British Journal of Mathematical and Statistical Psychology
the british psychological society

Look-ahead content balancing method in variable-length computerized classification testing

Xiao Li[1]*, Jinming Zhang[1] and Hua-hua Chang[2]

[1]Department of Educational Psychology, University of Illinois at Urbana-Champaign, Illinois, USA

[2]Educational Psychology & Research Methodology, College of Education, Purdue University, West Lafayette, Indiana, USA

IF: 2.388

Reporter: Yingshi Huang

Introduction

- Computerized classification testing (CCT)
  - **statistical constraints:** max FI at current $\hat{\theta}$ …
  - **non-statistical constraints:** content balancing, exposure control …

A

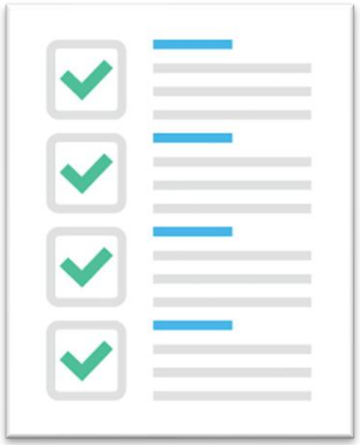
B

C

D

25% each area ×  items

= 15 items



variable-length computerized classification testing (VL-CCT)

proposing **two feasible methods** gains control over content coverage in **VL-CCT** programs

Introduction

- the constraint CAT method (Kingsbury & Weiss, 1983)
- the modified multinomial model method (Chen & Ankenman, 2004)
- the modified constraint CAT method (Leung, Chang, & Hau, 2000)
- the maximum priority index (MPI) (Cheng & Chang, 2009)
- the content-weighted item selection index (CWI) (Huo, 2009)

Methods - Content balancing item selection

How to assemble a test?

- Maximum Fisher information method

$$I_j(\theta) = \frac{(1-c_j)a_j^2 e^{a_j(\theta-b_j)}}{[1+e^{a_j(\theta-b_j)}]^2\{1-c_j+c_j[1+e^{a_j(\theta-b_j)}]\}}$$

- Maximum priority index

$I_j(\theta) \times$ items' contribution towards meeting constraints

$$\mathrm{PI}_j = I_j \prod_{k=1}^{K} (\omega_k f_k)^{c_{jk}}$$

Cheng & Chang, 2009 *BMSP*

Methods

- Maximum priority index (MPI)

$$\mathrm{PI}_j = I_j \prod_{k=1}^{K} (\omega_k f_k)^{c_{jk}}$$

$$C = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

	k_1	k_2	k_3	k_4
item 1	0	1	0	0
item 2	0	0	1	0
item 3	1	0	0	0
…	…	…	…	…
item J	0	0	0	1

$$f_k = \frac{X_k - x_k}{X_k}$$

$$= \frac{15 - \text{(items have been selected)}}{15}$$

k_1	A	ω_1
k_2	B	ω_2
k_3	C	ω_3
k_4	D	ω_4

~~25%~~ each area

- especially for exposure control

$$f_{kr} = \frac{r - n/N}{r}$$

? $l_k \le \mu_k \le u_k$

- **first phase**

$$f_k = \frac{(l_k - x_k)}{l_k}$$

all the lower bounds will be met at the end of the first phase

- **second phase**

$$f_k = \frac{(u_k - x_k)}{u_k}$$

Cheng & Chang, 2009 *BMSP*

Methods

- Content-weighted item selection index (CWI)
  - two-stage a-stratified method

1. first phase: fixed-length testing course

$\min|\hat{\theta} - b|$ until min test length

$$CWI_1 = \frac{l_k}{l_k - x_k + 1} \mid \hat{\theta} - b \mid$$

2. second phase: variable-length testing part

$\min|\hat{\theta} - b|$ until threshold

&

circularly increasing / decreasing

strata1 < strata2 < strata3 < strata4

$$CWI_2 = \frac{u_k}{u_k - x_k + 1} \mid \hat{\theta} - b \mid$$

Ca ➡	item1	item2	item3	item4
Cd ➡	item4	item3	item2	item1

Huo, 2009 *DOCTORAL DISSERTATION*

Methods

- MPI and CWI
  - upper bound

$$u_k = U \times u_k\%$$

	test	content area 1 (20%)	content area 2 (20%)	content area 3 (30%)	content area 4 (30%)
Lower bound	20	4	4	6	6
Upper bound	40	8	8	12	12

can be much larger than the actual ones

➡ **“step size”** + existing number of selected items

1. constant *S*
2. adaptive *S*

Methods

- Look-ahead content balancing (LA-CB): based on MPI
  - first phase $f_k = \frac{l_k - x_k}{l_k}$
  - second phase $f_k = \frac{(u_k - x_k)}{u_k}$

