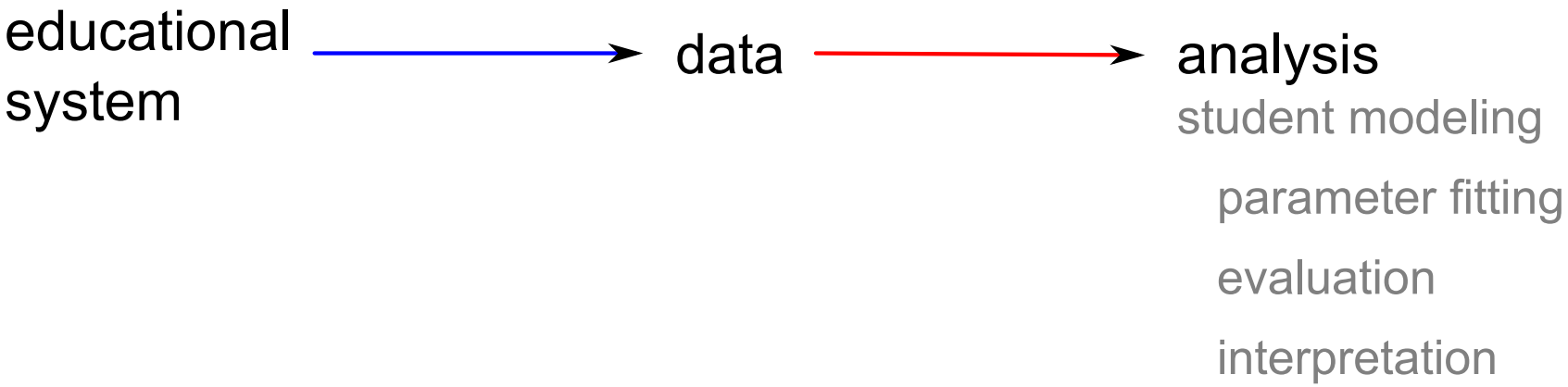
Impact of Data Collection on Interpretation and Evaluation of Student Models

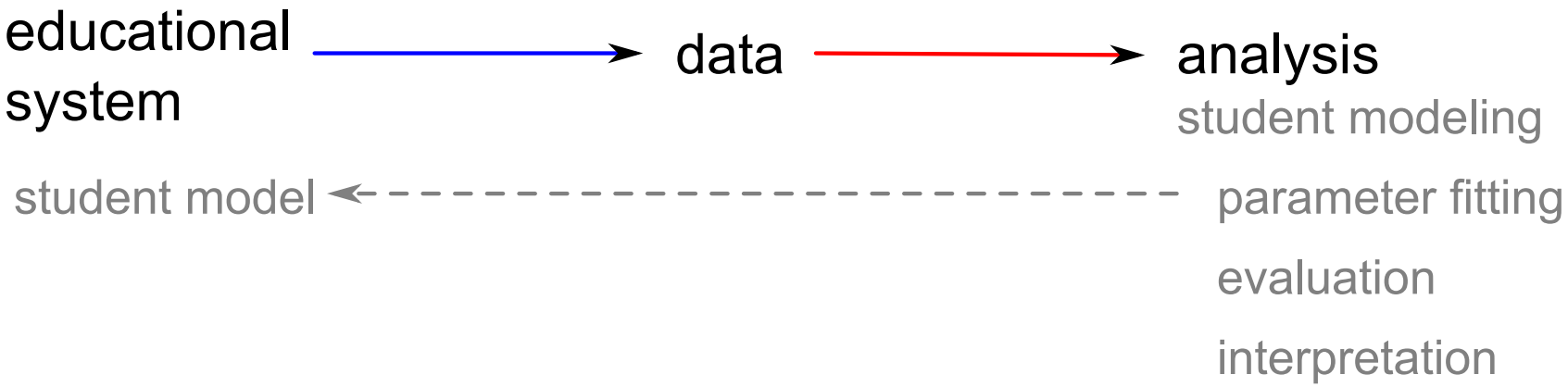
Radek Pelánek, Jiří Řihák, Jan Papoušek

Adaptive Learning
Research group, Masaryk university Brno

LAK 2016

educational system
data
analysis


educational system
data
analysis
student modeling
parameter fitting
evaluation
interpretation


educational system
data
analysis
student modeling
student model
parameter fitting
evaluation
interpretation

