Supporting Instructional Decisions with Group-level, Standard-Specific Inferences

S. Thomas Christie, Carson Cook, Garron Gianopulos

Current Dilemma

3x / year

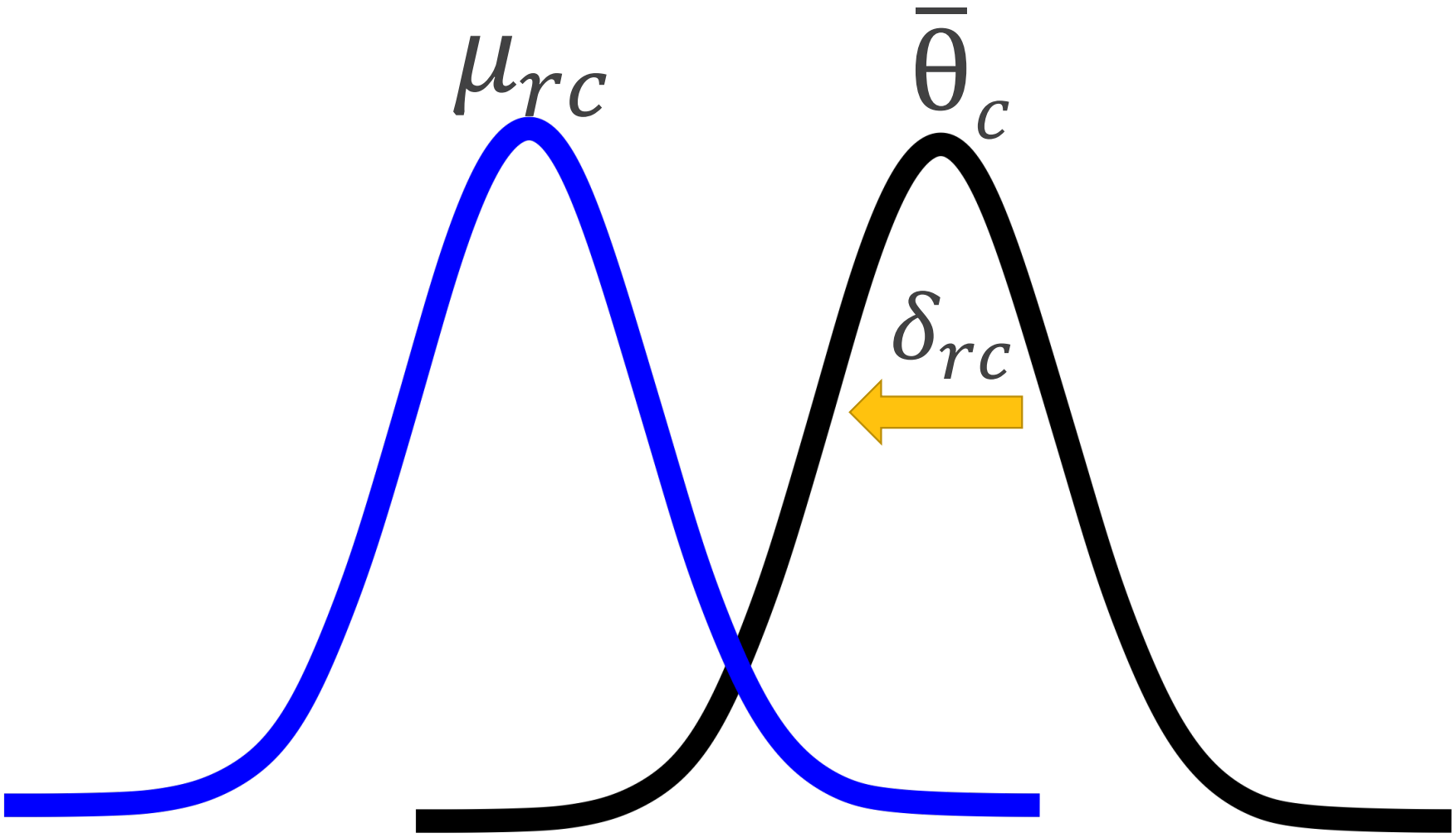
How to make classroom decisions with scores?

Granular sub-scores?

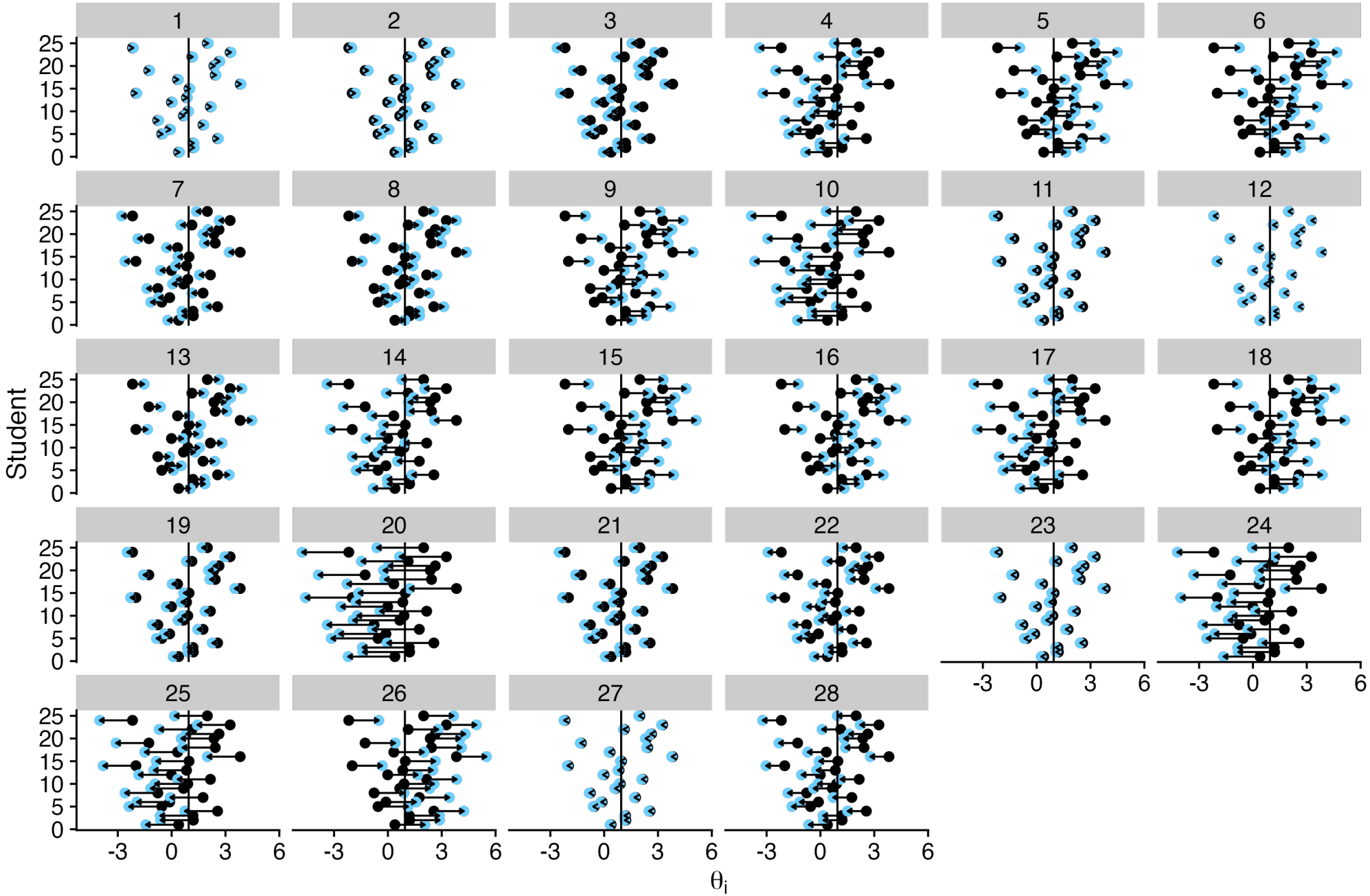
- **Sub-scores:**
  - 8+ items to produce a sub-score with sufficient reliability
  - Are often still not very granular, e.g. ‘Literature’ or ‘Geometry’
  - Are often highly correlated with overall score
- **However…**
  - All else being equal, greater granularity is more useful
  - Instructional decisions are often made at the level of a **class**

