

Pauci sed moni: An Item Response Theory approach for shortening tests

Ottavia M. Epifania^{1,2}, Pasquale Anselmi¹ & Egidio Robusto¹

¹ University of Padua, Padova (IT)

² Catholic University of the Sacred Heart, Milano (IT)

ottavia.epifania@unipd.it

SCAN ME



Introduction

Item Response Theory (IRT) is the theoretical framework often used for shortening existing tests. IRT models describe the probability of observing a response as a function of the characteristics of respondent p (i.e., the latent trait level θ) and the characteristics of item s . IRT models provide detailed information on how well each item measures a certain θ level (i.e., *item information function*, *IIF*). Two types of short forms can be created by exploiting the *IIF*s:

- Adaptive short forms:** *Ad-hoc* tests for each person (i.e., Computerized Adaptive Testing, CAT). The items administered to each respondent vary according to the responses that this respondent gave to the previously administered items \rightarrow the information is maximized for each level of θ (i.e., each respondent)
Issue: *Different short test forms for each respondent \rightarrow Unfair assessments in recruitment or admissions tests*
- Static short forms:** Static tests equal for all respondents (i.e., only the items from the full-length test that provide the highest information are included in the short form) \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 θ s*

Aim

New IRT-based procedures for the development of short test forms combining the advantages of adaptive short test forms (i.e., tailoring the tests to different θ levels) and those of static short forms (i.e., being equal for all respondents).

The new item selection procedures are based on the definition of trait levels of interest (i.e., θ' targets, denoted as θ') \rightarrow The items that best assess the trait levels represented by the θ' targets (i.e., optimal items with highest *IIF*s for each θ') are included in the short form.

Item Response Theory and information functions

This illustration is based on the 2-parameter logistic model (2PL) for dichotomous responses:

$$P(x_{ps} = 1 | \theta_p, b_s, a_s) = \frac{\exp[a_s(\theta_p - b_s)]}{1 + \exp[a_s(\theta_p - b_s)]} \quad (1)$$

where $P(x_{ps} = 1 | \theta_p, b_s, a_s)$ is the probability of respondent p to respond correctly to item s given the ability (θ) of p and difficulty (b) and discrimination (a) of s . The *Item Characteristics Curves* (ICCs) of three items with same difficulty but different discriminations are illustrated in Figure 1a. The *item information function* (*IIF*) informs about the precision with which the item measures the abilities θ s. In the 2PL model, the *IIF* is obtained as:

$$IIF = a^2[P(\theta)(1 - P(\theta))], \quad (2)$$

where $P(\theta)$ is the probability of a respondent with a certain θ of responding correctly to an item, and $1 - P(\theta)$ is their probability of responding incorrectly to the same item. The *IIF*s of the items depicted in Figure 1a are illustrated in Figure 1b.

The *test information function* (*TIF*) is obtained by summing the *IIF*s across items (*test information function*, $TIF = \sum_{s=1}^S IIF_s$, Figure 1c).

