Effect of θ Estimation Method and Starting Value on the Recovery of θ

Rick Guyer

Assessment Systems Corporation

David J. Weiss

University of Minnesota

Purpose

- Guyer (2008) examined the effect of misfit on the recovery of θ in Computerized Adaptive Testing (CAT)
  - Observed that Weighted Maximum Likelihood (WML) was sensitive to initial item difficulty when misfit was present
  - Expected a Posteriori (EAP) estimation provided less biased θ estimates in the presence of misfit

Purpose

- **Present study**
  - Determine if sensitivity of WML to initial item difficulty replicates when no misfit is introduced
  - Vary starting θ value and generating θ
  - Compare alternatives to Maximum Likelihood (MLE) when the response pattern is not mixed

Method

- **Independent Variables:**
  - θ estimation method:
    - Expected a Posteriori (EAP)
    - Weighted maximum likelihood (WML)
    - Maximum likelihood (MLE) with
      - Arbitrary θs until the response pattern is mixed
        - » Starting θ value incremented by 1
      - WML estimation until the response pattern is mixed
      - EAP estimation until the response pattern is mixed
  - Generating θ
    - −2, −1, 0, 1, 2

Method

- **Independent Variables:**
  - Starting θ estimate
    - $-2, -1, 0, 1, 2$
- **Dependent Variables**

Bias

$$\frac{\sum_{i=1}^{N} \hat{\theta}_i - \theta}{N}$$

Empirical SE

$$\sqrt{\frac{\sum_{i=1}^{N} \hat{\theta}_i - \bar{\hat{\theta}}^2}{N}}$$

RMSE

$$\sqrt{\frac{\sum_{i=1}^{N} \hat{\theta}_i - \theta^2}{N}}$$

Procedure

- Monte Carlo Design
  - Item parameters for the 300 items generated using the following distributions:

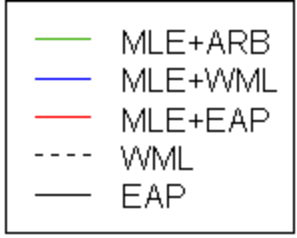
    a ~ log-normal(–.223, 0.2), b ~ U[–3.5, 3.5], c ~ N(.20, .02)

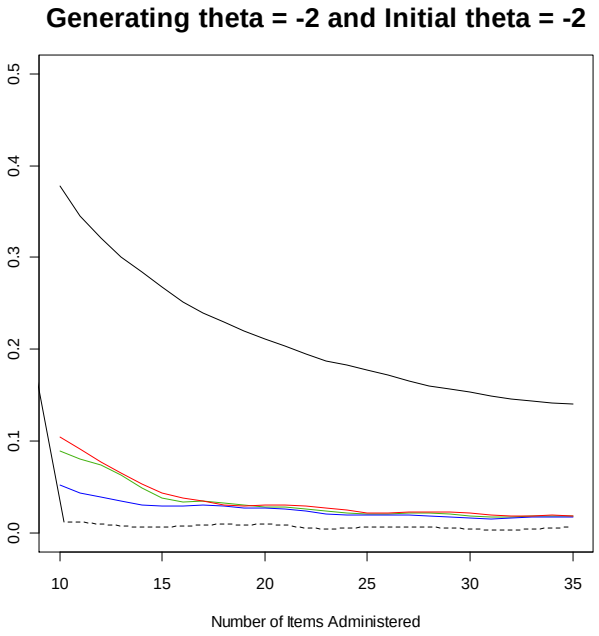
    In the logistic metric, the mean of a was 0.82 with an SD = .15
  - 1,000 replications performed for each cell
  - Item responses were generated according to the 3PL model