Note on Student Models

BKT Bayesian Knowledge Tracing

PFA Performance Factor Analysis

Elo modified Elo Rating System

Our Focus

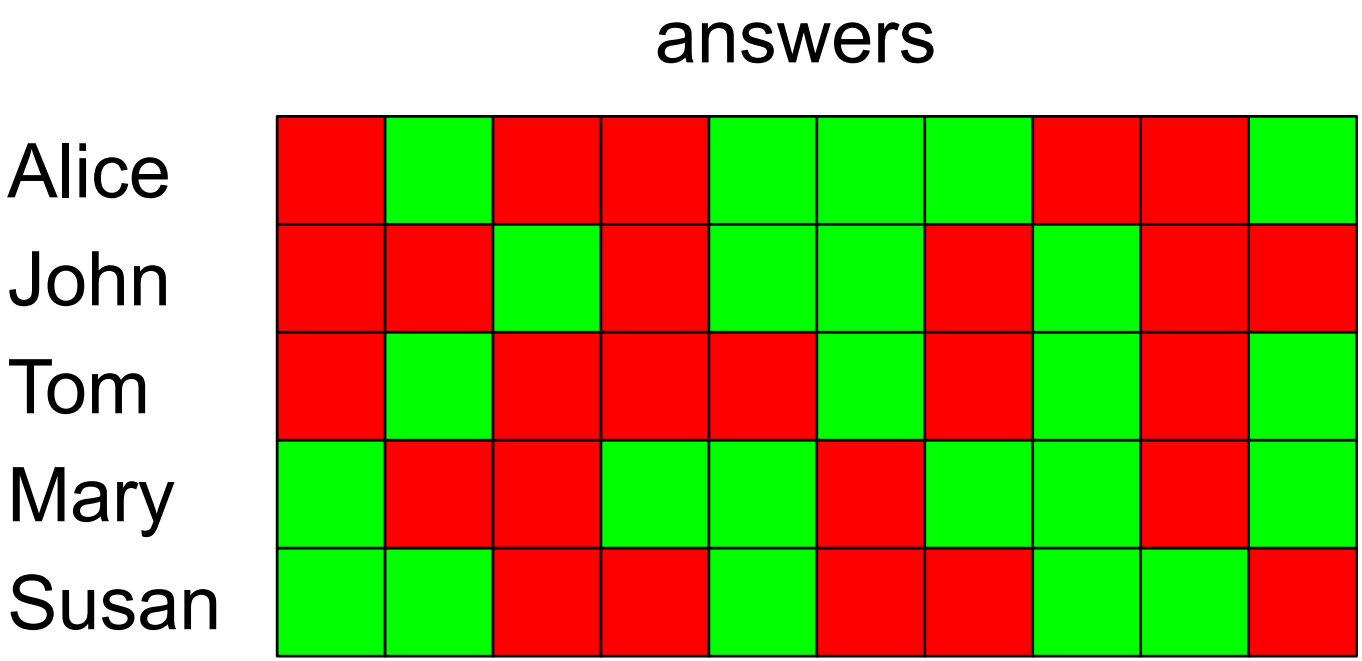
Issues explored:

- mastery attrition, number of answers per student
- item ordering and selection

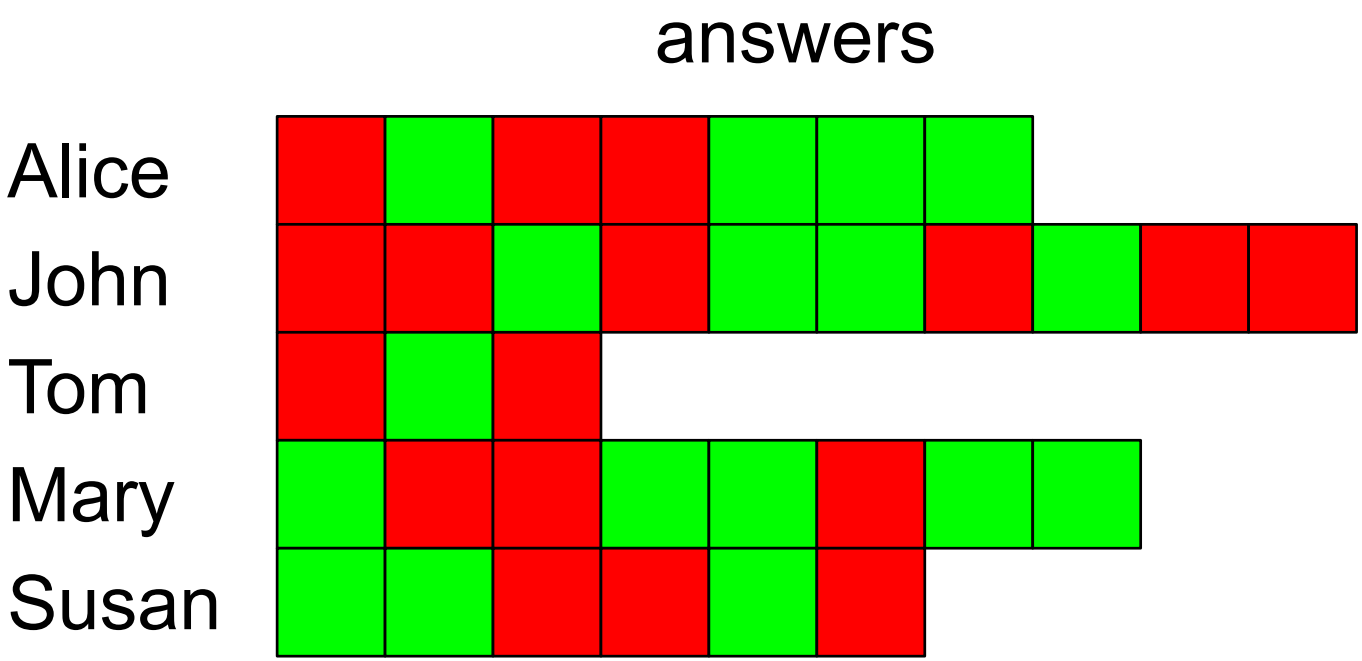
Methods used:

- simulated data
- real data (adaptive practice of geography)

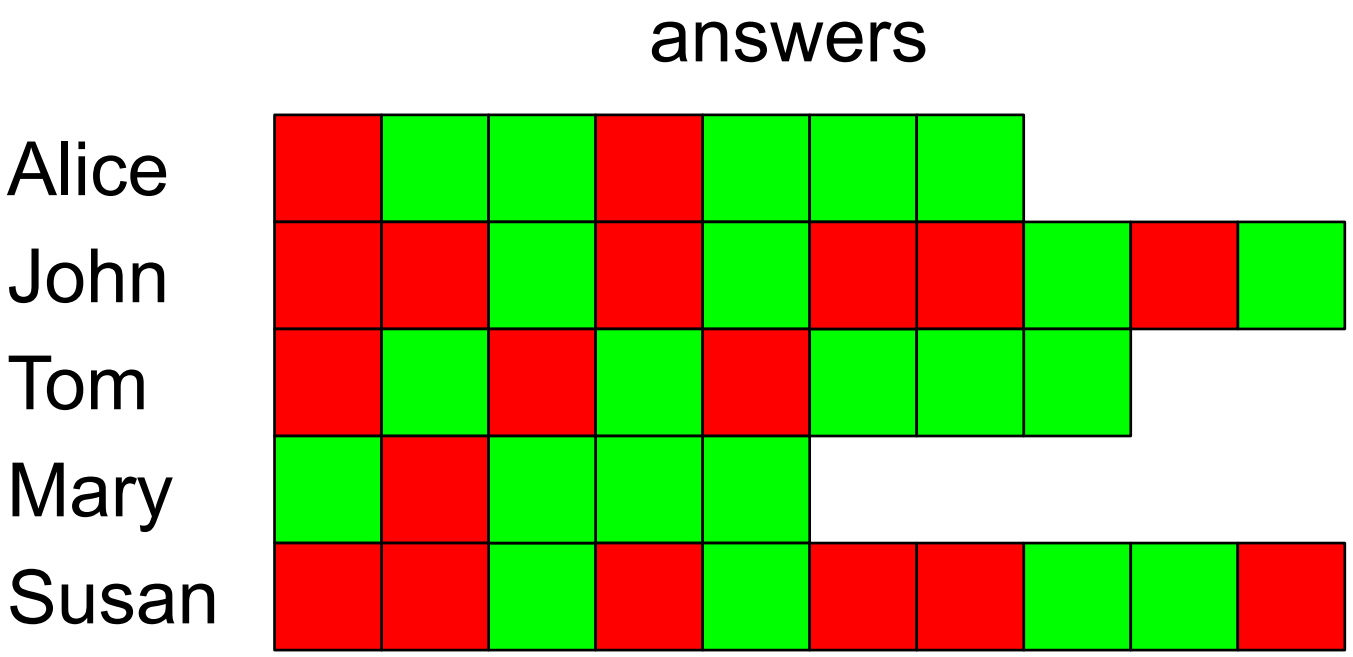
Full Data



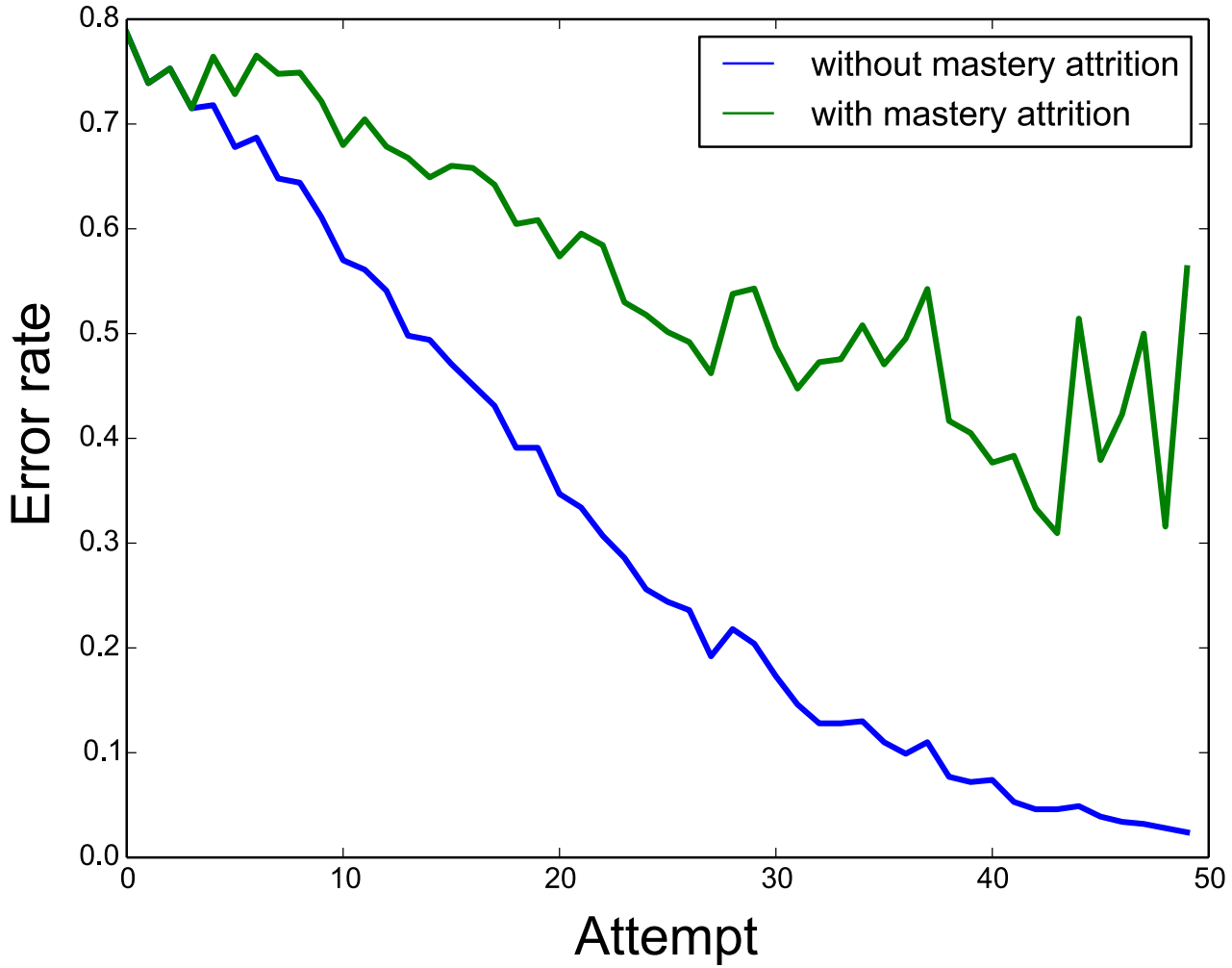
Different Number of Answers



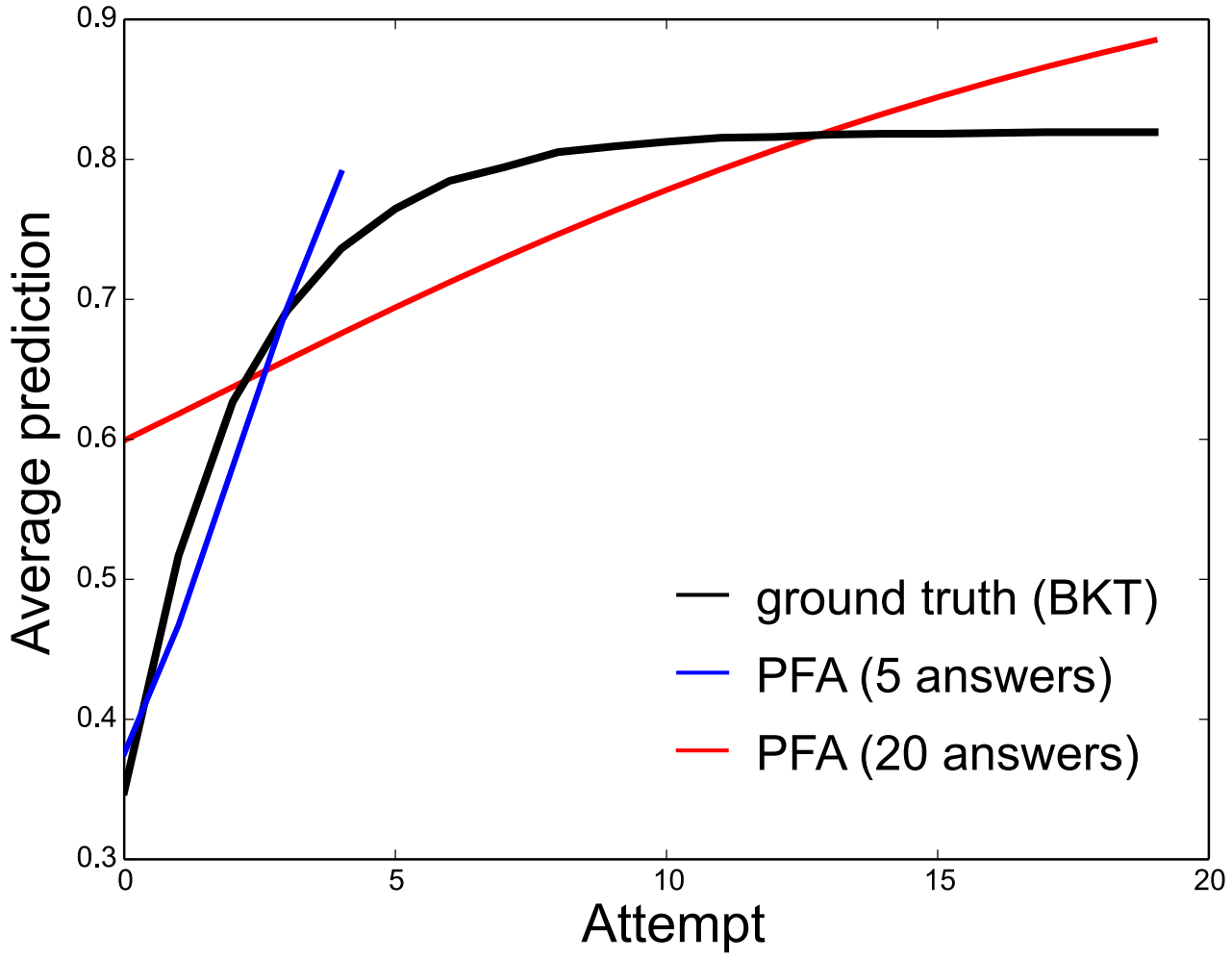
Mastery Learning



Mastery Attrition



Number of Answers – Impact on Parameter Fitting



Parameter Fitting and Mastery Attrition