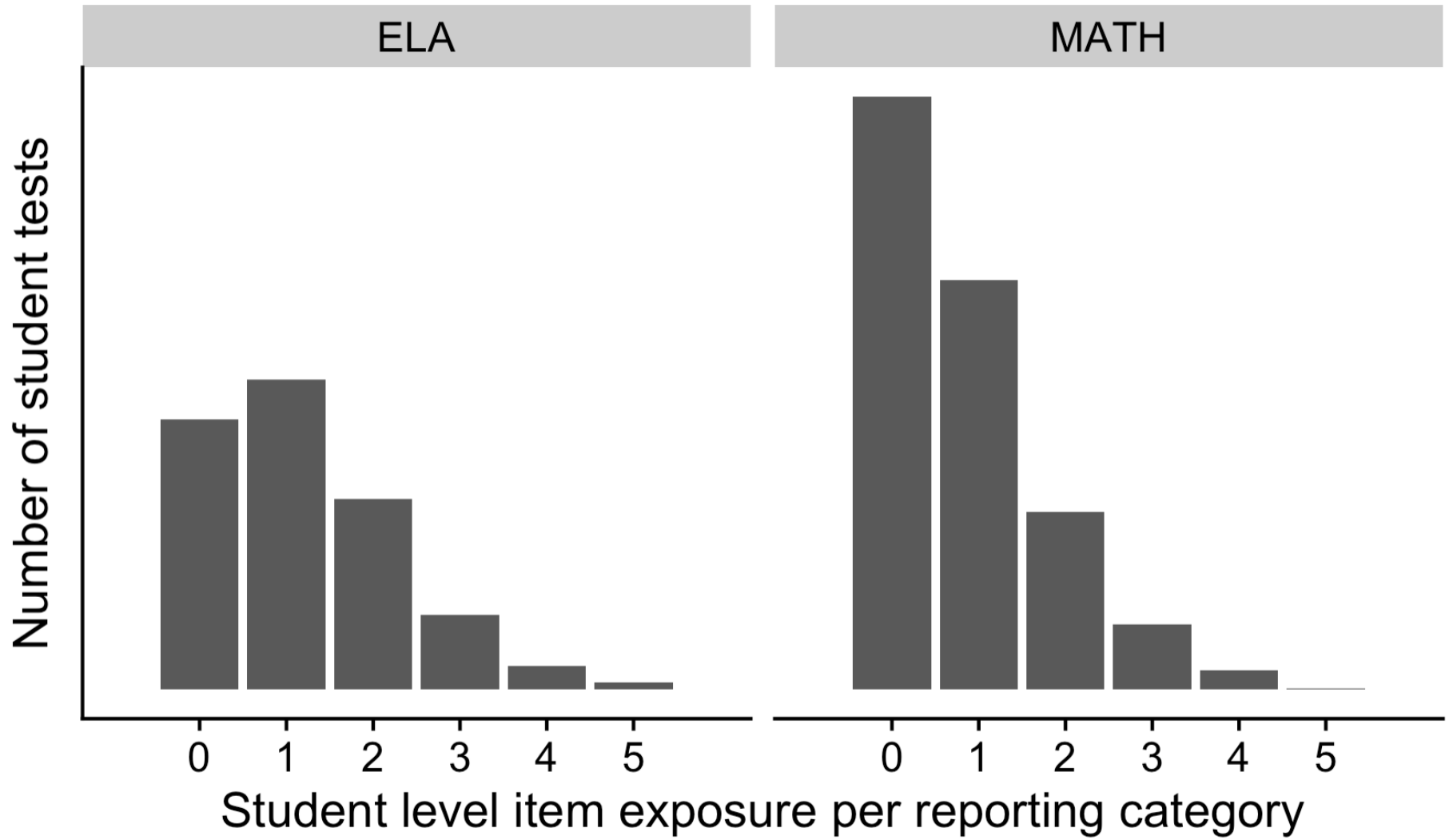
Research Questions

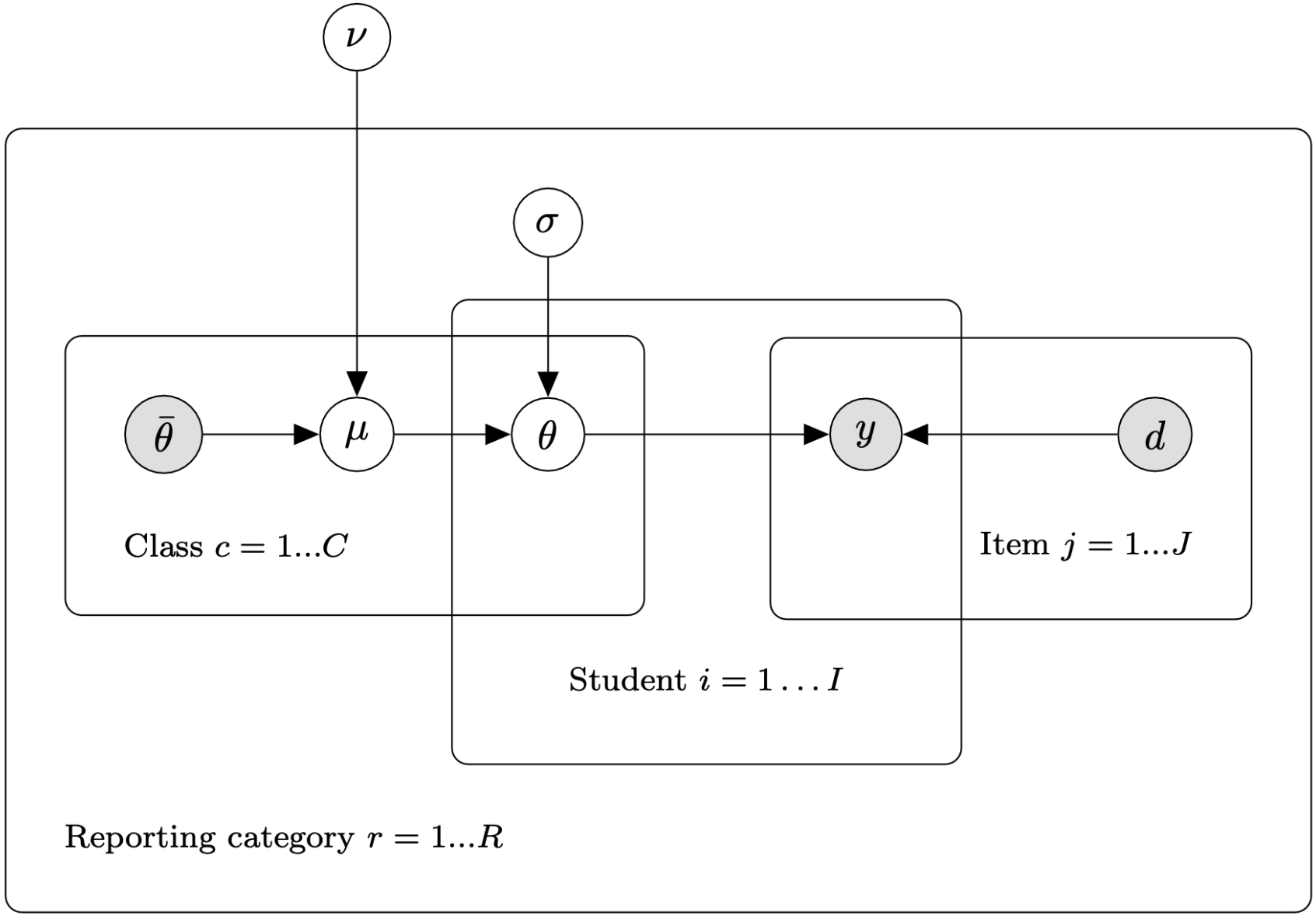
- Can we produce valid group-level sub-scores?
- What are the requirements in terms of:
  - Class size?
  - # of items per reporting category?
  - Effect size?
- Can we detect class-level effects in historical tests?