➡ the total test length (TL) ≤ the maximum test length (*U*)

What would be the remaining length?

$$x_k + S \times u_k\% \leq U \times u_k\%$$

$$1 \leq S \leq U - \sum_{k=1}^{K} x_k$$

- LA-CB-C: with constant step size

How to be adaptive?

S^{constant} ➡ a constant integer within the range

Methods

- LA-CB-A: with adaptive step size

$$1 \leq S \leq U - \sum_{k=1}^{K} x_k$$

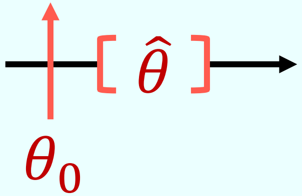
  - inspired by the ability confidence interval (ACI)

$$\hat{\theta} - Z_{\alpha} \times \text{SEM} < \theta < \hat{\theta} + Z_{\alpha} \times \text{SEM} \rightarrow SD(\hat{\theta}) \rightarrow \frac{1}{\sqrt{\sum_{j=1}^{n} I_j(\theta)}}$$

the normal deviate for a 100(1 - α)% CI

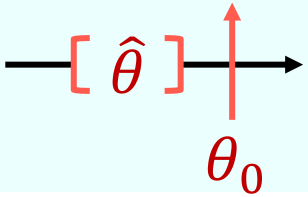
$$\hat{\theta} - Z_{\alpha} \frac{1}{\sqrt{\sum_{j=1}^{k} I_j(\theta)}} < \theta < \hat{\theta} + Z_{\alpha} \frac{1}{\sqrt{\sum_{j=1}^{k} I_j(\theta)}}$$

<pass>



$$\theta_0 = \hat{\theta} - Z_{\varepsilon} \frac{1}{\sqrt{\text{FI}_0}}$$

<fail>



$$\theta_0 = \hat{\theta} + Z_{\varepsilon} \frac{1}{\sqrt{\text{FI}_0}}$$

Methods

- LA-CB-A

$$\theta_0 = \hat{\theta} \pm Z_\varepsilon \frac{1}{\sqrt{FI_0}}$$

$$\theta_0 - \hat{\theta} = \pm Z_\varepsilon \frac{1}{\sqrt{FI_0}}$$

$$\left[\theta_0 - \hat{\theta}\right]^2 = \left[Z_\varepsilon \frac{1}{\sqrt{FI_0}}\right]^2$$

$$FI_0 = \left[\frac{Z_\varepsilon}{\theta_0 - \hat{\theta}}\right]^2$$

the remaining length!

$$\frac{FI_0 - \sum_{j=1}^{k} I_j(\hat{\theta})}{\max(I_{\text{unselected}})} < \text{no. of items} < \frac{FI_0 - \sum_{j=1}^{k} I_j(\hat{\theta})}{\min(I_{\text{unselected}})}$$

as the look-ahead upper bound *U*

$$S_0^{\text{adaptive}} = \frac{FI_0 - \sum_{j=1}^{k} I_j(\hat{\theta})}{\max(I_{\text{unselected}})} - \sum_{k=1}^{K} x_k$$

in a relatively early stage, not yet close to FI_0

$$S^{\text{adaptive}} = \min\{S_0^{\text{adaptive}}, S^{\text{constant}}\}$$

$$f_k = \frac{(u_k - x_k)}{u_k} = \frac{S \times u_k\%}{x_k + S \times u_k\%}$$

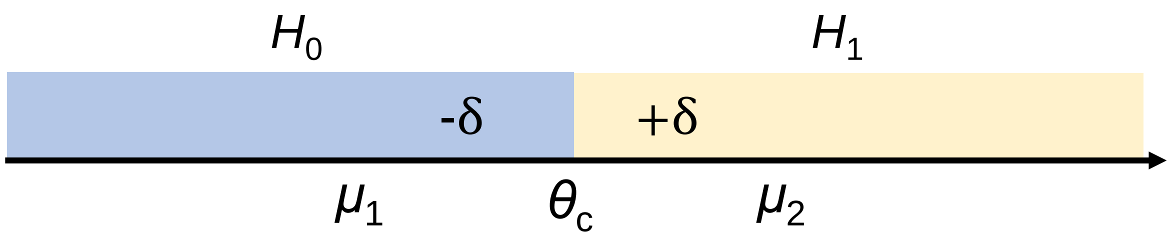
S^{constant}

S^{adaptive}

Methods - Classification

How to stop?

