Don't say CAT: New Item Response Theory approaches for developing short test forms

Ottavia M. Epifania[1,2], Pasquale Anselmi[1], Egidio Robusto[1]
ottavia.epifania@unipd.it
[1]Univeristy of Padova
[2]Catholic University of the Sacred Heart

September 30th 2022, Padova

XXX Annual Conference of the Italian Psychology Association (AIP)

CAT

CAT

Computerized **A**daptive **T**esting

Item Response Theory and short test forms

ADAPTIVE SHORT FORMS: *Ad-hoc* tests for each person $\rightarrow$ The information is maximized for each level of θ (i.e., for each respondent) $\rightarrow$ (**CAT**: Computerized Adaptive Testing)

STATIC SHORT FORMS: Static tests equal for all respondents $\rightarrow$ The information is maximized across θ levels (i.e., across all respondents)

Item Response Theory and short test forms

ADAPTIVE SHORT FORMS: *Ad-hoc* tests for each person → The information is maximized for each level of θ (i.e., for each respondent) → (**CAT**: Computerized Adaptive Testing)

Issue

Different short test forms for each respondent → Potential fairness issues in assessments, e.g. for recruitment

STATIC SHORT FORMS: Static tests equal for all respondents → The information is maximized across θ levels (i.e., across all respondents)

Item Response Theory and short test forms

ADAPTIVE SHORT FORMS: *Ad-hoc* tests for each person $\rightarrow$ The information is maximized for each level of θ (i.e., for each respondent) $\rightarrow$ (**CAT**: Computerized Adaptive Testing)

Issue

Different short test forms for each respondent $\rightarrow$ Potential fairness issues in assessments, e.g. for recruitment

STATIC SHORT FORMS: Static tests equal for all respondents $\rightarrow$ The information is maximized across θ levels (i.e., across all respondents)

Issue

Not being tailored to any θ level of interest $\rightarrow$ Potentially more items are needed to cover a wide range of θs

Aim

New IRT-based procedures for shortening tests

Aim

New IRT-based procedures for shortening tests

Equal for all respondents

Aim

New IRT-based procedures for shortening tests

Equal for all respondents

Tailored to specific levels of the latent trait

Item Response Theory

2-PL Model

$$P(x_{pj} = 1|\theta_p, b_j, a_j) = \frac{exp[a_j(\theta_p - b_j)]}{1 + exp[a_j(\theta_p - b_j)]}$$

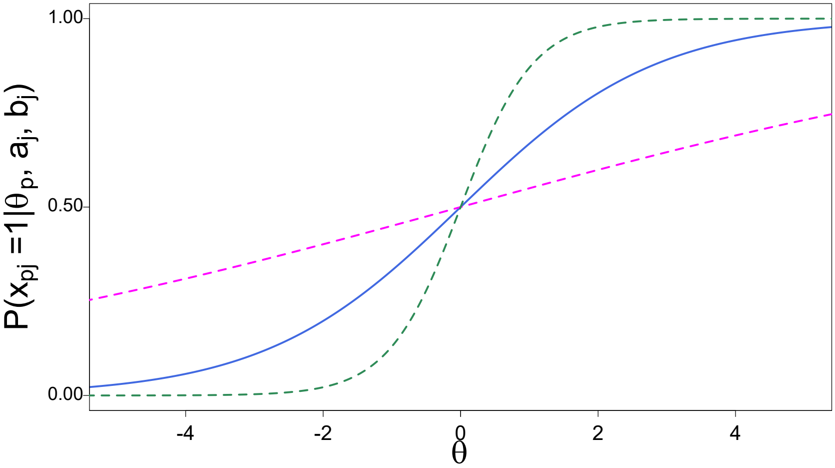
where:

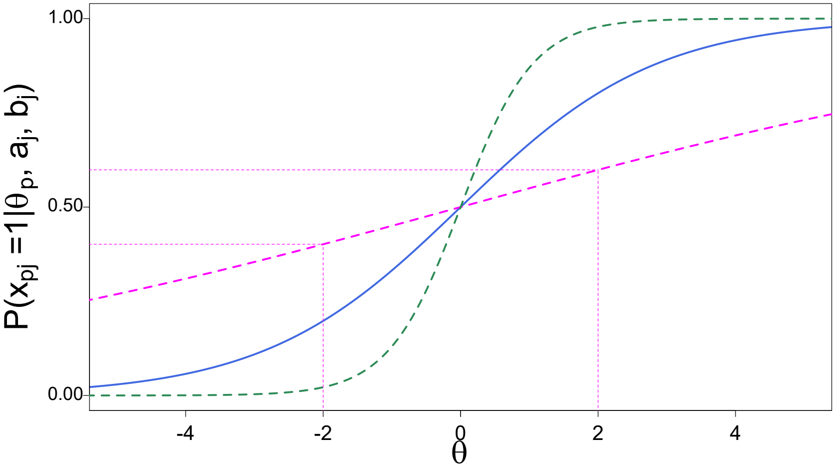
$P(x_{pj} = 1)$: Probability of a correct response to item j by respondent p

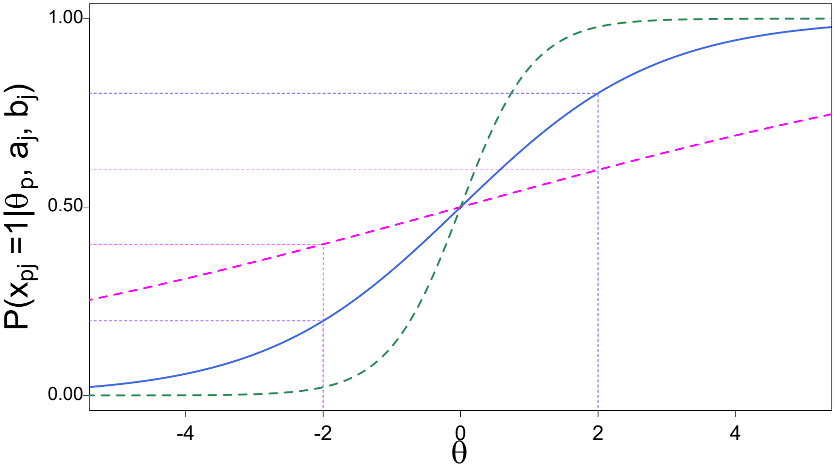
θ_p: Ability of respondent p

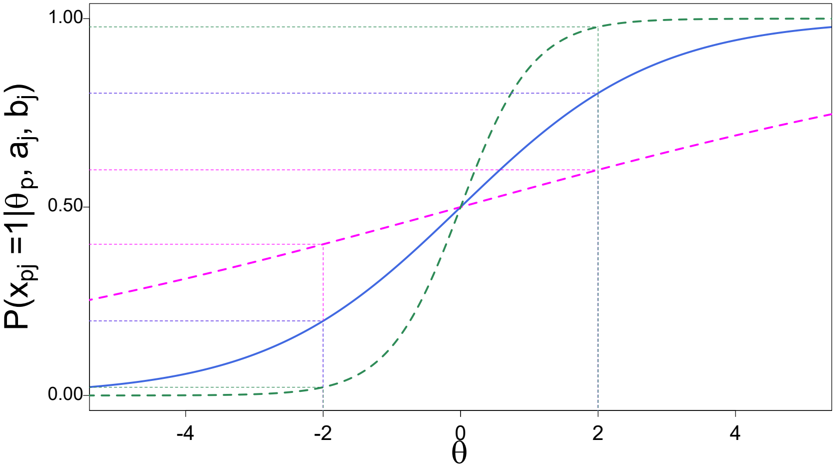
b_j: Difficulty of item j

a_j: Discrimination of item j


1.00
0.50
0.00
-4
-2
0
2
4
θ
P(x_pj =1|θ_p, a_j, b_j)