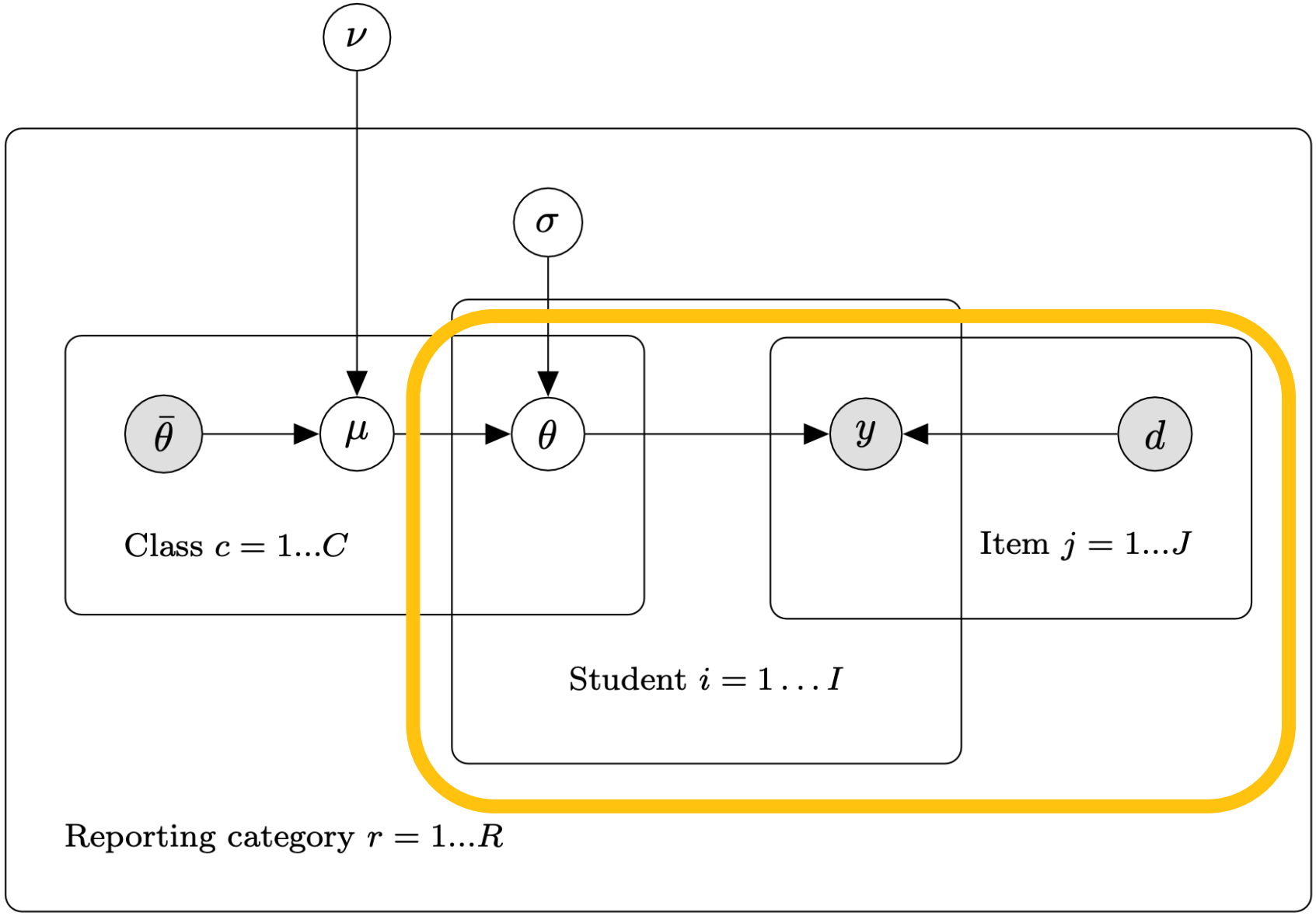

μ_{rc}
$\bar{\theta}_c$
δ_{rc}
Distribution of sub-scores
Distribution of overall scores

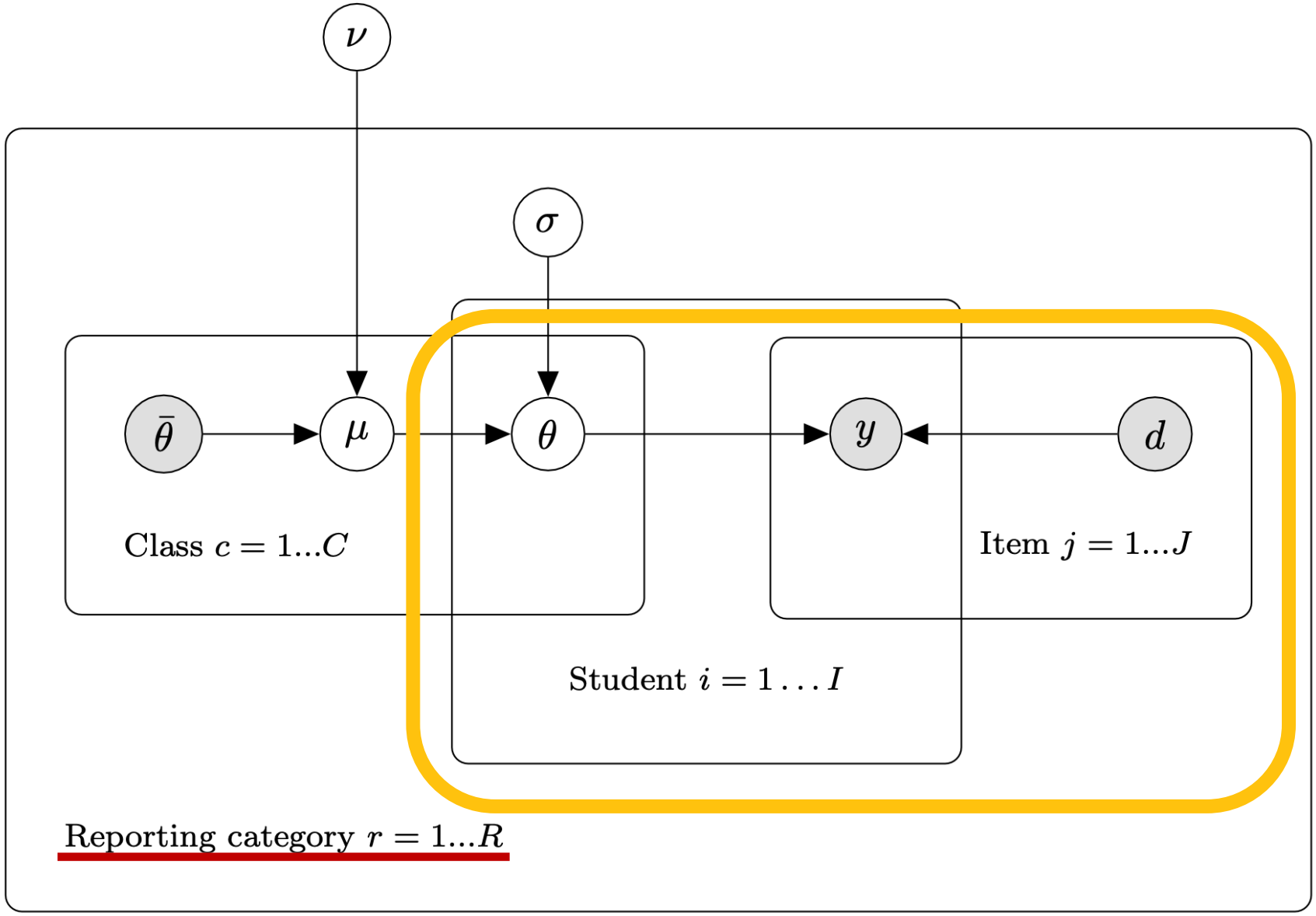
Simulated effects

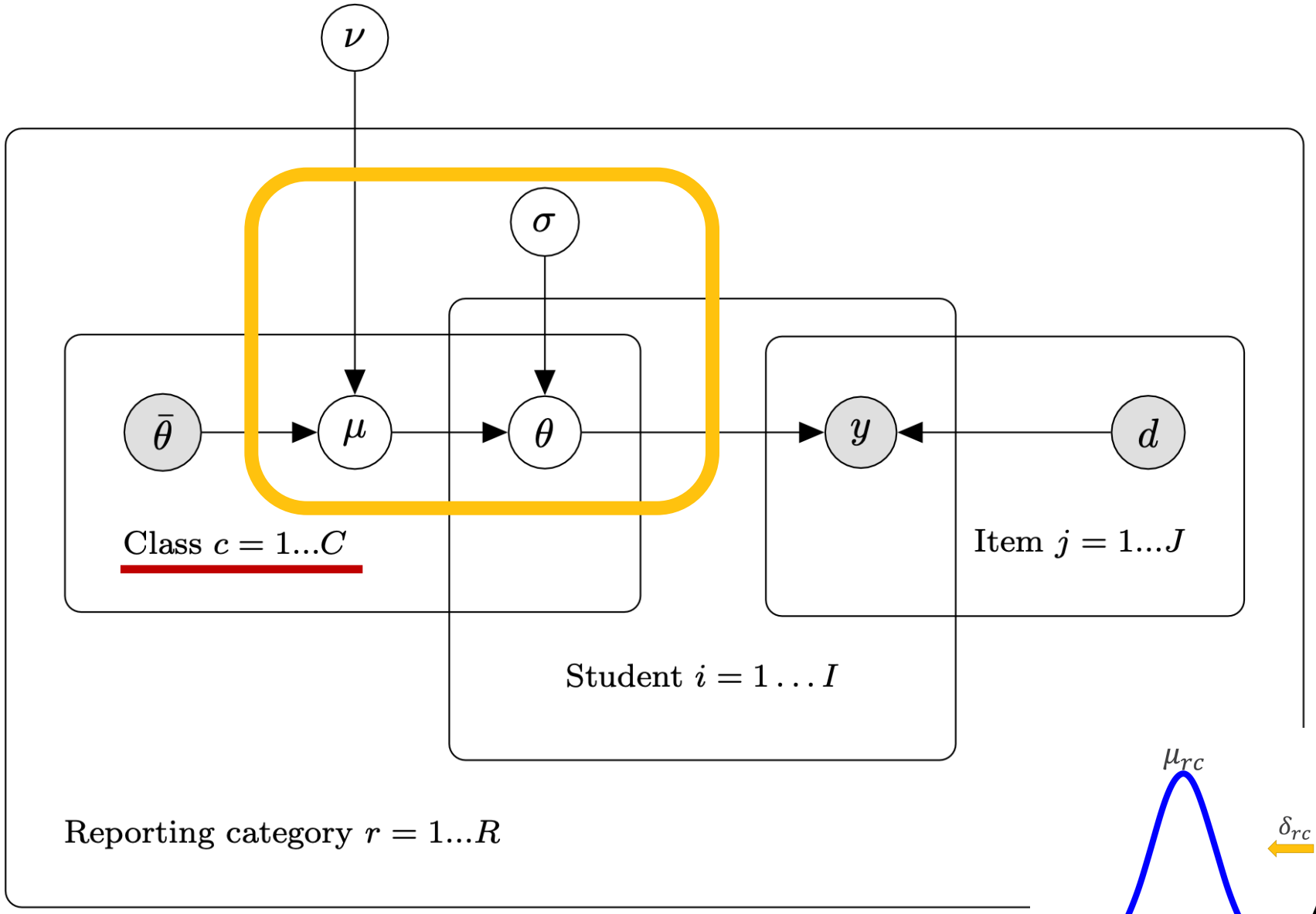


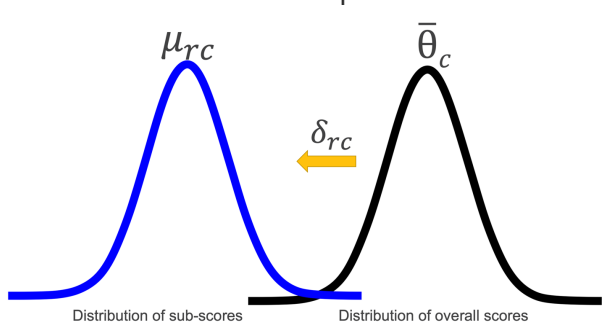

ELA
MATH
Number of student tests
0
1
2
3
4
5
0
1
2
3
4
5
Student level item exposure per reporting category

