Ramsay-Curve Differential Item Functioning

Carol M. Woods

Washington University in St. Louis

IRT-LR-DIF (Thissen, Steinberg, & Gerrard, 1986; Thissen, Steinberg, & Wainer, 1988, 1993) is popular for testing items for differential item functioning.

✓ 2-group IRT w/ likelihood ratio tests between nested models
✓ Some items presumed DIF-free: anchors vs studied items
✓ F-group mean & SD estimated; R-group mean = 0, SD = 1

Typically, the latent variable (θ) is assumed to be normal for both groups. However, if θ is skewed for one of the groups but presumed normal:

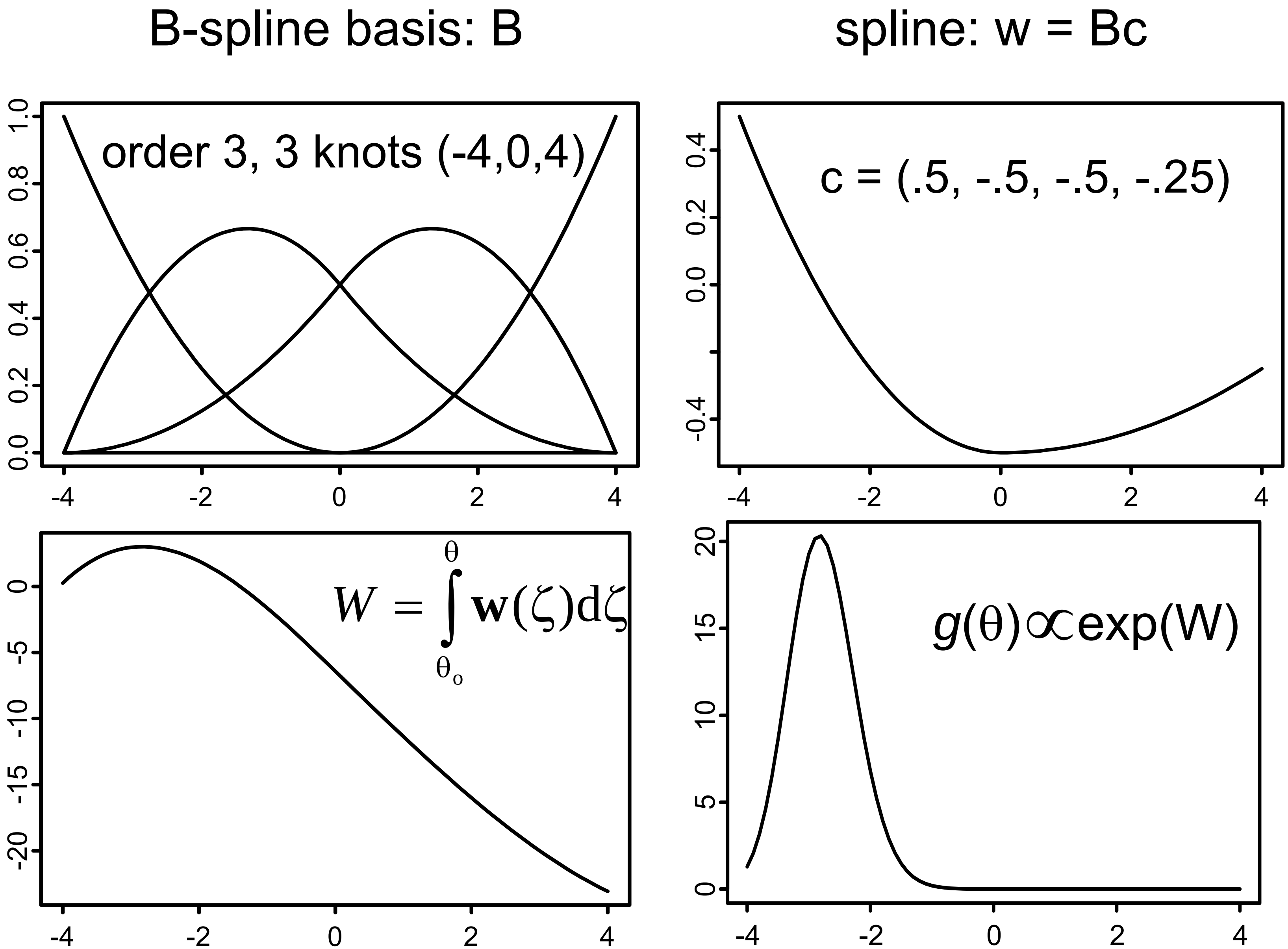
- inflated Type I error for LR-DIF tests
- inaccurate item parameter estimates
- inaccurate estimates of F-group mean & SD

RC-DIF-1: IRT-LR-DIF testing with estimation of the F-group latent density, $g_F(\theta)$.

