New Item Selection Designs for Computerized Classification Test

Yingshi Huang[1] He Ren[1] Ping Chen[1*]

[1]Collaborative Innovation Center of Assessment for Basic Education Quality,
Beijing Normal University

NCME Annual Conference

Two problems faced by CCT's item selection method

- Computerized classification testing (CCT):
  - divide students into different groups (e.g., "pass" or "fail")

- Maximize Fisher information at the cut score θ_c (MFC):

$$i_k = \underset{j \in R_{k-1}}{\operatorname{argmax}} FI_j(\theta_c)$$

→ high risk of item leaking
long and uneven test-taking time

- Timed-MFC by Sie et al. (2015):

$$i_k = \underset{j \in R_{k-1}}{\operatorname{argmax}} \frac{FI_j(\theta_c)}{E\left(T_j \middle| \hat{\tau}_{k-1}\right)}$$

→ strong preference for the low time density items

Purpose: develop new item selection methods
✓ balanced item bank usage + short and stable test-taking time

How to achieve a balanced item usage?

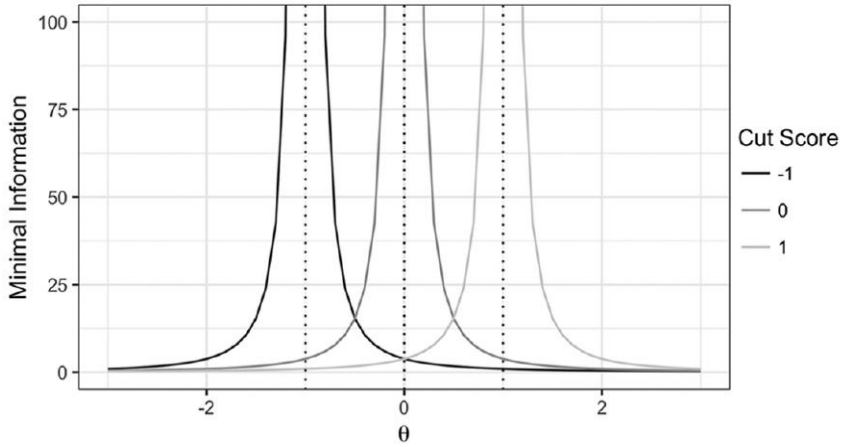
What if we relax the compulsory rule? (select the most informative items in all situations)

- Stage adaptive item selection method (SAI)

items		ability level
high information	→	"near-cut"
slightly low information	→	"away-cut"

$r\left[FI_j(\theta_c)\right] \in [0,1]$
(percentile rank function)

$$s = \left|\frac{\hat{\theta} - \theta_c}{range\ of\ \hat{\theta}}\right| = \frac{\left|\hat{\theta} - \theta_c\right|}{8}$$



[Luo et al., 2018 *Applied Psychological Measurement*]

measure the similarity between r and s:

$$SAI = \exp\left\{-\left|r\left[FI_j(\theta_c)\right] + w \times s - 1\right|\right\}$$

(w: weighting parameter ≥ 0)

$$i_k = \underset{j \in R_{k-1}}{\operatorname{argmax}}\, SAI$$

- the *"near-cut"* positions require more information
- while the *"away-cut"* positions require less

How to shrink the deviation and cost of test-taking time?

- Modified timed-MFC: combine the idea proposed by Choe et al. (2018)

$$i_k = \underset{j \in R_{k-1}}{\operatorname{argmax}} \frac{FI_j(\theta_c)}{\left|E\left(T_j \middle| \hat{\tau}_{k-1}\right) - v\right|}$$

  - $\hat{\tau}_{k-1}$ examinee's speed after $k-1$ items
  - β_j time density of item j
  - α_j time discriminating power of item j
  - v centering parameter ≥ 0

