theta_2^{(2)}$	$\theta_3^{(2)}$
.50	0.38	0.37	0.38	0.41	0.40	0.41
.75	0.39	0.39	0.39	0.43	0.43	0.42

Higher-Order Ability: Varying λ

❶ Combined Calibration:

MSE

λ	$\xi^{(1)}$	$\xi^{(2)}$
.80	0.26	0.29
.90	0.19	0.22

Correlation

λ	$\xi^{(1)}$	$\xi^{(2)}$
.80	.86	.86
.90	.90	.91

❷ Separate Calibration:

MSE

λ	$\xi^{(1)}$	$\xi^{(2)}$
.80	0.29	0.31
.90	0.20	0.23

Correlation

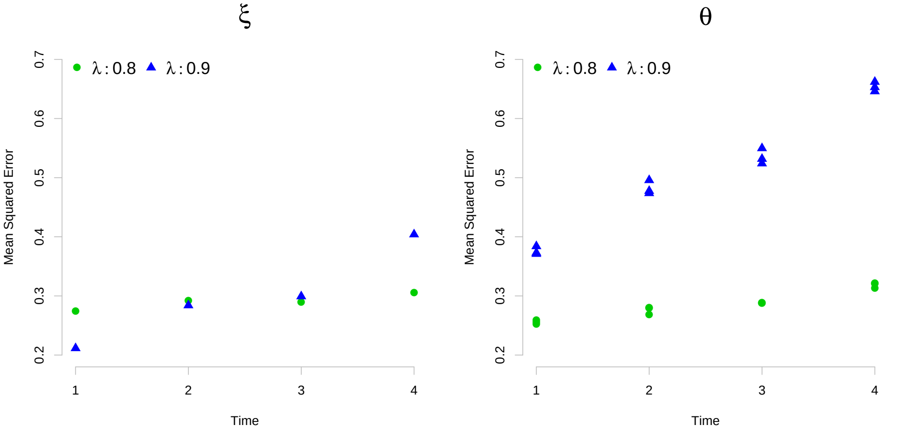
λ	$\xi^{(1)}$	$\xi^{(2)}$
.80	.84	.84
.90	.89	.89

Overall Procedure: $T = 4$

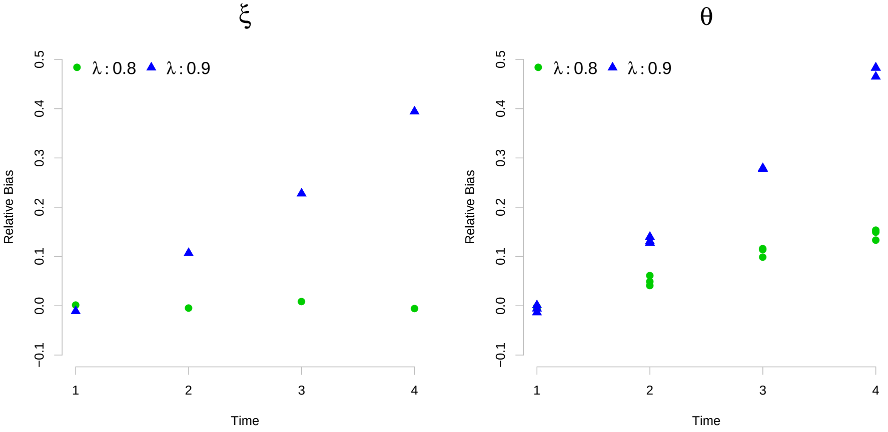
For the $T = 4$ simulation, we did the following:

1. Generated response matrix of all persons to all items across all time using `R`.
2. Separately estimated item and person parameters at each time point using Mplus.
   - Linked parameters at $t \geq 2$ to parameters at $t = 1$.
3. Estimated slope/intercept using `lme4` in `R`.
   - `lmer( xi ~ time + (time | person) )`
4. Repeated each condition 25 times.

