EXTRA SLIDES

Model 2: Latent Regression LLTM + e

$$logit(\Pr(Y_{ip} = 1 \mid \theta_p) = \theta_p - \delta_i \quad (1)$$

$$\delta_i = \sum_{k=0}^{K} \beta_k X_{ki} + \epsilon_i \quad \epsilon_i \sim N(0, \sigma^2_{\epsilon_i}) \quad (2)$$

$$\theta_p = \sum_{j=0}^{J} \vartheta_j Z_{pj} + \epsilon_p \quad \epsilon_p \sim N\left(0, \sigma^2_{\epsilon_p}\right) \quad (3)$$

Indices:
p = person
i = item
j = person covariate
k = item/text feature

→ **Inclusion of ε_i in (2) relaxes the strict assumption of LLTM**

→ **However this model (cross classified, random effect model) takes a very long time to converge.**

Model 3: Two-Stage Estimation of Latent Regression LLTM + e (Furr, 2017, a random-effects meta-regression model)

$$logit(\Pr(Y_{ip} = 1 \mid \theta_p) = \theta_p - \delta_i \quad (1)$$

$$\delta_i = \sum_{k=0}^{K} \beta_k X_{ki} + \boxed{u_i} + \epsilon_i \qquad (2)$$

$$u_i \sim N\left(0, \widehat{var}(\hat{\delta}_i)\right)$$

$$\epsilon_i \sim N\left(0, \sigma^2_{\epsilon_i}\right)$$

u_i is residual related to uncertainty in the estimated item difficulty $\hat{\delta}_i$

u_i has known variance ($\widehat{var}(\hat{\delta}_i)$), which is the square of estimated std error of $\hat{\delta}_i$

$$\theta_p = \sum_{j=0}^{J} \vartheta_j Z_{pj} + \epsilon_p \quad \epsilon_p \sim N\left(0, \sigma^2_{\epsilon_p}\right) \quad (3)$$

