Rasch model for tests with speediness-items and achievement-items under B-GLIRT framework

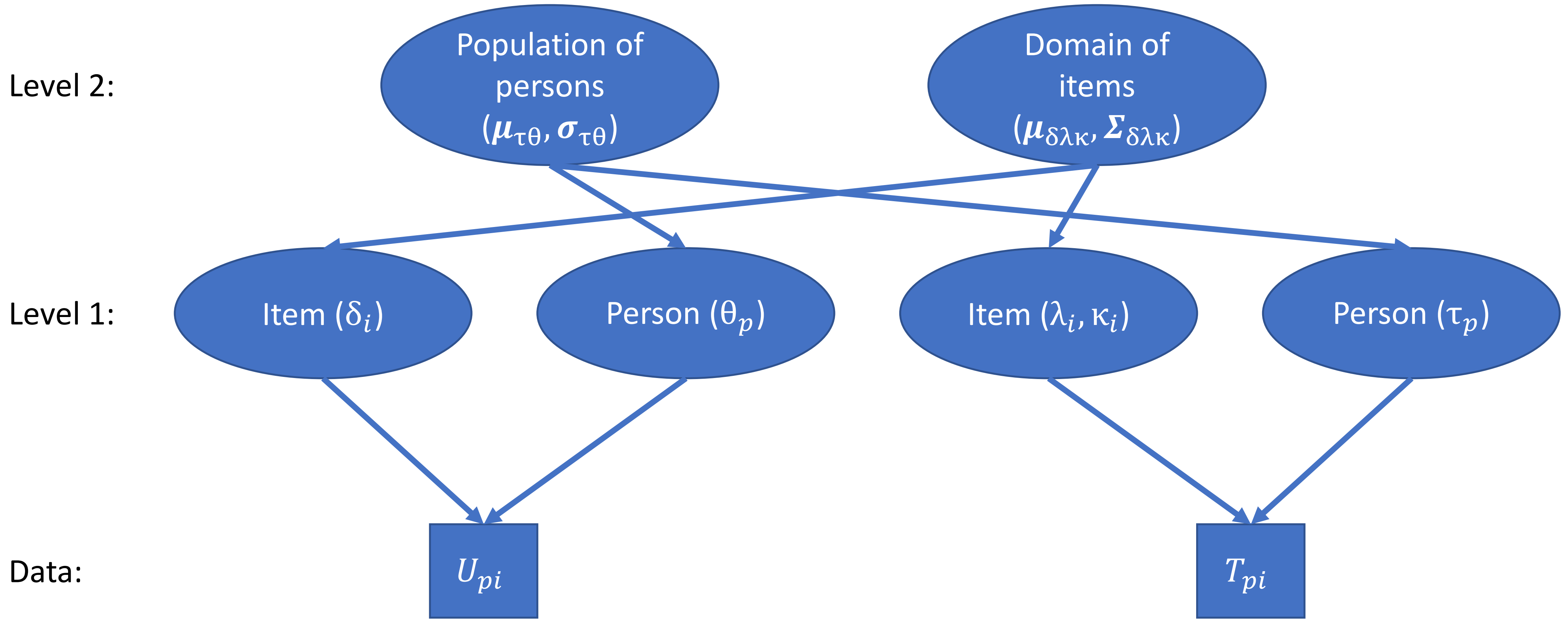
Denis Federiakin

Center for Psychometrics and Measurement in Education,

Institute of Education, NRU HSE

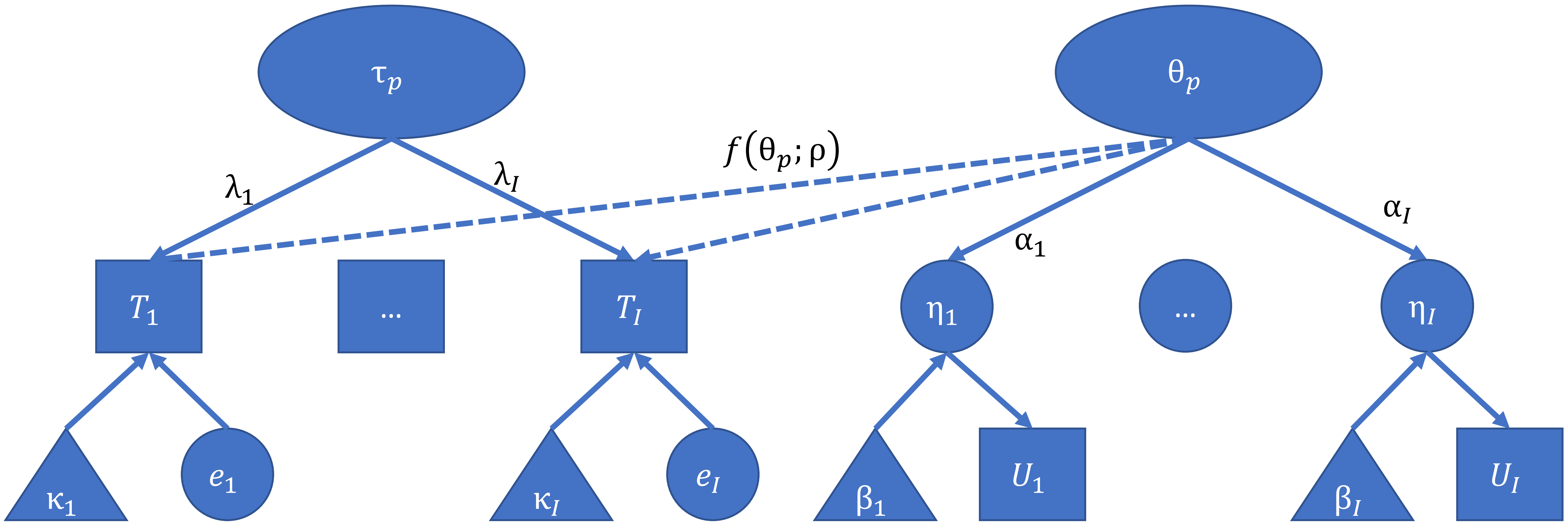
Moscow, 2020

Conceptual framework for item response time and item response accuracy (van der Linden, 2007)



Bivariate generalized linear IRT framework (B-GLIRT) (Molenaar, Tuerlinckx, van der Maas, 2015a)

$$\begin{cases} g[E(U_i)] = \alpha_i\theta + \beta_i \\ E(\ln T_i) = \kappa_i + \lambda_i\tau + f(\theta;\rho) \end{cases}$$



Cross-relation function $f(\theta;\rho)$

- Cross-relation function can have many forms (including non-linear) and describes the exact form of relations between speediness and ability
- Manipulating the cross-relation function, it is possible to show that many IRT-models for response time and response accuracy can be described as special cases of the framework (e.g., van der Linden (2007), Fox, et al. (2007), Roskam (1987), Wang and Hanson (2005))
- Nonetheless, ρ is one of the key parameters according to the framework of van der Linden (2007) – it describes correlation between speediness and ability: $f(\theta;\rho) = \rho\lambda_i\theta$

The problem