- R-group latent density, $g_R(\theta)$, assumed N(0,1)
- designed for variables which are approximately normal for mainstream group but possibly not for some F-groups
  - maybe F-group more heterogeneous than R-group

  e.g., θ = English proficiency
  R-group = native English speakers
  F-group = nonnative English speakers
  (various languages and cultures)

Ramsay-curve = linear combination of B-splines + operations so result is a pdf (Ramsay, 2000; Silverman, 1982).



The RC-DIF-1 Procedure

e.g., 2PL model: $$\mathrm{T}(\mathrm{u}_{ig} = 1) = \frac{1}{1 + e^{[-\mathrm{a}_{ig}(\theta - \mathrm{b}_{ig})]}}$$

u = item response; a = discrimination; b = difficulty
i = item; = g = group

<u>For each studied item:</u>

1) Fit 16 R-curves to $g_F(\theta)$ assuming $a_F = a_R$, $b_F = b_R$ for all anchors & the studied item (other items ignored)

- order and number of knots for B-splines: 2, 3, 4, 5

2) Select one best R-curve using Hannan-Quinn criterion

- HQ = -2LogL + 2m(log(log(N)))
- m = parameters, N = people

3) Fit the model permitting $a_F \neq a_R$ and $b_F \neq b_R$ for this studied item with $g_F(\theta)$ = HQ best R-curve.

4) Test DIF using -2loglikelihoods from the two models that both use $g_F(\theta)$ = HQ best R-curve

H_o: $a_F = a_R$ and $b_F = b_R$
H_a: $a_F \neq a_R$, $b_F \neq b_R$, or both

RC-DIF-1 is implemented in C++.

Simulation study done to evaluate performance:

Method

N_R = 1,500; N_F = 500; $\bar{\theta}_R = 0; \bar{\theta}_F = -0.4$; both SDs = 1

24 binary items; 2PL model

a_R ~ N(μ = 1.7, σ = 0.3) truncated at 1.2 and 4

b_R ~ N(μ = 0, σ = 1) truncated at ± 2

12 DIF-free anchor items:	$a_F = a_R$	$b_F = b_R$
4 items with nonuniform DIF:	$a_F = a_R - \delta$	$b_F = b_R + \gamma$
4 items with uniform DIF:	$a_F = a_R$	$b_F = b_R + \gamma$
4 studied items without DIF:	$a_F = a_R$	$b_F = b_R$