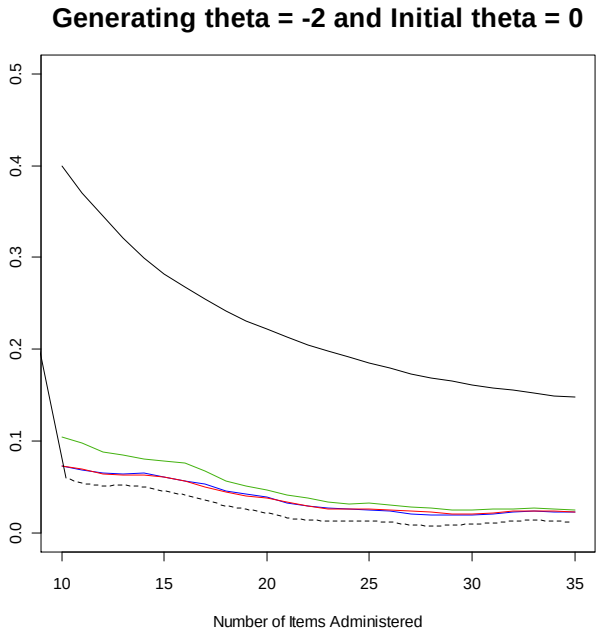
Procedure

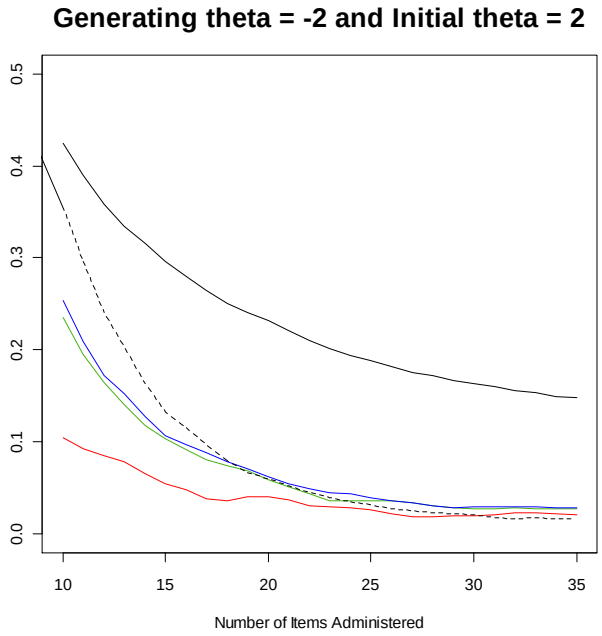
- CAT
  - The program R was used for the CAT simulation in this study
  - Maximum information item selection was used for all conditions in this study
  - Dependent variables were calculated after 10-35 items were administered in the CAT