- Between-person speed-accuracy trade off suggests that the more time is spent on decision-making, the more accurate the decision is (traditional ability-tests) (Goldhammer, 2015)
  - Nonetheless, some tests contain many simple items, and a respondent should solve them as quickly as possible (speediness-tests) (e.g., Cheung, & Yang, 2020)
- Since $f(\theta; \rho)$ is the same for all items, it reflects an assumption that all items are either speediness-kind or ability-kind
  - Moreover, it provides rather limited information on item-level time-parameters (e.g., $\rho\lambda_i\theta$ or $\rho\alpha_i\theta$)
- Unconstrained simplistic factor analytical model for response times (Fox, et al., 2007) has not been studied yet under B-GLIRT formulation (Molenaar, Tuerlinckx, van der Maas, 2015b)

Current study

- This study aims to (i) investigate Rasch-models possible within B-GLIRT framework, and (ii) investigate a subset of simplistic Rasch-models which provide extended information on item-level parameters
- The further outline:
  - Discussion of the Rasch models within B-GLIRT framework
  - Simulation study of the parameter recovery
  - An example of EE test for higher education students

Rasch models within B-GLIRT framework

$$\begin{cases} g[E(U_i)] = \alpha\theta + \delta_i \\ E(\ln T_i) = \kappa_i + \lambda_i\tau + f(\theta; \xi_i) \end{cases}$$