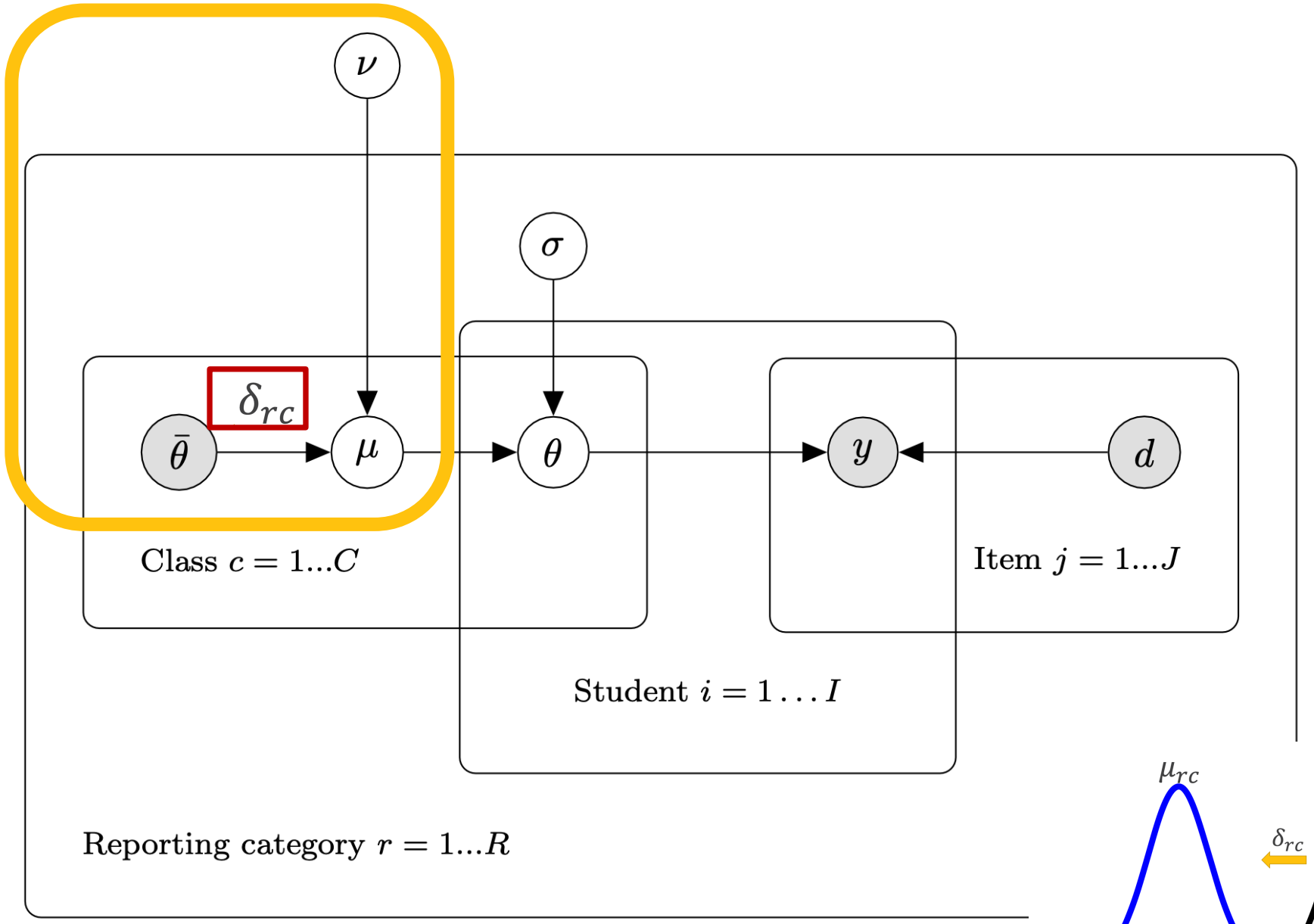

ν
σ
$\bar{\theta}$
μ
θ
y
d
Class $c = 1...C$
Item $j = 1...J$
Student $i = 1 \ldots I$
Reporting category $r = 1...R$

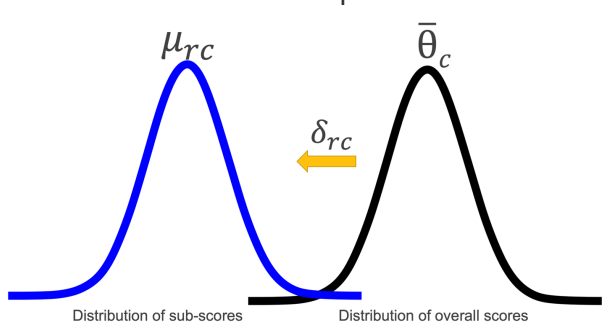

ν
σ
$\bar{\theta}$
μ
θ
y
d
Class $c = 1...C$
Item $j = 1...J$
Student $i = 1 \ldots I$
Reporting category $r = 1...R$


ν
σ
$\bar{\theta}$
μ
θ
y
d
Class $c = 1...C$
Item $j = 1...J$
Student $i = 1 \ldots I$
Reporting category $r = 1...R$


ν
σ
$\bar{\theta}$
μ
θ
y
d
Class $c = 1...C$
Item $j = 1...J$
Student $i = 1 \ldots I$
Reporting category $r = 1...R$

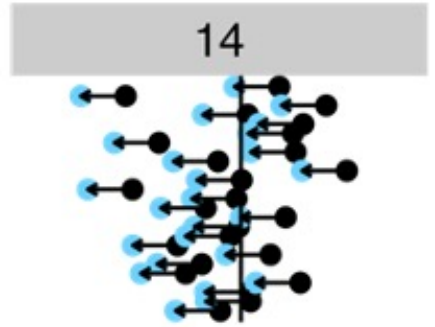

μ_{rc}
$\bar{\theta}_c$
δ_{rc}
Distribution of sub-scores
Distribution of overall scores


ν
σ
δ_{rc}
$\bar{\theta}$
μ
θ
y
d
Class $c = 1...C$
Item $j = 1...J$
Student $i = 1 \ldots I$
Reporting category $r = 1...R$


μ_{rc}
$\bar{\theta}_c$
δ_{rc}
Distribution of sub-scores
Distribution of overall scores