Average Bias

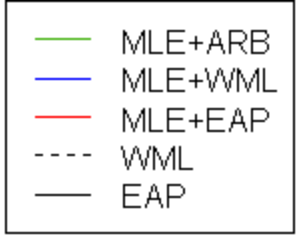




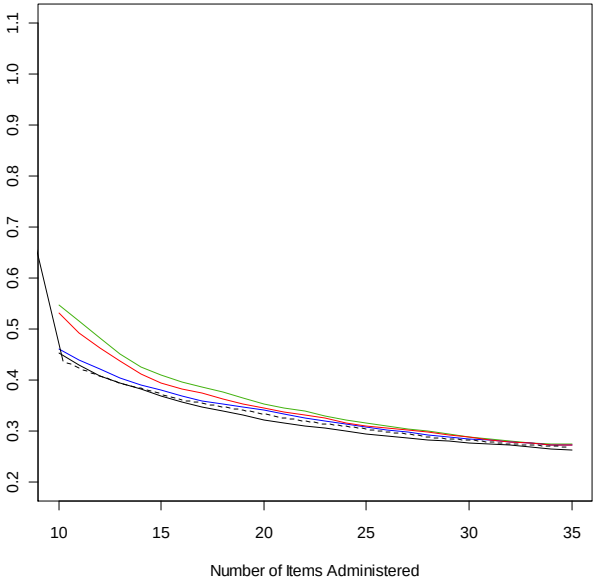




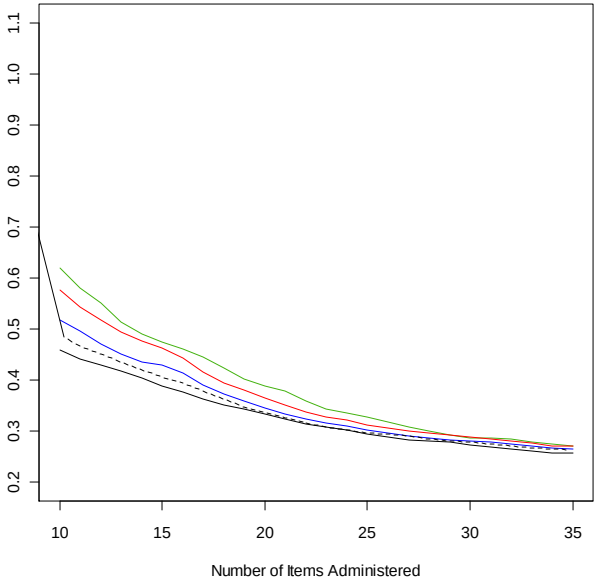
Empirical SE



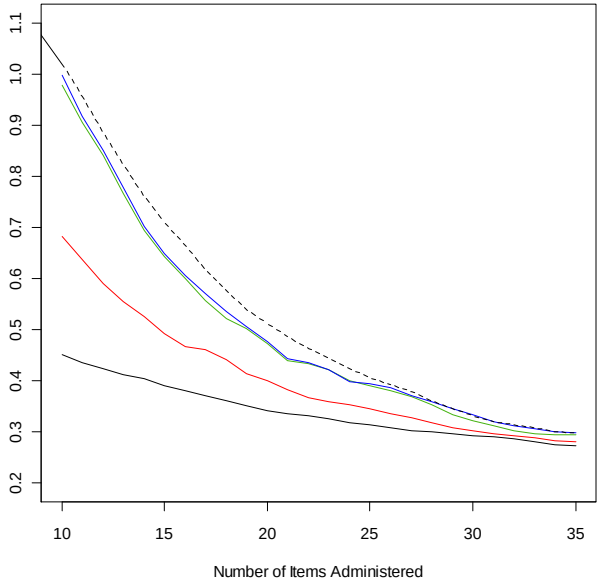
Generating theta = -2 and Initial theta = -2



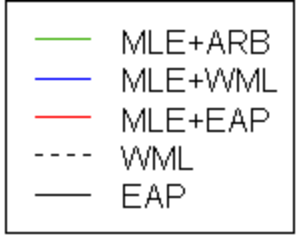
Generating theta = -2 and Initial theta = 0



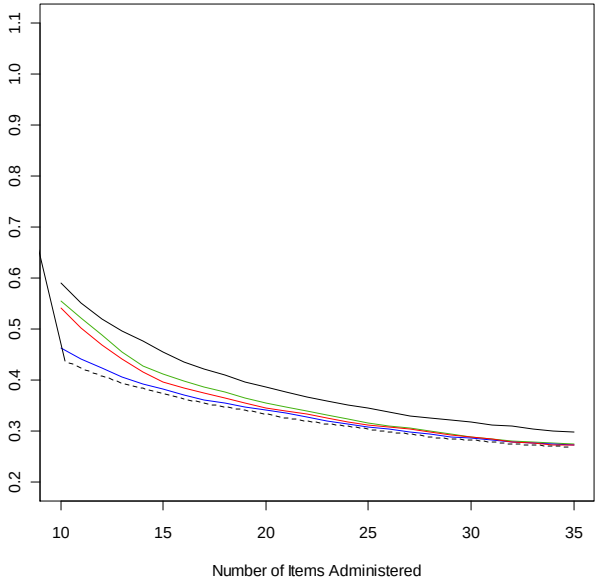
Generating theta = -2 and Initial theta = 2



