British Journal of Mathematical and Statistical Psychology

the british psychological society

Editor's Choice

A hierarchical latent response model for inferences about examinee engagement in terms of guessing and item-level non-response



Esther Ulitzsch[1]*, Matthias von Davier[2] and Steffi Pohl[1]

[1]Methods and Evaluation/Quality Assurance, Freie Universität Berlin, Germany

[2]National Board of Medical Examiners, Philadelphia, Pennsylvania, USA

Reporter: Yingshi Huang

Introduction

- large-scale assessments (LSAs)
  - examinees actively try to determine the correct answer
  - low-stake testing: **disengagement**

- ➢ randomly guessing
- ➢ answering items perfunctorily
- ➢ generating no response at all

➡

Purpose: provide a generalized modelling framework to identify disengagement

(guessing responses + omissions + response times)

Previous approaches for disengaged behaviour

- Guessing and perfunctory answers
  - Response-time-based scoring techniques
  - Model-based approaches

- Omissions
  - Response-time-based scoring techniques
  - Model-based approaches

Previous approaches for disengaged behaviour

- Guessing and perfunctory answers **(RT-based)**
  - RTs below a certain **threshold**

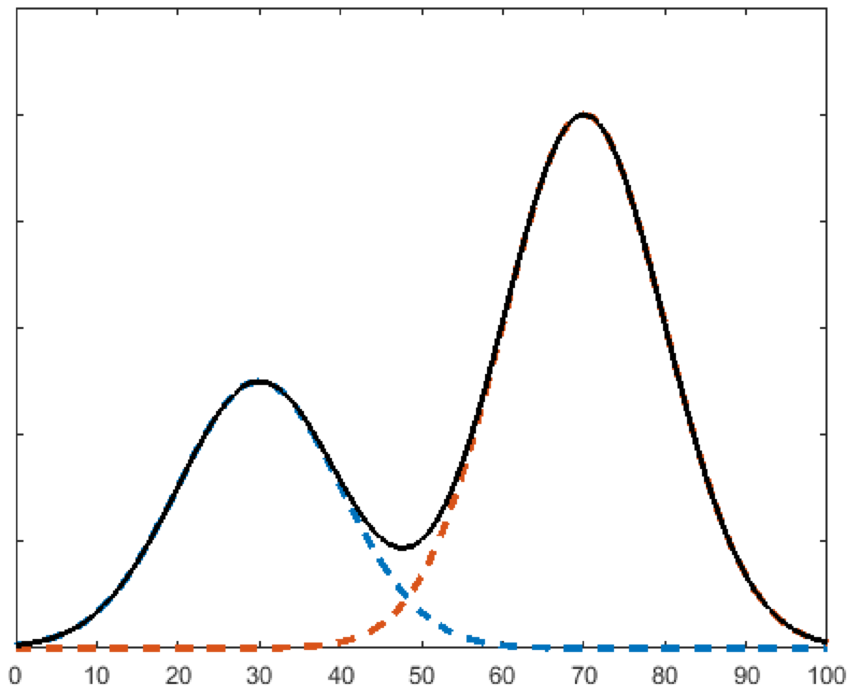
1. define a common threshold for **all items**

   the minimum time needed to engage

2. **item-specific** thresholds

   10% of the average time

   bimodal RT distributions for a distinctive gap



Previous approaches for disengaged behaviour

- Guessing and perfunctory answers **(RT-based)**
  - RTs below a certain **threshold**

1. define a common threshold for **all items**

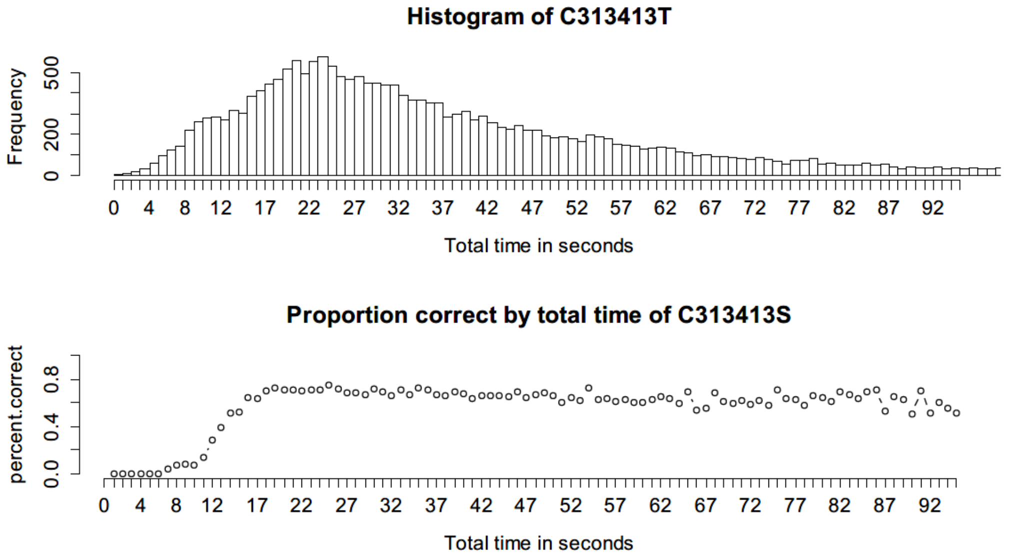
   the minimum time needed to engage

2. **item-specific** thresholds

   10% of the average time

   bimodal RT distributions for a distinctive gap

   RT distributions jointly with the conditional proportion correct



Goldhammer et al., 2016 *OECD Education Working Papers*

Previous approaches for disengaged behaviour

- Guessing and perfunctory answers **(model-based)**
  - apply mixture modelling techniques

two different processes: solution behaviour and rapid guessing behaviour

1. customary item response theory (IRT) models

   solution behaviour: examinee ability and item difficulty

$$P(Y_{ij}=1|\Delta_{ij}=1,a_j,b_j,c_j)=c_j+(1-c_j)\frac{\exp[a_j(\theta_i-b_j)]}{1+\exp[a_j(\theta_i-b_j)]}$$

   rapid guessing processes: contain no information on ability

$$P(Y_{ij}=1|\Delta_{ij}=0)=g_j$$

2. different lognormal distributions

$$T_{ij}^{\text{obs}}=(1-\Delta_{ij})T_{ij}+\Delta_{ij}C_{ij}$$

Wang & Xu, 2015 *BJMSP*

Previous approaches for disengaged behaviour

- Assumptions and limitations
  - Response-time-based scoring techniques
    1. heuristic and might considerably disagree in the rate
    2. coded as missing and therefore ignored when estimating ability

  - Model-based approaches

    with strong assumptions

    1. mixing proportions

       varying mixing proportions at the **item level**: items vary & examinees constant

       varying mixing proportions at the **examinee level**: items constant & examinees vary

       ✓ examinee characteristics: academic ability or achievement goals

       ✓ item characteristics: response format or position

    2. dependency between ability and engagement

➡ **vary at the item-by-examinee level + joint modelling disengagement and ability**

Previous approaches for disengaged behaviour

- Omissions
  - Response-time-based scoring techniques

    distinguish item omissions occurring due to processes **different** from and **similar** to those operating when examinees generate **(engaged) responses**

    1. remarkably short RTs: skipped it without trying to solve

    2. RTs that do not notably differ from RTs associated with (wrong) observed responses: occurred for skill-related reasons

    ✓ the Programme for the International Assessment of Adult Competencies (PIAAC):

    5-s scoring rule

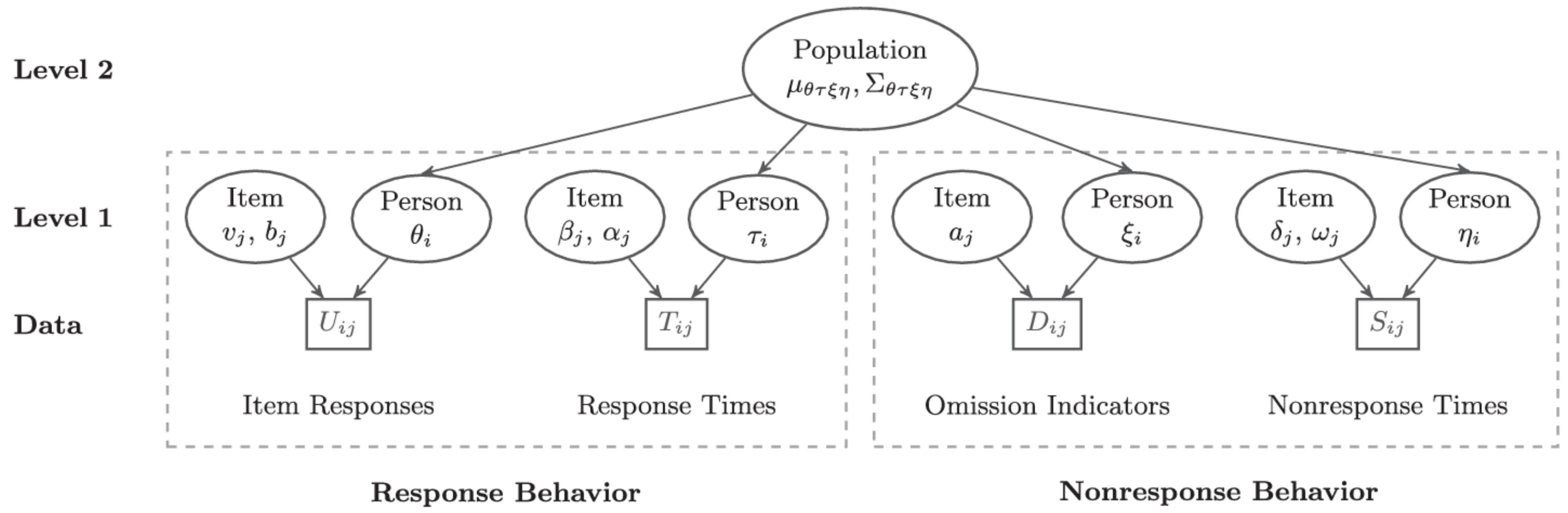
Previous approaches for disengaged behaviour