data generated and fitted by BKT model

model	P_{init}	P_{learn}	P_{slip}	P_{guess}
ground truth	0.25	0.08	0.12	0.30
fitted model (20 answers)	0.27	0.08	0.10	0.27
fitted model (mastery learning)	0.72	0.23	0.52	0.15

Mastery Attrition and Model Comparison

data generated using the logistic function:
$\sigma(\theta + k \cdot 0.1)$, where $\theta \sim \mathcal{N}(-0.4, 2)$

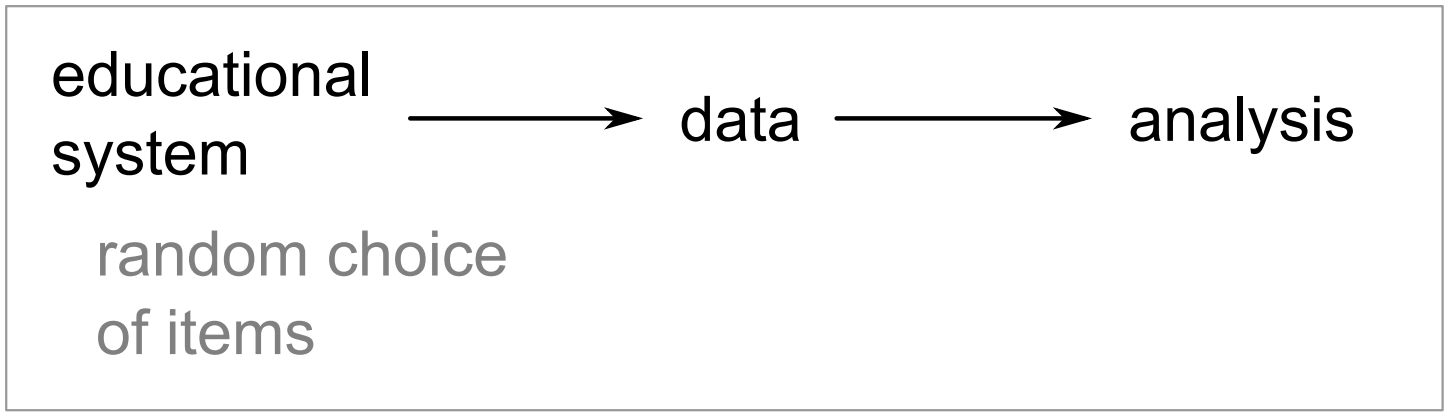
constant number of answer	PFA better than BKT
mastery learning	BKT better than PFA

Main Message

"data collection" may influence results more than "modeling"

Adaptive Choice of Items

- adaptivity, personalization
- select items of appropriate difficulty
    - 75 % chance of correct answer