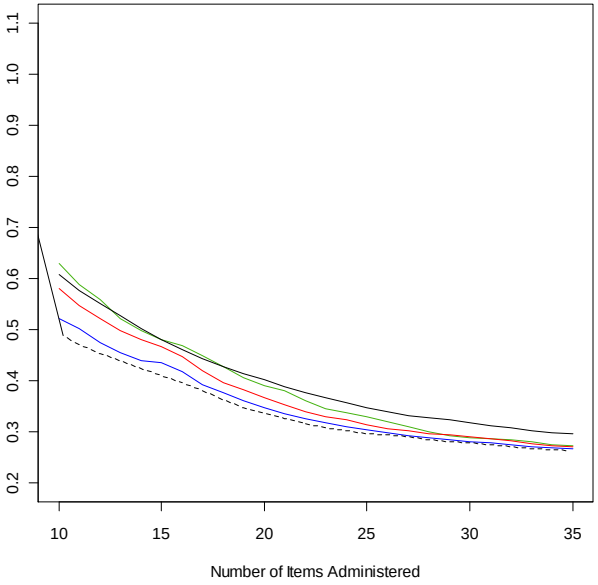
RMSE



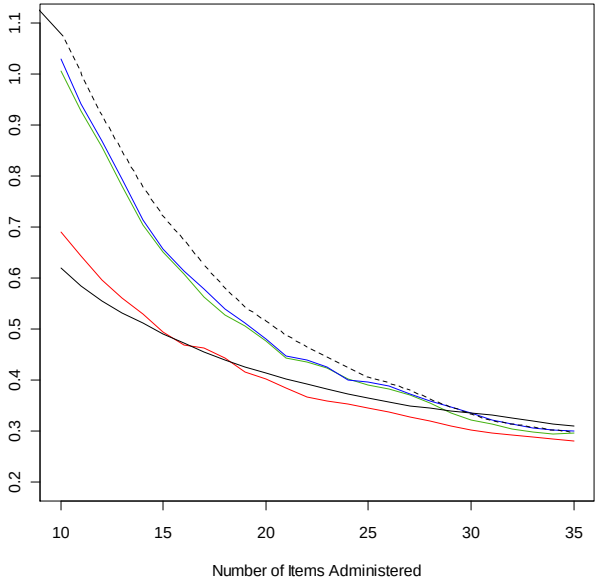
Generating theta = -2 and Initial theta = -2



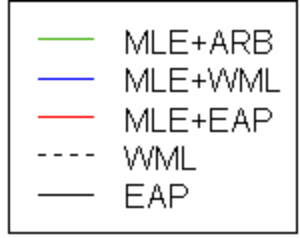
Generating theta = -2 and Initial theta = 0



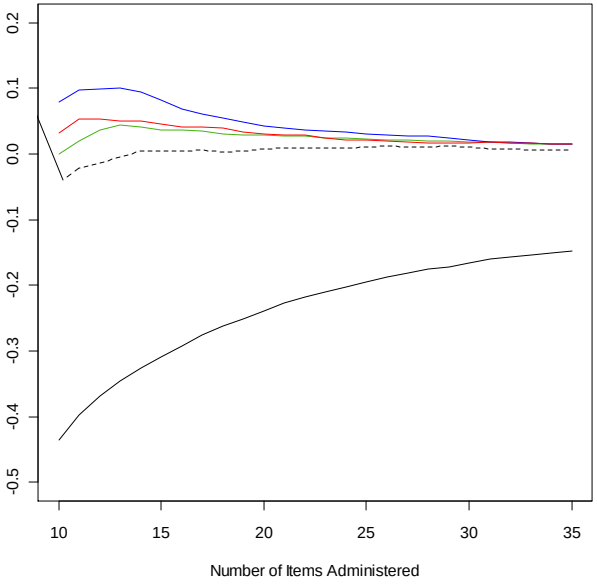
Generating theta = -2 and Initial theta = 2



