# The Impact of Within-Template Subset Effects

Quinn N Lathrop

University of Notre Dame
**Dr. Ying "Alison" Cheng**
Psychological and Educational
Measurement Lab

Pearson
**Dr. John Behrens**
Center for Digital Data, Analytics
and Adaptive Learning

October 11, 2013

# What are templates?

Templates generate items/tasks during computerized assessment.

- It's a small step from digital assessment to algorithmically generated items
- The data got ahead of the psychometrics...

Templates contain:

- a question form
- distributions for all variables in the question form

# Examples of Templates

Question Form:

"What is $X + Y$?"

Distributions:

$$f_X(x) = 1/5, x \in \{1, 2, 3, 4, 5\}$$

$$f_Y(y) = 1/6, y \in \{3, 4, 5, 6, 7, 8\}$$

Question Form:

"What is the average of $x_1, x_2, x_3, x_4, x_5$?"

Distributions:

$$x_{1-5} \sim Binom(40, .5)$$

# Our Motivating Template

Question Form:

"What is the probability of rolling a $X$ on a $Y$-sided die?"

Distributions:

$$f_Y(y) = 1/5, \quad y \in \{6, 8, 10, 12, 20\}$$
$$f_X(x) = 1/y, \quad x \in \{1, 2, ..., y\}$$

Correct strategy:

$$\frac{1}{y}$$

An incorrect strategy:

$$\frac{x}{y}$$

For a subset ($x = 1$), students can use the *wrong* strategy and still get the correct answer!

## Model within-template differences with multi-level IRT?

- Albers (1995)
- Glas and van der Linden (2003)
- Johnson Sinharay (2005)

## Model within-template differences with covariates?

- Fischer (1973)
- de Boeck and Wilson (2004)

## Both?

- Lathrop (???)

## Neither?

- everybody already does...

# Sampling Design

Each person responds to the same templates

- $p = 1, 2, ..., N$ for persons and $t = 1, 2, ..., T$ for templates

Each template has some number of items

- $t_i = 1, 2, ..., t_I$

The items within a template may be grouped by a design matrix

- For subsets, we will use a dummy variable where $X_{t_i} = 1$ if $t_i$ is in the subset and 0 otherwise

When person $p$ is assigned template $t$, item $t_i$ is randomly drawn from available items

- The response is $Y_{pt_i} \sim \text{Bernoulli}(\eta_{pt_i})$

# Four Models

**2P-T**

$$\eta_{pt} = \alpha_t \times (\theta_p - \mu_t)$$

- the "neither" option, just a template level IRT model

**2P-TX**

$$\eta_{pt_i} = \alpha_t \times (\theta_p - \mu_t + \lambda_t X_{t_i})$$

- adds a covariate, $\lambda_t$ to explain differences contained in $X_{t_i}$

**2P-R**

$$\eta_{pt_i} = \alpha_t \times (\theta_p - \beta_{t_i})$$
$$\beta_{t_i} \sim \mathcal{N}_1(\mu_t, \sigma_t)$$

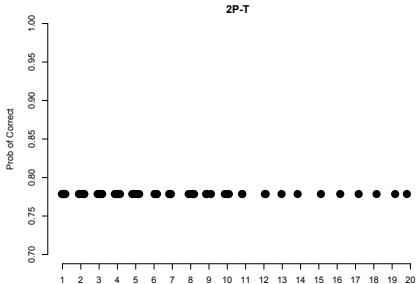
- multi-level model

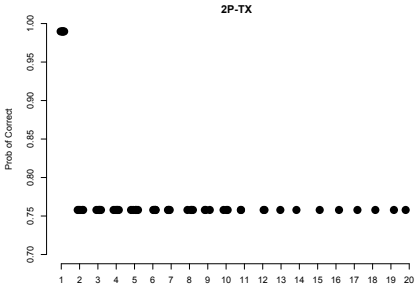
**2P-RX**

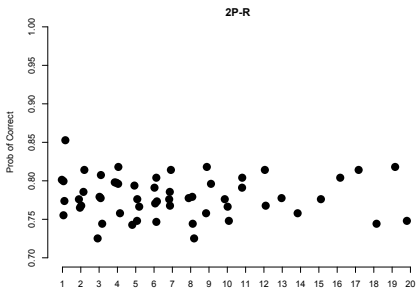
$$\eta_{pt_i} = \alpha_t \times (\theta_p - \beta_{t_i} + \lambda_t X_{t_i})$$
$$\beta_{t_i} \sim \mathcal{N}_1(\mu_t, \sigma_t)$$

