Modeling item-order effects:

in search of a general framework

Rianne Janssen

K.U. Leuven

Presentation-order effect:
first conclusions, then the argumentation

Item-order effects

- Diverse set of phenomena
- ‘Measurement of change’ and ‘change of the measure’
- Need for an overall framework

Diverse set of phenomena

- Wide range of research fields
- Different type of results

Educational measurement

- Guertin (1954): arithmetic subtest of Wechsler-Bellevue scale

- Hambleton & Traub (1974):
  - Easy to difficult
  - Difficult to easy
    - Lower performance
    - More stress (higher heart rate)

- Practice and fatigue effects (Bejar,1985; Kingston & Dorans, 1982)

Clinical psychology

- Disinhibition scale of Sensation Seeking Scales Form V (Weinberger et al., 2006):
  - order of item groups (alcohol, sex, ambiguous)
- Beck Depression Inventory (Dahlstrom, Brooks, & Peterson, 1990):
  - Original order: from least to most pathological
  - Reversed order
  - Random order
- Intimate partner violence (Ramirez & Strauss, 2006):
  - Order according to ‘culturally recognized behavior patterns’
  - Random order

Neuropsychology

- Investigation of practice effects
  - Test-retest
  - Within tests (e.g., Collie et al., 2003)

Personality assessment

- Increased consistency (and reliability) of responses (Knowles,1988; Knowles & Byers, 1996; Hamilton & Shuminsky,1990):
  - increased activation of self-schema during test administration
  - Meaning clarification
- Cf. social survey research: Socratic effect (Jagodzinski et al., 1987): within and across test administrations

Different (type of) results

- Difficulty:
  - Increased:
    - fatigue effects
  - Decreased:
    - Practice effects, higher disclosure rate
- Reliability/discrimination:
  - Increased
    - Socratic effect, increased self-activation
  - Decreased
    - E.g., switch towards guessing strategy at end of test

Item-order effects

- Long history
- Different research fields
- Different (type of) results

Item-order effects

- Diverse set of phenomena
- ‘Measurement of change’ and ‘change of the measure’
- Need for an overall framework

Modeling

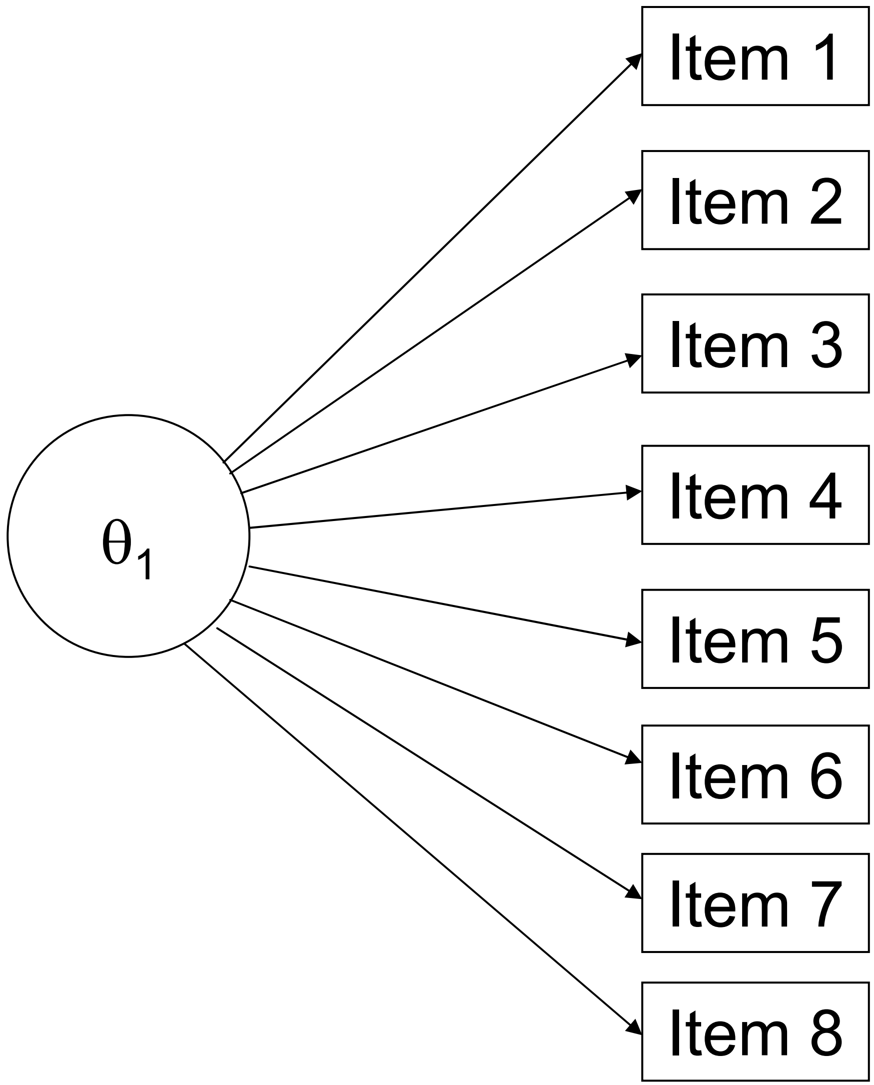
- Different research designs and models available