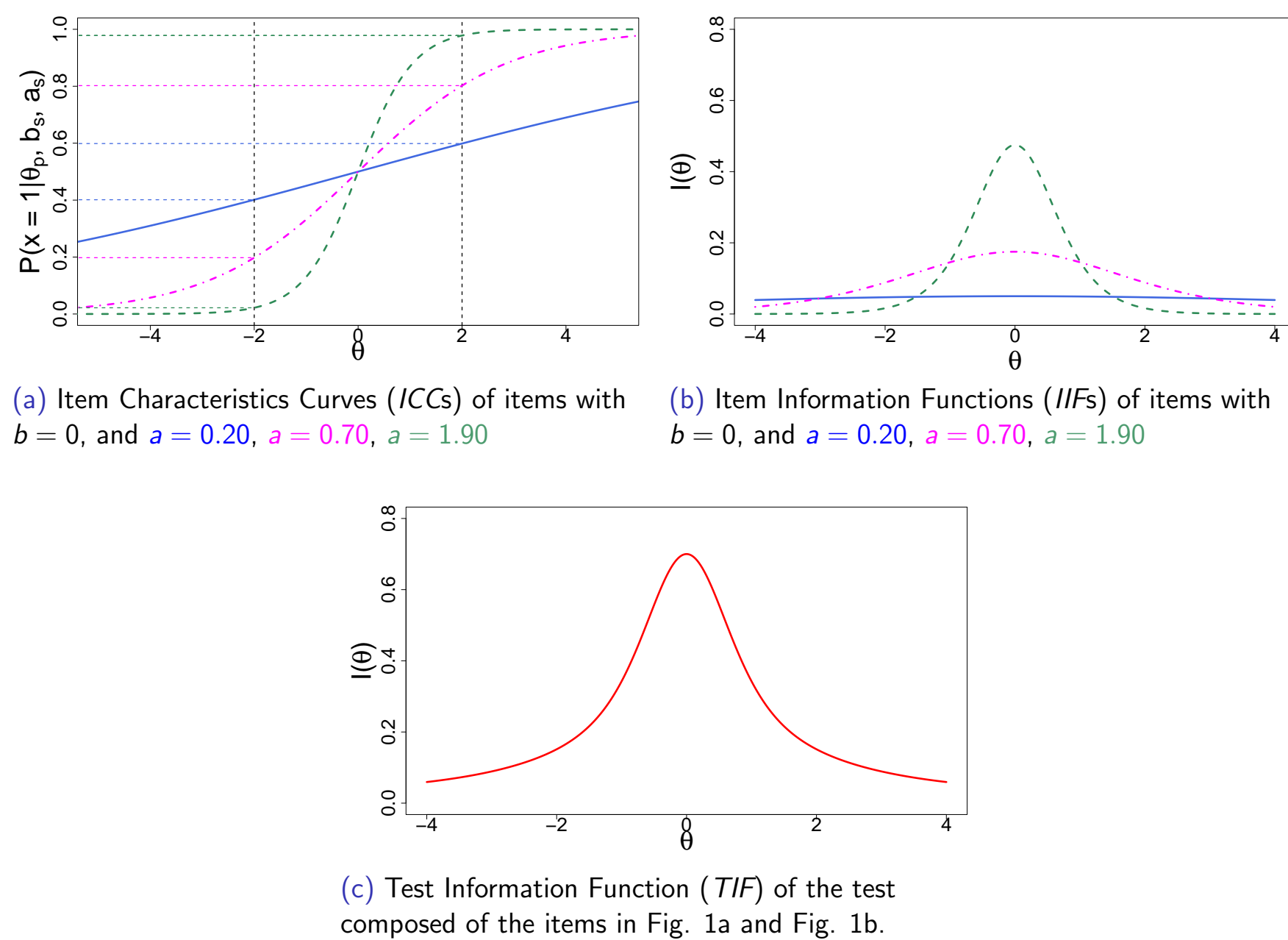


Figure 1: 2-PL and information functions

Item Selection Procedures

- Benchmark:** The N items with the highest *IIF*s are selected from the full-length test to be included in the static short form, where N is the desired length of the short form (Benchmark Procedure, BP).
- Procedures based on θ' :**
 - Cluster:** The latent trait is clustered in N clusters, where n is the number of items to be included in the short form. The centroids of the clusters are the θ' (Unequal Intervals Procedure, UIP).
 - Intervals:** The latent trait is segmented into $N + 1$ intervals. Each interval is defined by $[\theta'_{n-1}; \theta'_n]$. The θ' s are obtained by averaging between the lower and upper bound of each interval to avoid that the first and the last θ' s correspond to the minimum and maximum θ values (Equal Intervals Procedure, EIP).
- Random:** Items are randomly selected from the full-length tests (RP).

Development of a 5-item short form from a 10-item full-length test:

Typical procedure				θ' -based procedures					
item	b	a	IIF	item	θ_1	θ_2	θ_3	θ_4	θ_5
1	-2.51	1.68	0.10	1	0.07	0.12	0.12	0.07	0.03
2	-2.43	0.25	0.02	2	0.02	0.11	0.32	0.25	0.06
3	-2.28	1.62	0.13	3	0.02	0.02	0.01	0.01	0.01
4	-0.67	0.71	0.11	4	0.01	0.01	0.06	0.71	0.45
5	-0.66	0.44	0.05	5	0.02	0.03	0.03	0.04	0.03
6	0.50	1.19	0.27	6	0.45	0.46	0.06	0.01	0.01
7	0.64	0.50	0.06	7	0.03	0.05	0.06	0.06	0.04
8	0.72	0.33	0.03	8	0.57	0.38	0.04	0.01	0.01
9	1.72	0.39	0.03	9	0.04	0.05	0.05	0.04	0.03
10	2.12	1.98	0.16	10	0.02	0.02	0.03	0.03	0.02

Method

Comparison between the item selection procedures:

- Benchmark procedure (BP)
- Unequal Intervals Procedure (UIP)
- Equal Interval Procedure (EIP)
- Random Procedure (RP)

in the development of 10, 30, 50, 70, 90 items test short forms from a 100-item full-length test (For each short test form, RP randomly selects the items 10 times).