- Omissions
  - Model-based approaches

an additional manifest or latent variable: the examinees' propensity to omit items

✓ Response: $p(u_{ij}=1)=\frac{\exp(\theta_i - b_j)}{1+\exp(\theta_i - b_j)}$

✓ Omission: $p(d_{ij}=1)=\frac{\exp(\xi_i - a_j)}{1+\exp(\xi_i - a_j)}$



Ulitzsch et al, 2019 *MBR*

Previous approaches for disengaged behaviour

- Assumptions and limitations
  - Response-time-based scoring techniques
    1. the probability of solving an omitted item is zero
    2. item omissions are ignorable
  - Model-based approaches
    - ✓ allow to assess how examinee ability relates to the probability of omitting responses

      ↓

      jointly model disengaged behaviour and ability

      omission: disengaged behaviour

      observed responses: solution behaviour (examinees do not omit while engaged)

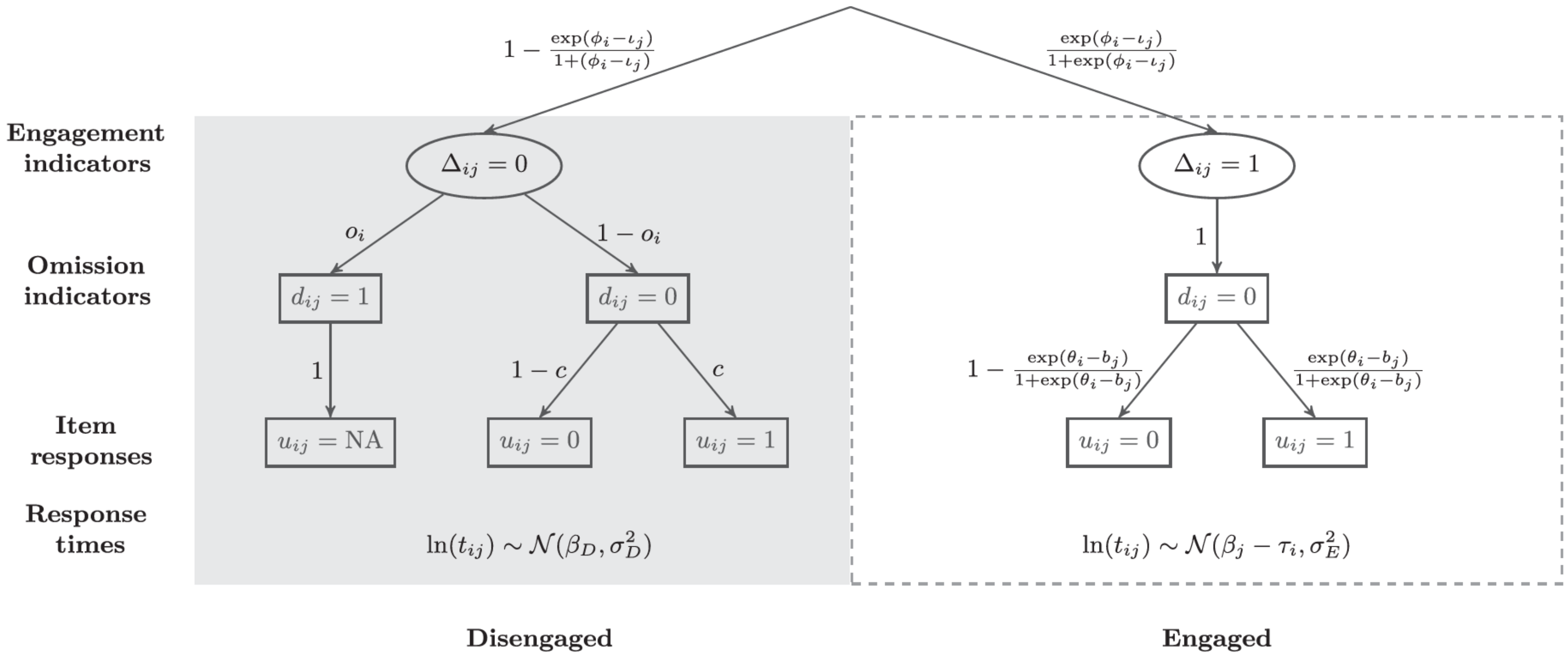
      ↓

      all item omissions / observed responses to stem from the same data-generating processes

Proposed model

1. item-by-examinee specific
2. disengagement: guessing & omission
3. jointly model with ability

- the speed-accuracy + engagement (SA+E) model



Proposed model

- disengaged behaviour ($\Delta_{ij} = 0$)

1. Omission:

$$p(d_{ij} = 1|\Delta_{ij} = 0) = o_i = \frac{\exp(\gamma_0 + \gamma_1\theta_i + \gamma_2\tau_i)}{1 + \exp(\gamma_0 + \gamma_1\theta_i + \gamma_2\tau_i)}$$

2. Response:

$$p(u_{ij} = 1|\Delta_{ij} = 0) = c$$

3. Response time:

$$\ln(t_{ij}|\Delta_{ij} = 0) \sim \mathcal{N}(\beta_D, \sigma_D^2)$$

$$\beta_j = \beta_D + \beta_j^*, \quad \text{where } \beta_j^* \geq 0$$

- engaged behaviour ($\Delta_{ij} = 1$)

1. Omission:

$$p(d_{ij} = 1|\Delta_{ij} = 1) = 0$$

2. Response:

$$p(u_{ij} = 1) = \frac{\exp(\theta_i - b_j)}{1 + \exp(\theta_i - b_j)}$$

3. Response time:

$$\ln(t_{ij}|\Delta_{ij} = 1) \sim \mathcal{N}(\beta_j - \tau_i, \sigma_E^2)$$

offset parameter: how much longer examinees need to engage with the item

Proposed model

- Higher-order models

$$p(\Delta_{ij} = 1) = \frac{\exp(\phi_i - \iota_j)}{1 + \exp(\phi_i - \iota_j)}$$

engagement: whether examinees tend to approach items engagedly

engagement difficulty: how easily examinees interact with an item engagedly

  - person parameters (setting the expectations to zero):

$$\boldsymbol{\mu}_{\mathcal{P}} = (\mu_\phi, \mu_\theta, \mu_\tau)$$

$$\boldsymbol{\Sigma}_{\mathcal{P}} = \begin{pmatrix} \sigma_\phi^2 & \sigma_{\phi\theta} & \sigma_{\phi\tau} \\ \sigma_{\phi\theta} & \sigma_\theta^2 & \sigma_{\theta\tau} \\ \sigma_{\phi\tau} & \sigma_{\theta\tau} & \sigma_\tau^2 \end{pmatrix}$$

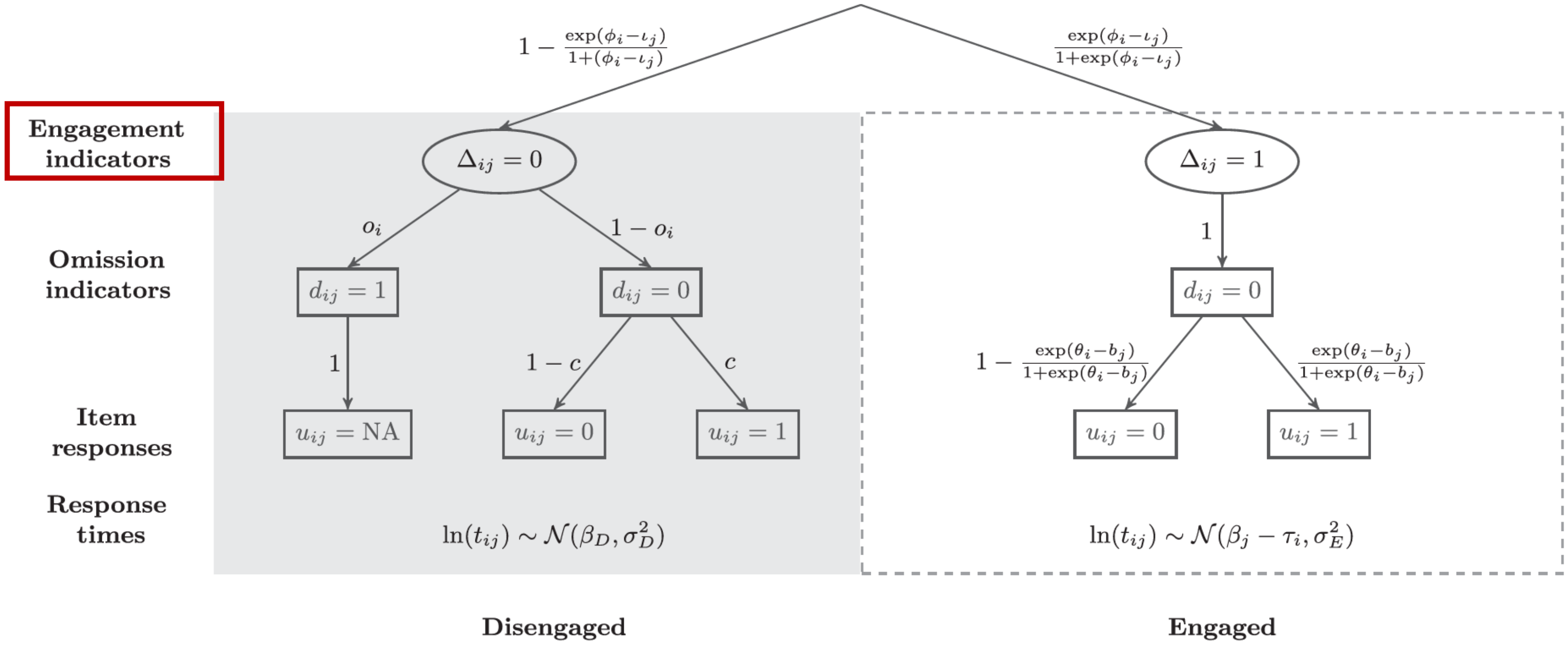
  - item parameters are modelled as fixed effects