- Item-order effects need item response modeling

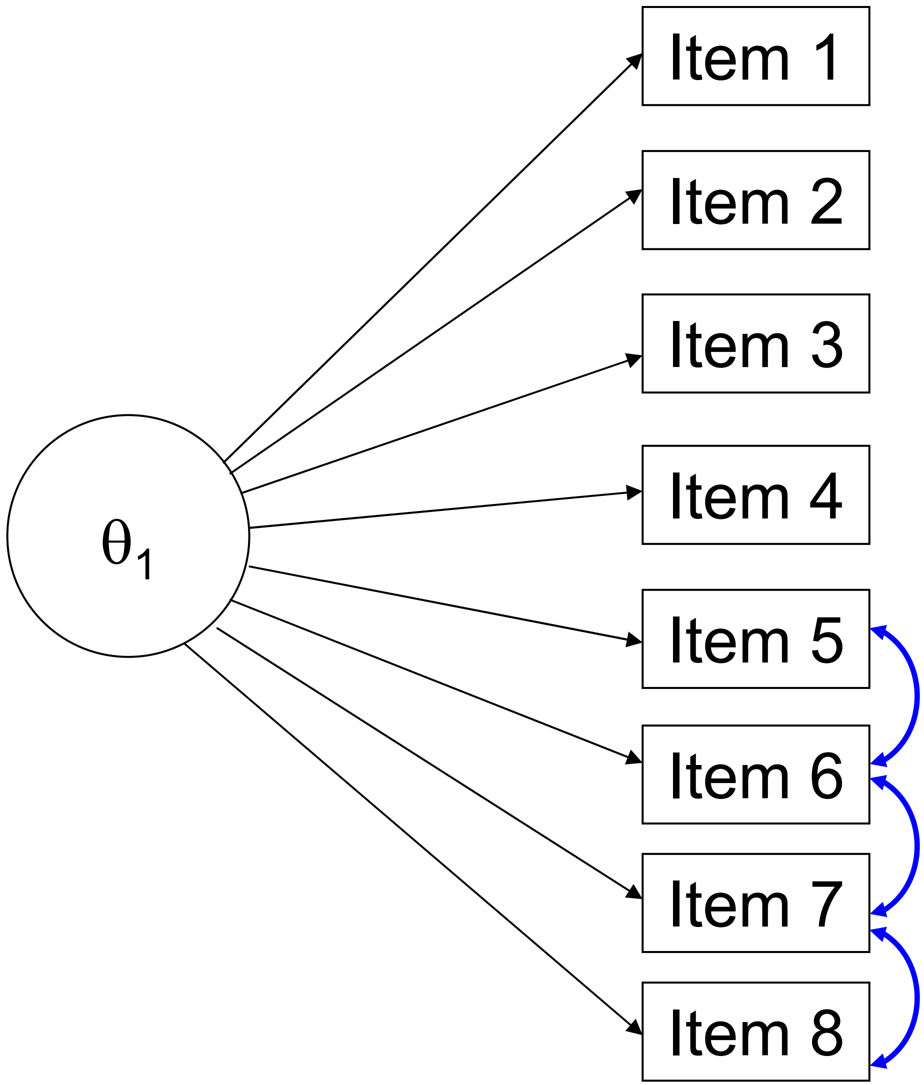
Research design I:

- Single-test administration
- Focus on person side (measurement of change)
- Models from SEM and IRT

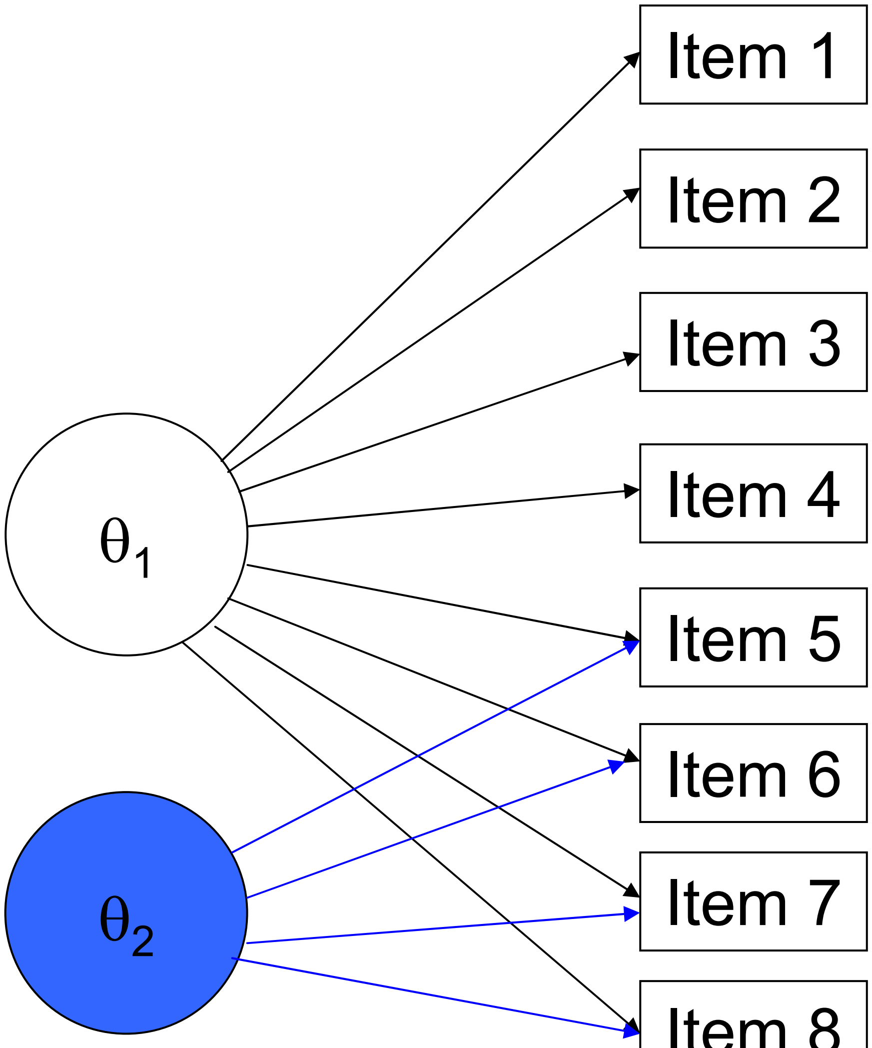
Example of SEM (e.g., Schweizer, 2009)



Example of SEM (ctd.)



Example of SEM (ctd.)



Example from IRT

- Dynamic Rasch model (Verguts & De Boeck, 2000):

$$\log it\left[P\left(X_{pi}=1\right)\right]=\theta_p+t_{pi}\gamma-\beta_i$$

  - t_{pi} is the number of correct answers for person *p* up to item *i*−1.
  - γ is the learning parameter

Dynamic Rasch model (ctd.)

- More general form (Verhelst & Glas, 1993):

$$\log it\left(P\left(X_{pi}=1\right)\right)=\theta_p+\boxed{(k-1)\gamma}-\beta_i$$

no individual differences in “learning” or overall effect of item order (test length)

Dynamic Rasch model (ctd.)