- from $E\left(T_j \middle| \hat{\tau}_{k-1}\right) \to 0$ to $E\left(T_j \middle| \hat{\tau}_{k-1}\right) \to v$

$$E\left(T_j \middle| \hat{\tau}_{k-1}\right) = e^{\beta_j - \tau + 1/(2{\alpha_j}^2)} = v$$

$$\beta_j + \frac{1}{2{\alpha_j}^2} = \hat{\tau}_{k-1} + lnv$$

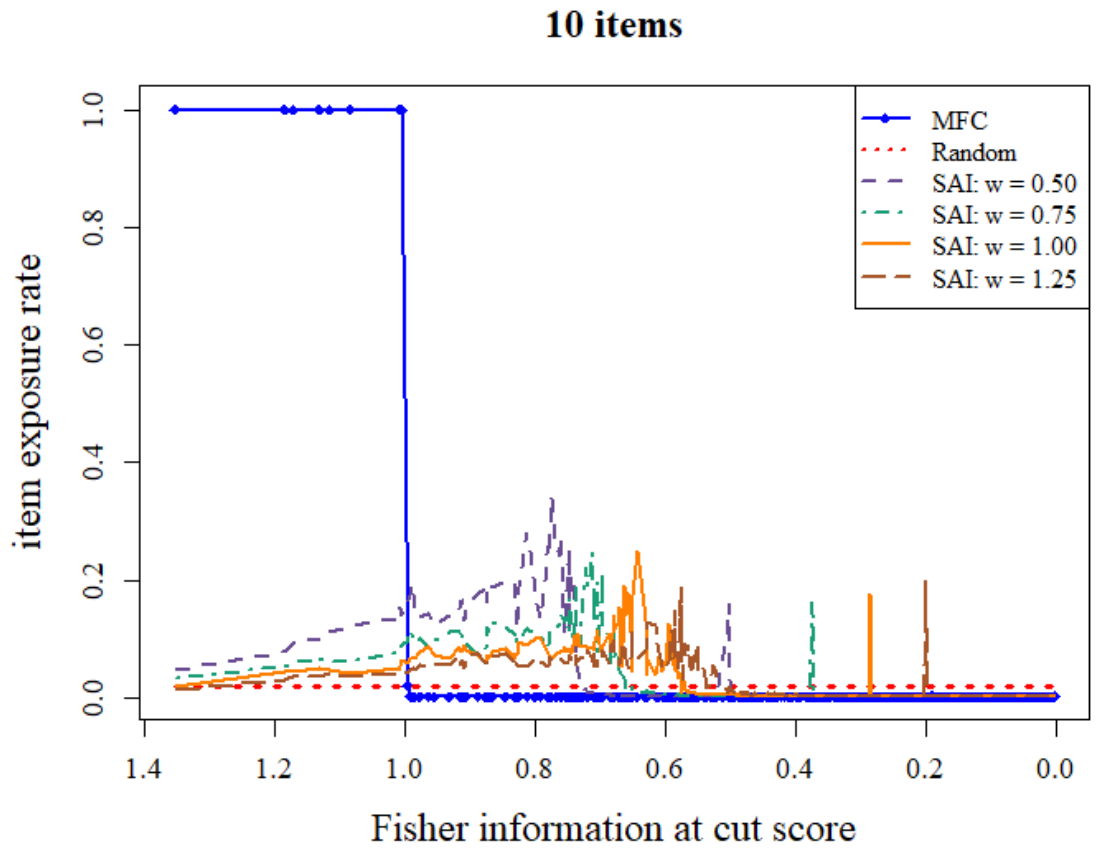
- increase the impact of examinee's speed