constant predictor poor predictions

advanced model good predictions

big difference

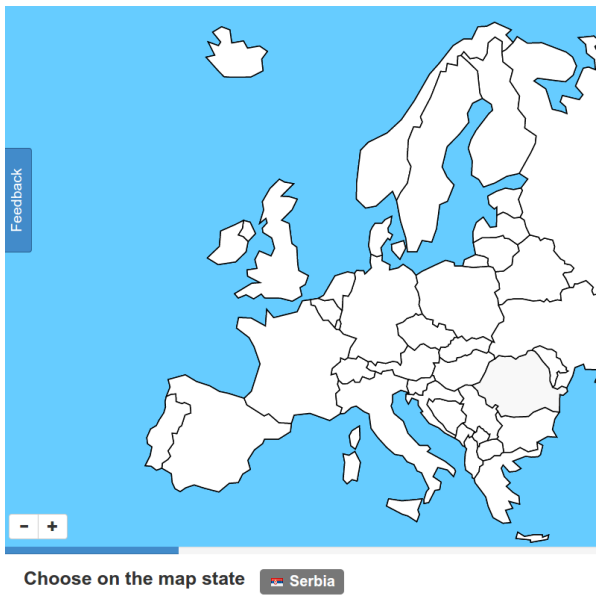
constant predictor good predictions

advanced model good predictions

small difference

Data from Experiment

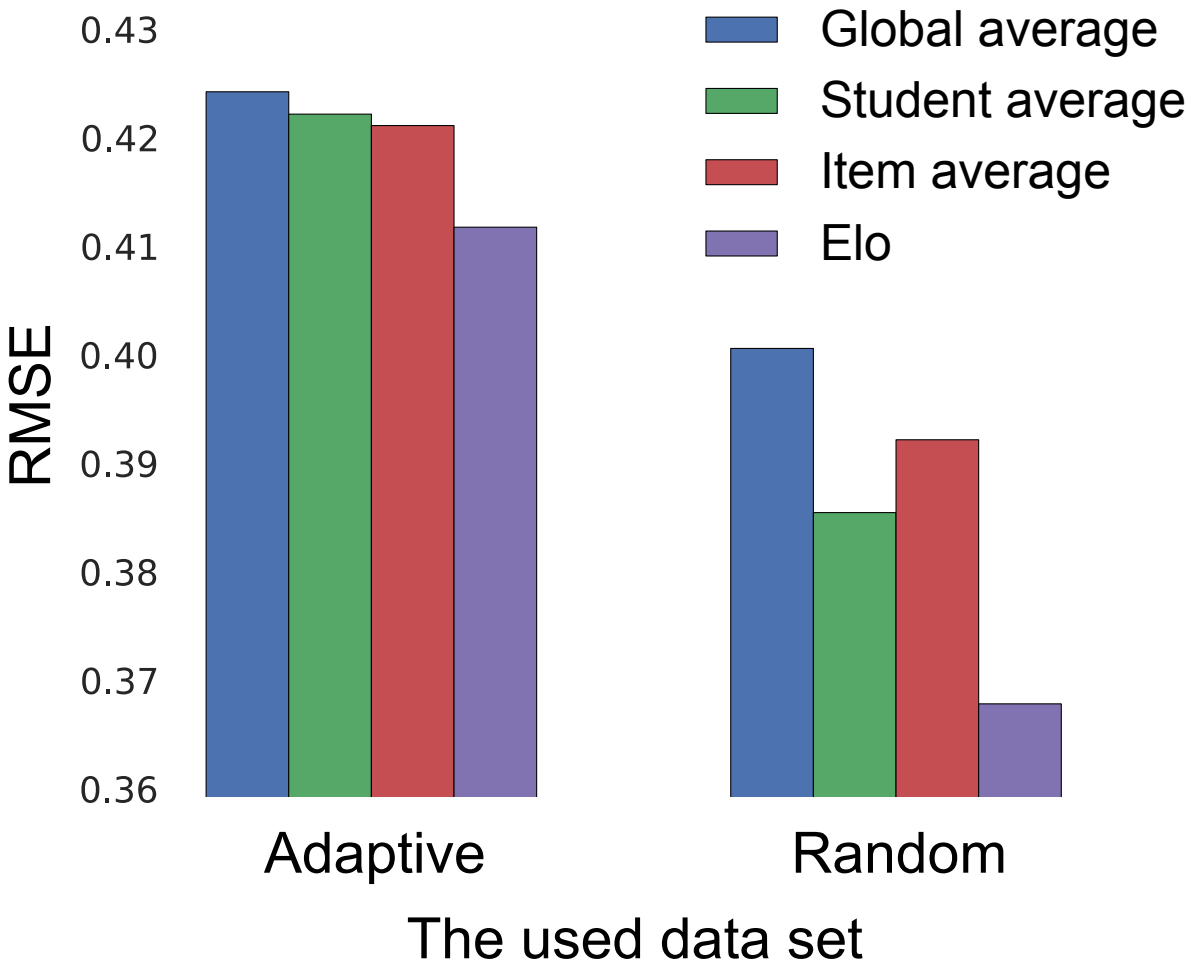
`outlinemaps.org`



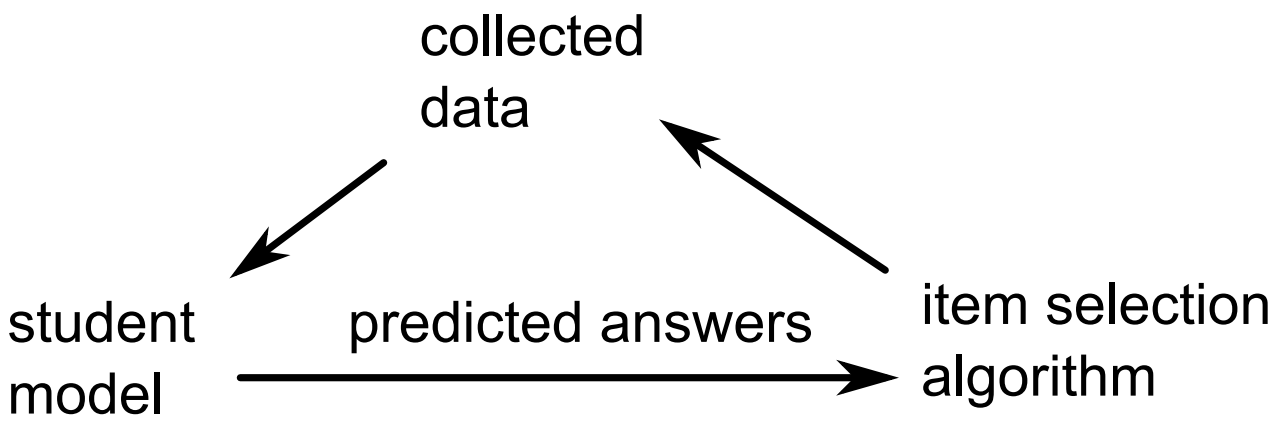
randomized trial, comparison of adaptive vs random choice of items