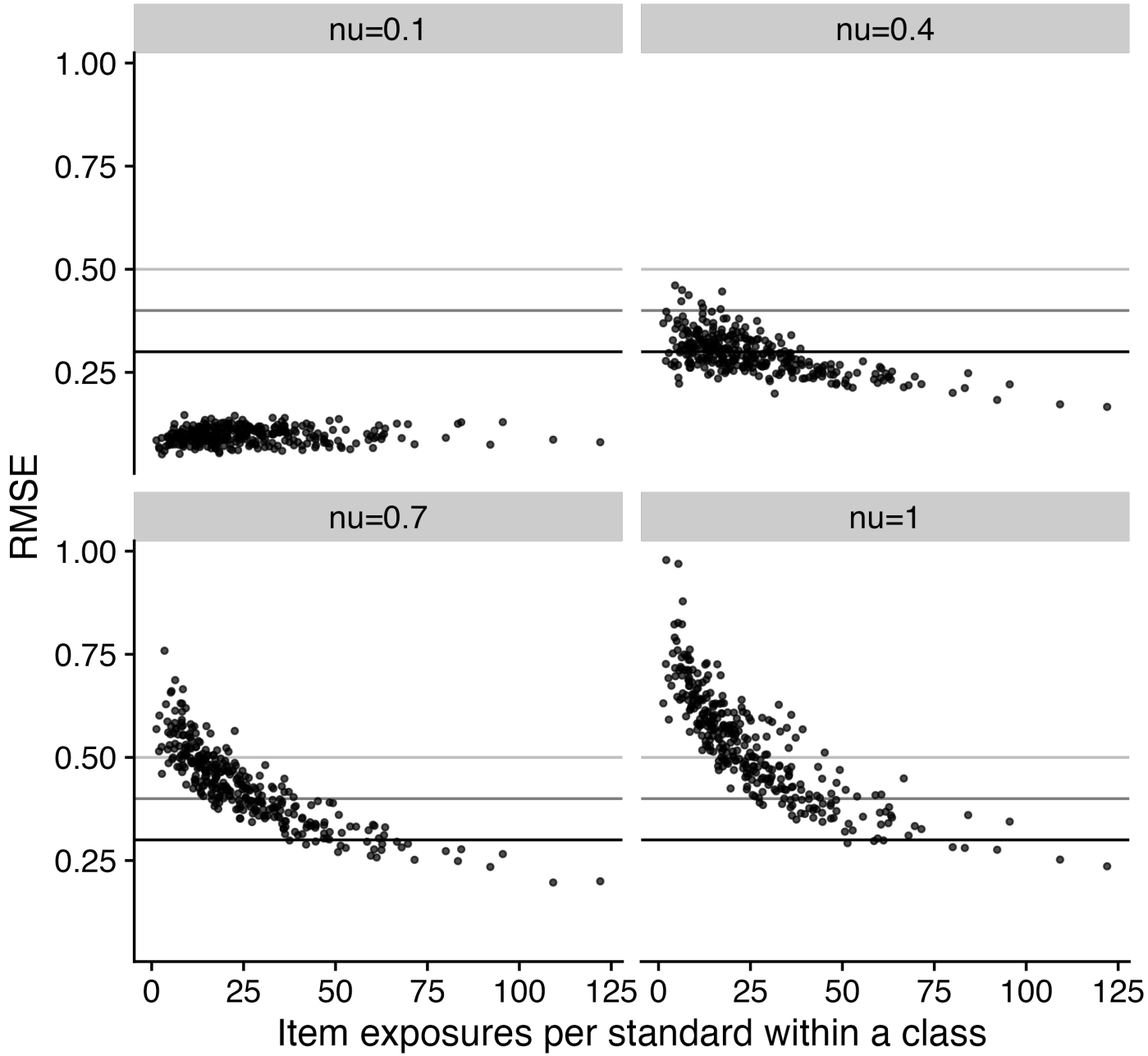
Two Studies

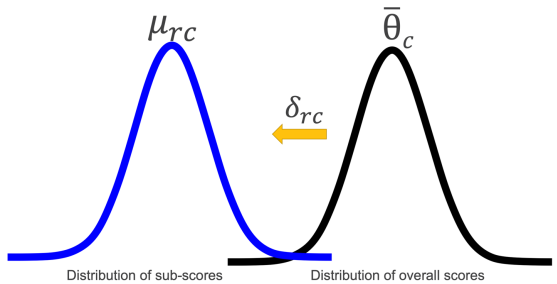
1. Simulate group-level effects of various sizes
   - $\boldsymbol{\nu}$ = .1, .4, .7, 1
   - simulate new responses to historical tests
   - attempt parameter recovery



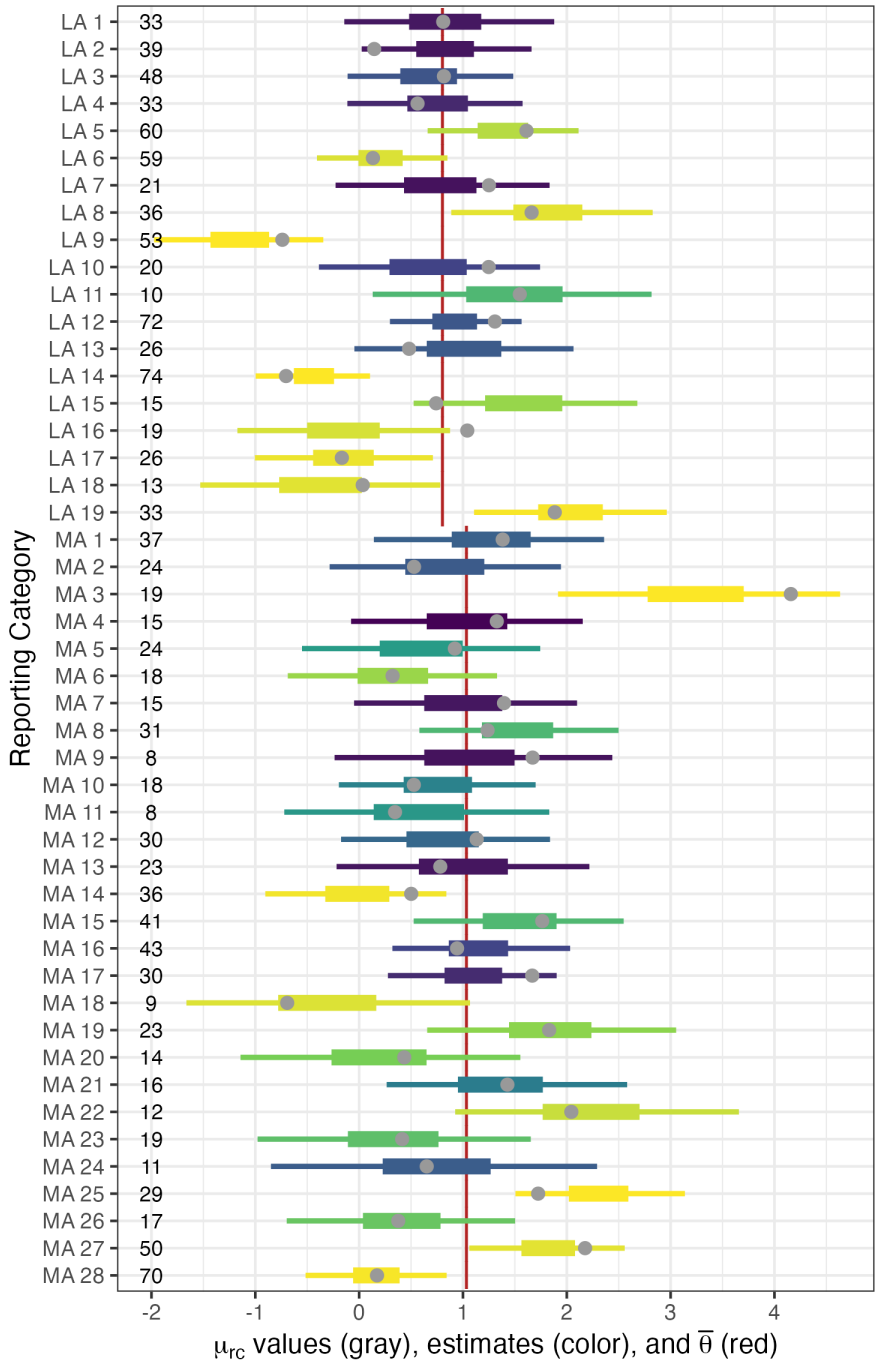
2. Estimate $\boldsymbol{\nu}$ from historical data
   - what effect sizes are present?
   - can we recover parameters?