1.00
0.50
0.00
-4
-2
0
2
4
θ
P(x_pj = 1|θ_p, a_j, b_j)


1.00
0.50
0.00
-4
-2
0
2
4
θ
$P(x_{pj} = 1|\theta_p, a_j, b_j)$


1.00
0.50
0.00
P(x_pj =1|θ_p, a_j, b_j)
-4
-2
0
2
4
θ

Information functions

Item Information Function

$$IIF_j = a_j^2[P(\theta)(1 - P(\theta))]$$

Information functions

Item Information Function

$$IIF_j = a_j^2[P(\theta)(1 - P(\theta))]$$

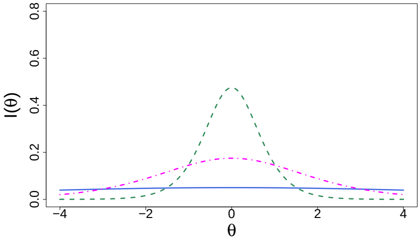


Figure 1: $a = 0.20$, $a = 0.70$, $a = 1.90$, $b = 0$

Information functions

Item Information Function

$$IIF_j = a_j^2[P(\theta)(1 - P(\theta))]$$

Test Information Function

$$TIF = \sum_{j=1}^{J} IIF_j$$

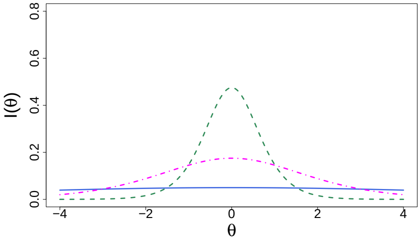


Figure 1: $a = 0.20$, $a = 0.70$, $a = 1.90$, $b = 0$

Information functions

Item Information Function

$$IIF_j = a_j^2[P(\theta)(1 - P(\theta))]$$

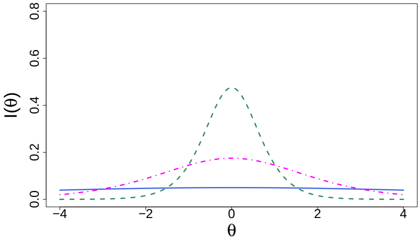


Figure 1: $a = 0.20$, $a = 0.70$, $a = 1.90$, $b = 0$

Test Information Function

$$TIF = \sum_{j=1}^{J} IIF_j$$

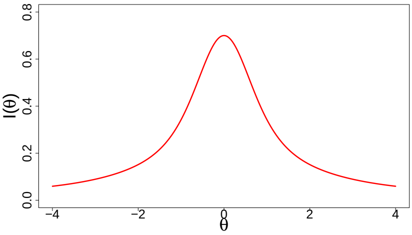


Figure 2: TIF $= IIF_1 + IIF_2 + IIF_3$

Benchmark procedure

Selected items → items with the highest *IIF*s

e.g.: 3-item short form from 10-item full-length test

item	b	a	IIF
1	−0.67	0.71	0.08
2	0.50	1.19	0.15
3	−2.43	0.25	0.01
4	2.12	1.98	0.24
5	1.72	0.39	0.03
6	−2.28	1.62	0.19
7	0.64	0.50	0.05
8	−2.51	1.68	0.19
9	−0.66	0.44	0.04
10	0.72	0.33	0.02

Benchmark procedure

Selected items → items with the highest *IIF*s

e.g.: 3-item short form from 10-item full-length test

item	b	a	*IIF*
4	2.12	1.98	0.24
8	−2.51	1.68	0.19
6	−2.28	1.62	0.19
2	0.50	1.19	0.15
1	−0.67	0.71	0.08
7	0.64	0.50	0.05
9	−0.66	0.44	0.04
5	1.72	0.39	0.03
10	0.72	0.33	0.02
3	−2.43	0.25	0.01