Proposed model

- The proposed model's likelihood

$$\mathcal{L} = \prod_{i=1}^{N}\prod_{j=1}^{K}\Big(p(\Delta_{ij}=1|\phi_i,\iota_j)(1-d_{ij})p(u_{ij}|\theta_i,b_j)f(t_{ij}|\tau_i,\beta_j,\sigma_E^2)+$$

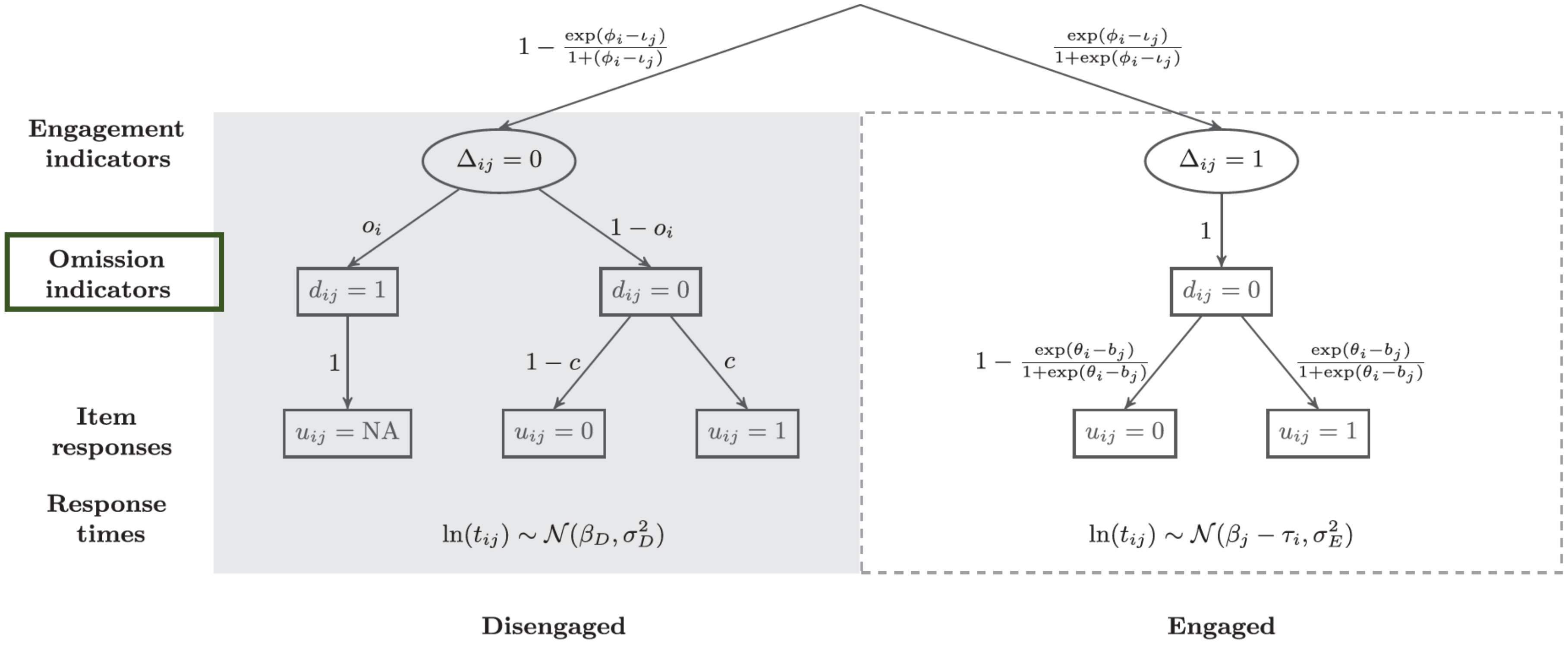
$$(1-p(\Delta_{ij}=1|\phi_i,\iota_j))p(d_{ij}|\gamma,\theta_i,\tau_i)p(u_{ij}|c)^{(1-d_{ij})}f(t_{ij}|\beta_D,\sigma_D^2)\Big)\cdot g(\boldsymbol{\phi},\boldsymbol{\theta},\boldsymbol{\tau}|\boldsymbol{\mu}_{\mathcal{P}},\boldsymbol{\Sigma}_{\mathcal{P}})$$



Proposed model

- The proposed model's likelihood

$$\mathcal{L} = \prod_{i=1}^{N}\prod_{j=1}^{K}\Big(p(\Delta_{ij}=1|\phi_i,\iota_j)(1-d_{ij})p(u_{ij}|\theta_i,b_j)f(t_{ij}|\tau_i,\beta_j,\sigma_E^2)+$$
$$(1-p(\Delta_{ij}=1|\phi_i,\iota_j))p(d_{ij}|\gamma,\theta_i,\tau_i)p(u_{ij}|c)^{(1-d_{ij})}f(t_{ij}|\beta_D,\sigma_D^2)\Big)\cdot g(\boldsymbol{\phi},\boldsymbol{\theta},\boldsymbol{\tau}|\boldsymbol{\mu}_{\mathcal{P}},\boldsymbol{\Sigma}_{\mathcal{P}})$$



Prior distributions

- person parameter variance–covariance matrix:

$$\Sigma_{\mathcal{P}} = \text{diag}(S_{\mathcal{P}})\Omega_{\mathcal{P}}\text{diag}(S_{\mathcal{P}})$$

standard deviations

correlation matrix

✓ circumvent the dependencies between variances and correlations inherent to inverse Wishart priors

$\Omega_{\mathcal{P}}$: LKJ prior with shape 1 (a uniform distribution)

$S_{\mathcal{P}}$: half Cauchy priors with location 0 and scale 5

- item parameters:

ι_j

b_j

β_j^* β_D ➡ normal priors with mean 0 and standard deviation 10

γ

σ_E σ_D ➡ half Cauchy priors with location 0 and scale 5

c ➡ beta priors with $B(1,1)$

```
// prior person parameter
PersPar~ multi_normal(Zero,SigmaP);
sigmaP ~  cauchy(0,5);
correlP ~ lkj_corr(1);
// prior item parameter
iota  ~ normal(0, 10);
b  ~ normal(0, 10);
diffbeta  ~ normal(0, 10);
pCorrNE ~ beta(1,1);
gamma0 ~ normal(0,10);
gamma1 ~ normal(0,10);
gamma2 ~ normal(0,10);
muC ~ normal(0, 10);
sigmaE ~  cauchy(0,5);
sigmaD ~  cauchy(0,5);
```

Parameter recovery

- Simulation purpose

1. whether true parameter values can satisfactorily be recovered under realistic conditions

2. identify boundary conditions concerning the sparseness of information on examinee disengagement for the detection

Parameter recovery

- Data generation (the SA+E model)
  - the number of **examinees** (per item): 250, 500, 1000
  - the number of **items**: 10, 20
  - the rate of **disengaged behaviour**: 5%, 10%
  - the percentage of **omissions** (opposed to guessing): 10%, 50%, 90%

  - variances of ϕ, θ, and τ: 3.50, 1.00, and 0.05

    $\text{cor}(\phi,\theta)$: 0.55

    $\text{cor}(\phi,\tau)$: 0.20

    $\text{cor}(\theta,\tau)$: -0.40

  - item parameters:

    $\{\iota_0 + 0.5l\}_{l=1}^{5}$ for engagement difficulties (ι_0=-5, -4.25)

    $\{3 + 0.25l\}_{l=1}^{5}$ for difficulties

    $\{-1 + 0.5l\}_{l=1}^{5}$ for time intensities

    $c = .25$ for guessing

  - the logistic regression parameters:

    $\gamma_\tau = -10 \quad \gamma_\theta = -1$

    the intercept

    $\gamma_0 = 3 \quad \gamma_0 = -3 \quad \gamma_0 = 0$

  - logarithmized disengaged RTs

    $\beta_D = 3 \quad \sigma_D^2 = 0.15$

    engaged RTs

    $\sigma_E^2 = 1.95$

Parameter recovery

- Estimation procedure
  - the No-U-Turn sampler (an adaptive form of Hamiltonian Monte Carlo sampling)
  - each data set: four Markov chain Monte Carlo (MCMC) chains
  - 10,000 iterations each chain
    (first 5,000 employed as warm-up)

- Evaluation indexes
  - potential scale reduction factor (PSRF) values
  - effective sample sizes (ESSs)