- 1000 respondents p 100 items s :
- Three θ distributions:
- 1. Normal distribution $p \sim \mathcal{N}(0, 1)$
 - 2. Positive skewed distribution $p \sim \text{Beta}(1, 100)$ (linearly transformed to obtain negative values)
 - 3. Uniform distribution $p \sim \mathcal{U}(-3, 3)$
- $b \sim \mathcal{U}(-3, 3)$
 $a \sim \mathcal{U}(0.40, 2)$

Results

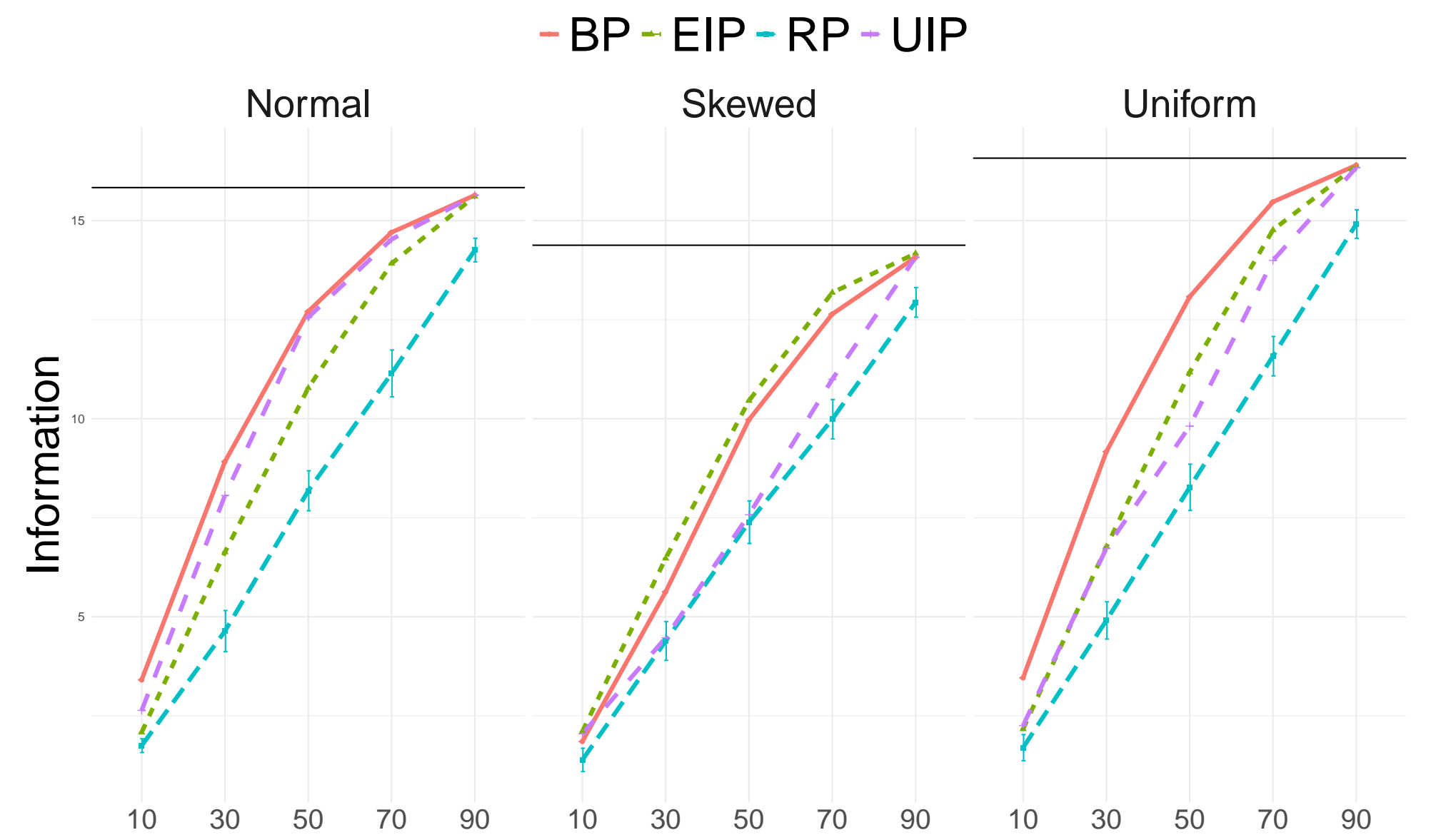


Figure 2: Overall information of the short test forms

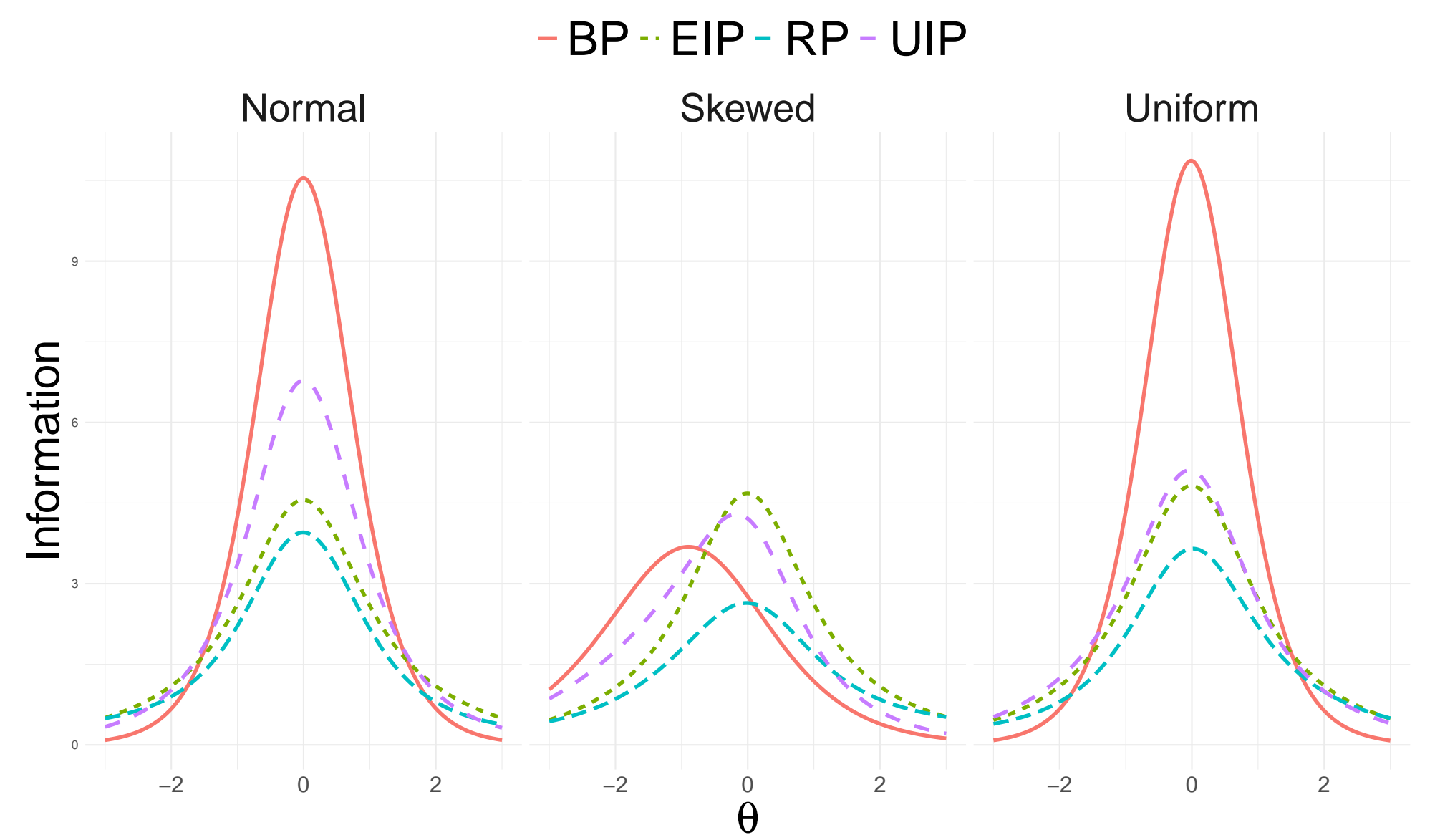


Figure 3: Detailed information of the short test forms

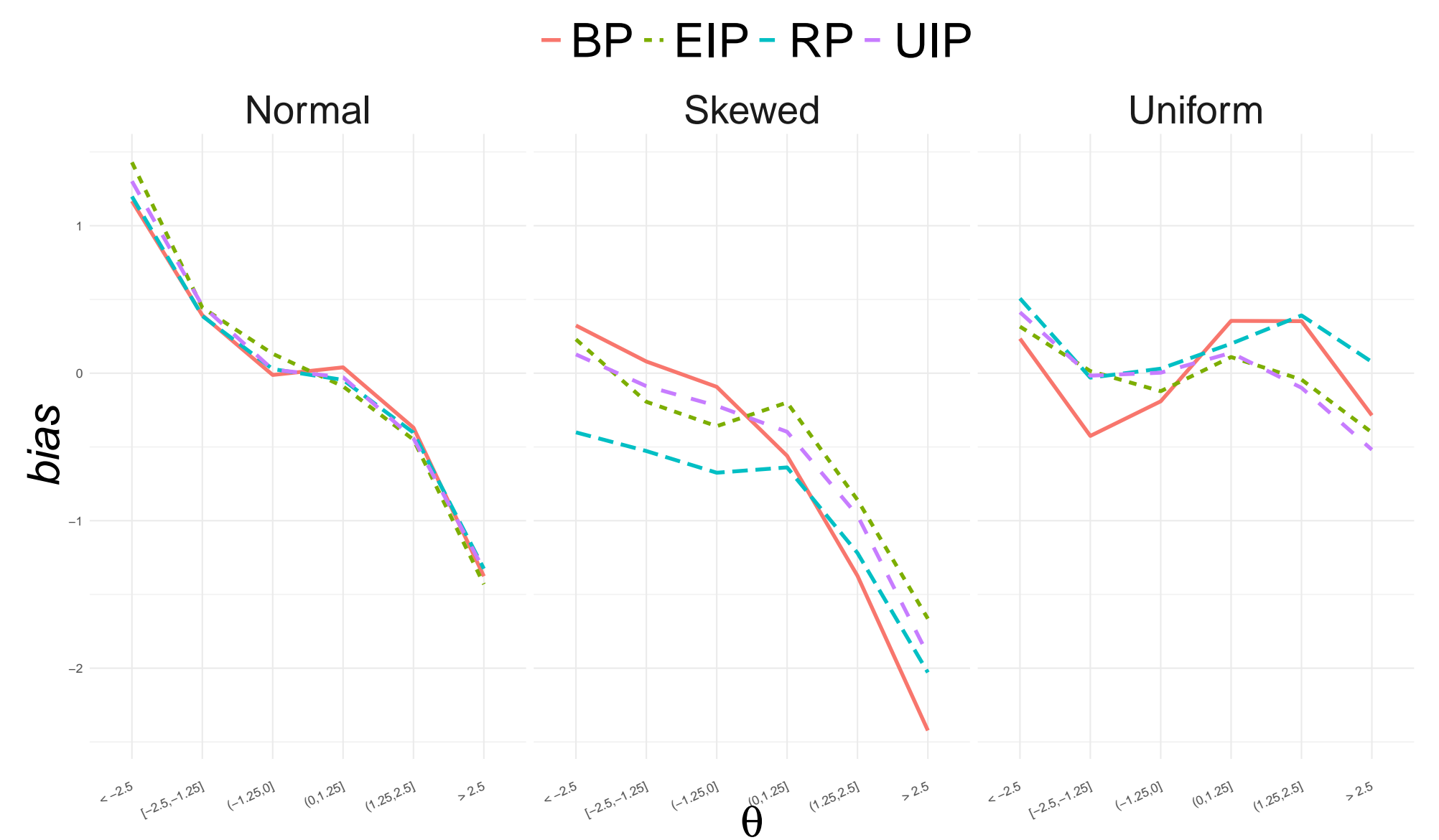


Figure 4: Bias for different group of θ

Discussion

- Different methods for different θ distributions
- Better performance of θ -based procedures on the extreme ends of the distributions
- By considering the θ' in the item selection procedures \rightarrow not the highest information but best coverage of the entire latent trait