- The sequential probability ratio test (SPRT)



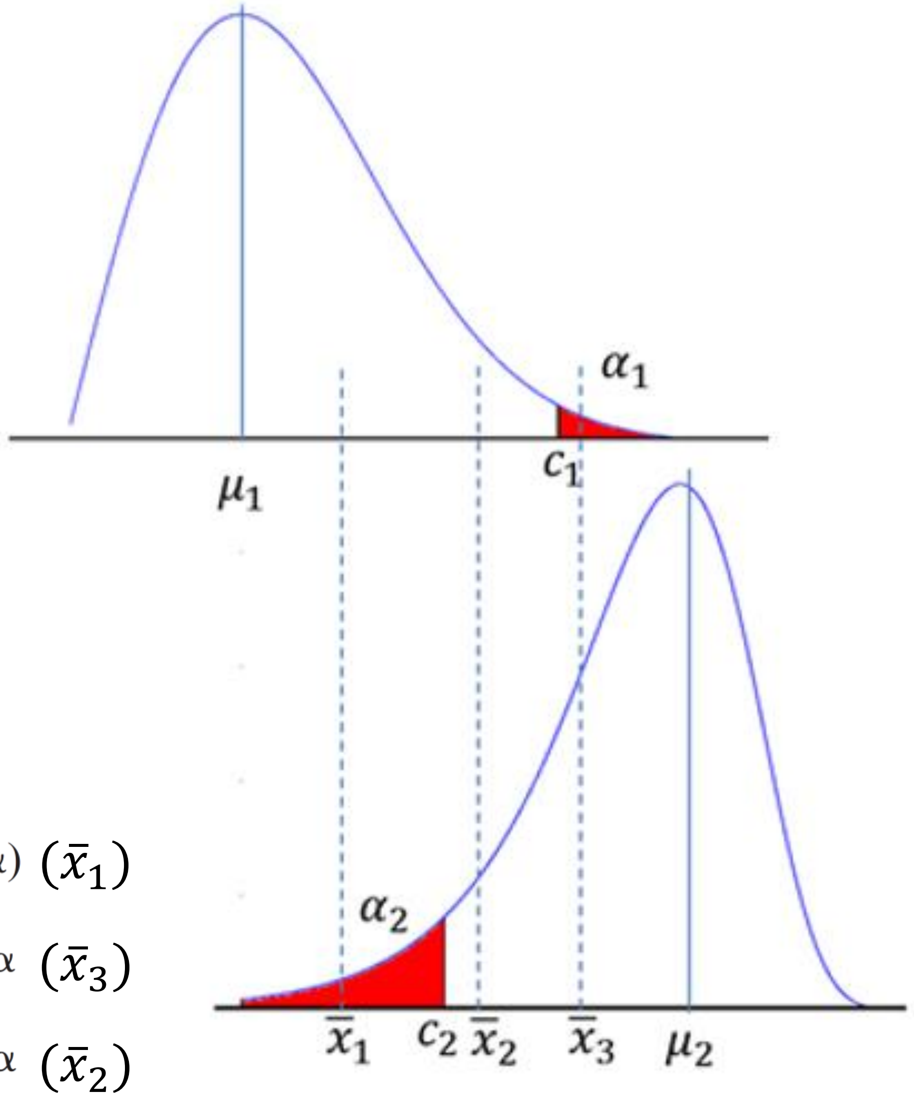
$$L(\theta;\mathbf{x}) = \prod_{i=1}^{k} p_i(\theta)^{x_i}\left[1 - p_i(\theta)\right]^{1-x_i}$$

$$LR(\theta_c+\delta;\theta_c-\delta) = \frac{L(\theta_c+\delta;\mathbf{x})}{L(\theta_c-\delta;\mathbf{x})}$$

– ability below the cutting point if $LR(\theta_c+\delta;\theta_c-\delta) \le \beta(1-\alpha)$ $(\bar{x}_1)$

– ability above the cutting point if $LR(\theta_c+\delta;\theta_c-\delta) \ge (1-\beta)\alpha$ $(\bar{x}_3)$