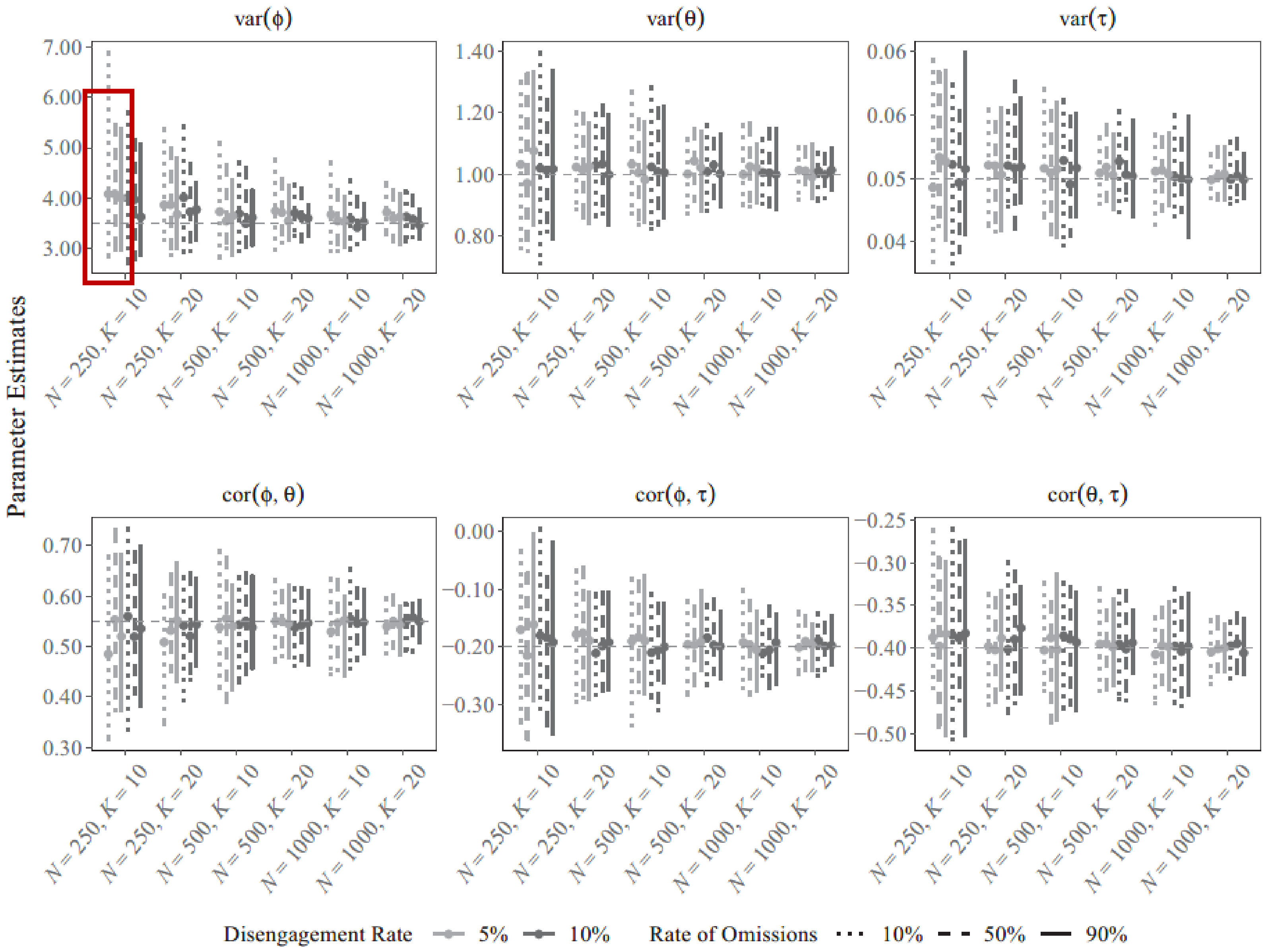
Results

Table 1. Proportions of replications with PSRF values < 1.10 and ESS > 400 for all parameters after 10,000 iterations

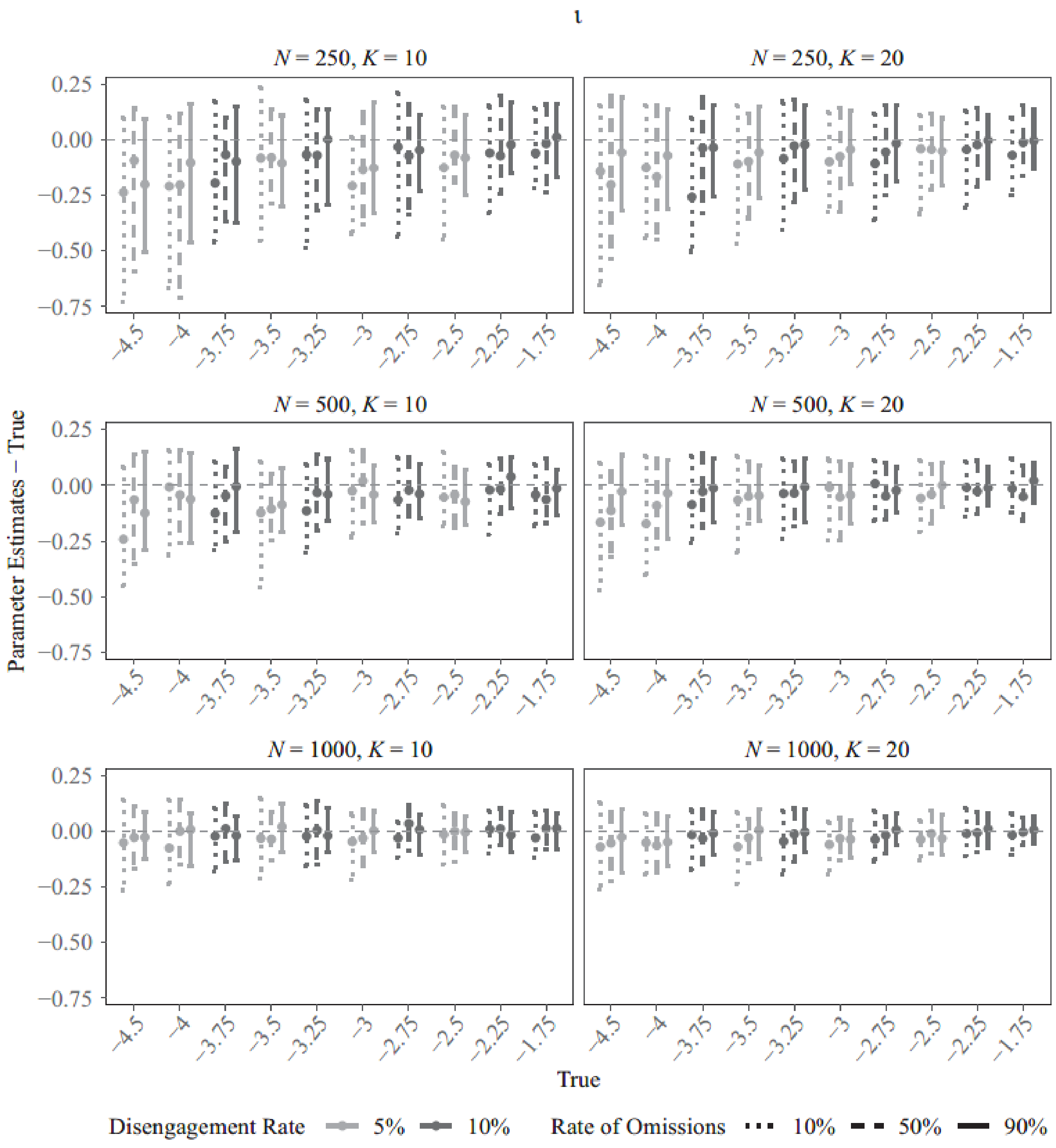
N	*K*	Disengaged (%)	Omitted (%)	PSRF < 1.10	ESS > 400
250	10	5	10	1.00	.94
			50	.84	.74
			90	.96	.96
		10	10	.92	.82
			50	.86	.84
			90	.92	.92
	20	5	10	.96	.96
			50	1.00	1.00
			90	.96	.96
		10	10	1.00	1.00
			50	.98	.98
			90	1.00	1.00
500	10	5	10	.92	.88
			50	.98	.94
			90	.94	.94
		10	10	.96	.96
			50	.98	.98
			90	.82	.82
	20	5	10	1.00	1.00
			50	.98	.98
			90	.94	.94
		10	10	1.00	1.00
			50	.98	.98
			90	.94	.92
1,000	10	5	10	1.00	1.00
			50	.96	.96
			90	.88	.86
		10	10	.98	.98
			50	.98	.98
			90	.92	.92
	20	5	10	.98	.98
			50	1.00	1.00
			90	1.00	1.00
		10	10	.98	.98
			50	.98	.98
			90	.98	.98

Note. Omissions give the percentage of item omissions on disengaged behaviour.
N = number of examinees; *K* = number of items.