$\delta; \gamma$ randomly determined = .3, .4, .5, .6, or .7

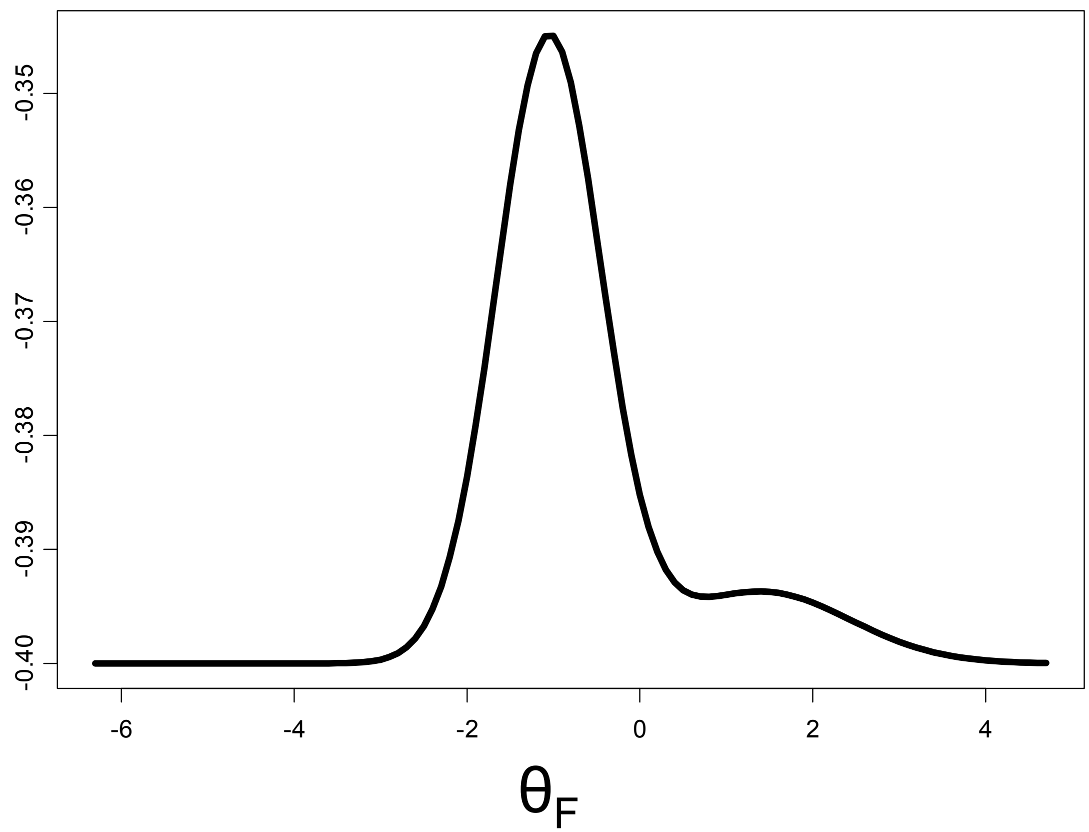
(equally likely values; $\delta \neq \gamma$ except by chance)

100 replications

True latent density

$\theta_R \sim N(0,1)$

$\theta_F \sim$ either $N(-0.4,1)$ or skewed:



RC-DIF-1: max order and # knots = 5

quadrature: -5.5, 5.5 incr: .1 (111 points)

Results

Misspecification of $g_F(\theta)$:

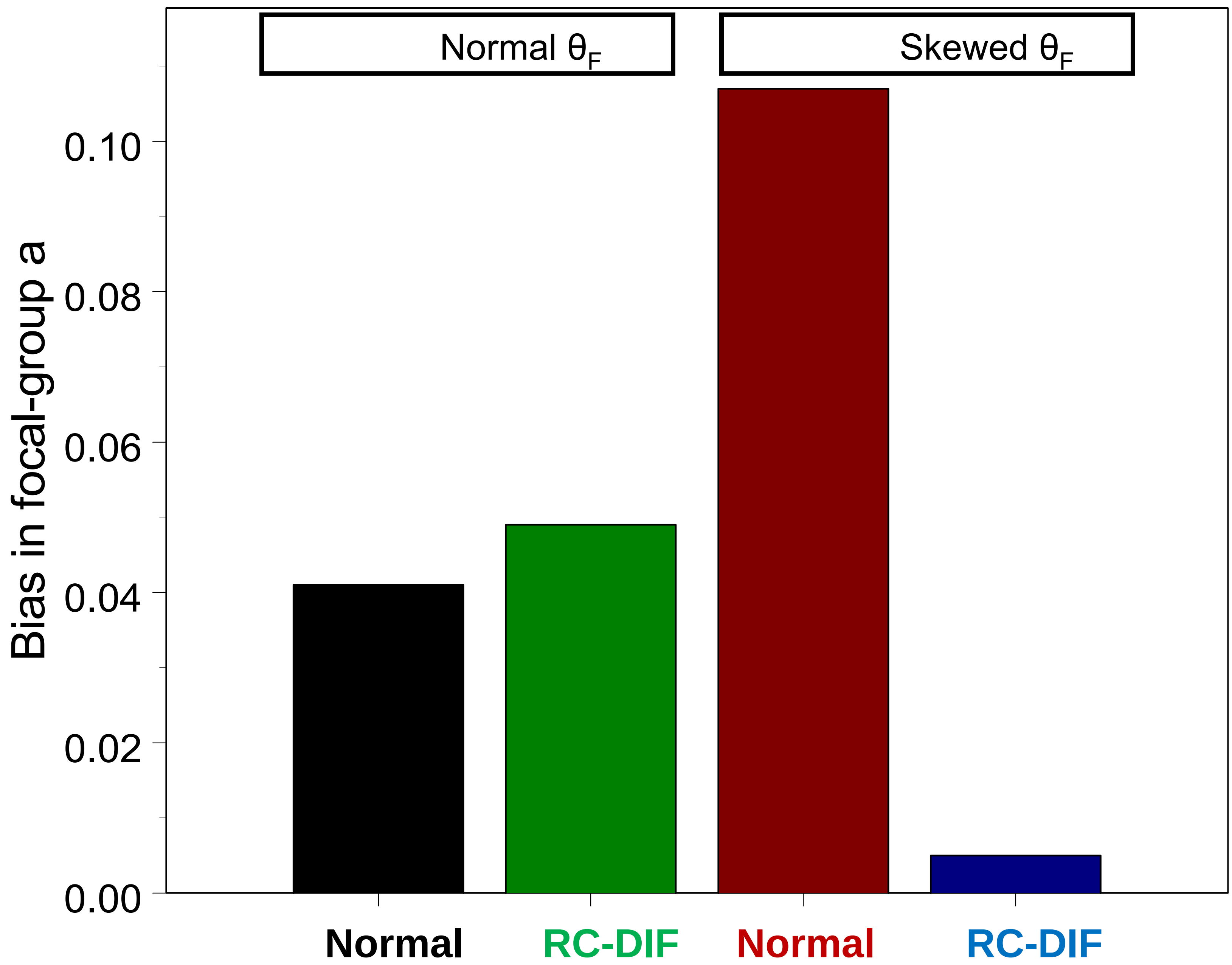
has little effect on R-group item parameters