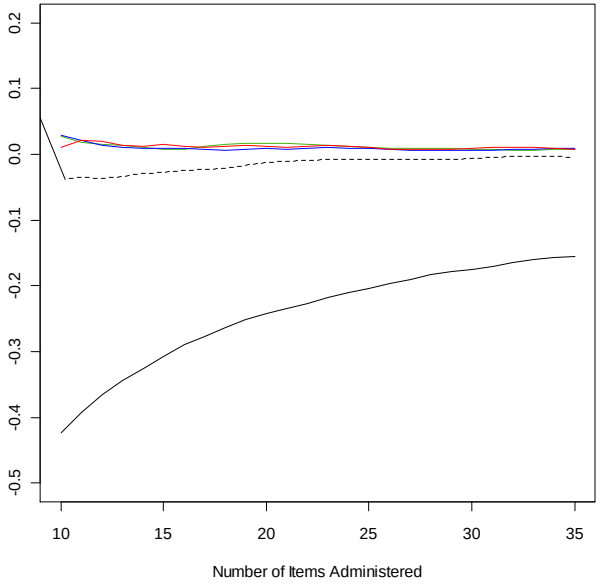
Average Bias



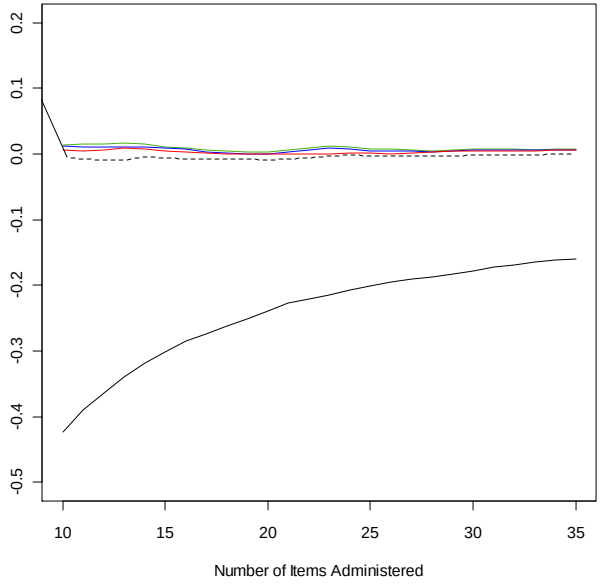
Generating theta = 2 and Initial theta = -2



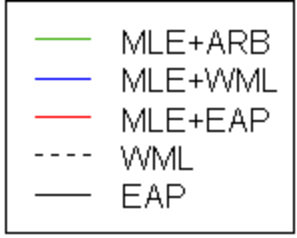
Generating theta = 2 and Initial theta = 0



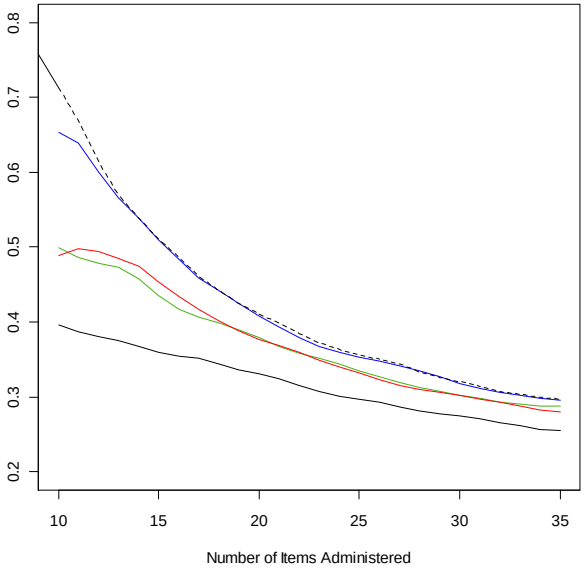
Generating theta = 2 and Initial theta = 2