- Key point: ξ_i provides more information on item-level (Fox, et all., 2007) by sacrificing information on test-level (no ρ)
- Originally proposed forms of $f(\theta; \rho)$:
  - Linear and polynomial function of ability (θ^k)
  - Linear and polynomial function of interaction between ability and speediness ($\theta^k\tau^k$) (see Bolsinova, Molenaar, 2018)
  - Polynomial function of difference between person ability and item intercept ($(\alpha_i\theta + \beta_i)^k$)
  - Their combinations, other functions (e.g., Ranger and Kuhn, 2012) and their fixed counterparts (Roskam, 1987)
- In Rasch-modelling the third form of $f(\theta; \rho)$ is not a simple option
- Therefore, further we discuss only the first two forms of $f(\theta; \xi_i)$

Design of the simulation study 1

- Constants:
  - $I = 20$, all dichotomous
  - $P = 2000$
  - $f(\theta; \xi_i) = \xi_i \theta$
  - $\boldsymbol{\kappa} = \boldsymbol{\delta} =$ equally spaced vector of length I from -2 to 2
- Conditions
  - $\rho = (0, 0.25, 0.5)$
- 100 replications for each condition:
  - $\xi^2 \sim Uniform(\min = 0.16, \max = 0.49)$
  - $\lambda^2 \sim Uniform(\min = 0.16, \max = 0.49)$
  - $(\tau, \theta) \sim MVN(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu} = (0,0), \boldsymbol{\Sigma} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$
- Statistics used for the analysis of simulations:
  - $Bias = \frac{\sum_{r=1}^{R} \widehat{\eta_r}}{R} - \eta$
  - $RMSE = \sqrt{\frac{\sum_{r=1}^{R} (\widehat{\eta_r} - \eta)^2}{R}}$
  - $MAD = \frac{\sum_{r=1}^{R} |\widehat{\eta_r} - \eta|}{R}$
  - Linear Pearson correlation

The model:

$$\begin{cases} g[E(U_i)] = \alpha\theta + \delta_i \\ E(\ln T_i) = \kappa_i + \lambda_i \tau + \xi_i \theta \end{cases}$$

Results of the simulation study 1

Statistics	$\rho = 0$	$\rho = 0.25$	$\rho = 0.5$
Item difficulty vs. true values (δ_i)	Bias = 0.003 RMSE = 0.067 MAD = 0.053	Bias < 0.001 RMSE = 0.065 MAD = 0.051	Bias = 0.002 RMSE = 0.063 MAD = 0.050
Loading of ability on response times vs. true values (ξ_i)	RMSE = 0.023 MAD = 0.017	RMSE = 0.163 MAD = 0.158	RMSE = 0.324 MAD = 0.317
Loading of speediness on response times vs. true values (λ_i)	RMSE = 0.018 MAD = 0.014	RMSE = 0.028 MAD = 0.023	RMSE = 0.087 MAD = 0.082
Ability estimates vs. true values (θ_p)	Cor = 0.940 (SD = 0.012)	Cor = 0.940 (SD = 0.011)	Cor = 0.947 (SD = 0.009)
Speediness estimates vs. true values (τ_p)	Cor = 0.960 (SD = 0.015)	Cor = 0.930 (SD = 0.012)	Cor = 0.847 (SD = 0.015)
RMSEA	0.003 (SD = 0.003)	0.003 (SD = 0.003)	0.004 (SD = 0.003)

Reliability

Omega coefficient for correlated measurement errors (Bentler, 1972, 2009) is calculated on linear CFA model. Expected a Posteriori (EAP; Bock, Mislevy, 1982) and Warm's Weighted Maximum Likelihood (WLE; Warm, 1989) estimates were calculated on IRT model.

Conditions	Reliability	$\rho = 0$	$\rho = 0.25$	$\rho = 0.5$
Without time	EAP	0.769 (0.007)	0.769 (0.007)	0.771 (0.007)
	WLE	0.753 (0.006)	0.753 (0.007)	0.755 (0.007)
	Omega for Ability	0.754 (0.007)	0.754 (0.008)	0.756 (0.008)
With time	Omega for Ability	0.961 (0.005)	0.974 (0.003)	0.981 (0.002)
	Omega for Speediness	0.969 (0.006)	0.967 (0.007)	0.957 (0.009)