but produces inaccurate:

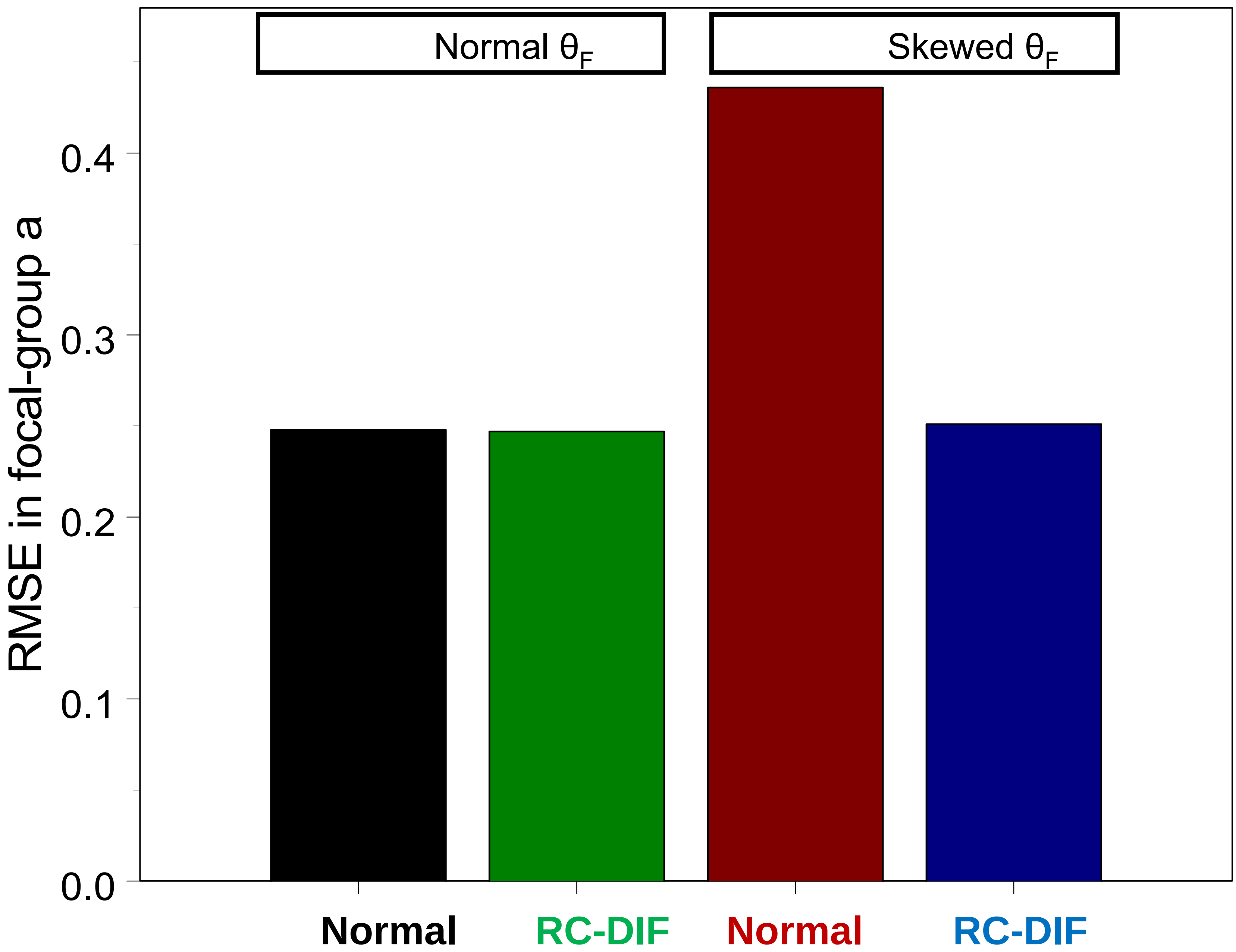
- F-group item parameters (thus DIF effects)
- estimates of $\bar{\theta}_F$
- false alarm rates for DIF tests