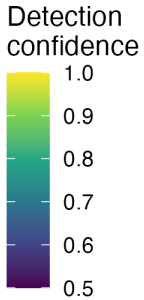
Study 1: RMSE for μ_{rc} estimates

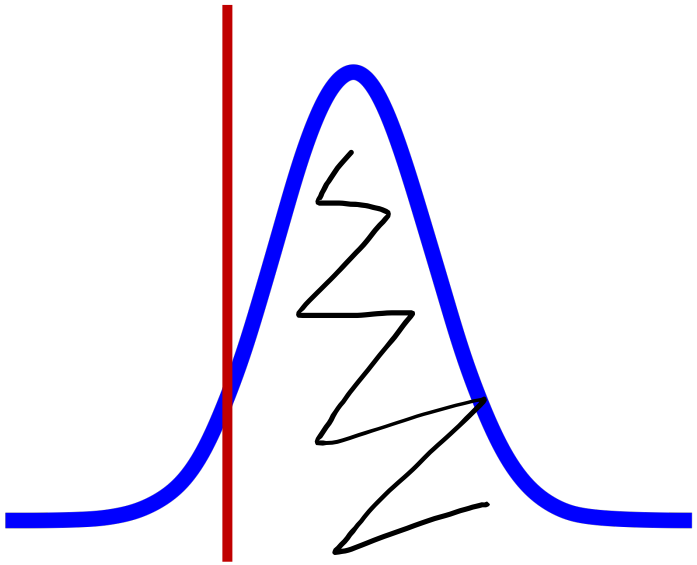




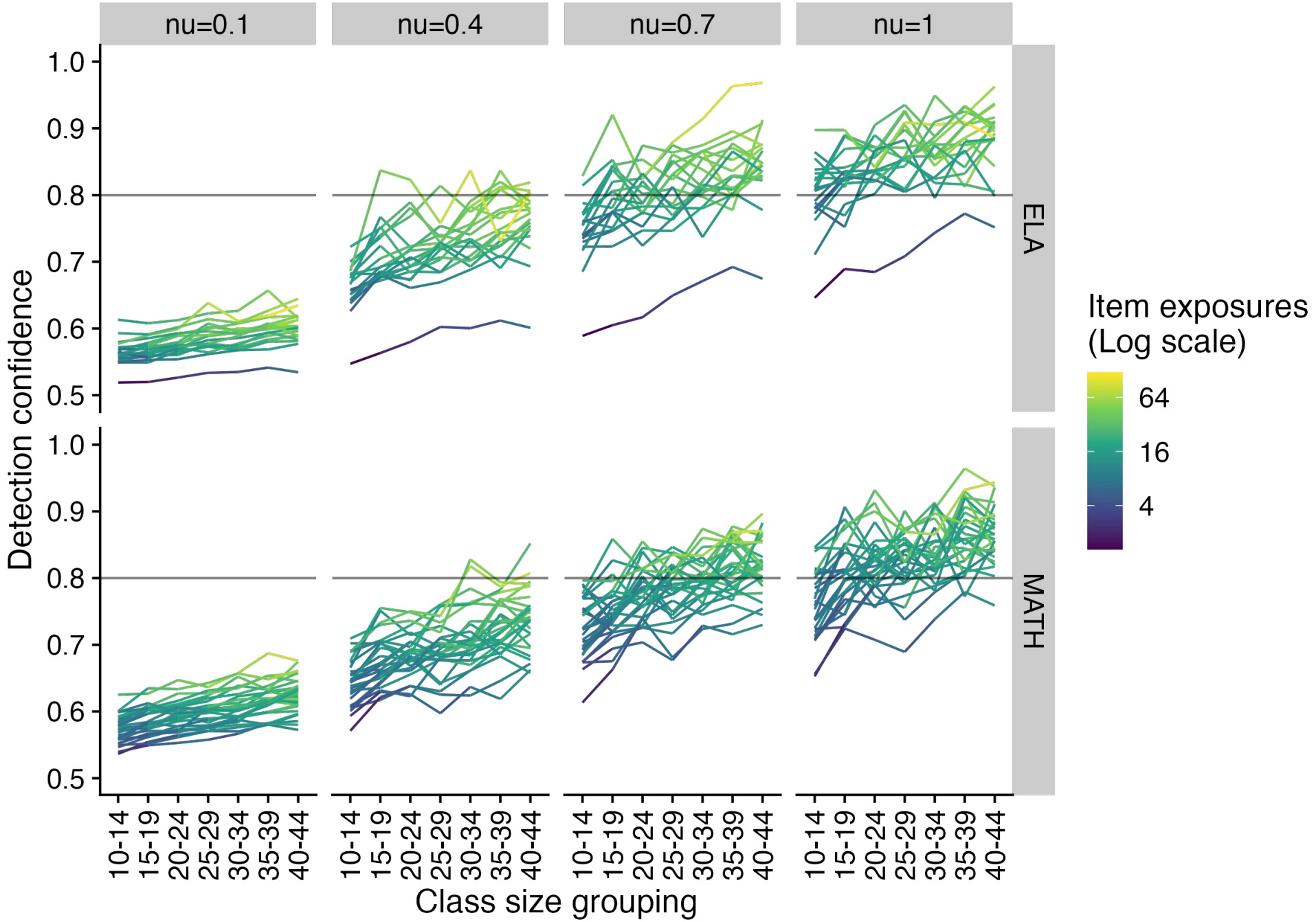
Study 1: Estimates of μ_{rc} for a simulated class of 30 students



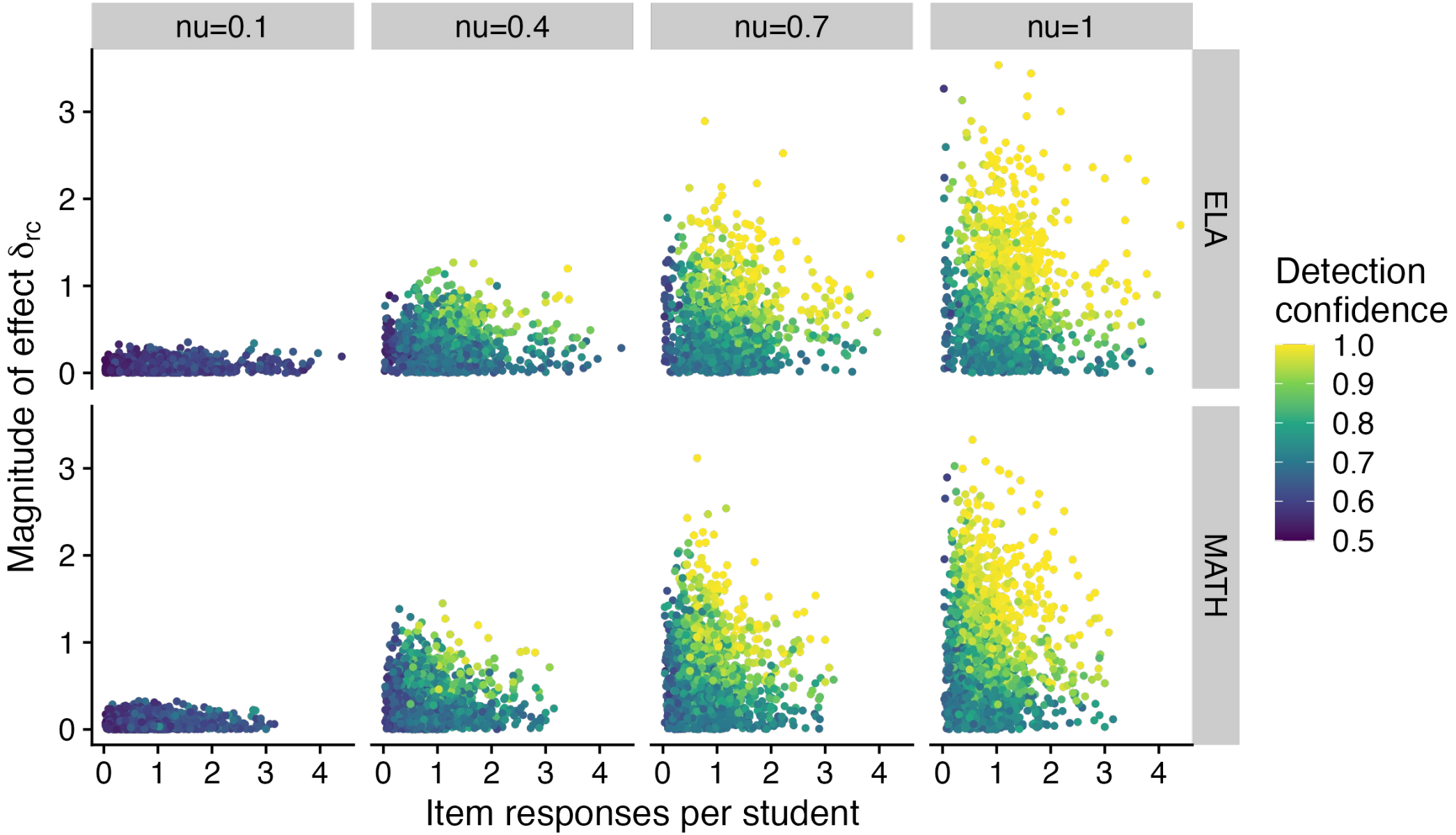




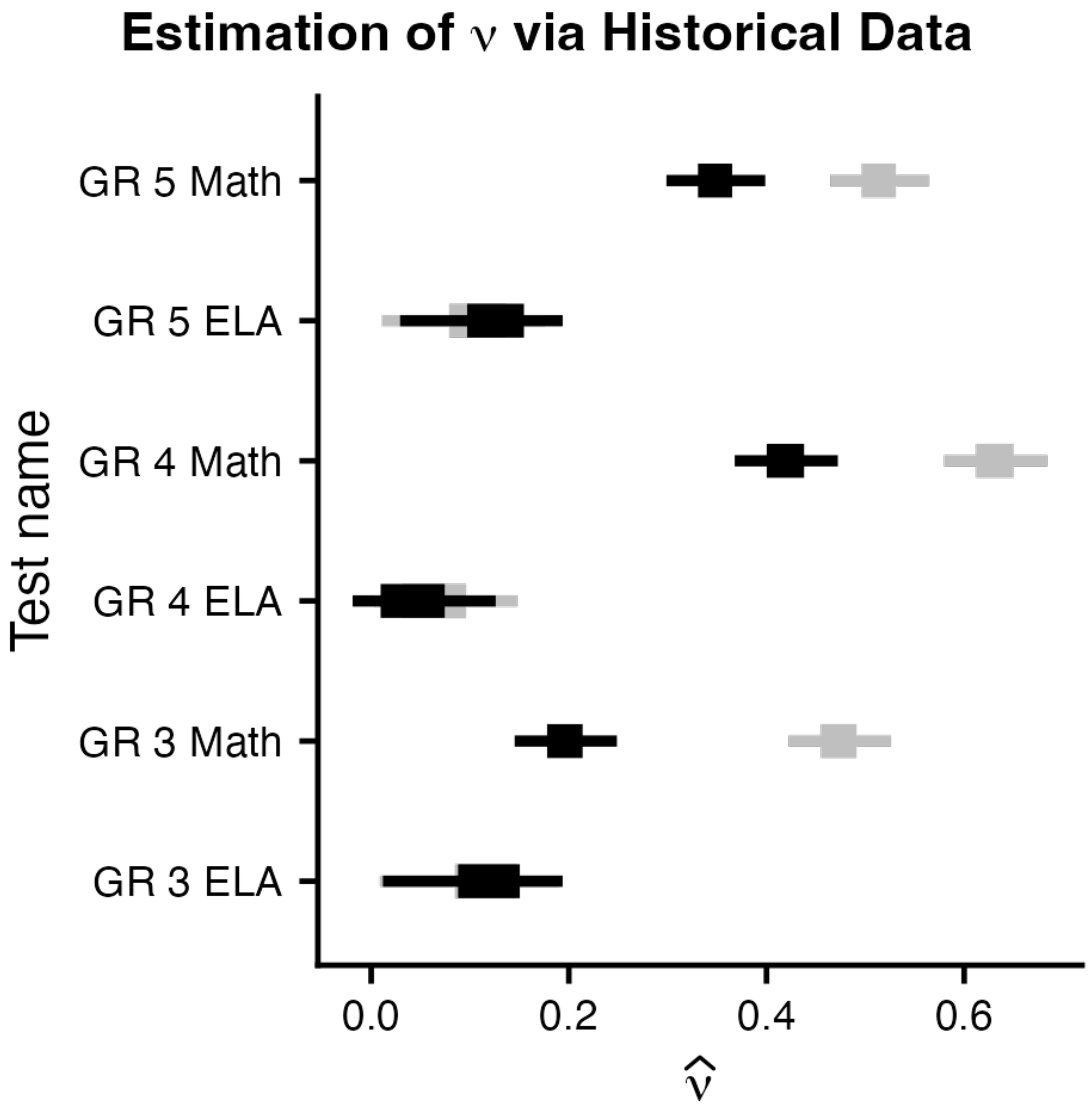
Study 1: Detecting an effect for a given RC