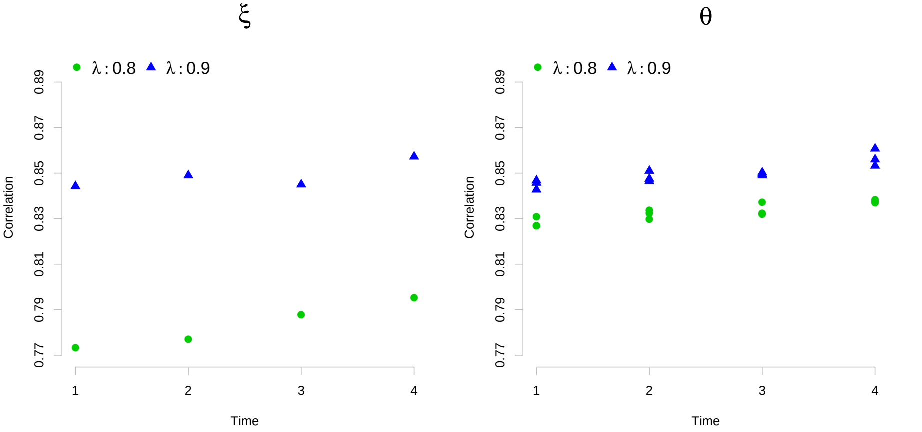
Varying λ: $K = 3$ MSE



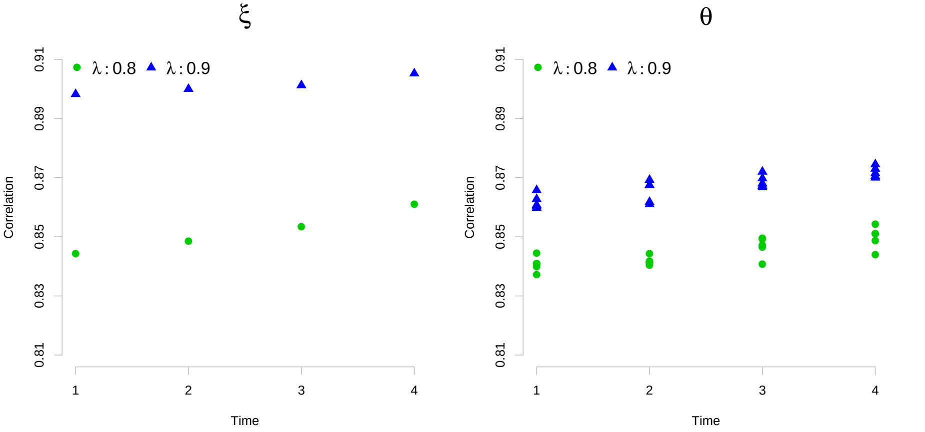
Varying λ: $K = 3$ Bias



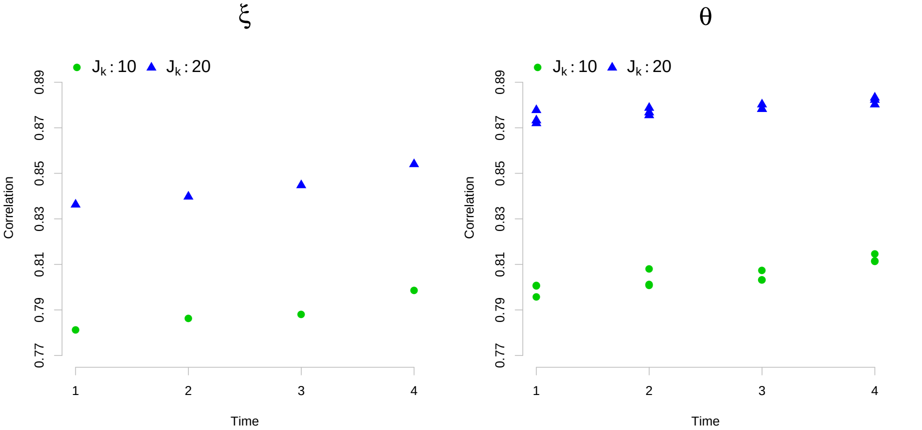
Varying λ: $K = 3$ Correlation



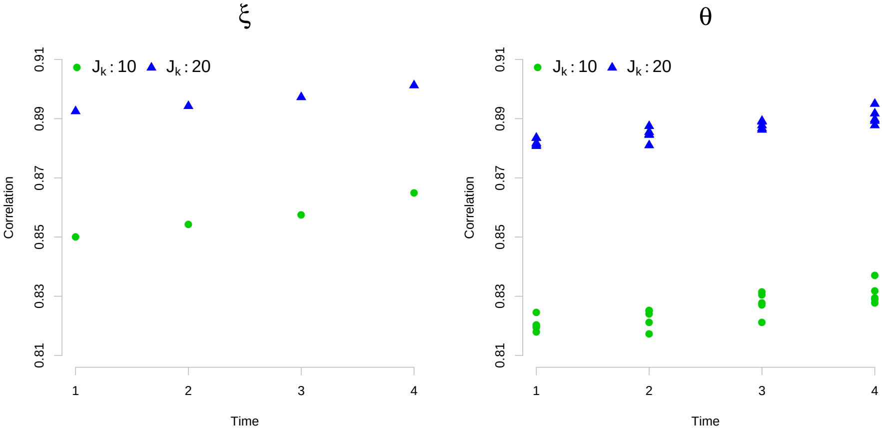
Varying λ: $K = 5$ Correlation