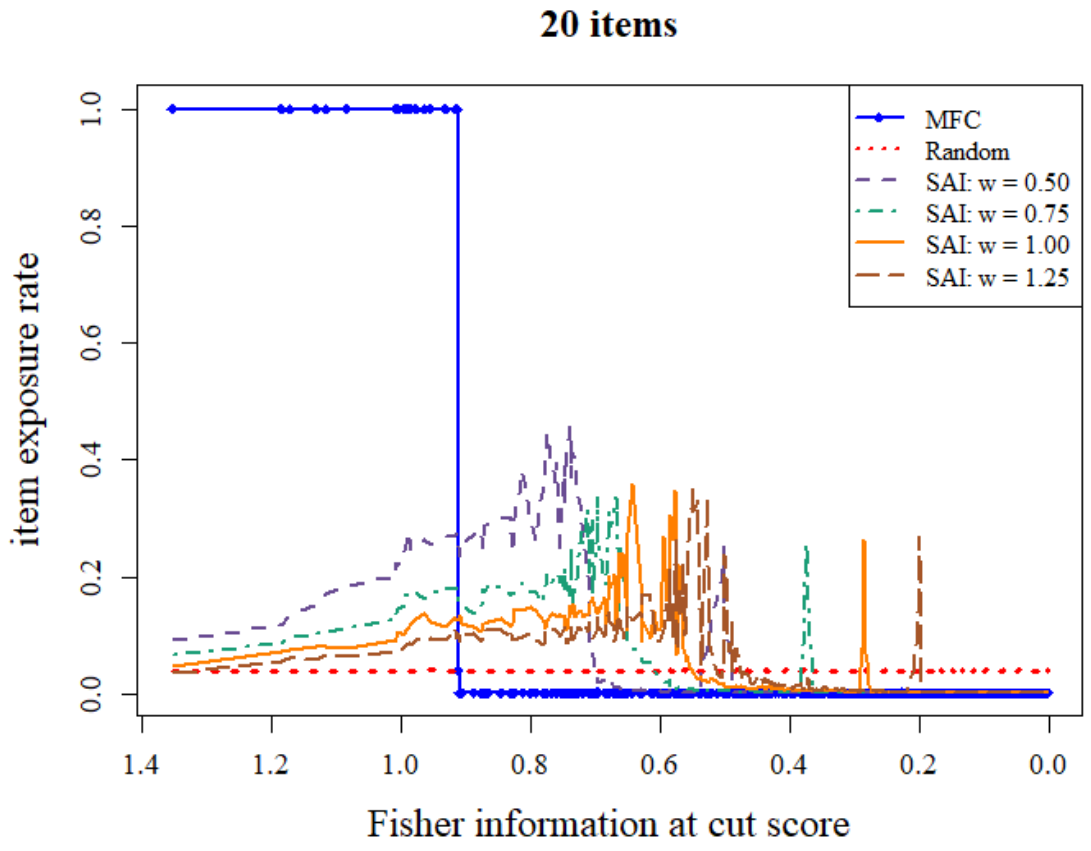
- Put forward the timed-SAI method:

$$i_k = \underset{j \in R_{k-1}}{\operatorname{argmax}} \frac{SAI}{\left|E\left(T_j \middle| \hat{\tau}_{k-1}\right) - v\right|}$$

Can SAI counterbalance the item usage?

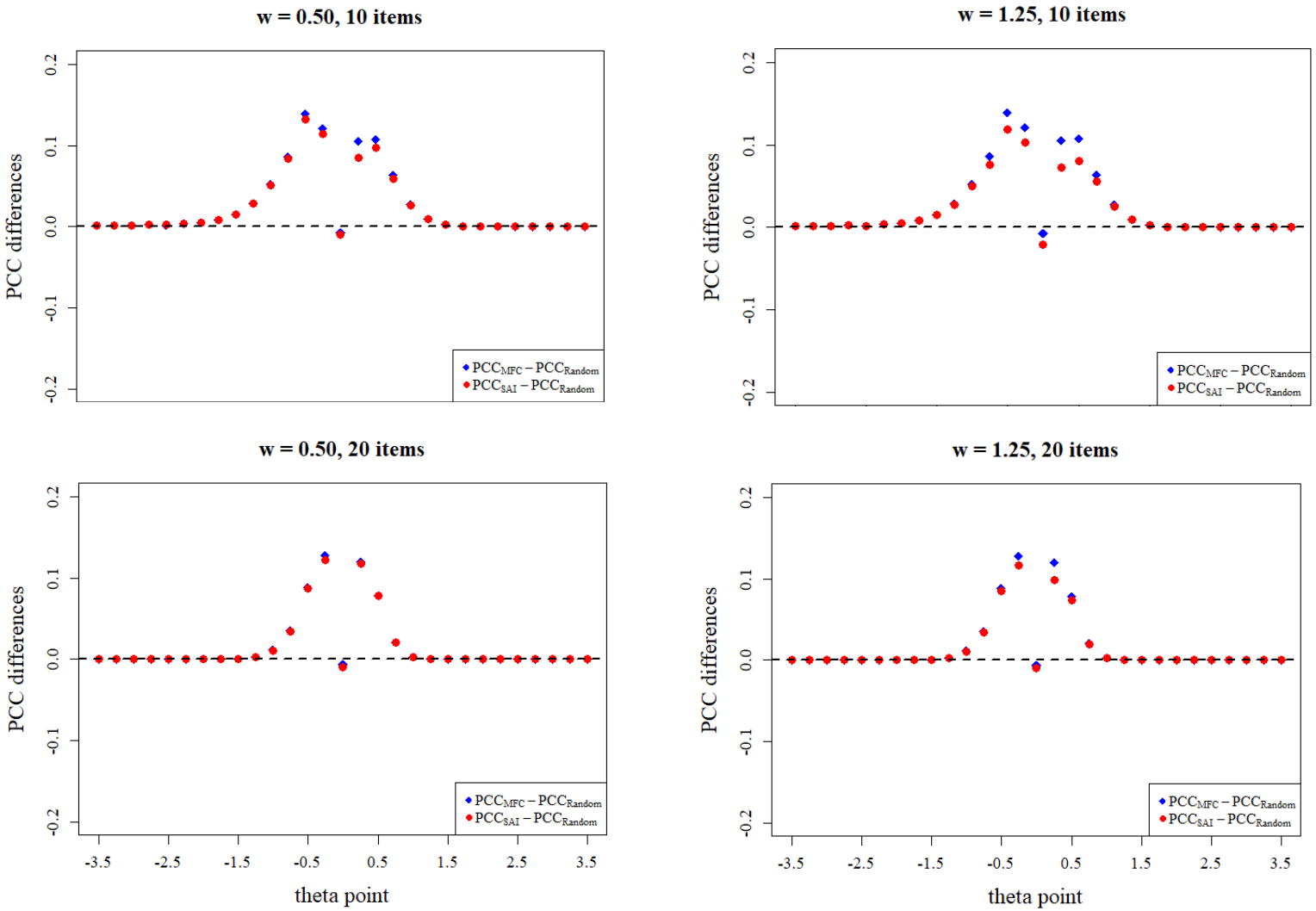
- ✓ The item exposure rates comparison among the random selection method, MFC, and SAI