θ-target procedures

Selected items $\rightarrow$ items with highest *IIF*s in respect to θ targets (θ')

e.g.: 3-item short form from 10-item full-length test

	θ'_1	θ'_2	θ'_3
item	-2.67	0.01	2.67
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			

θ-target procedures

Selected items → items with highest *IIF*s in respect to θ targets (θ')

e.g.: 3-item short form from 10-item full-length test

item	θ'_1 -2.67	θ'_2 0.01	θ'_3 2.67
1	0.04	0.12	0.08
2	0.09	0.33	0.03
3	0.01	0.01	0.02
4	0.73	0.06	0.01
5	0.04	0.03	0.02
6	0.01	0.06	0.59
7	0.05	0.06	0.03
8	0.01	0.04	0.69
9	0.03	0.05	0.04
10	0.02	0.03	0.02

θ-target procedures

Selected items $\rightarrow$ items with highest *IIF*s in respect to θ targets (θ')

e.g.: 3-item short form from 10-item full-length test

item	θ'_1 -2.67	θ'_2 0.01	θ'_3 2.67
1	0.04	0.12	0.08
2	0.09	0.33	0.03
3	0.01	0.01	0.02
4	0.73	0.06	0.01
5	0.04	0.03	0.02
6	0.01	0.06	0.59
7	0.05	0.06	0.03
8	0.01	0.04	0.69
9	0.03	0.05	0.04
10	0.02	0.03	0.02

θ-target procedures

Selected items → items with highest *IIF*s in respect to θ targets (θ')

e.g.: 3-item short form from 10-item full-length test

item	θ'_1 −2.67	θ'_2 0.01	θ'_3 2.67
1	0.04	0.12	0.08
2	0.09	0.33	0.03
3	0.01	0.01	0.02
4	0.73	0.06	0.01
5	0.04	0.03	0.02
6	0.01	0.06	0.59
7	0.05	0.06	0.03
8	0.01	0.04	0.69
9	0.03	0.05	0.04
10	0.02	0.03	0.02

θ-target procedures

Selected items → items with highest *IIF*s in respect to θ targets (θ')

e.g.: 3-item short form from 10-item full-length test

item	θ'_1 -2.67	θ'_2 0.01	θ'_3 2.67
1	0.04	0.12	0.08
2	0.09	0.33	0.03
3	0.01	0.01	0.02
4	0.73	0.06	0.01
5	0.04	0.03	0.02
6	0.01	0.06	0.59
7	0.05	0.06	0.03
8	0.01	0.04	0.69
9	0.03	0.05	0.04
10	0.02	0.03	0.02

θ-target procedures

Selected items $\rightarrow$ items with highest *IIF*s in respect to θ targets (θ')

e.g.: 3-item short form from 10-item full-length test

item	θ'_1 -2.67	θ'_2 0.01	θ'_3 2.67
1	0.04	0.12	0.08
2	0.09	0.33	0.03
3	0.01	0.01	0.02
4	0.73	0.06	0.01
5	0.04	0.03	0.02
6	0.01	0.06	0.59
7	0.05	0.06	0.03
8	0.01	0.04	0.69
9	0.03	0.05	0.04
10	0.02	0.03	0.02

θ-target procedures

Selected items $\rightarrow$ items with highest *IIF*s in respect to θ targets (θ')

e.g.: 3-item short form from 10-item full-length test

item	θ'_1 -2.67	θ'_2 0.01	θ'_3 2.67
1	0.04	0.12	0.08
2	0.09	0.33	0.03
3	0.01	0.01	0.02
4	0.73	0.06	0.01
5	0.04	0.03	0.02
6	0.01	0.06	0.59
7	0.05	0.06	0.03
8	0.01	0.04	0.69
9	0.03	0.05	0.04
10	0.02	0.03	0.02