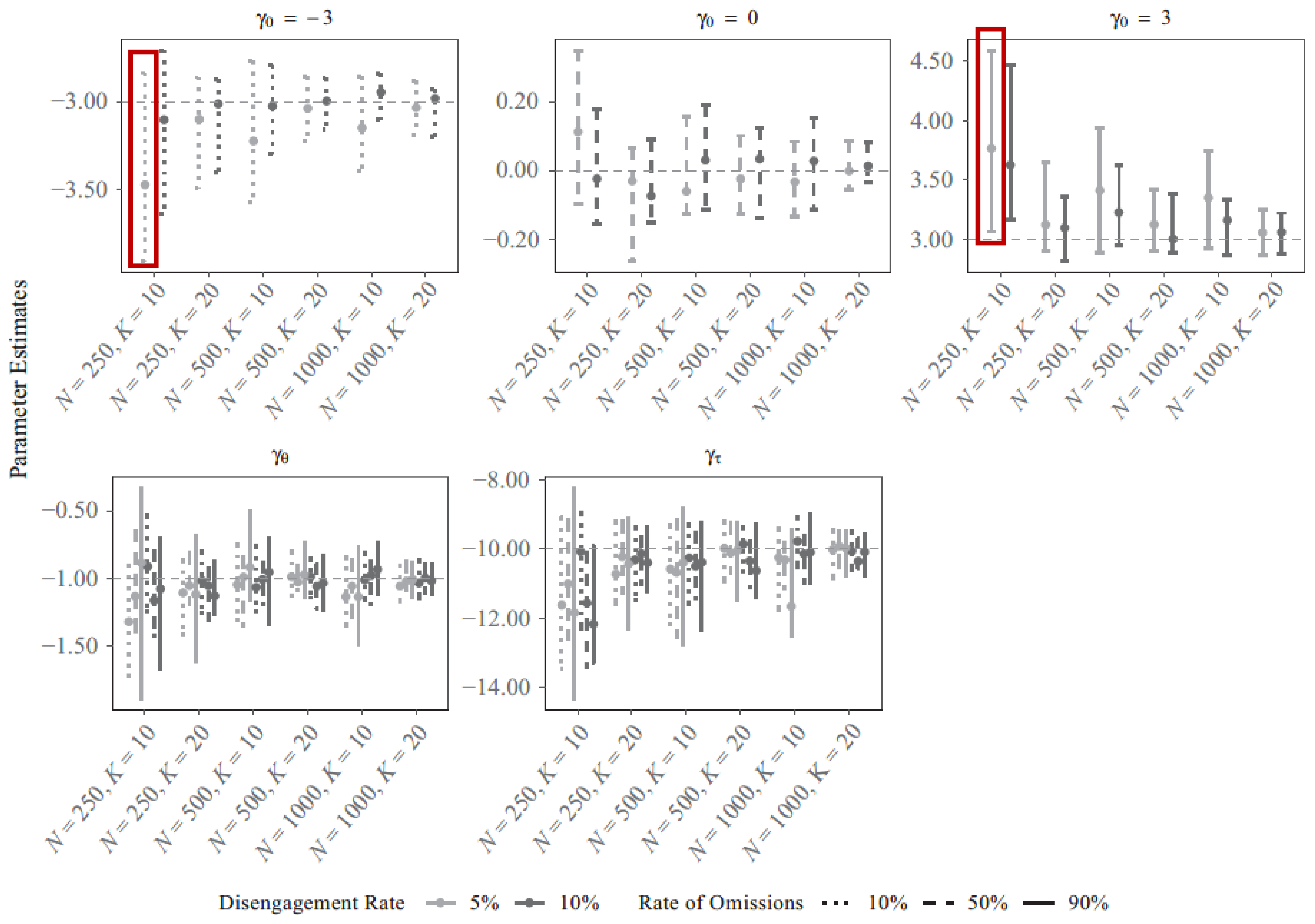
Results


var(ϕ)
var(θ)
var(τ)
cor(ϕ, θ)
cor(ϕ, τ)
cor(θ, τ)
Parameter Estimates
N = 250, K = 10
N = 250, K = 20
N = 500, K = 10
N = 500, K = 20
N = 1000, K = 10
N = 1000, K = 20
Disengagement Rate 5% 10%
Rate of Omissions 10% 50% 90%

Results


ι
N = 250, K = 10
N = 250, K = 20
N = 500, K = 10
N = 500, K = 20
N = 1000, K = 10
N = 1000, K = 20
0.25
0.00
−0.25
−0.50
−0.75
−4.5
−4
−3.75
−3.5
−3.25
−3
−2.75
−2.5
−2.25
−1.75
Parameter Estimates − True
True
Disengagement Rate
5%
10%
Rate of Omissions
10%
50%
90%

Results

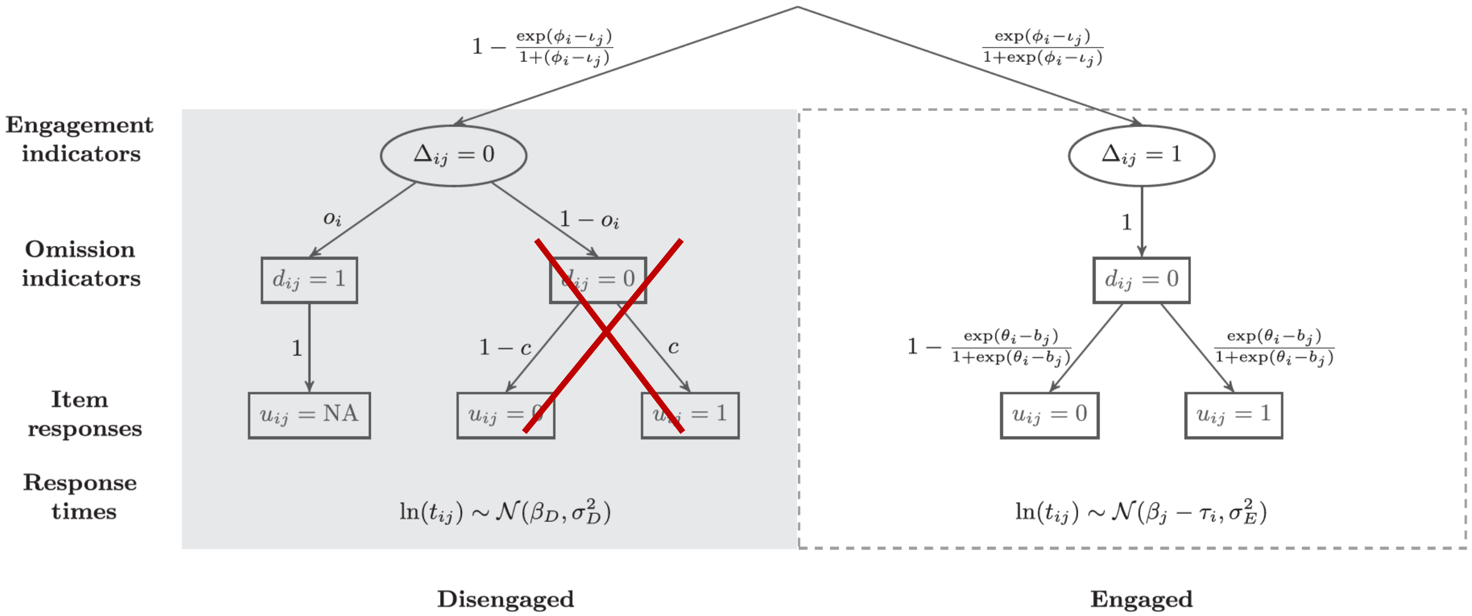


γ0 = −3
γ0 = 0
γ0 = 3
γθ
γτ
Parameter Estimates
N = 250, K = 10
N = 250, K = 20
N = 500, K = 10
N = 500, K = 20
N = 1000, K = 10
N = 1000, K = 20
Disengagement Rate 5% 10%
Rate of Omissions 10% 50% 90%

Illustrating the model

- Simulation purpose
  - how the SA+E model differs conceptually from current approaches
    (disengagement rate: 10%; omissions: 50%)

1. assume all observed responses to stem from engaged response processes:
   the speed-accuracy + omission (SA+O) model (Ulitzsch et al., 2019)

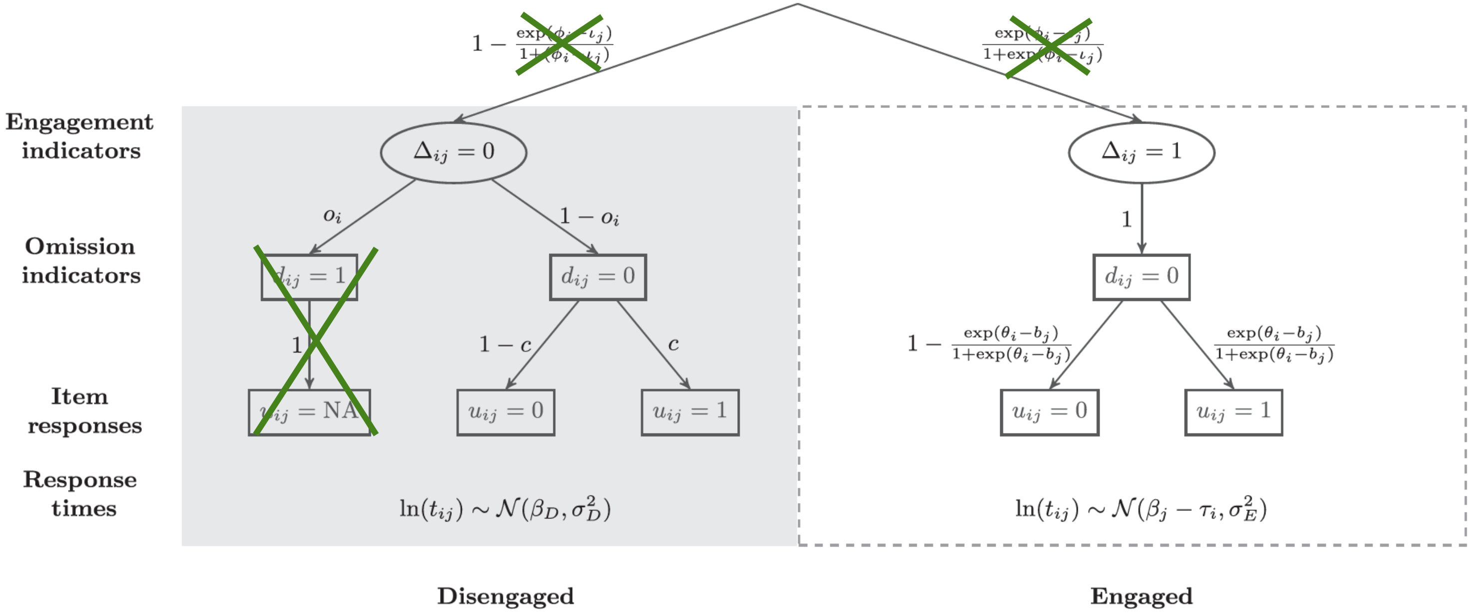


Illustrating the model

- Simulation purpose
  - how the SA+E model differs conceptually from current approaches
    (disengagement rate: 10%; omissions: 50%)