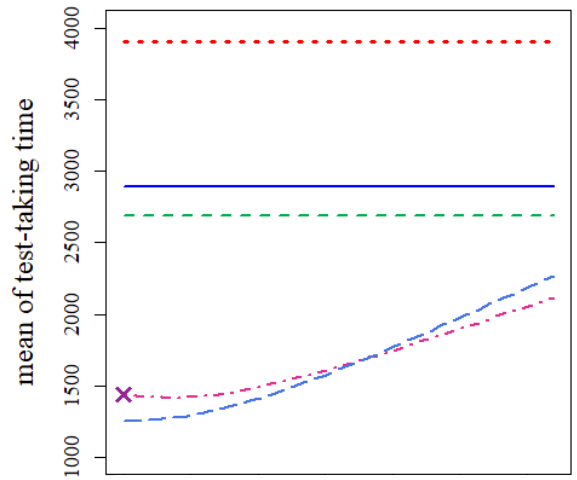
Is the balanced item exposure at a sacrifice of accuracy?

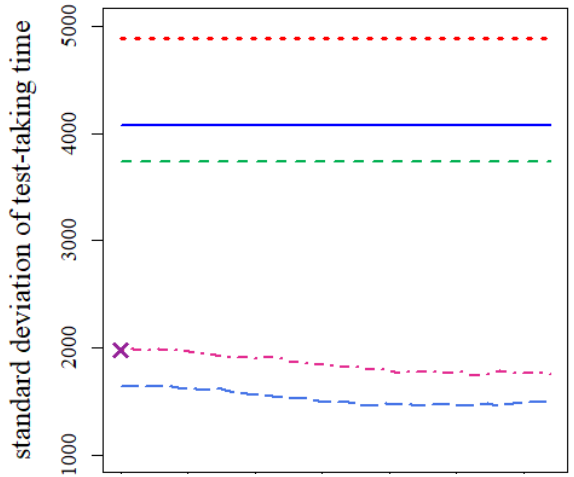
- ✓ The differences in percentage of correct classification (PCC) with the random method as baseline

