Measuring Similarity of Educational Items Using Data on Learners' Performance

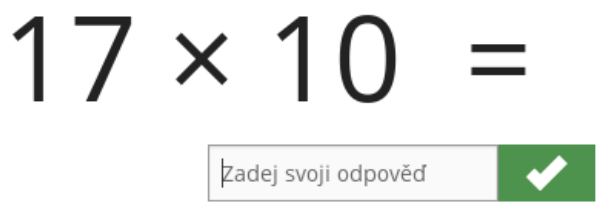
Jiří Řihák, Radek Pelánek

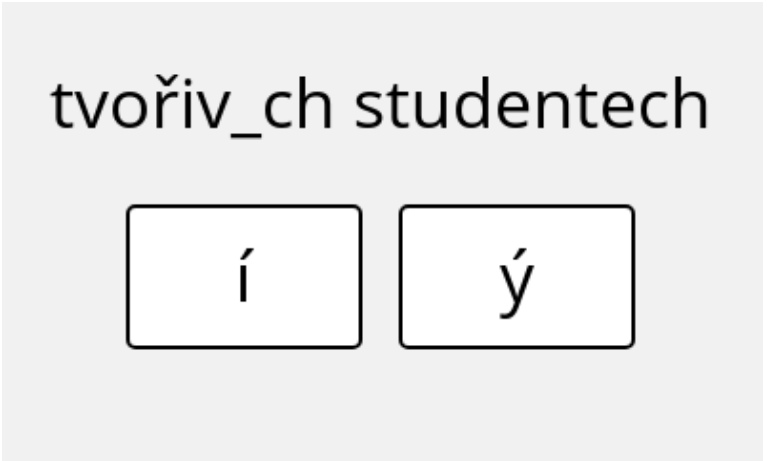
Masaryk University Brno

Adaptive learning

Adaptive practice systems

- **items** — simple questions
- practice — rapid sequence of items





Adaptive Learning

Motivation

Large pool of items

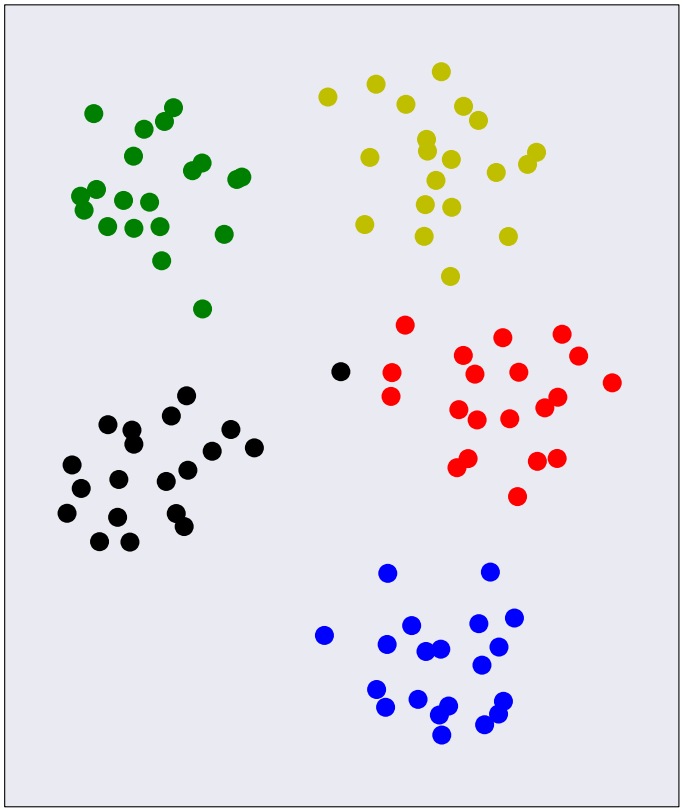
- How to organize these items?
- What knowledge components should be used?
- Are there some anomalies?
- ...

Adaptive Learning

Motivation

Large pool of items

- **clustering**
- visualization
- outlier detection
- ...