- Example of a multidimensional version (cf. SEM-example):

$$\log it\left[P\left(X_{pi} = 1\right)\right] = \theta_p + (k-1)\gamma_p - \beta_i$$

Item-order effects in single-test administration

- Focus on person side
- Measurement of change (and most often individual differences) during test or towards end of test
- Effect of test length
- Not ‘change of the measure’

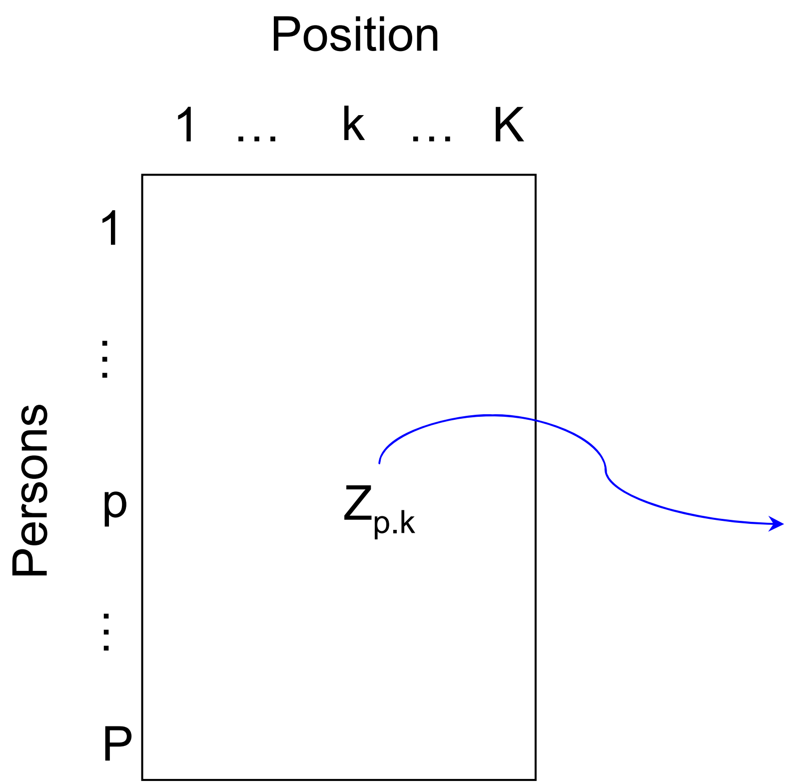
Research design II

- Between-group test design
  - Two groups
    - group 1: 1 → 2 → 3 → 4 → 5
    - group 2: 5 → 4 → 3 → 2 → 1
  - Randomized Latin-square design
    - Group 1: 3 → 5 → 2 → 1 → 4
    - Group 2: 5 → 2 → 1 → 4 → 3
    - Group 3: 2 → 1 → 4 → 3 → 5
    - Group 4: 1 → 4 → 3 → 5 → 2
    - Group 5: 4 → 3 → 5 → 2 → 1
- Focus on item side (and ‘change of the measure’)