– administer another item if $\beta(1-\alpha) < LR(\theta_c+\delta;\theta_c-\delta) < (1-\beta)\alpha$ $(\bar{x}_2)$



Methods

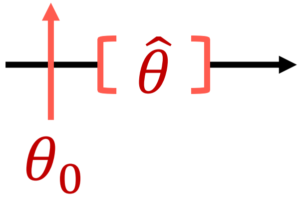
- Ability confidence interval (ACI)

$$\hat{\theta} - Z_\alpha \times \text{SEM} < \theta < \hat{\theta} + Z_\alpha \times \text{SEM}$$

Z_α: the normal deviate for a 95% CI

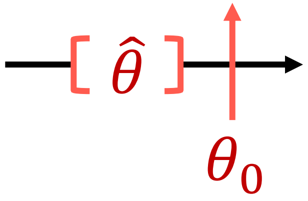
SEM: $SD(\hat{\theta}) \rightarrow \frac{1}{\sqrt{\sum_{j=1}^{n} I_j(\theta)}}$

  - 95% CI is above the cut-off score θ_0



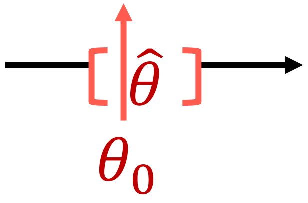
<pass>

  - 95% CI is below the cut-off score θ_0



<fail>

  - 95% CI is equal to or within the cut-off score θ_0



<continue>

Simulation studies

- Study 1.
  to choose a preferable classification method (**ACI or SPRT** methods) with LA-CB-C

- Study 2.
  whether the **LA-CB-C** method controls content constraints better than the **existing MPI and CWI** methods

- Study 3.
  whether the **LA-CB-A** method further improves the content balancing performance on top of the **LA-CB-C** method

Simulation studies

- Data generation
  - Item pool structure
    - $a \rightarrow$ 0.5, 1.0, 1.5, 2.0
    - $b \sim N(0, 1)$
    - $c \sim U(0, 0.25)$
    - 3PLM with 400 items
    - 4 content areas, 100 items/area, 25%
    - weights are all = 10
    - [minimum, maximum] test lengths = [28, 60]
    - constraint k = 1, 2, 3, 4 (bounded = [7, 15])
    - exposure rate < 0.2
    - weight = 100
  - Examinee generation
    - 2000 examinees with $\theta \sim N(0, 1)$

Simulation studies

- Data generation

  - Model settings

    - $\delta = 0.2$, $\alpha = \beta = 0.05$
    - cut-off score $\theta_0 = 0$
    - step size $S = [3, 20]$, i.e., 18 integral values

$$S^{\text{actual}} = \max\left\{1, \min\left\{S, U - \sum_{k=1}^{K} x_k\right\}\right\}$$

LA-CB-C: $S = S^{\text{constant}}$

LA-CB-A: $S = S^{\text{adaptive}}$

    - the first three items are always selected randomly

selected from the two best items:

maximized priority index / Fisher information

minimized weighted index …

Simulation studies

- Evaluation criteria

1. Classification accuracy

- Classification error rate (CER)
- Type I error rate (Type I ER)
- Type II error rate (Type II ER)
- Mean square error:

$$\mathrm{MSE} = \frac{\sum_{i=1}^{N} (\hat{\theta}_i - \theta_i)^2}{N}$$

2. Content balancing

- The average of a test:

$$\bar{V} = \frac{\sum_{i=1}^{N} V_i}{N}$$

- The average of a step size:

$$\text{Average } \bar{V} = \frac{\sum_{p=P_0}^{P} \bar{V}_p}{P - P_0 + 1}$$

3. Exposure control

- The maximum item exposure rate
- The proportion of over-exposed items (exposure rate > 0.2)
- The proportion of unused items
- Observed vs. expected exposure rates (ER):

$$\chi^2 = \sum_{j=1}^{K} \frac{(\mathrm{ER}_j - \overline{\mathrm{ER}})^2}{\overline{\mathrm{ER}}}$$