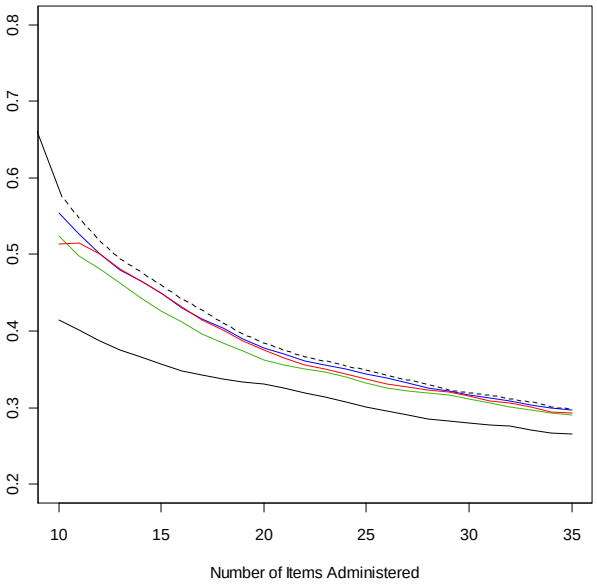
Empirical SE



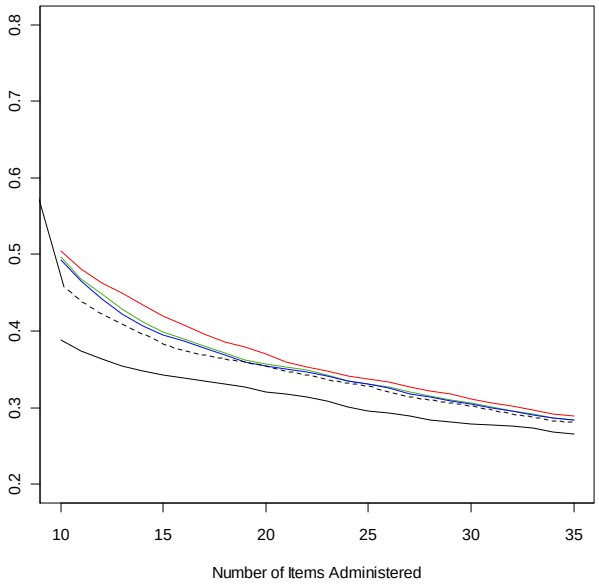
Generating theta = 2 and Initial theta = -2



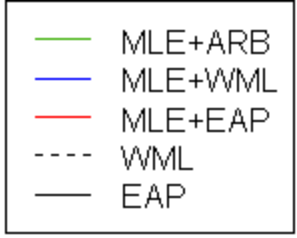
Generating theta = 2 and Initial theta = 0



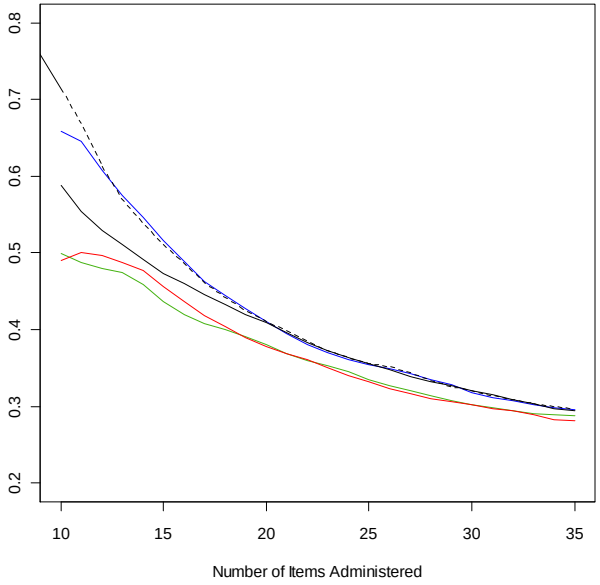
Generating theta = 2 and Initial theta = 2