Modeling in CTT

- r between item position and r_{it}

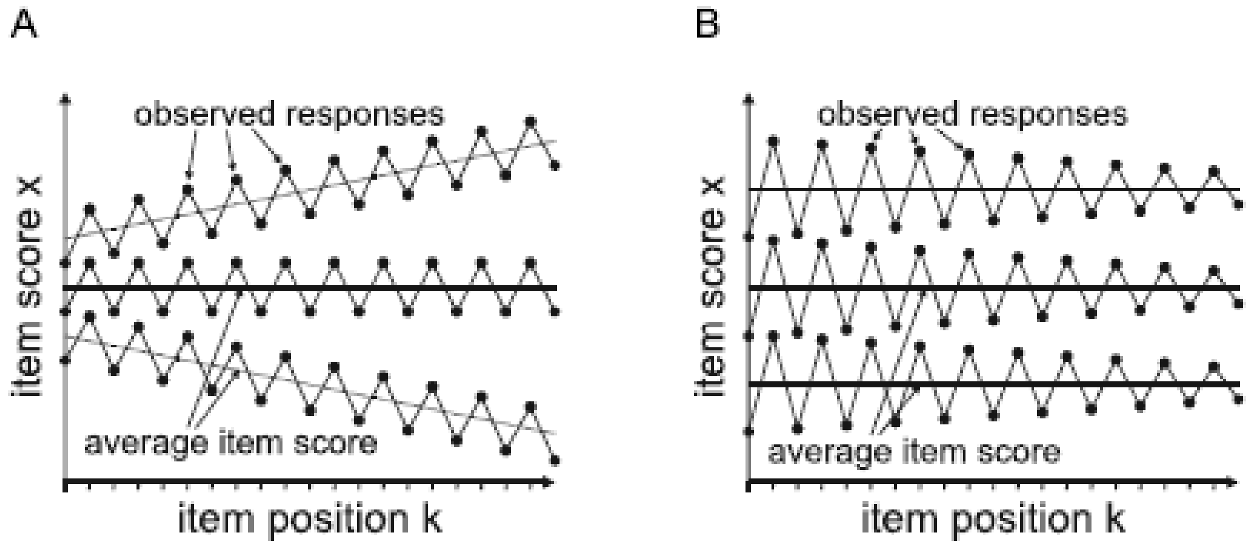
Randomized Latin-square design



$$z_{p.k} = \frac{x_{pik} - \bar{x}_{.i.}}{s_x}$$

Confirmatory analysis of item reliability trends (CAIRT)

Hartig, Hölzel, & Moosbrugger, 2007



Between-subjects

Increased true score variance

Within-subject

decreased error score variance

Confirmatory analysis of item reliability trends (CAIRT)

Hartig, Hölzel, & Moosbrugger, 2007

$$z_{p.k} = \lambda_k \theta_p + \varepsilon_{pk}$$

with $\sigma^2_{\varepsilon_k} = \sigma^2_{\varepsilon_1} + (k-1)\gamma_\varepsilon$ decreased error score variance

$$\lambda_k = \lambda_1 + (k-1)\gamma_\lambda$$

hence $z_{p.k} = \lambda_1 \theta_p + (k-1)\gamma_\lambda \theta_p + \varepsilon_{pk}$

Increased true score variance

Research design II:
IRT models

Steinberg (1994)

- questioning the generality of ‘measuring changes the measure’
- the reliability shift in personality measures may be explained by more specific item-context effects
  - item content (ambiguity)
  - serial position
- use IRT instead of CTT

Item-order effects as DIF

- Cf. Lord’s definition: Depending on the group (condition of test administration), you have a different ICC
- Examples:
  - Steinberg (1994): Samejima’s model for graded item responses
  - Janssen and Kebede (2009): Rasch and 2PL-model

Modeling item-order effects in the Rasch model

		dummy coding	effect coding
item *i*	order 1	β_i	$\beta_i + \delta_i$
	order 2	$\beta_i + \delta_i$	$\beta_i - \delta_i$

Modeling item-order effects in 2PL

2PL – DIF on difficulty

$$\operatorname{logit}(\pi_{pik}) = \begin{cases} \alpha_i\left(\theta_p - (\beta_i + \delta_i)\right) & order\,1 \\ \alpha_i\left(\theta_p - (\beta_i - \delta_i)\right) & order\,2 \end{cases}$$