4. Test efficiency

- The average across various examinees:

$$\overline{\mathrm{TL}} = \frac{\sum_{i=1}^{N} \mathrm{TL}_i}{N}$$

Results – Study 1

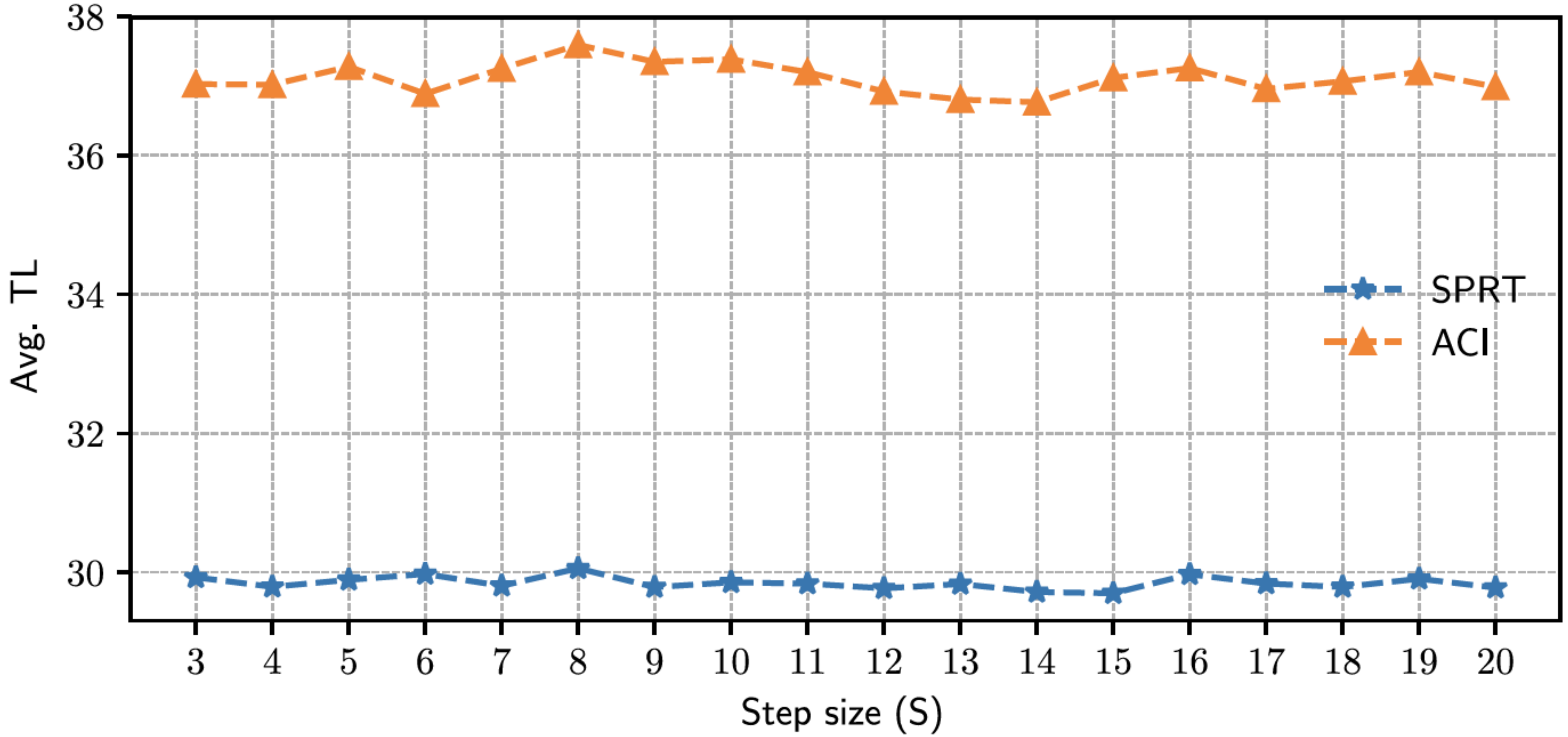


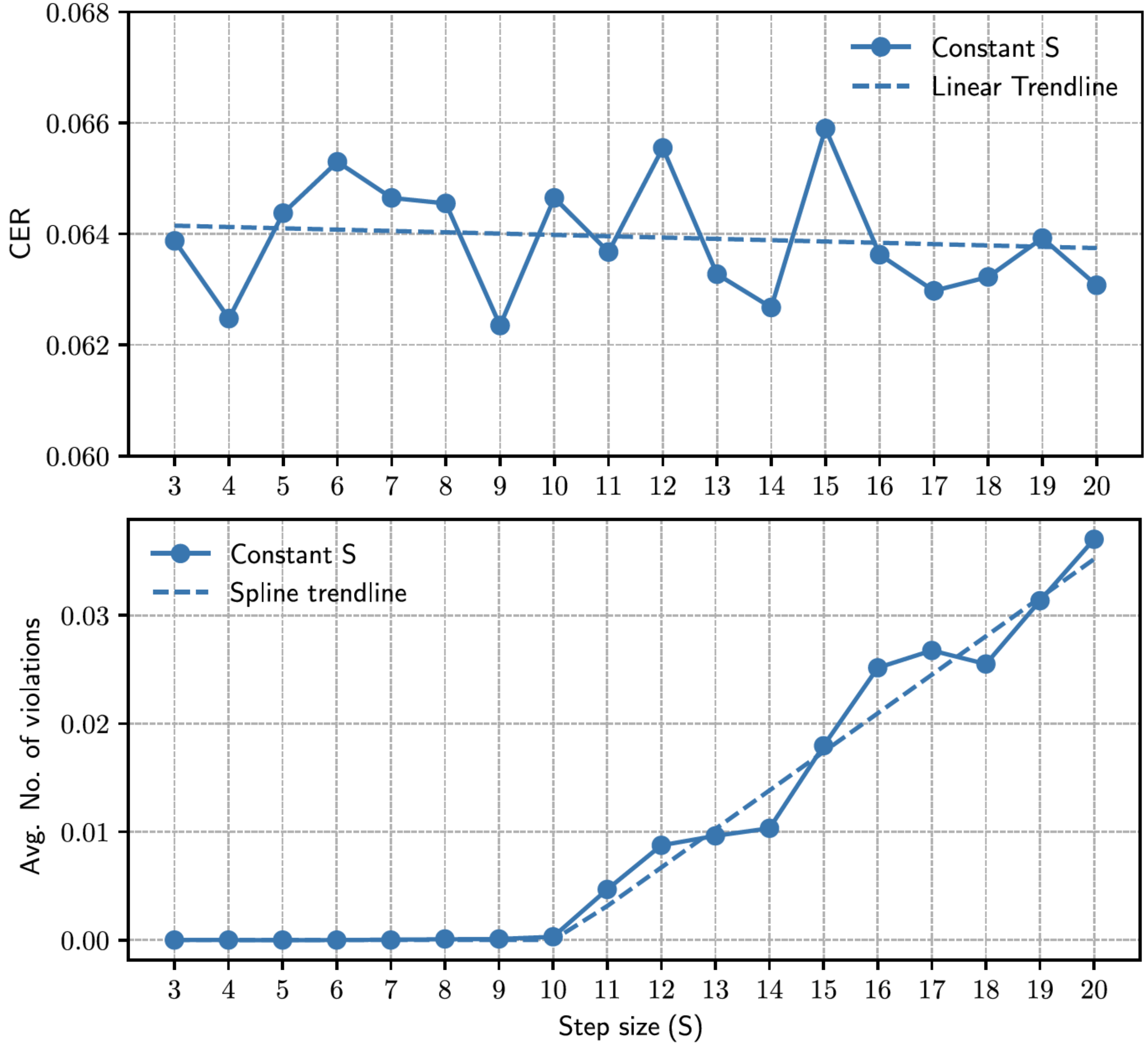
Figure 1. Average test length (TL) of sequential probability ratio test (SPRT) and ability confidence interval (ACI). [Colour figure can be viewed at wileyonlinelibrary.com]

Results – Study 1

Table 1. Overall performance of sequential probability ratio test (SPRT) and ability confidence interval (ACI) classification methods

Methods	SPRT	ACI
Avg. test length TL	29.85	37.11
Grand avg. violated constraints $\bar{V}$	0.011	0.017
Average classification error rate	0.063	0.060
Average Type I error rate	0.033	0.030
Average Type II error rate	0.030	0.030

Results – Study 2



Results – Study 2

Table 2. Classification error rates (ER) and mean square error (MSE) of the LA-CB-C method and three other methods for 18 step sizes *S*