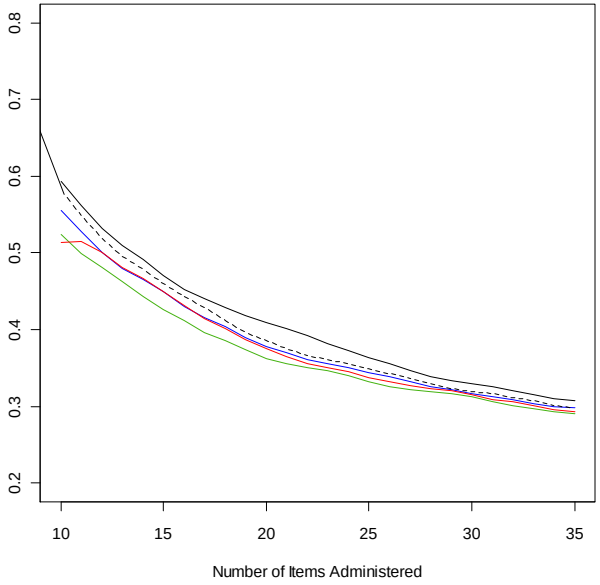
RMSE



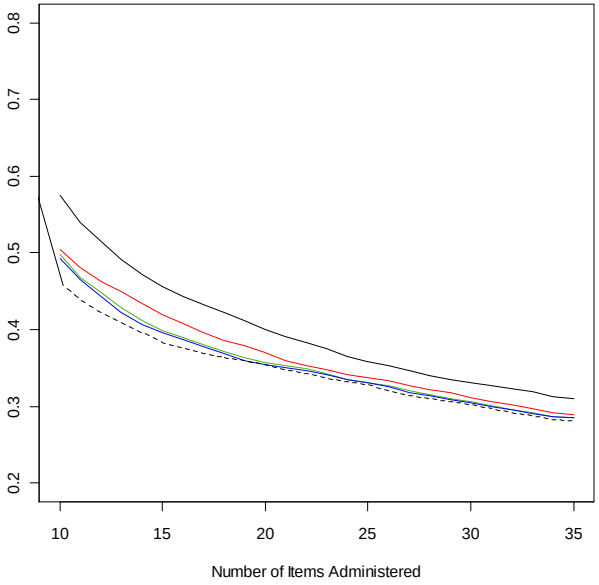
Generating theta = 2 and Initial theta = -2