2PL – DIF on discrimination

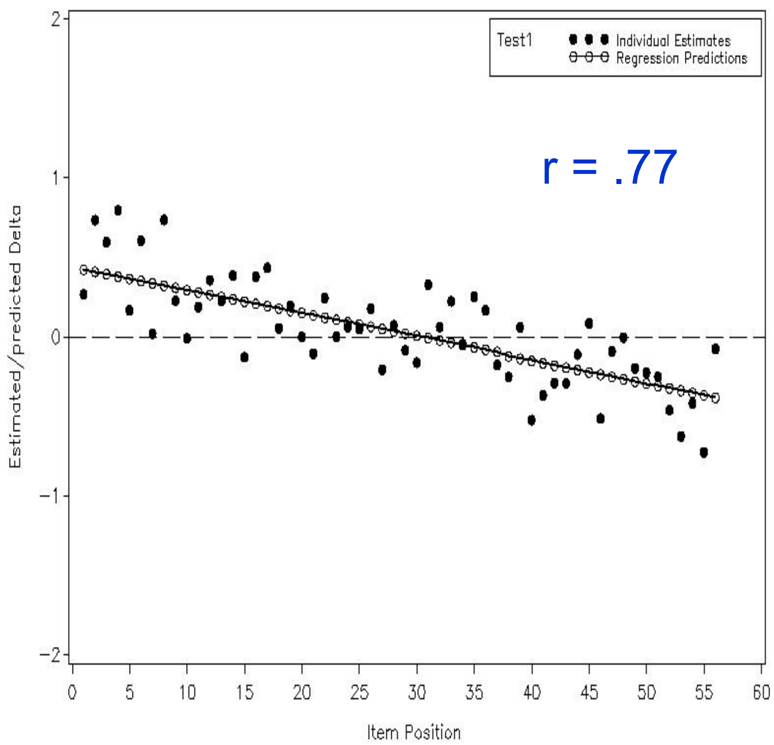
$$\operatorname{logit}(\pi_{pik}) = \begin{cases} (\alpha_i + \xi_i)\left(\theta_p - \beta_i\right) & order\,1 \\ (\alpha_i - \xi_i)\left(\theta_p - \beta_i\right) & order\,2 \end{cases}$$

Example of application

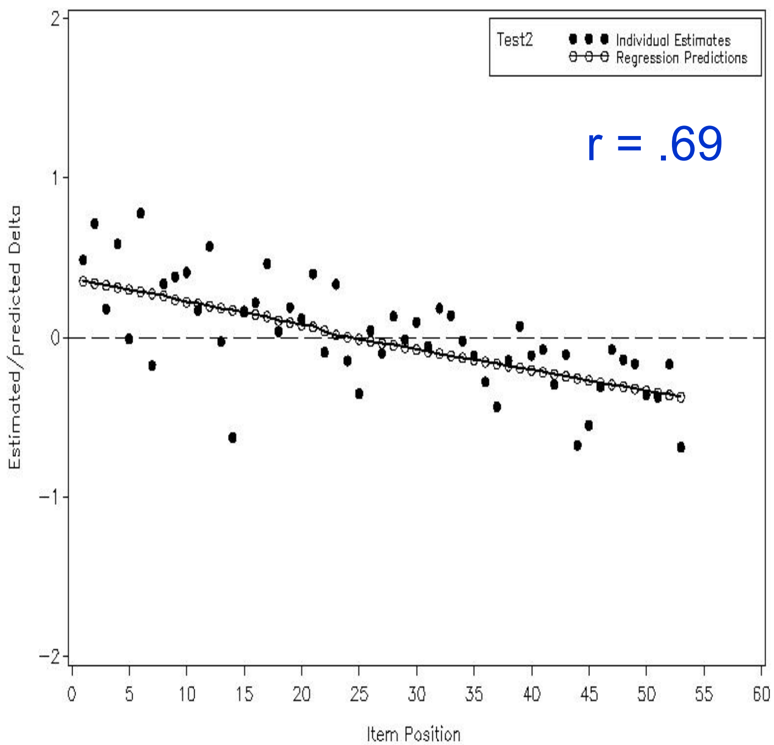
(Janssen and Kebede, 2009)

Estimates of item-order effect ($-\delta_i$) in 2PL with DIF on difficulty

Test 1



Test 2



Summarizing item-order effects

(Janssen and Kebede, 2009)

DIF summarized by position of item

DIF on difficulty $\delta_i = \gamma_\delta k$

DIF on discrimination $\xi_i = \gamma_\xi k$

$\Rightarrow$ *Using the LLTM to estimate item-order effects*

Modeling item-order effects by the LLTM (Kubinger, 2008)

elementary operation j	1	2	3	4	5	6	7	8
virtual item i	item root A	item root B	item root C	item root D	position 1 within test	position 2 within test	position 3 within test	position 4 within test
1	1				1			
2		1				1		
3			1				1	
4				1				1
5	1							1
6		1					1	
7			1			1		
8				1	1			
9	1					1		
10		1						1
11			1		1			
12				1			1	
13	1						1	
14		1			1			
15			1					1
16				1		1		

Figure 4:

The LLTM's matrix of weights ((q_{ij})) for the case: 4 item roots and four different sequences of presentation – each position within item presentation is hypothesized as having its own special effect.