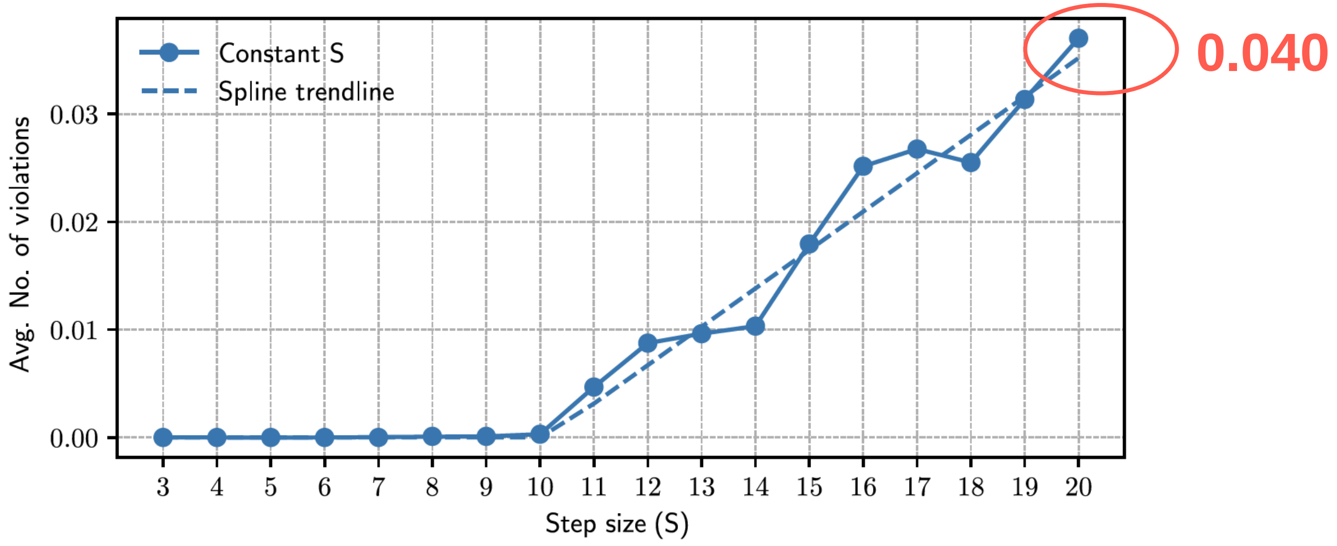
S	CER	Type I ER	Type II ER	MSE
3	0.064	0.034	0.030	0.074
4	0.062	0.033	0.030	0.074
5	0.064	0.033	0.031	0.075
6	0.065	0.034	0.031	0.075
7	0.065	0.033	0.031	0.075
8	0.065	0.033	0.032	0.074
9	0.062	0.033	0.030	0.075
10	0.065	0.034	0.031	0.076
11	0.064	0.032	0.032	0.075
12	0.066	0.034	0.031	0.075
13	0.063	0.033	0.030	0.075
14	0.063	0.033	0.030	0.075
15	0.066	0.034	0.032	0.075
16	0.064	0.033	0.031	0.074
17	0.063	0.033	0.030	0.075
18	0.063	0.033	0.030	0.075
19	0.064	0.032	0.032	0.075
20	0.063	0.033	0.030	0.075
LA-CB-C average	0.064	0.033	0.031	0.075
Maximum priority	0.062	0.032	0.030	0.074
Content-weighted	0.078	0.040	0.039	0.118
Maximum information	0.054	0.032	0.022	0.054
Randomized	0.084	0.042	0.042	0.254

Results – Study 2

Table 5. Summary of content constraint violations ($\bar{V}$)

Measures	Average $\bar{V}$	Max $\bar{V}$	Min $\bar{V}$
LA-CB-C	0.0110	0.0370	0
Maximum priority*	0.0540	–	–
Content-weighted*	2.2295	–	–
Maximum information*	8.1380	–	–
Randomized*	7.0230	–	–

Note. This table summarizes the statistics of $\bar{V}$ for 18 step sizes. Methods with (*) do not include step sizes to make item selections and therefore maximum and minimum $\bar{V}$ are not applicable.



Results – Study 2

Table 3. Overall exposure control indices

Methods	LA-CB-C	Maximum priority	Content-weighted	Maximum information	Randomized
Max. exposure rate	0.178	0.175	0.166	0.532	0.100
Over-exposed (%)	0	0	0	3.2	0
Never exposed (%)	0	0	0	0	0
χ^2	20.297	20.228	4.009	83.041	0.153

Results – Study 3

Table 4. LA-CB-A classification error rates (ER) and mean square error (MSE) for different step sizes *S*