Indices:
- p = person
- i = item
- j = person covariate
- k = item/text feature

Add interaction terms

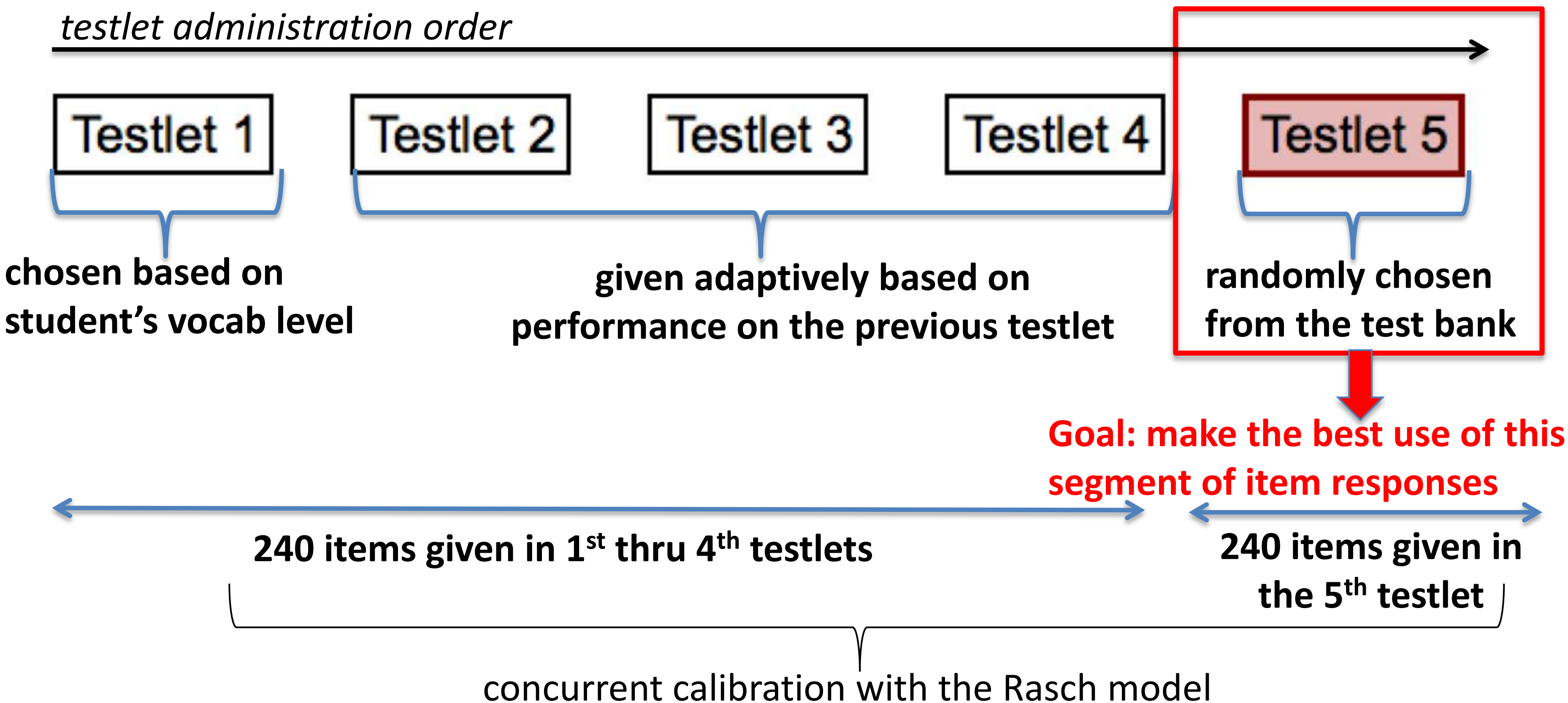
→ Between the reader factor (vocabulary level) and the text- and item-predictors to explore the effect modification.

Level	Testlet 1	Testlet 2	Testlet 3	Testlet 4
1	*Blobfish*	*Frog*	*Sand Dollar*	*Sea Star*
2	*Bread*	*Corn*	*Gum*	*Rice*
3	*Cave Art*	*Dance*	*Rain Sticks*	*Singing Bowls*
4	*Dragon Boats*	*Kite Fighting*	*Stick Juggling*	*Tribal Masks*
5	*History of Sports*	*Injury in Sports*	*Science in Sports*	*Technology in Sports*
6	*Black Holes*	*Comets*	*Solar Flares*	*Space Junk*
7	*High-Speed Boats*	*Mars Rovers*	*Self-Driving Cars*	*Unmanned Planes*
8	*Dream Catchers*	*Fish Rubbing*	*Paper Cutting*	*Shadow Puppets*
9	*Fracking*	*Red Tide*	*Sailing Stones*	*Wildlife Crossings*
10	*Busby Berkeley*	*Duke Ellington*	*John Williams*	*Martha Graham*
11	*Bonsai*	*Dadaism*	*Kabuki*	*Mehndi*
12	*Coral Reefs*	*Deserts*	*Grasslands*	*Tundras*

= link testlets (response data from 1st thru 5th testlets included)

= non-link testlets (response data only from the 5th testlests included)

Each student took 5 testlets in a testing session



6 link testlets were selected which had items **with least discrepancies in difficulty** between 1st-4th testlets vs. the 5th testlet

Model Comparisons

Pseudo-R^2 / Fit Index (Embretson, 1983)

$$\Delta^2 = \frac{\ln L_0 - \ln L_m}{\ln L_0 - \ln L_s}$$

where:
lnL= log-likelihood

<u>indices:</u>
0 = null model (constant difficulty for all items)
m = model to be evaluated (difficulty of item features are estimated)
s = saturated model (difficulty of all items are estimated)

AIC (Akaike, 1974) and BIC (Schwarz, 1978) were also used.