θ-target procedures

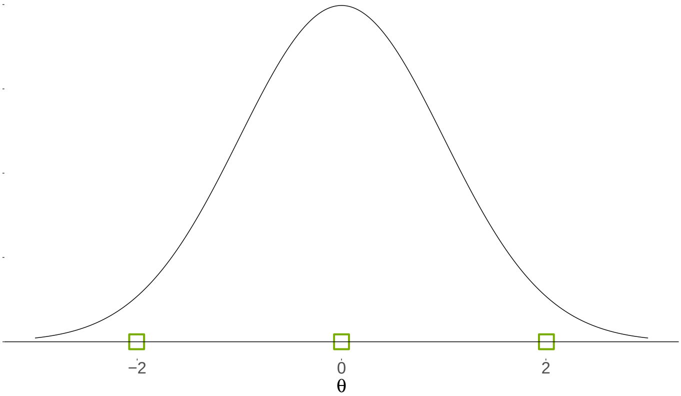
Selected items $\rightarrow$ items with highest *IIF*s in respect to θ targets (θ')

e.g.: 3-item short form from 10-item full-length test

item	θ'_1 -2.67	θ'_2 0.01	θ'_3 2.67
1	0.04	0.12	0.08
2	0.09	0.33	0.03
3	0.01	0.01	0.02
4	0.73	0.06	0.01
5	0.04	0.03	0.02
6	0.01	0.06	0.59
7	0.05	0.06	0.03
8	0.01	0.04	0.69
9	0.03	0.05	0.04
10	0.02	0.03	0.02

Segmenting the latent trait

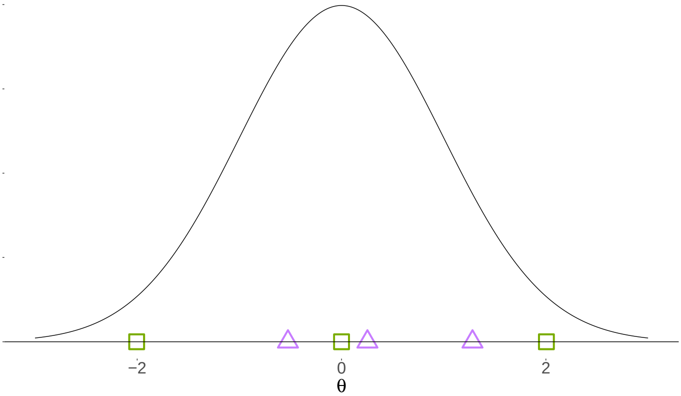
Segmenting the latent trait



Equal Intervals Procedure

Equal segmentation

Segmenting the latent trait



Equal Intervals Procedure
Equal segmentation

Unequal Intervals Procedure
Clustering

Comparison between the item selection procedures:

- **Benchmark procedure (BP)**: The N items with the highest *IIF*s are selected from the full-length test
- **Equal Intervals Procedure (EIP)**: The N items that maximize the information for each θ' obtained by dividing the latent trait into equal intervals are selected
- **Unequal Intervals Procedure (UIP)**: The N items that maximize the information for each θ' obtained by clustering the latent trait are selected
- **Random Procedure (RP)**: N items are randomly selected from the full-length tests

10, 30, 50, 70, 90-item short test forms from a 100-item full-length test

1000 respondents p

1. Normal distribution
   $p \sim \mathcal{N}(0,1)$
2. Positive skewed distribution
   $p \sim Beta(1,100)$ (linearly transformed to obtain negative values)
3. Uniform distribution
   $p \sim \mathcal{U}(-3,3)$

100 items j:

- $b \sim \mathcal{U}(-3,3)$
- $a \sim \mathcal{U}(0.40,2)$

An overall look

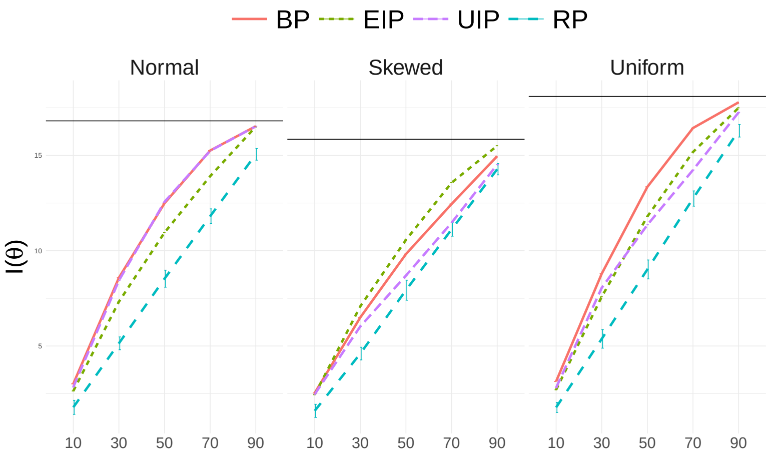


Figure 3: Overall Information of the short test forms

A closer look

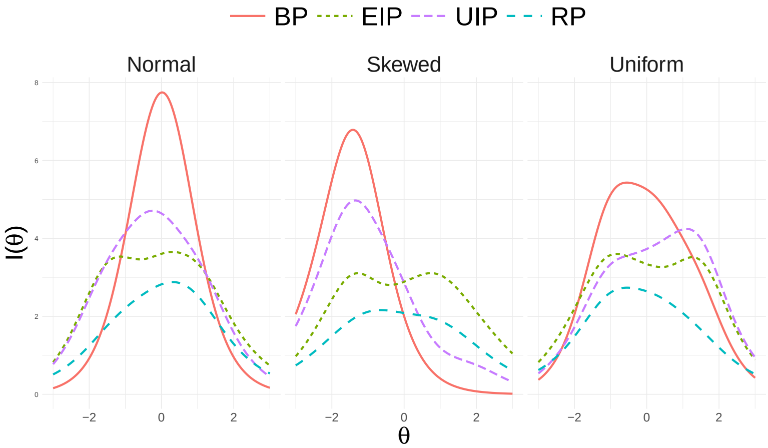


Figure 4: TIF of the 10-item short test form

An even closer look

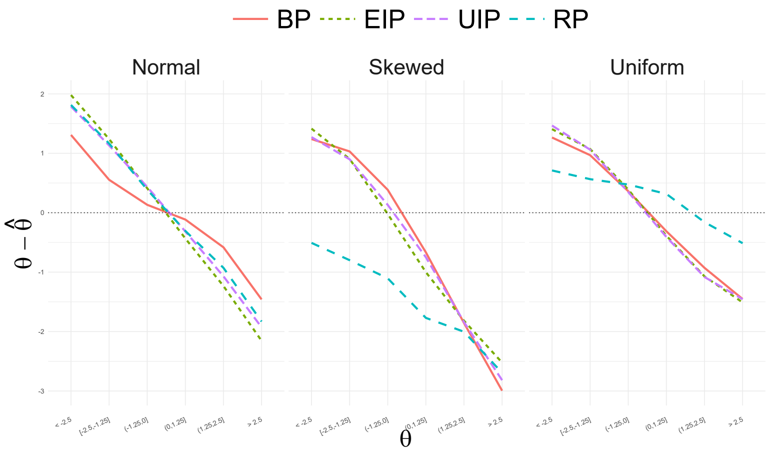


Figure 5: *bias* $= \theta - \hat{\theta}$ of the 10-item short test form

Good!

There's no "one-fits-all" solution

The θ distribution is a key element

Good!

There's no "one-fits-all" solution

The θ distribution is a key element

..but work is still needed

Real life applications are missing

The CAT is missing

Thank you!

ottavia.epifania@unipd.it