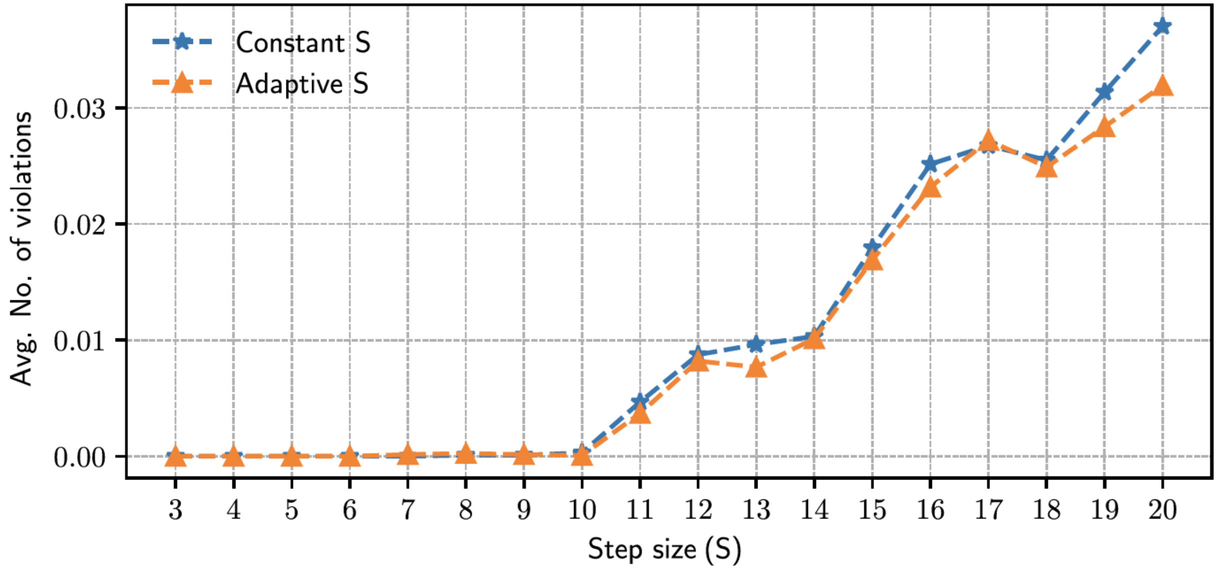
Constant *S*	CER	Type I ER	Type II ER	MSE
3	0.065	0.034	0.031	0.076
4	0.063	0.032	0.031	0.076
5	0.064	0.034	0.030	0.076
6	0.064	0.033	0.031	0.075
7	0.062	0.032	0.030	0.075
8	0.064	0.033	0.031	0.074
9	0.065	0.033	0.031	0.076
10	0.063	0.033	0.031	0.073
11	0.065	0.033	0.031	0.074
12	0.062	0.032	0.030	0.074
13	0.066	0.034	0.032	0.075
14	0.063	0.034	0.029	0.074
15	0.063	0.032	0.031	0.076
16	0.064	0.033	0.031	0.074
17	0.064	0.033	0.031	0.074
18	0.062	0.032	0.030	0.075
19	0.064	0.033	0.030	0.074
20	0.064	0.033	0.031	0.075
LA-CB-A average	0.064	0.033	0.031	0.075
Maximum priority	0.062	0.032	0.030	0.074
Content-weighted	0.078	0.040	0.039	0.118
Maximum information	0.054	0.032	0.022	0.054
Randomized	0.084	0.042	0.042	0.254

Results – Study 3

Table 5. Summary of content constraint violations ($\bar{V}$)

Measures	Average $\bar{V}$	Max $\bar{V}$	Min $\bar{V}$
LA-CB-C	0.0110	0.0370	0
LA-CB-A	0.0102	0.0319	0
Maximum priority*	0.0540	–	–
Content-weighted*	2.2295	–	–
Maximum information*	8.1380	–	–
Randomized*	7.0230	–	–

Note. This table summarizes the statistics of $\bar{V}$ for 18 step sizes. Methods with (*) do not include step sizes to make item selections and therefore maximum and minimum $\bar{V}$ are not applicable.



Results – Study 3

Table 6. Overall exposure control indices

Methods	LA-CB-C	LA-CB-A
Maximum exposure rate	0.178	0.178
Over-exposed (%)	0	0
Never exposed (%)	0	0
χ^2	20.297	20.289

Conclusions and Future directions

- the LA-CB methods **perform better** than the CWI and MPI methods
  - controlling constraints, while still maintaining high classification accuracy
- **different stopping rules** can be evaluated and optimally determined
- integrated with the **shadow test approach**
- how the LA-CB methods work when integrated with **other item selection methods**

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

Results – Study 1