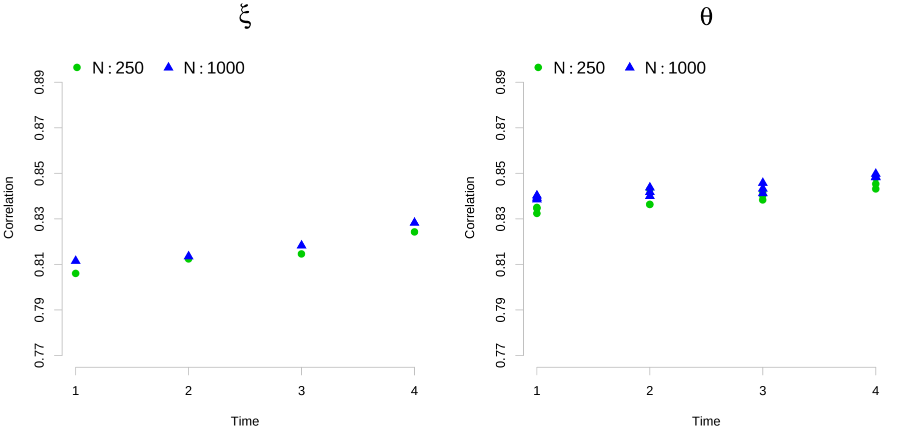
Varying J_k: $K = 3$ Correlation



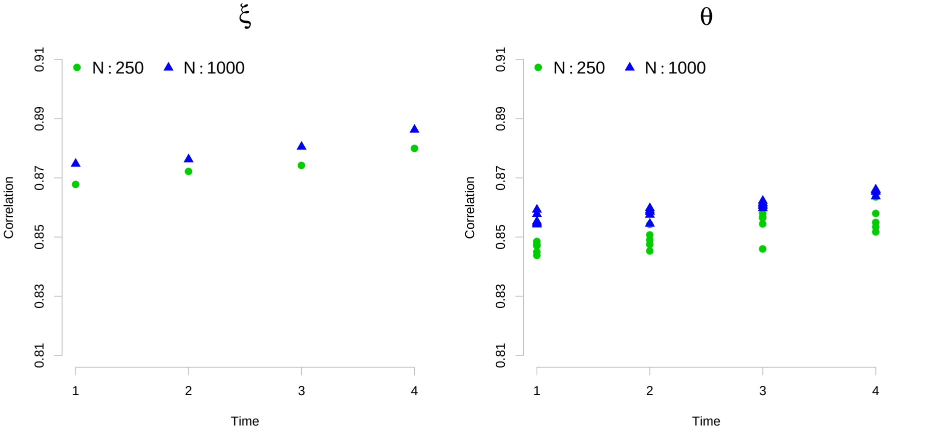
Varying J_k: $K = 5$ Correlation



Varying N: $K = 3$ Correlation



Varying N: $K = 5$ Correlation



Summary of Results

What are the answers to the following questions?