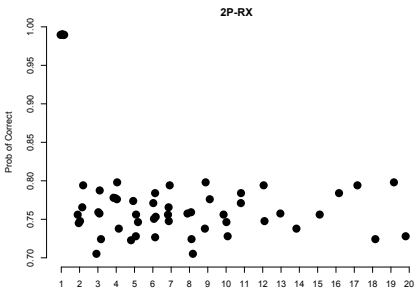
- the "both" option

# "What is the probability of rolling a $X$ on a $Y$-sided die?"



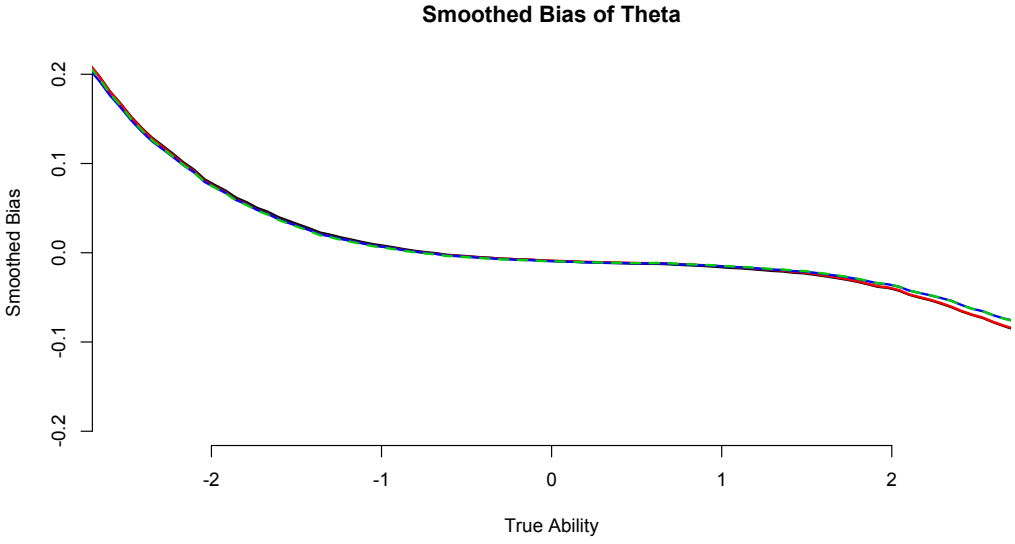



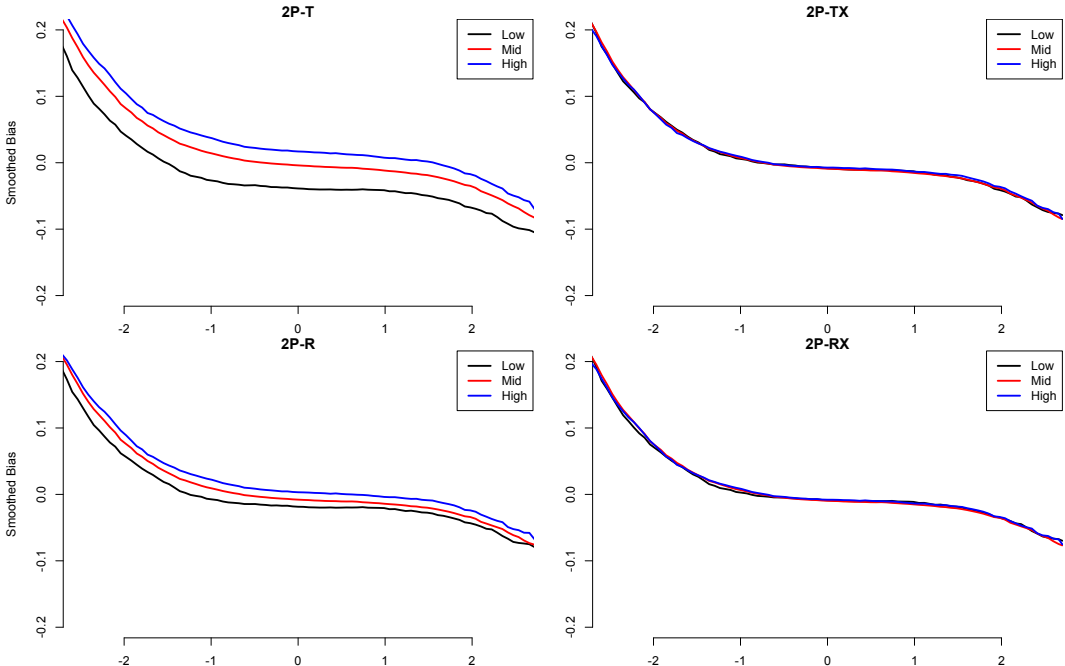



# How do subsets affect template parameters?

| Model | Discrimination | Difficulty | Covariate | Sigma |
|---|---|---|---|---|
| 2P-T | Biased | Biased | - | - |
| 2P-TX | Biased | Good | High Type I error | - |
| 2P-R | Good | Biased | - | Ok |
| 2P-RX | Good | Good | Good | Ok |


Smoothed Bias of Theta
Smoothed Bias
0.2
0.1
0.0
-0.1
-0.2
-2
-1
0
1
2
True Ability

| Model | Subsets | Mean Bias |
|---|---|---|
| 2P-T | Low | -.030 |
| | Mid | .005 |
| | High | .025 |
| 2P-TX | Low | .000 |
| | Mid | .000 |
| | High | .001 |
| 2P-R | Low | -.010 |
| | Mid | .001 |
| | High | .013 |
| 2P-RX | Low | .000 |
| | Mid | -.000 |
| | High | .001 |


2P-T
2P-TX
2P-R
2P-RX
Low
Mid
High
Smoothed Bias
0.2
0.1
0.0
-0.1
-0.2
-2
-1
0
1
2

# Summary

The model we specify affects our results...

The results we want affect the model we specify...

- For Discrimination, need* to fit an R model
- For Difficulty, need* to fit an X model
- For Ability, maybe should fit X model
- For Covariate, maybe should fit RX model

# Data Collection is Key

Many systems have thousands of templates each with potentially thousands of items.

- Is the item index being recorded (needed* for R models)?
- How do we organize the items by meaningful dimensions in $X$ (needed for X models)?

If we don't collect the data, we can't even begin to ask.