1. assume all observed responses to stem from engaged response processes:
   the speed-accuracy + omission (SA+O) model (Ulitzsch et al., 2019)

2. assume engagement to be unrelated to ability and item omissions to be ignorable:
   the mixture model (Wang & Xu, 2015)



Illustrating the model

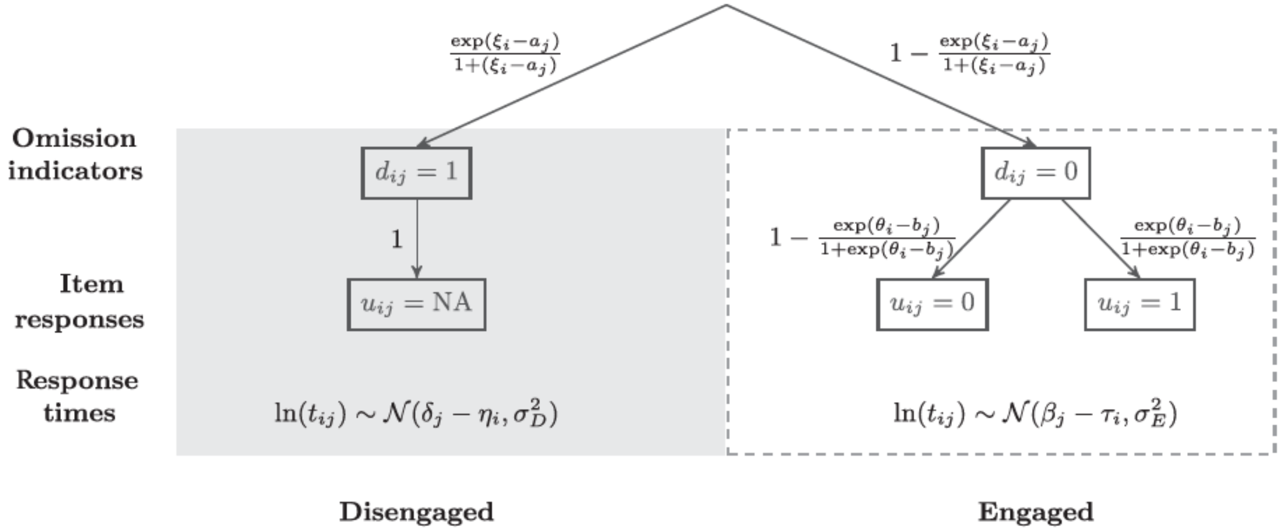


Figure 5. SA+O model by Ulitzsch et al. (2019).

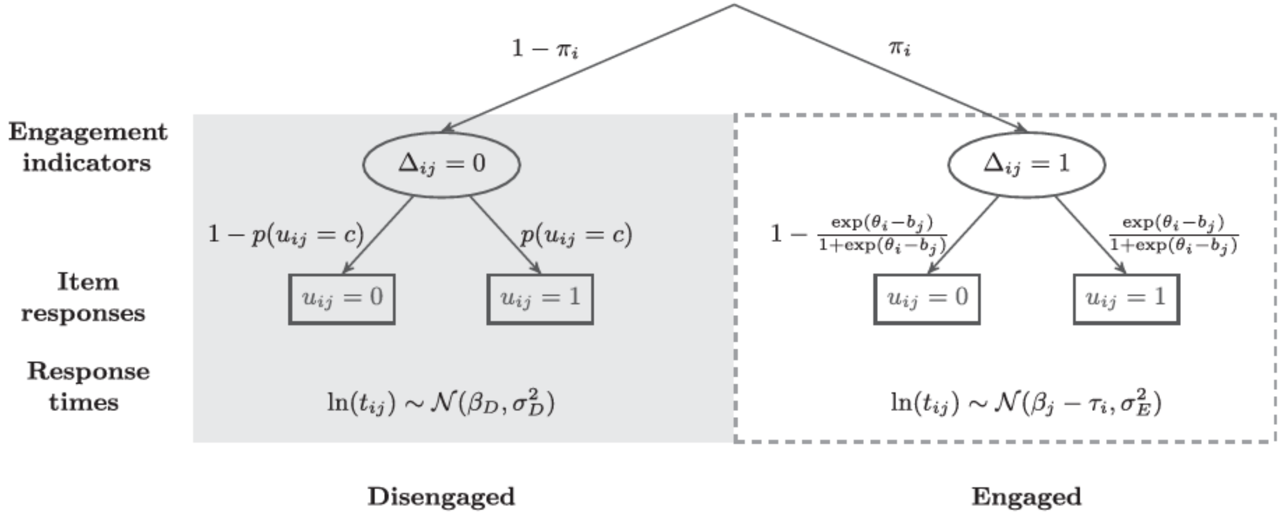
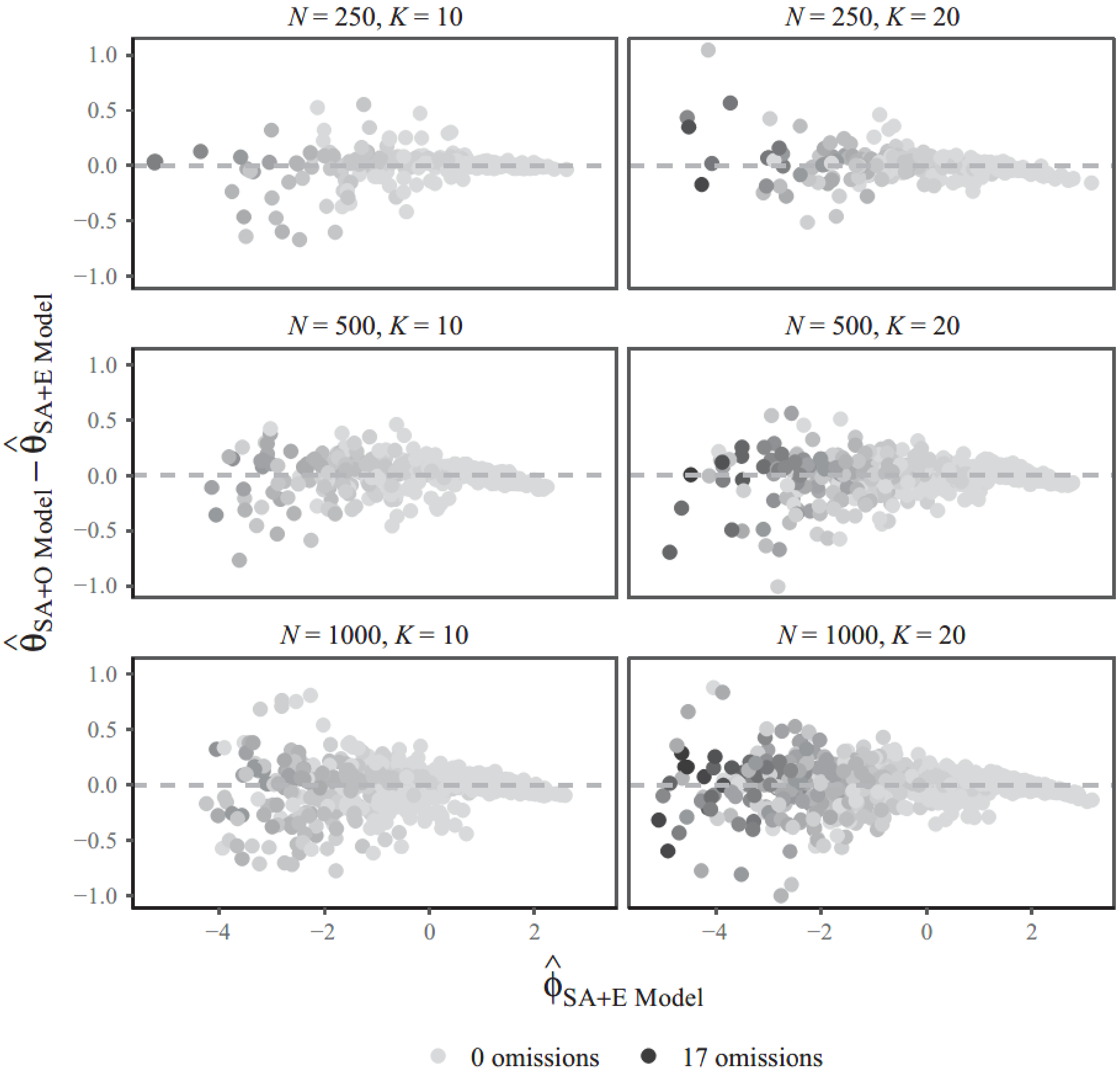
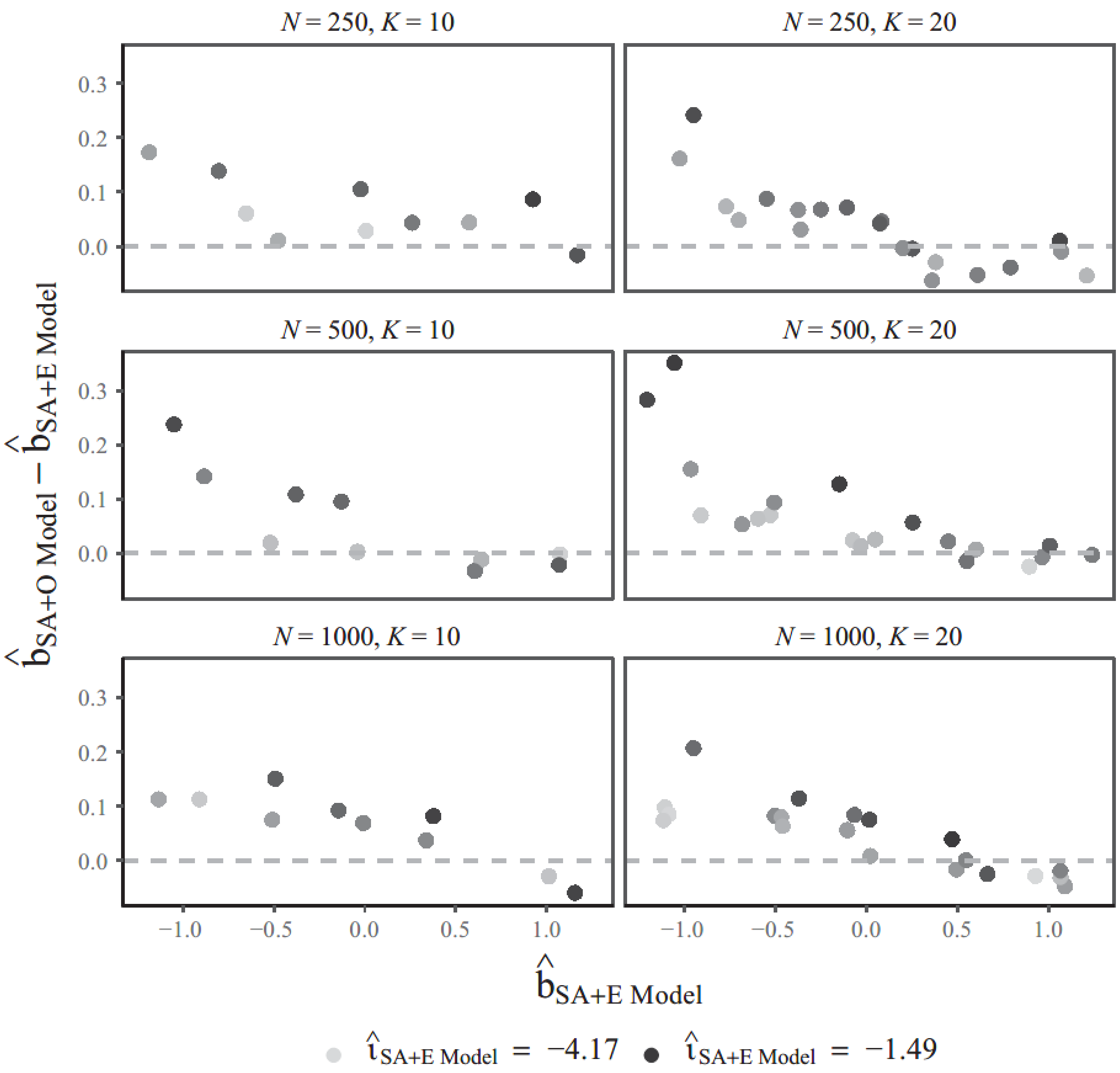