1. **Does including higher-order trends into the model sufficiently improve estimation precision?**
2. Is there a difference between three and five domain abilities in terms of accuracy?
3. Does bias/MSE change for higher-order/lower-order abilities when increasing time?
4. Can Mplus adequately capture changes in a higher-order IRT model over time?

Not really. The accuracy slightly improves but at a cost of extreme computational running times.

Summary of Results

What are the answers to the following questions?

1. Not really. The accuracy slightly improves but at a cost of extreme computational running times.
2. **Is there a difference between three and five domain abilities in terms of accuracy?**
3. Does bias/MSE change for higher-order/lower-order abilities when increasing time?
4. Can Mplus adequately capture changes in a higher-order IRT model over time?

Yes. The average estimated/true correlation is larger for $K = 5$, especially when considering higher-order abilities.

Summary of Results

What are the answers to the following questions?

1. Not really. The accuracy slightly improves but at a cost of extreme computational running times.
2. Yes. The average estimated/true correlation is larger for $K = 5$, especially when considering higher-order abilities.
3. **Does bias/MSE change for higher-order/lower-order abilities when increasing time?**
4. Can Mplus adequately capture changes in a higher-order IRT model over time?

Not really. The increase in bias as $T \uparrow$ is an artifact of scaling for λ close to 1.

Summary of Results

What are the answers to the following questions?

1. Not really. The accuracy slightly improves but at a cost of extreme computational running times.
2. Yes. The average estimated-true correlation is larger for $K = 5$, especially when considering higher-order abilities.
3. Not really. The increase in bias as $T \uparrow$ is an artifact of scaling for λ close to 1.
4. **Can Mplus adequately capture changes in a higher-order IRT model over time?**

Possibly! Both lower-order and higher-order domain abilities were precisely estimated regardless of condition.

Conclusions

How can we improve IRT trend analysis?

- Would MCMC or similar methods yield better estimates to intractable problems?
- Could we improve Mplus with two-stage estimation?
- Should the slopes/intercepts obtained from Mplus or `lmer` inform educational decisions?
- Do other models/methods/algorithms better capture the true change of ability over time?

Thank You!

Appendix: Higher-Order IRT Models

Item responses are connected to ability via a logit link.

$$p_{jk_j}(\theta_{ik}) = \Pr(Y_{ijk_j}|\theta_{ik}, a_{jk_j}, b_j) = \frac{1}{1+\exp[-a_{jk_j}(\theta_{ik}-b_j)]}, \qquad (1)$$

This is the standard 2PL IRT model:

1. a_{jk_j} represents the slope of item j.
2. b_j represents the location of item j.
3. θ_{ik} represents an individual's ability on domain k.

Mplus Assumptions: $T = 2$ and Combined

```
xi1 BY th1_1-th3_1*.8 (lamb);
xi2 BY th1_2-th3_2*.8 (lamb);
[th1_1-th3_1@0]; [th1_2-th3_2*0.2];
th1_1-th3_1@.5  th1_2-th3_2@.5;
[xi1@0]; [xi2*0.5]; xi1@1; xi2*1; xi1-xi2 WITH xi1-xi2;
```

Notes from the above Mplus code:

- Loadings were the same for all dimensions across all time.
- All domain variances set to .5.
- Higher-order means/variances fixed at $t = 1$.

Mplus Assumptions: $T = 2$ and Separate

```
xi1 BY th1_1-th3_1*.8 (lamb);
[th1_1-th3_1@0];
th1_1-th3_1@.5;
[xi1@0]; xi1@1;
```

Notes from the above Mplus code:

- Every time-point calibrated in identical fashion.
- Parameters at $t = 2$ linked to location/scale at $t = 1$.
- Loadings could not be the same across all time.