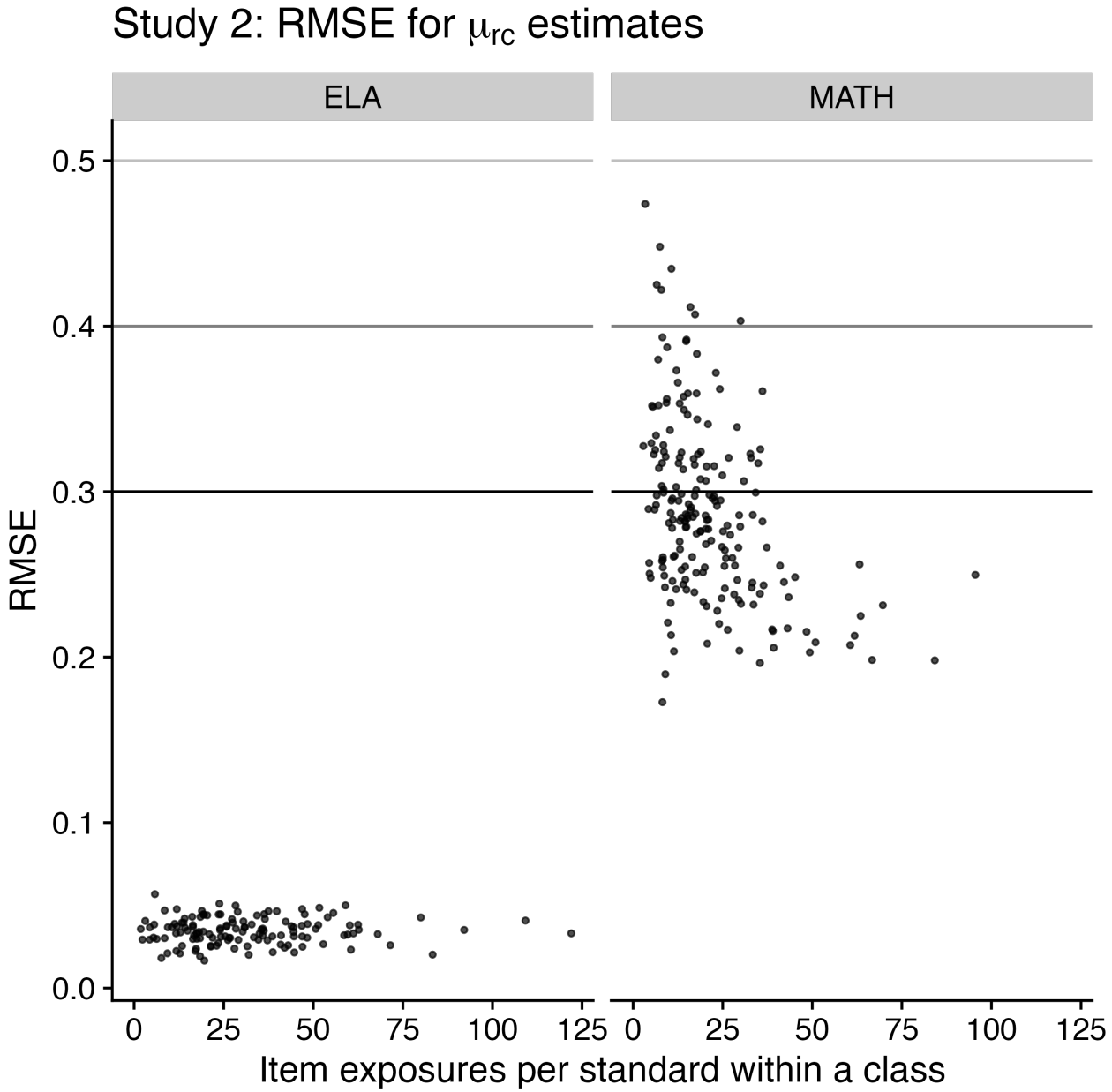


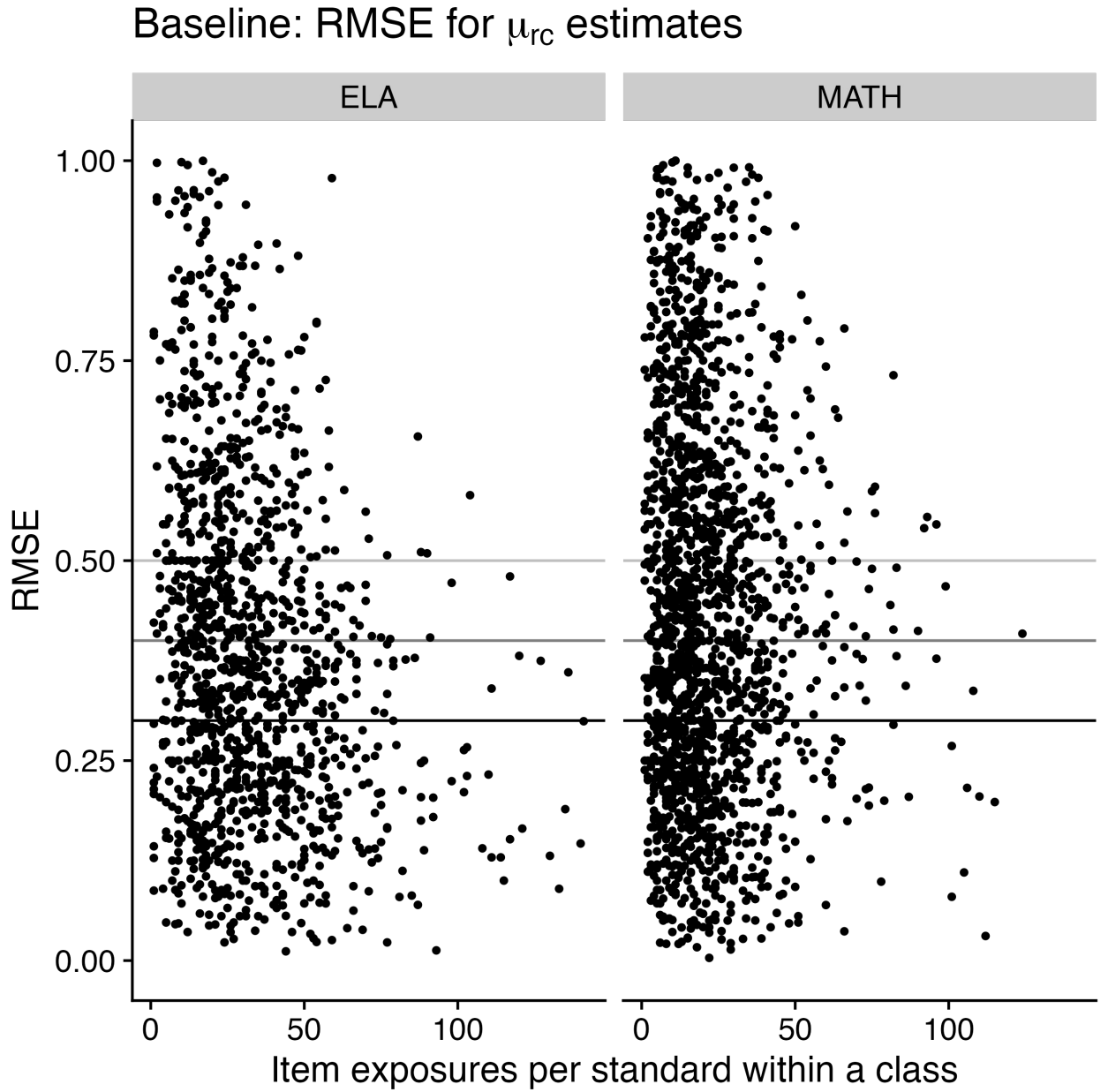
Study 1: Detecting individual effects



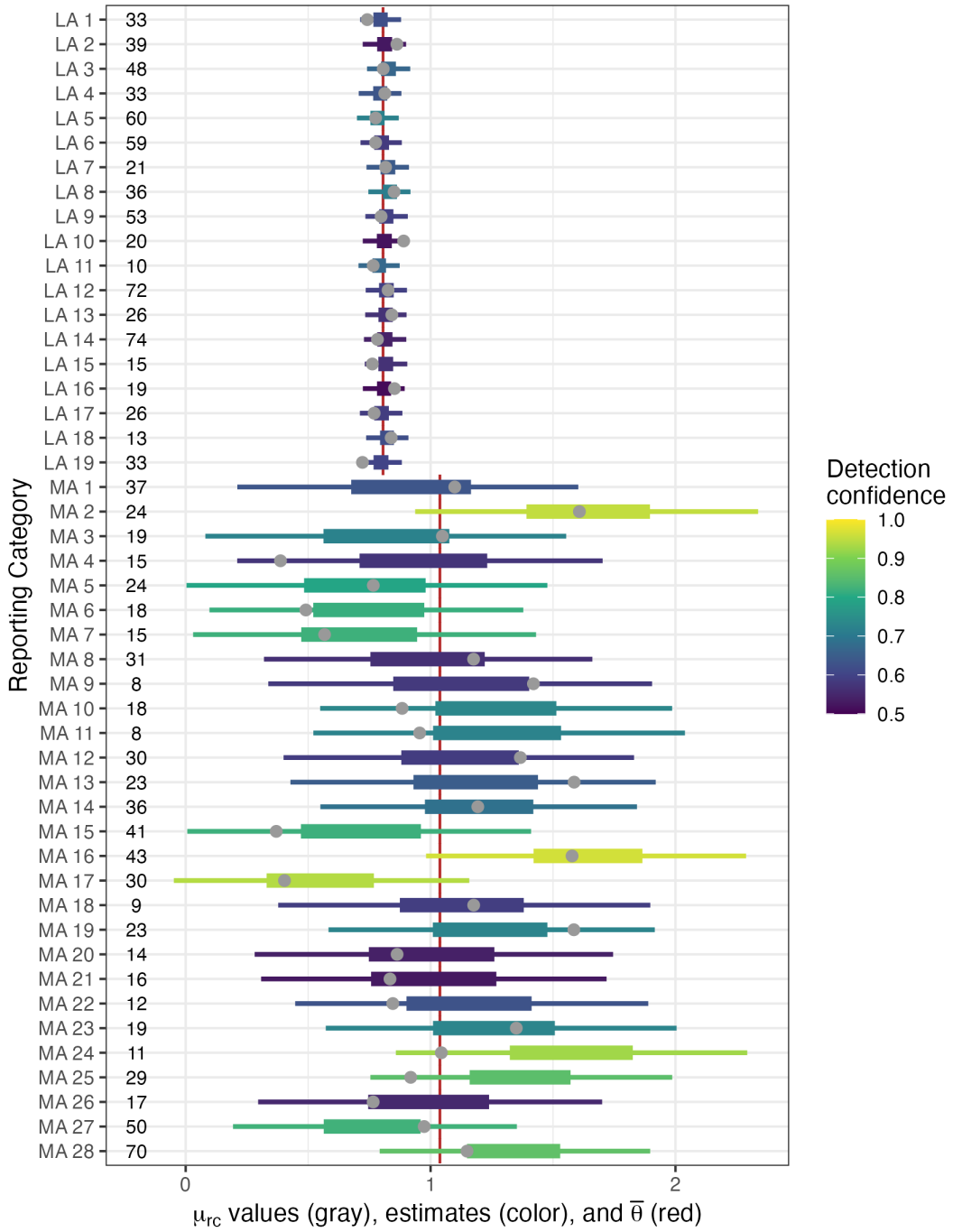
Study 2


Estimation of ν via Historical Data
Test name
GR 5 Math
GR 5 ELA
GR 4 Math
GR 4 ELA
GR 3 Math
GR 3 ELA
0.0
0.2
0.4
0.6
$\hat{\nu}$

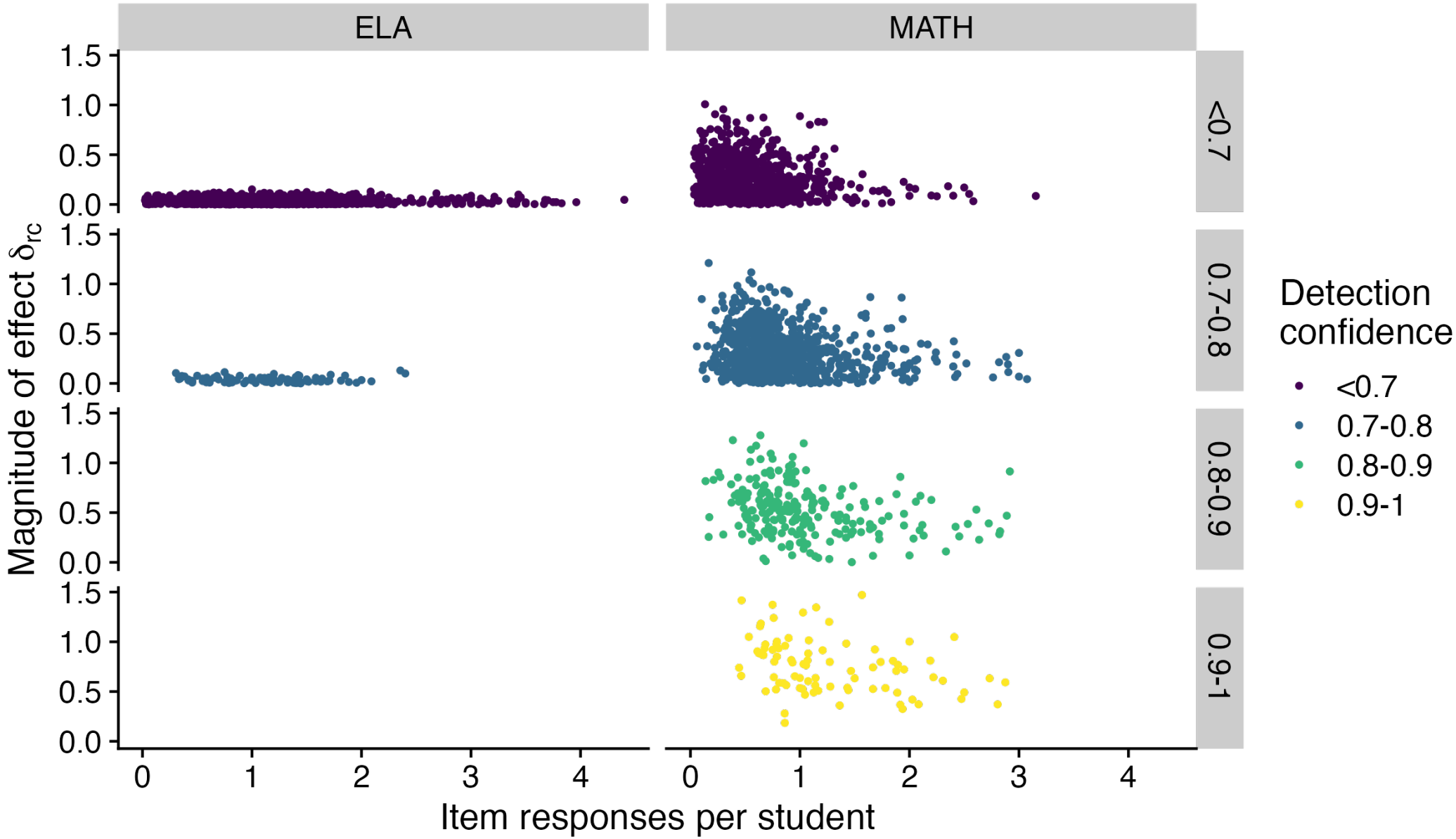

Study 2: RMSE for μ_{rc} estimates
ELA
MATH
RMSE
0.0
0.1
0.2
0.3
0.4
0.5
0
25
50
75
100
125
Item exposures per standard within a class


Baseline: RMSE for μ_{rc} estimates
ELA
MATH
RMSE
0.00
0.25
0.50
0.75
1.00
0
50
100
Item exposures per standard within a class

Study 2: Estimates of μ_{rc} for a simulated class of 30 students



Study 2: Detecting individual effects



Conclusions

- Class-level report for all RCs probably not useful
- Sub-scores → Alerting
- Moderate success in Math, especially given small # of items/student
- Consider:
  - lower granularity (more items)
  - altering item review for inclusion

Thank you!!!