Figure 6. Mixture model for identifying examinee engagement by Wang and Xu (2015).

Results

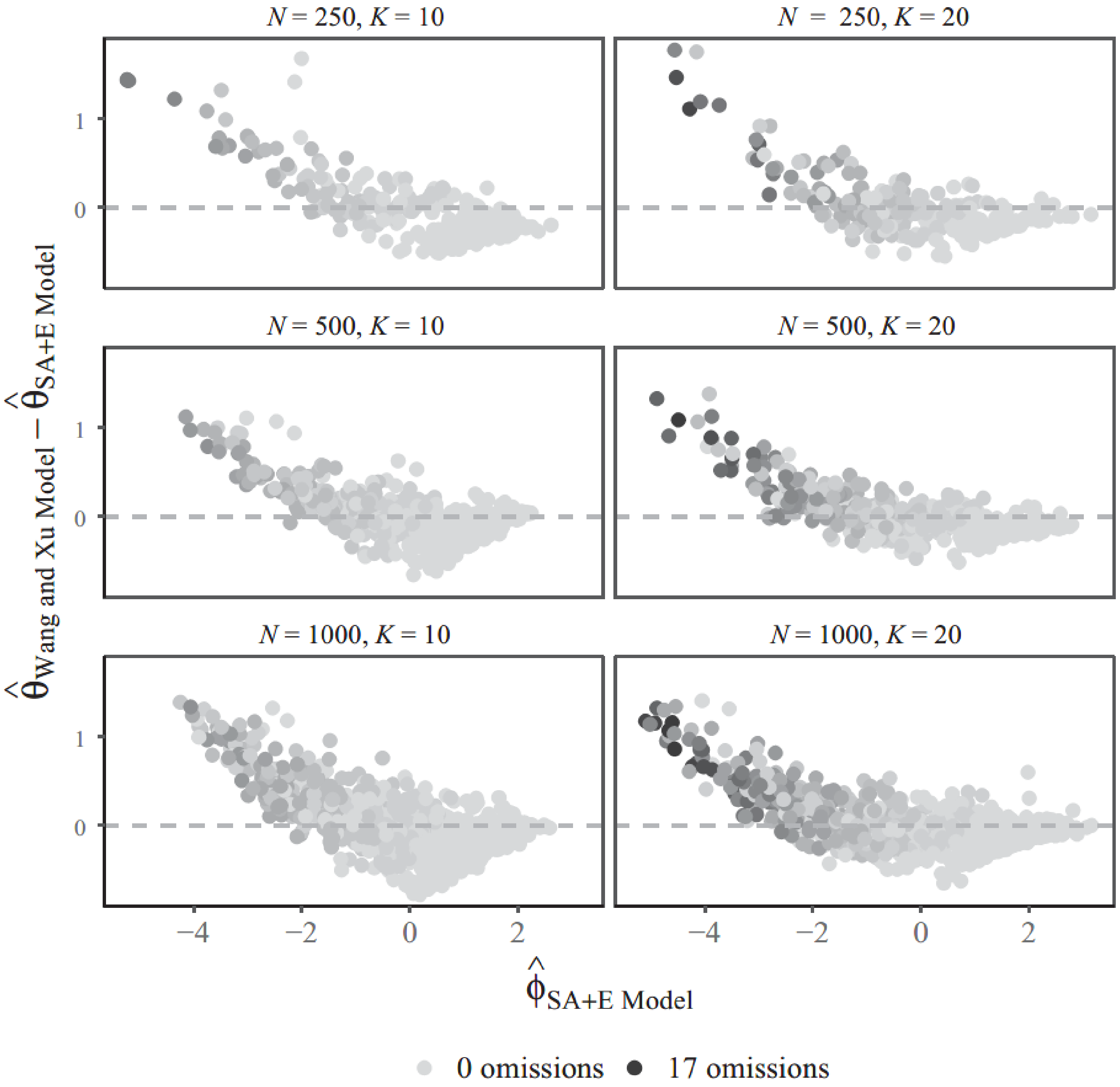

N = 250, K = 10
N = 250, K = 20
N = 500, K = 10
N = 500, K = 20
N = 1000, K = 10
N = 1000, K = 20
1.0
0.5
0.0
−0.5
−1.0
−4
−2
0
2
$\hat{\theta}_{SA+O\ Model} - \hat{\theta}_{SA+E\ Model}$
$\hat{\phi}_{SA+E\ Model}$
0 omissions
17 omissions