Generating theta = 2 and Initial theta = 0



Generating theta = 2 and Initial theta = 2



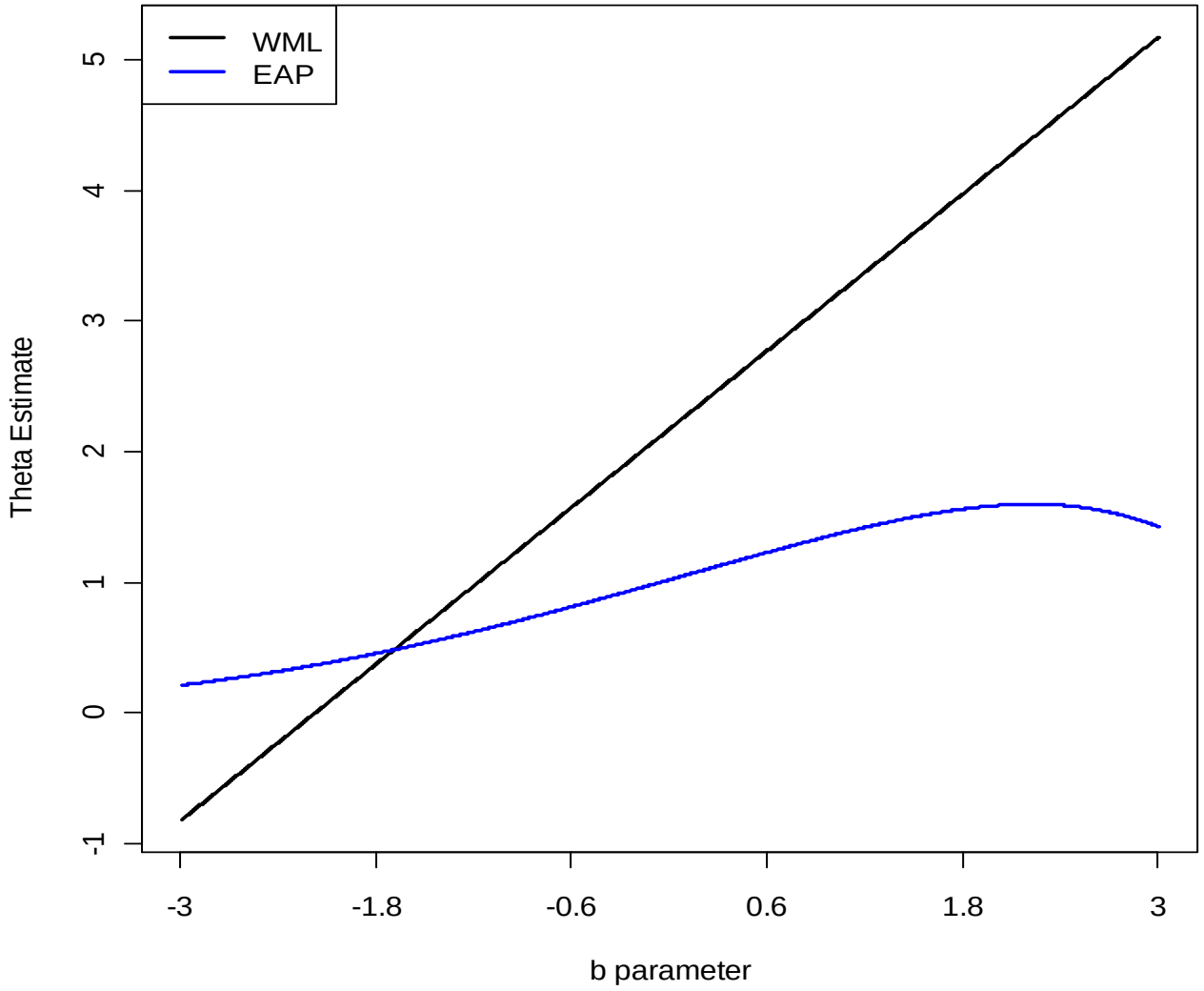
Results

- **Recovery of θ After 15 Items**
  - WML and MLE+WML had lowest bias and RMSE when generating θ and initial θ were equal
    - Bias/RMSE increased as difference between generating θ and initial θ increased
    - WML showed sensitivity to initial item difficulty
  - EAP had the largest bias but the lowest SEs – though bias decreased as generating θ approached the prior mean of 0

Results

- **Recovery of θ After 15 Items**
  - MLE+EAP had similar results as MLE+ARB except when generating $\theta = -2$ and starting $\theta = 2$
    - MLE+EAP had second lowest SE/RMSE when generating θ and initial θ differed by 3+ SD
- **Recovery of θ After 35 Items**
  - Bias of EAP remained far greater than the other four θ conditions when generating $\theta \neq 0$
  - EAP consistently had the lowest SEs
  - RMSEs of the five methods converged

Theta Estimates as a function of *b* for response pattern (1,1,1,1,1) with *a*D = 1 and *c* = 0.2



Conclusions

- **Alternatives to MLE for non-mixed response patterns**
  - EAP recommended when the response pattern is not mixed
  - WML not recommended due to its sensitivity to initial item difficulty
    - Especially problematic for high ability examinees
  - Arbitrary values also not recommended
- **WML is sensitive to initial item difficulty – early in the CAT**