RC-DIF-1 greatly improves accuracy of LR-DIF results.

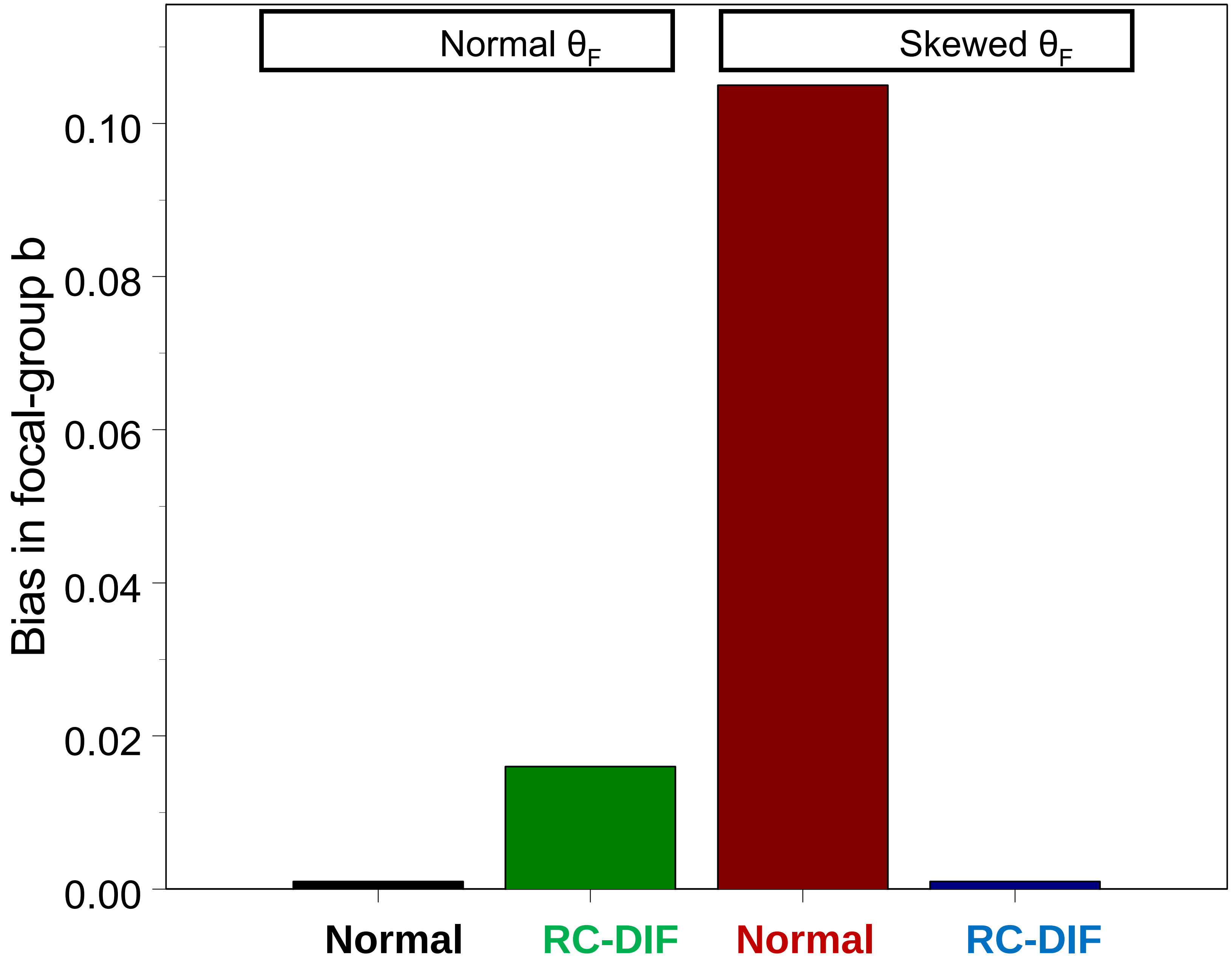
Bias: discrimination



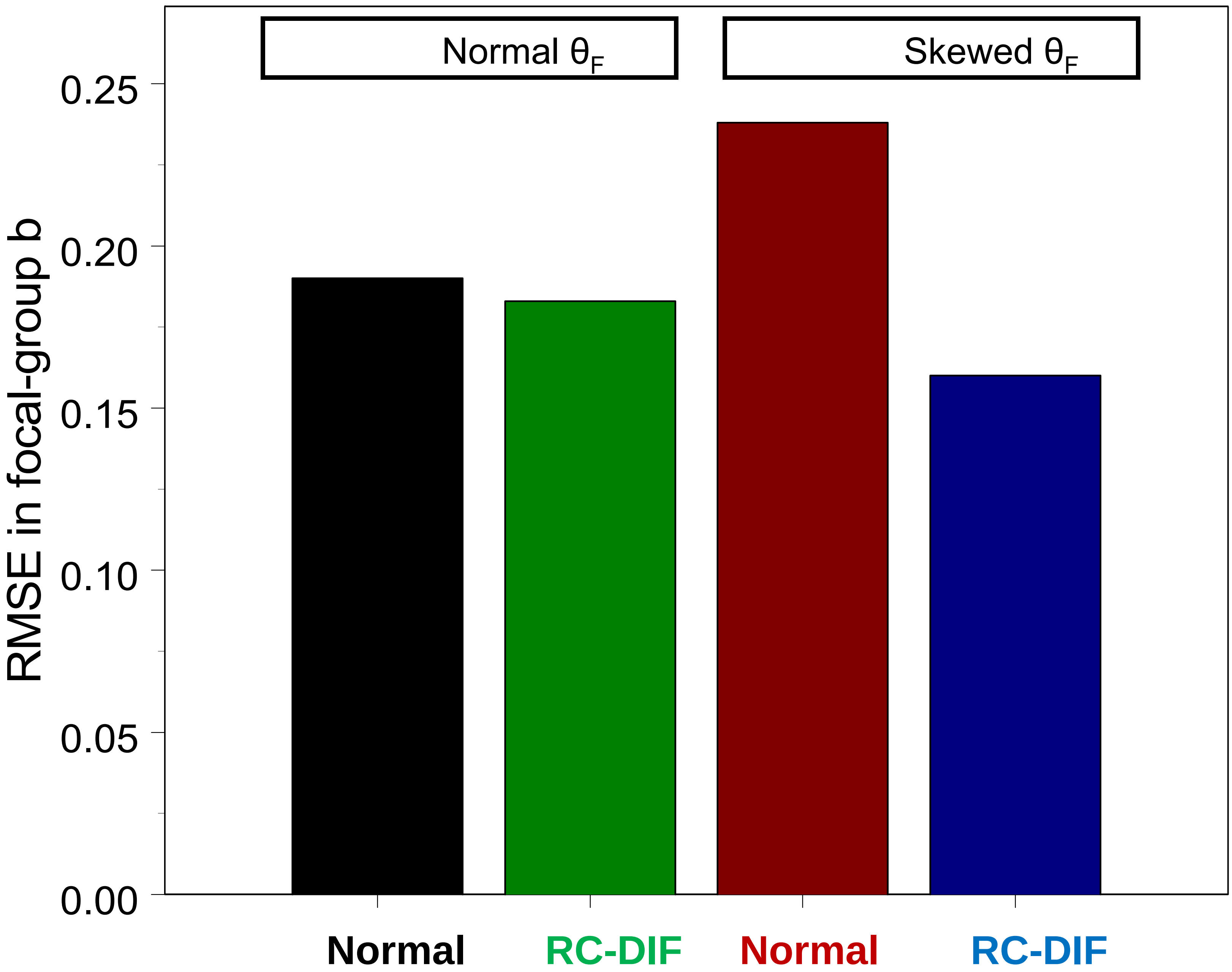
RMSE: discrimination



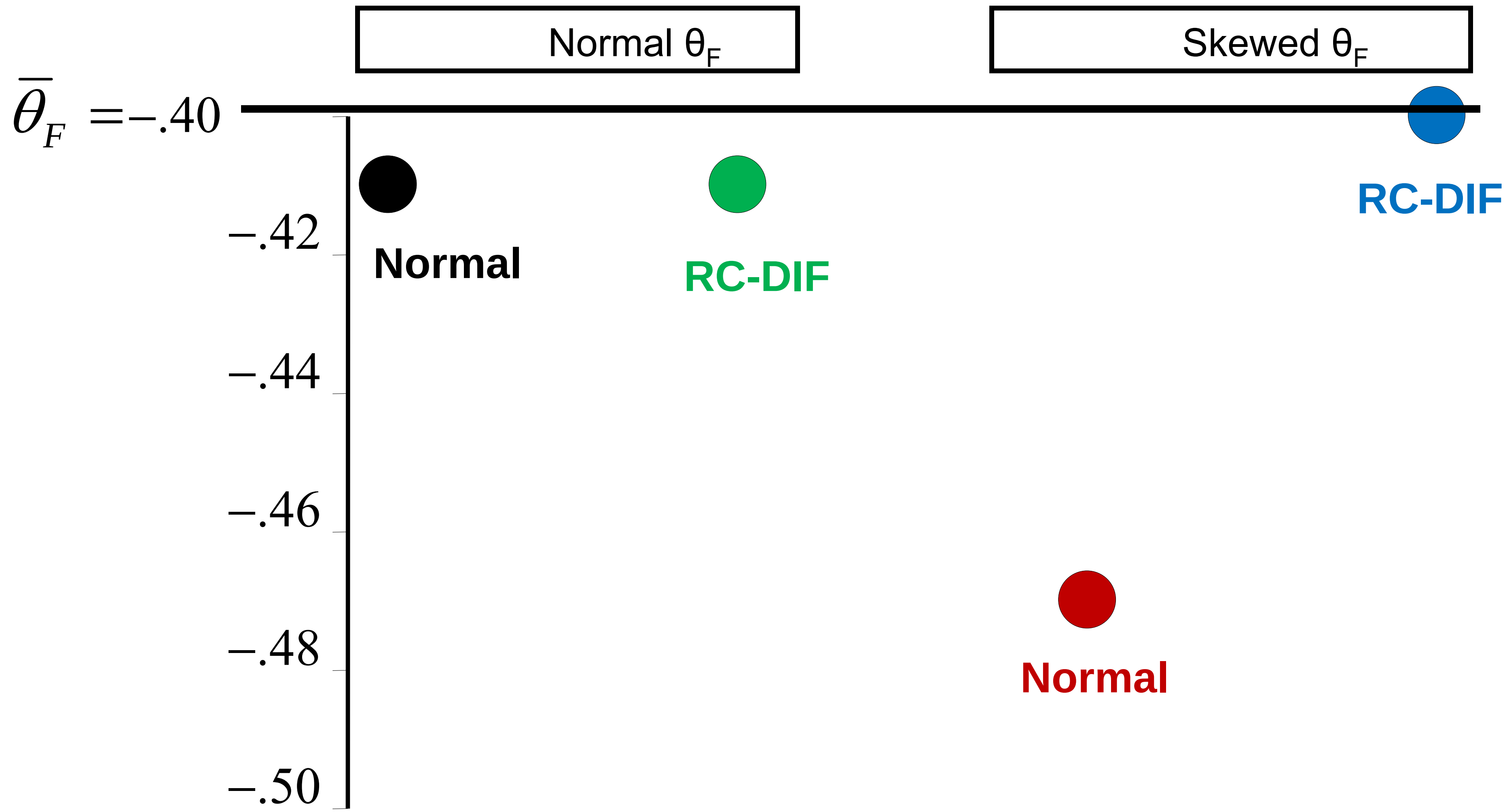