Results

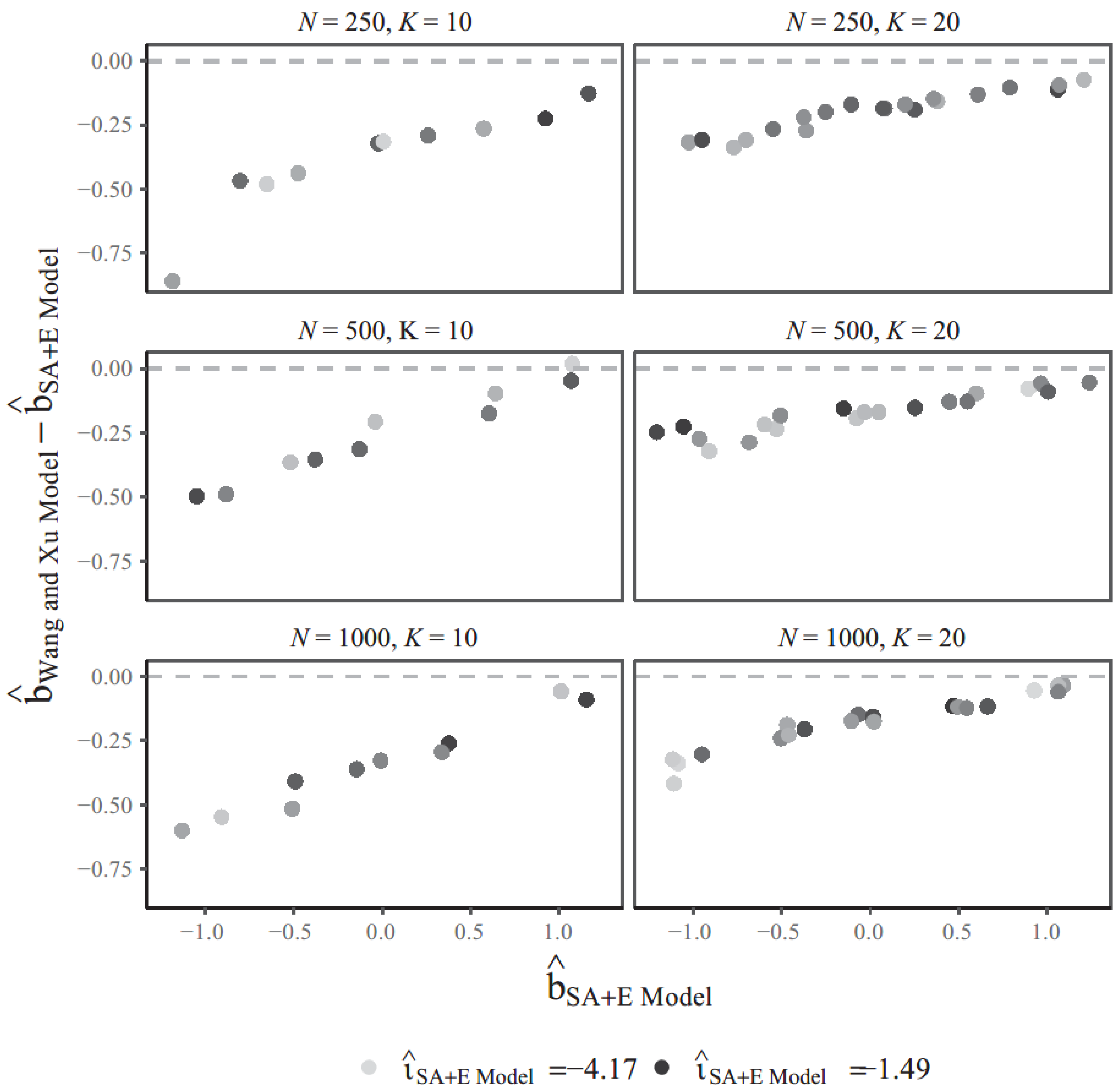


N = 250, K = 10
N = 250, K = 20
N = 500, K = 10
N = 500, K = 20
N = 1000, K = 10
N = 1000, K = 20
0.3
0.2
0.1
0.0
−1.0
−0.5
0.0
0.5
1.0
$\hat{b}_{SA+O\ Model} - \hat{b}_{SA+E\ Model}$
$\hat{b}_{SA+E\ Model}$
$\hat{\iota}_{SA+E\ Model} = -4.17$
$\hat{\iota}_{SA+E\ Model} = -1.49$

Results


N = 250, K = 10
N = 250, K = 20
N = 500, K = 10
N = 500, K = 20
N = 1000, K = 10
N = 1000, K = 20
1
0
−4
−2
0
2
θ̂Wang and Xu Model − θ̂SA+E Model
ϕ̂SA+E Model
0 omissions
17 omissions

Results



N = 250, K = 10
N = 250, K = 20
N = 500, K = 10
N = 500, K = 20
N = 1000, K = 10
N = 1000, K = 20
0.00
−0.25
−0.50
−0.75
−1.0
−0.5
0.0
0.5
1.0
$\hat{b}_{\text{Wang and Xu Model}} - \hat{b}_{\text{SA+E Model}}$
$\hat{b}_{\text{SA+E Model}}$
$\hat{\iota}_{\text{SA+E Model}} = -4.17$
$\hat{\iota}_{\text{SA+E Model}} = -1.49$

Empirical example

- To illustrate the use of the SA+E model
  - data from PISA 2015: Austrian subset (N = 844 examinees)
  - 12 items: OR + MC = 3 + 9
  - omission rate of 10.40%
  - item-level omission rates:
    from 0.04% for the MC item administered at position 1
    to 34.60% for the OR item administered at position 5

- Estimation and model checking
  - different item types
    item-type-specific probabilities correct when **guessing:** c_O and c_M
    item-type-specific regression **intercepts:** γ_{O0} and γ_{M0}
  - ignore not-reached items

Results

Table 2. Person parameter variances and correlations

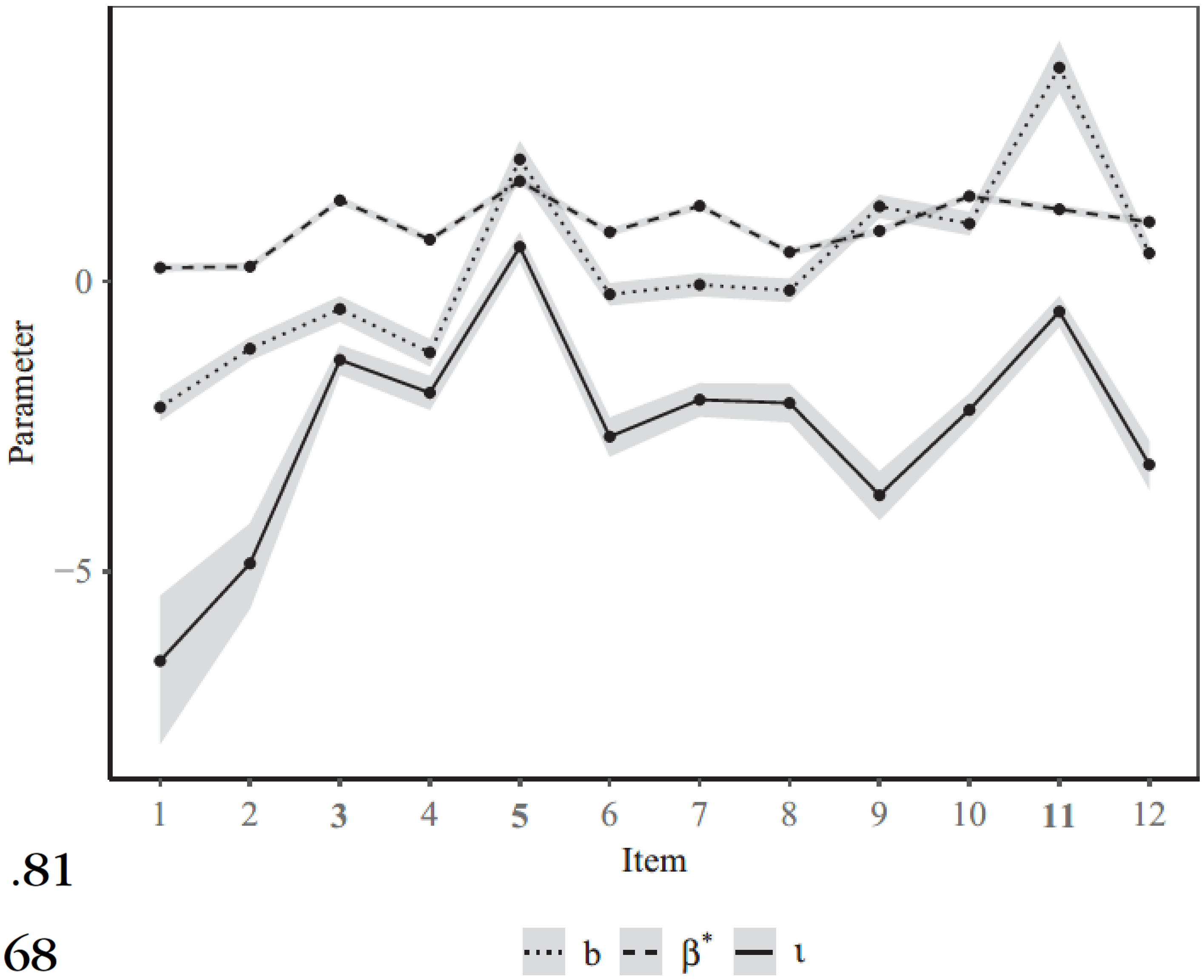
	ϕ	θ	τ
ϕ	3.25 [2.65, 3.93]		
θ	.59 [.50, .68]	1.47 [1.23, 1.74]	
τ	−.35 [−.44, −.25]	−.36 [−.45, −.26]	0.04 [0.03, 0.05]

Note. Highest density intervals are given in square brackets.
ϕ = engagement; θ = ability; τ = speed.

$\gamma_{M0} = -0.71$ [−0.94, −0.47], $\gamma_{O0} = 0.45$ [0.26, 0.67]

$\gamma_{\theta} = -0.74$ [−0.98, −0.52]; $\gamma_{\tau} = -4.79$ [−6.04, −3.70]

Results



$$\mathrm{cor}(\iota, \beta^*) = .81$$

$$\mathrm{cor}(\iota, b) = .68$$

Discussion

- Compared to RT-based scoring methods:
  - less strict assumptions concerning RT distributions
- Compared to previous model-based approaches:
  - allow disengaged behaviour to vary across both items and examinees
  - jointly model engagement and ability
- Applying the model to smaller data sets ($N < 500$ or $K < 20$) only when omission rates are high (at least 5%)

Limitations and future directions

1. allow for different omission mechanisms
2. non-stationarity of person engagement (by adding additional linear or nonlinear terms)
3. the probability of omitting: determined by other examinee- or item-specific factors such as demographic variables or item features
4. integrate research on modelling quitting behaviour
5. use demographic variables or personality to provide additional insight into possible reasons for examinee disengagement
6. implement more complex model instead of a Rasch-like model for response and RTs
7. consider the feasibility of maximum likelihood estimation

The End. Thanks for Listening!

beijing normal university

谢谢大家

多谢晒~

ありがとう

Danke

Merci

Reporter: Yingshi Huang