Evaluation of an Adaptive Practice System for Learning Geography Facts

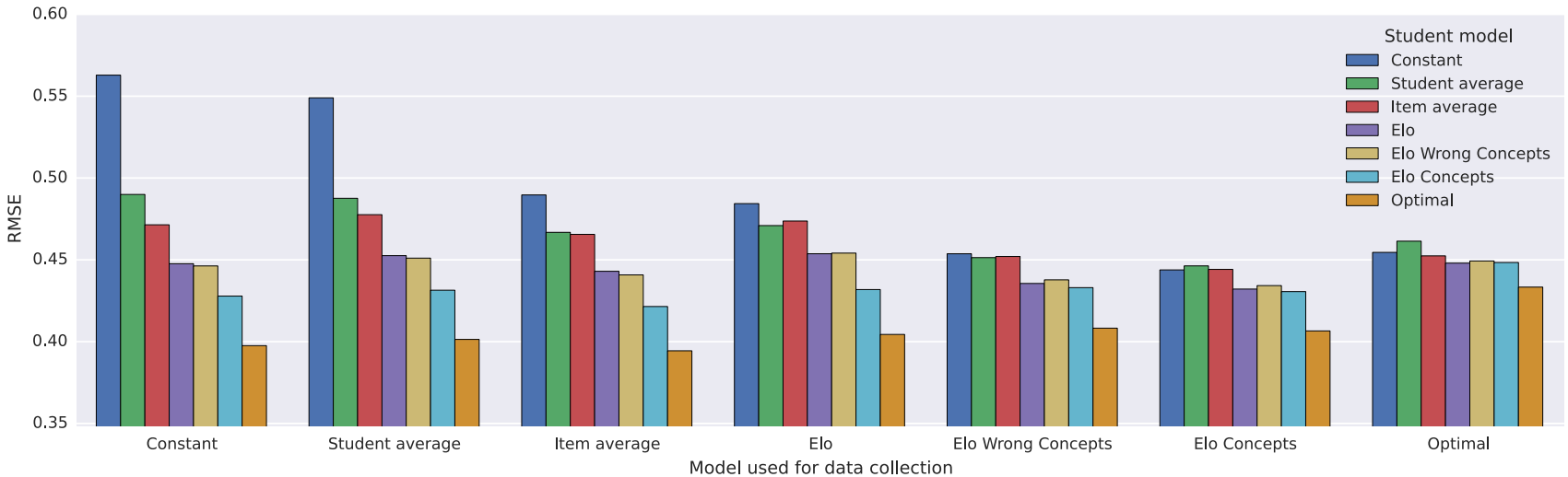
Impact on Model Evaluation



Feedback Loop



Feedback Impact



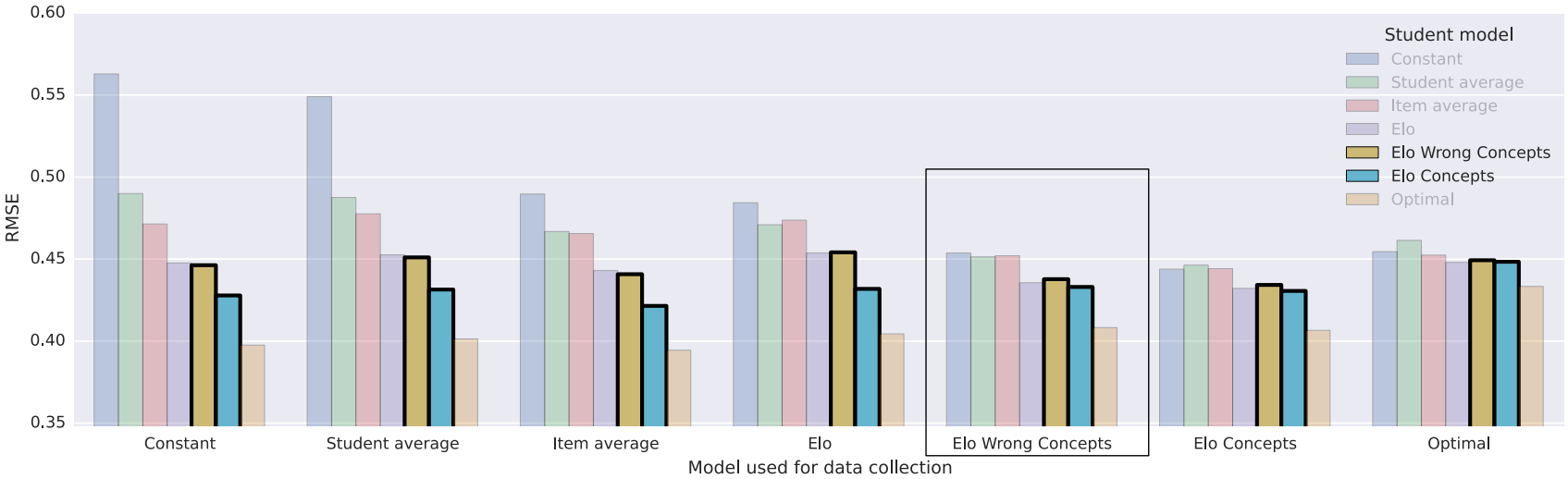
Feedback Impact

good adaptive educational systems

$\Downarrow$

small differences in predictive accuracy
of widely different models

Feedback Impact



wrong model used in adaptive educational systems

$\Downarrow$

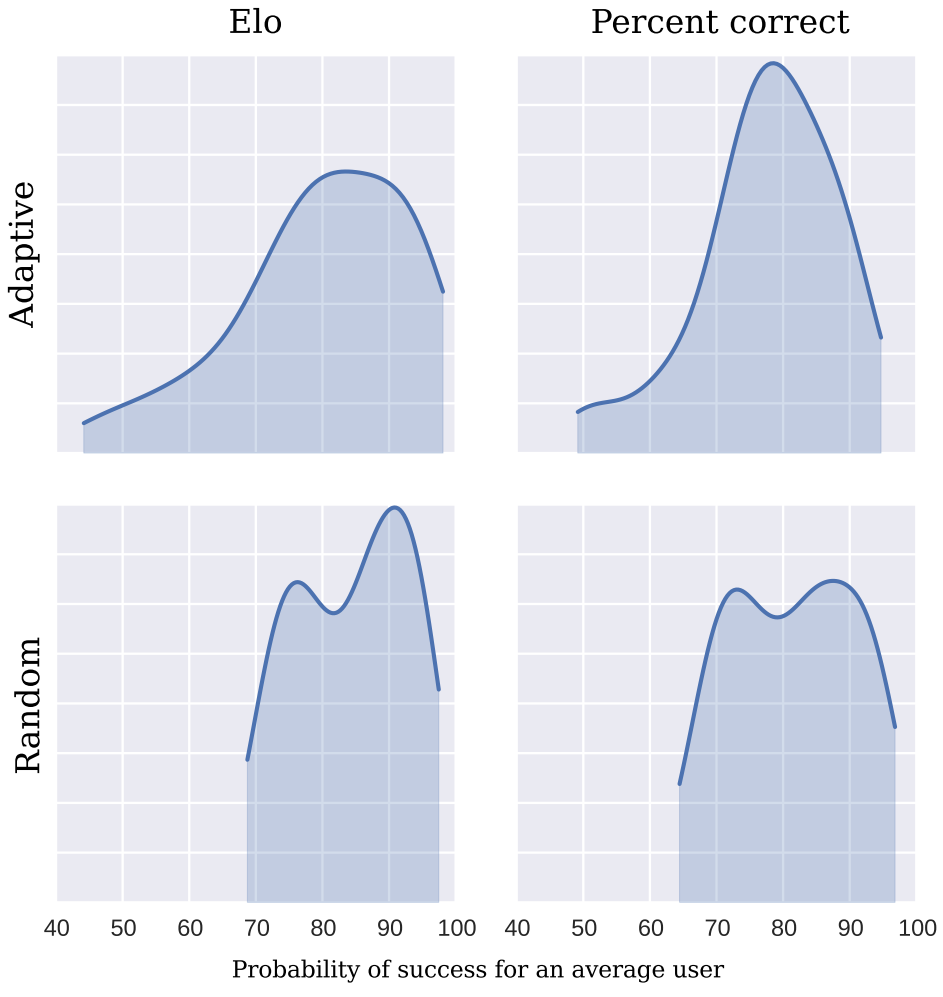
collected data insufficient to show deficiencies

Parameter Estimation

impact of adaptive choice of items on estimation of item difficulty

- two data collection methods:
  - random
  - adaptive
- two estimation methods:
  - naive – percent correct
  - student model – Elo

Parameter Estimation



Data Collection Matters

data collection influences:

- evaluation of models
  - size of differences in comparison metrics
  - sometimes even ordering of models
- interpretation of models
  - fitted parameter values
  - “discovery with models”

Consequences for Practice

1. publication of data sets
   - describe data collection mechanism
2. evaluation and interpretation of models
   - understanding of data
   - stability of results
3. data collection
   - controlled use of randomization

Questions

- What about your data?
- How were they collected?
- Does it matter? Does it influence your results?
- Are you sure there is no bias?