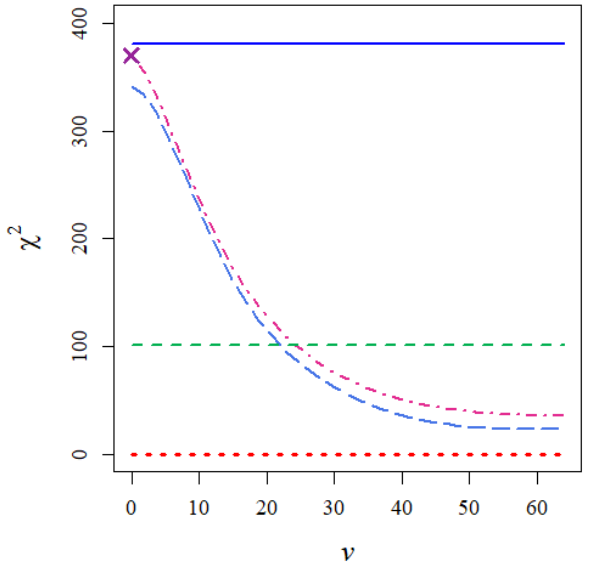


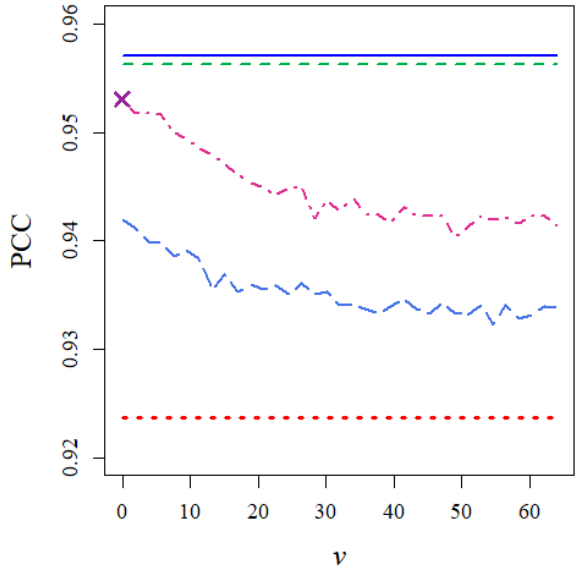
Can the new timed methods yield short and stable test time?

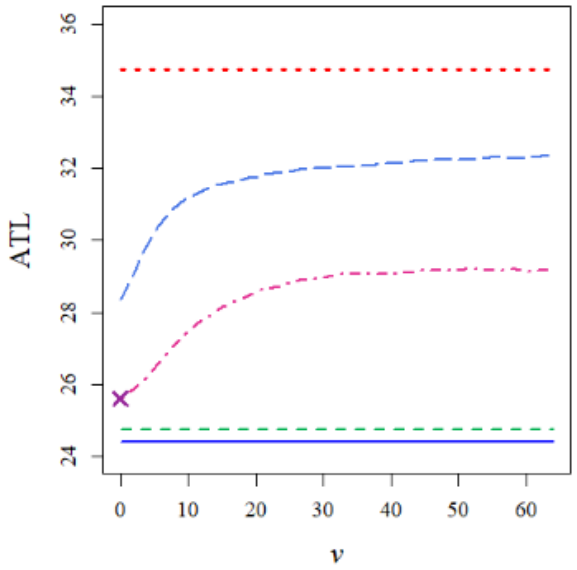
✓ Results of six item selection methods on all evaluation indicators with different v values

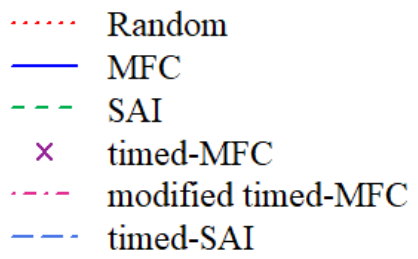












χ^2 measures the balance of item usage (the smaller, the better)

ATL = average test length

Take-home messages

- Can the stage adaptive method yields a balanced item bank usage?

– Yes, it give items with less information more opportunities to be selected.

- Can the new timed methods shrink the deviation and cost of test-taking time?

– Yes, the new methods gain the best time control achievement, but they lead to a slight extra cost in accuracy and test length.

- For practitioners:

– The implementation of SAI:

a $w \geq 1$ is recommended for high-stake tests

– For the two new timed methods:

design the v parameter according to the distance between the distribution of β_j and τ

Thanks for listening!

For any questions, please feel free to contact me:
h_yingshi@163.com