Design of the simulation study 2

- Constants:
  - $I = 20$, all dichotomous
  - $P = 2000$
  - $f(\theta; \xi_i) = \xi_i\theta$
  - $\boldsymbol{\kappa} = \boldsymbol{\delta} =$ equally spaced vector of length I from -2 to 2
  - $\rho = 0.25$
- Conditions:
  - $N(\xi_i < 0) = (5, 10)$
- 100 replications for each condition:
  - $\lambda^2 \sim Uniform(\min = 0.16, \max = 0.49)$
  - $(\tau, \theta) \sim MVN(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu} = (0,0), \boldsymbol{\Sigma} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$
- Statistics used for the analysis of simulations:
  - $Bias = \frac{\sum_{r=1}^{R} \widehat{\eta_r}}{R} - \eta$
  - $RMSE = \sqrt{\frac{\sum_{r=1}^{R} (\widehat{\eta_r} - \eta)^2}{R}}$
  - $MAD = \frac{\sum_{r=1}^{R} |\widehat{\eta_r} - \eta|}{R}$
  - Linear Pearson correlation

The studied model (see Fox, et al., 2007):

$$\begin{cases} g[E(U_{pi})] = \alpha\theta + \delta_i \\ E(\ln T_{pi}) = \kappa_i + \lambda_i\tau + \xi_i\theta \end{cases}$$

Constrained models (e.g., van der Linden, 2007):

$$\begin{cases} g[E(U_{pi})] = \alpha_i\theta + \beta_i \\ E(\ln T_{pi}) = \kappa_i + \lambda_i\tau - \rho\lambda_i\theta \end{cases}$$

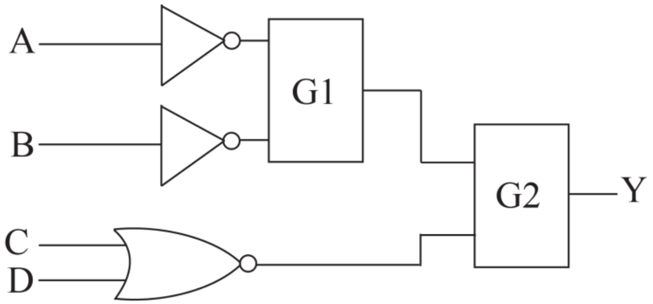
Results of the simulation study 2

Statistics	$N(\xi_i < 0) = 5$	$N(\xi_i < 0) = 10$
Item difficulty vs. true values (δ_i)	Bias < 0.001 RMSE = 0.065 MAD = 0.052	Bias = 0.003 RMSE = 0.062 MAD = 0.049
Loading of ability on response times vs. true values (ξ_i)	RMSE = 0.161 MAD = 0.156	RMSE = 0.164 MAD = 0.160
Loading of speediness on response times vs. true values (λ_i)	RMSE = 0.026 MAD = 0.022	RMSE = 0.026 MAD = 0.022
Ability estimates vs. true values (θ_p)	Cor = 0.987 (SD = 0.006)	Cor = 0.959 (SD = 0.006)
Speediness estimates vs. true values (τ_p)	Cor = 0.960 (SD = 0.007)	Cor = 0.930 (SD = 0.012)
RMSEA	0.003 (SD = 0.003)	0.003 (SD = 0.003)

Electrical Engineering test for higher education students

- 25 dichotomous items, 853 Russian and 1662 Chinese EE students

In the figure, the output *Y* is supposed to be $Y = AB + \bar{C}\bar{D}$. The gates G1 and G2 must be, respectively



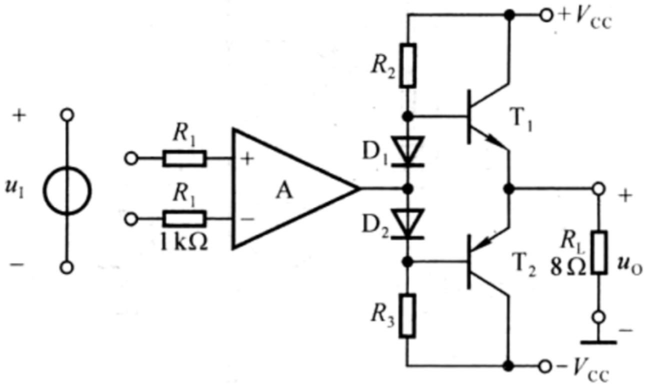
A. NOR, OR

B. OR, NAND

C. NAND, OR

D. AND, NAND

For the following circuit: VCC =15V, the saturation voltage of both T1 and T2 is |UCES|=1V, the maximal output voltage amplitude of integrated operational amplifier is ±13V, the threshold voltage of the diode is 0.7V.



Please determine the maximal output power of the circuit assuming that the amplitude of the input voltage is great enough.