Modeling item-order effects by the LLTM (Kubinger, 2008)

$$\operatorname{logit}\left[P\left(X_{pik}=1\right)\right]=\theta_p-\beta_i+\gamma_k$$

Item-order effects in between-group designs

- Focus on item side
- ‘change of the measure’: main effects of item order
- No individual differences

Item-order effects

- Diverse set of phenomena
- ‘Measurement of change’ and ‘change of the meaure’
- Need for an overall framework

A general framework?

Focus on	Single-test administration	Between-group design[a]
Person side	Individual differences	
Item side		Item main effects

Several type of effects :

Item-order effects need item response modeling

[a] or a repeated-measures design?

Assumption: no change in response process

- Speeded IRT model with gradual process change (Goegebeur, De Boeck, Wollack, & Cohen, 2008)

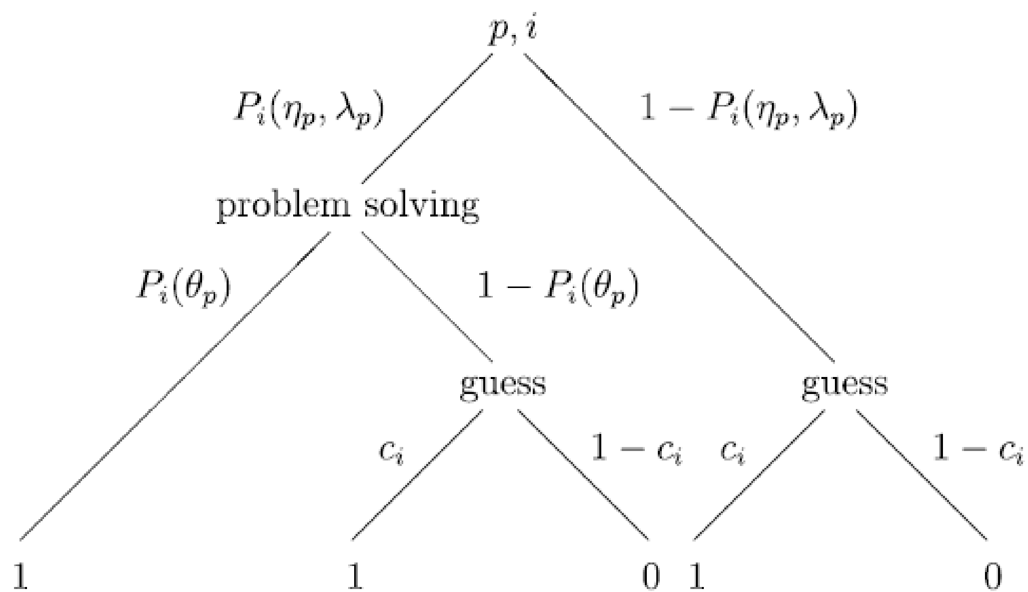


FIGURE 2.
Decision tree representation of speededness model.

IOE-modeling

(between-groups designs)

$$\operatorname{logit}(\pi_{pik}) = \alpha_{ik}(\theta_p - \beta_{ik})$$

Compare to 'null' model: $\operatorname{logit}(\pi_{pik}) = \alpha_i(\theta_p - \beta_i)$

IOE-modeling

(between-groups designs)

$$\log it(\pi_{pik}) = \alpha_{ik}(\theta_p - \beta_{ik})$$

Linear effect of item order

$$\log it(\pi_{pik}) = \underbrace{\left(\alpha_i + \gamma_\alpha (k-1)\right)}_{\text{Cf. CAIRT}}\left(\theta_p - (\underbrace{\beta_i + \gamma_\beta (k-1)}_{\text{Cf. dynamic Rasch model}})\right)$$

Multidimensional model (e.g., proneness to fatigue)

$$\log it(\pi_{pik}) = \left(\alpha_i + \gamma^\alpha (k-1)\right)\left(\theta_p - (\beta_i + \underbrace{\gamma_p^\beta}_{\text{Cf. RW-LLTM}} (k-1))\right)$$

Research questions

- Apply IOE-model
- Check computational feasibility
- Compare IOE-model with CAIRT
- Extend to other IRT models

Other (research) questions?