Bias: difficulty



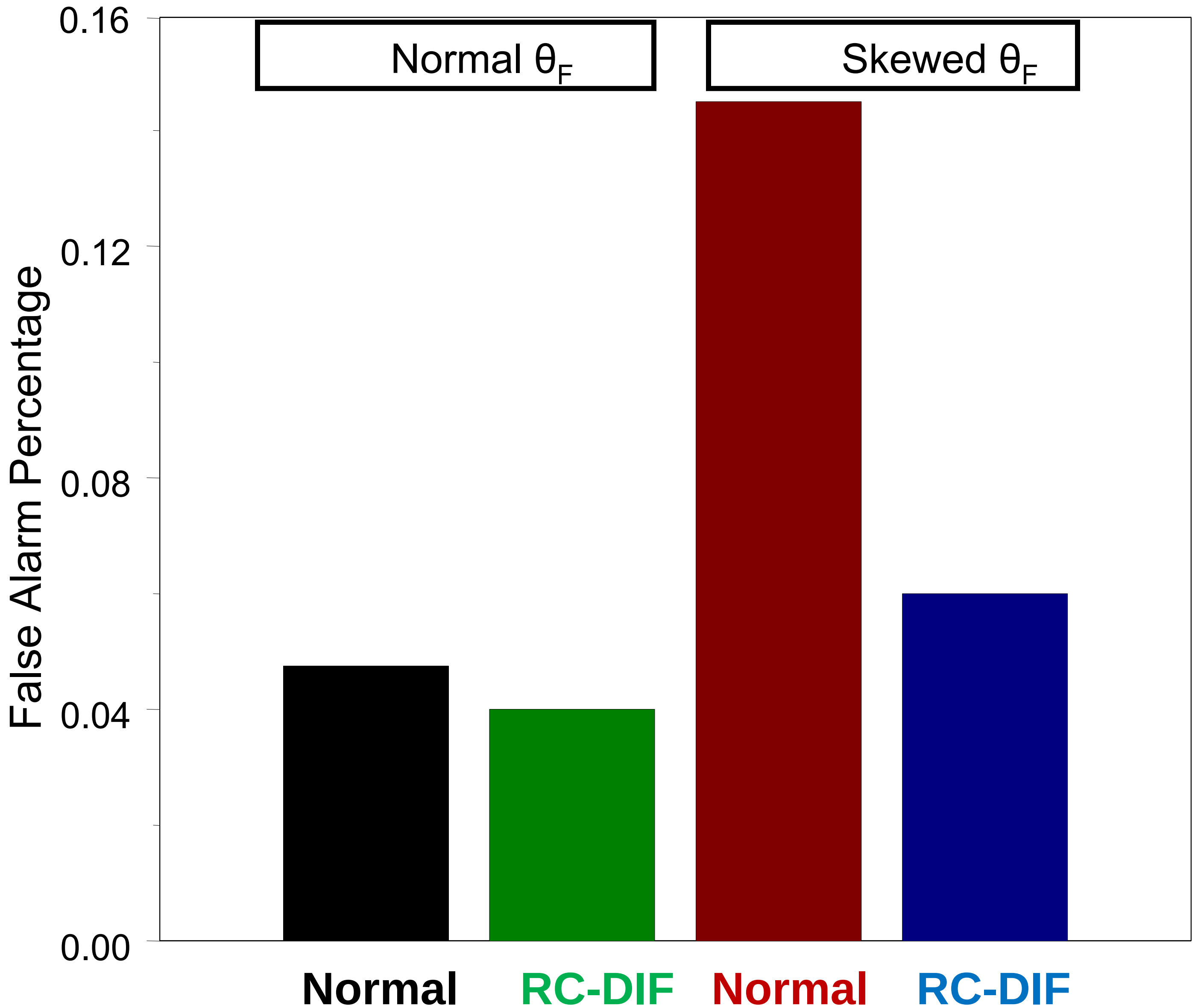
RMSE: difficulty



Estimated F-group mean



False alarms; α = .05



Percent out of 400 (4 DIF-free studied items x 100 reps)

Conclusions

RC-DIF-1 is a variation of IRT-LR-DIF that permits flexibility in the shape of $g_F(\theta)$.

When $g_F(\theta)$ is nonnormal and $g_R(\theta)$ is normal, RC-DIF-1 can improve (vs standard IRT-LR-DIF):

- F-group item parameters (thus DIF effects)
- estimates of $\bar{\theta}_F$
- false alarm rates for DIF tests

Future:

- ❑ Study performance of RC-DIF-1 more thoroughly
- ❑ Increase flexibility to permit nonnormality for both θ_F & θ_R