A. 15.6w

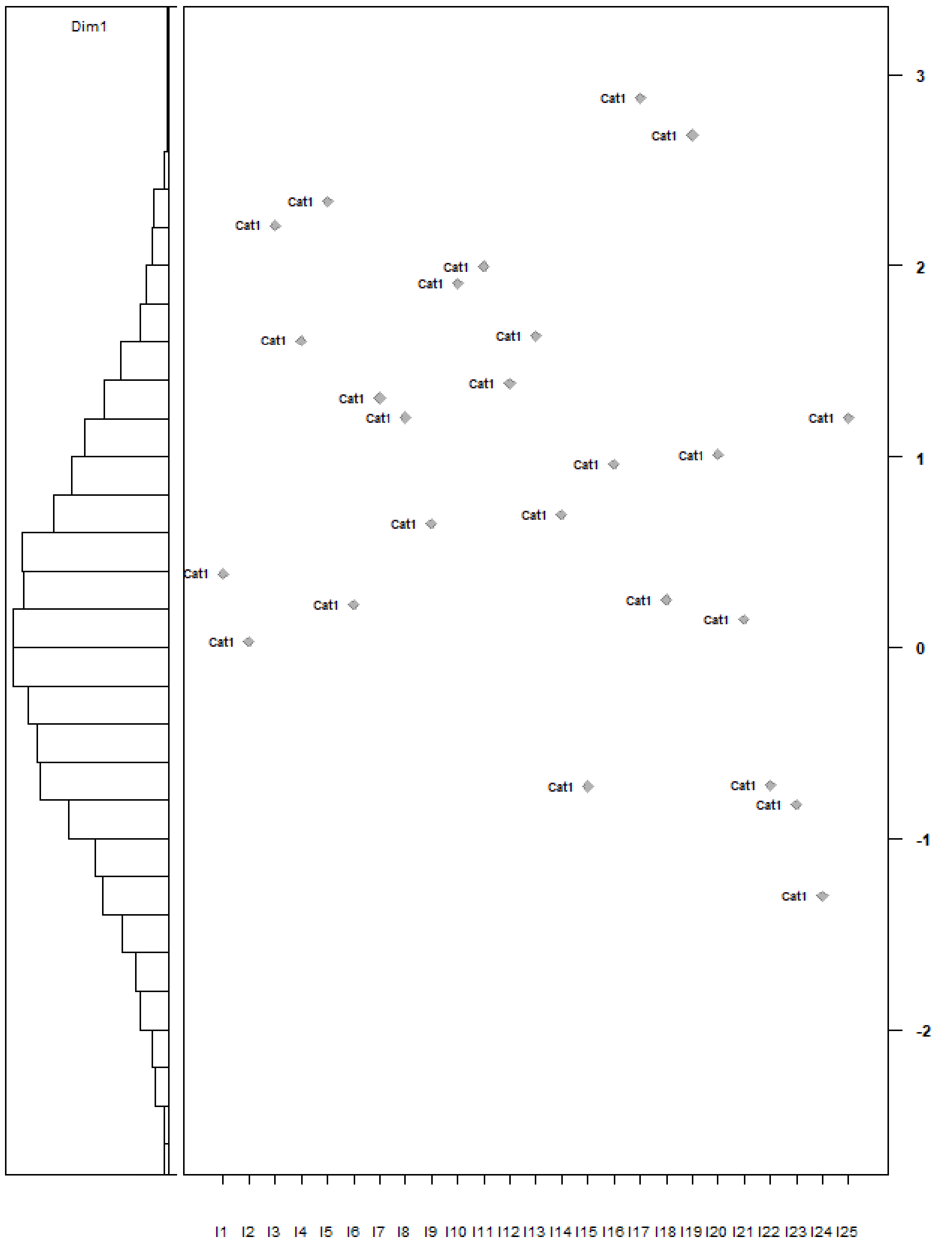
B. 18.5w

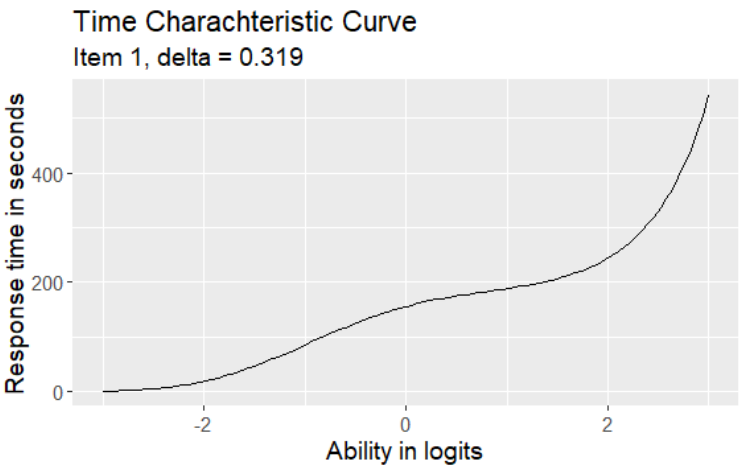
C. 10.6w

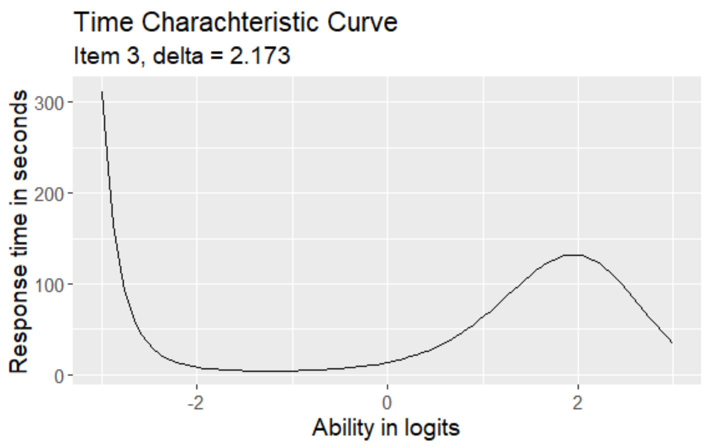
D. 8.1w

EE Results 1

$$\begin{cases} g[E(U_i)] = \alpha\theta + \delta_i \\ E(\ln T_i) = \kappa_i + \lambda_i\tau + \sum_{d=1}^{d=3} \xi_{di}\theta^d \end{cases}$$







EE Results 2

$$\begin{cases} g[E(U_i)] = \alpha\theta + \delta_i \\ E(\ln T_i) = \kappa_i + \lambda_i\tau + \sum_{d=1}^{d=3} \xi_{di}\theta^d + \sum_{t=1}^{t=2} \zeta_{ti}\theta^t\tau \end{cases}$$

Item	Interaction 1 ($\theta_p\tau_p$)	Interaction 2 ($\theta_p^2\tau_p$)	Ability (θ_p)	Ability 2 (θ_p^2)	Ability 3 (θ_p^3)
T1	-0.314	0	0.550	-0.079	-0.054
T2	-0.129	0	0.725	-0.245	0
T3	-0.236	0	1.533	0	-0.228
T4	-0.137	0	1.044	-0.190	-0.153
T5	-0.367	0	0.606	-0.112	-0.059
T6	-0.378	-0.123	0.702	-0.163	0
T7	-0.286	-0.063	0.601	-0.201	-0.034
T8	-0.191	-0.150	0.904	-0.201	-0.087
T9	0	0	0.461	-0.243	0
T10	0	0	0.668	-0.215	0
T11	-0.167	0	1.068	-0.179	-0.131
T12	-0.194	0	1.245	-0.158	-0.182
T13	0	0	1.261	-0.232	-0.199
T14	0	0	0.918	-0.222	-0.146
T15	-0.236	0.058	0.328	-0.208	-0.044
T16	-0.360	0	0.561	-0.166	-0.044
T17	-0.227	-0.052	0.631	-0.155	0
T18	0	0	0.656	-0.284	0
T19	0	0	0.610	-0.110	-0.078
T20	-0.202	0	0.462	-0.202	0
T21	-0.175	-0.070	0.715	-0.202	-0.046
T22	-0.114	0	0.651	-0.244	-0.034
T23	0	0	0.527	-0.306	0
T24	0	0	0.272	-0.179	0
T25	0	0	1.030	-0.224	-0.150

Index	The model without interactions	The model with interactions
AIC	169360.578	168662.308
BIC	170386.663	169979.894

Conclusion

- A set of simplistic Rasch models for item response accuracy and item response time can be proposed within B-GLIRT framework, but it limits possible cross-relation functions
- Distribution of single cross-relation function over the set of item-specific functions results in sufficient model fit and provides more insight on response process on item-level
  - However, constraining ability and speediness to be orthogonal can be inefficient for measuring speediness parameter
- Increasing correlation of ability and speediness results in increased loading of item response time on ability, but barely affects other estimates