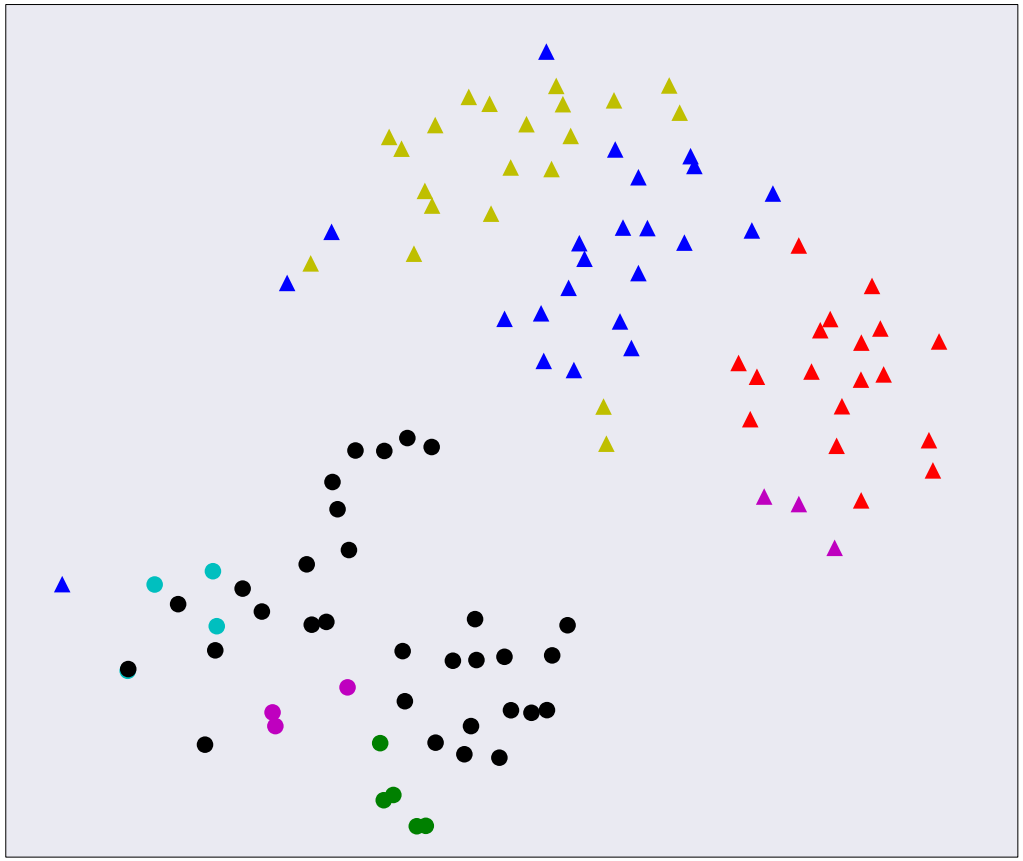


Adaptive Learning

Motivation

Large pool of items

- clustering
- **visualization**
- outlier detection
- ...

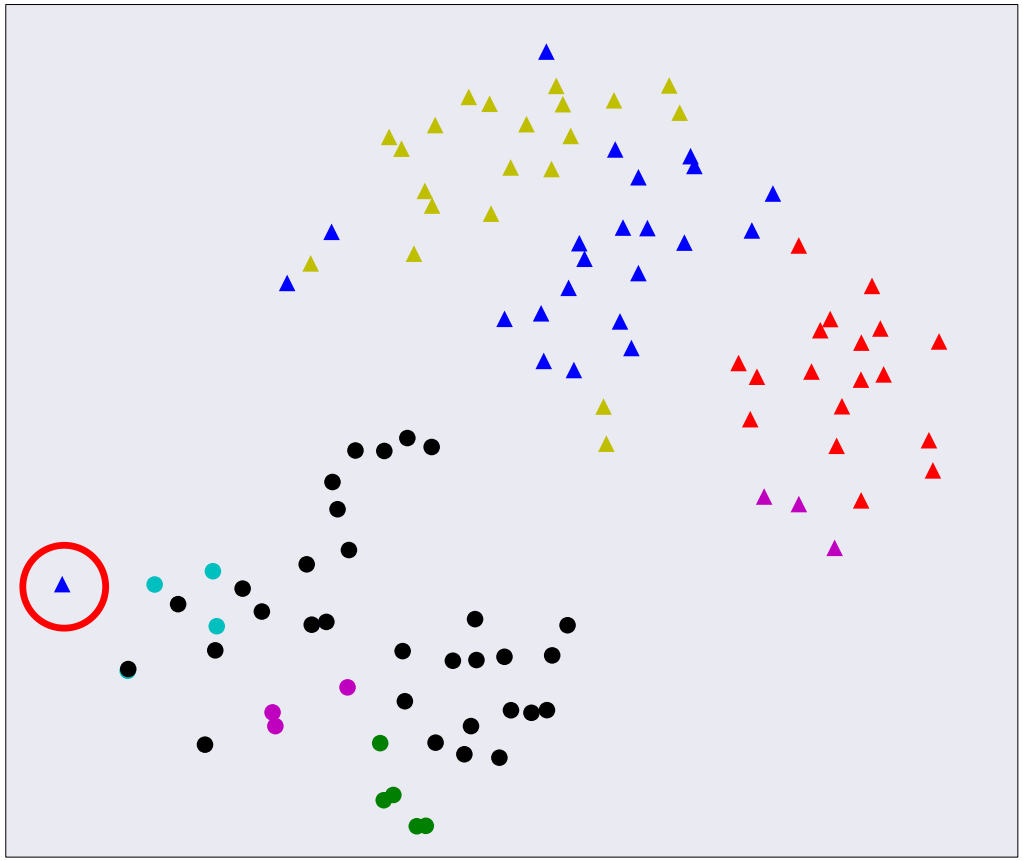


Adaptive Learning

Motivation

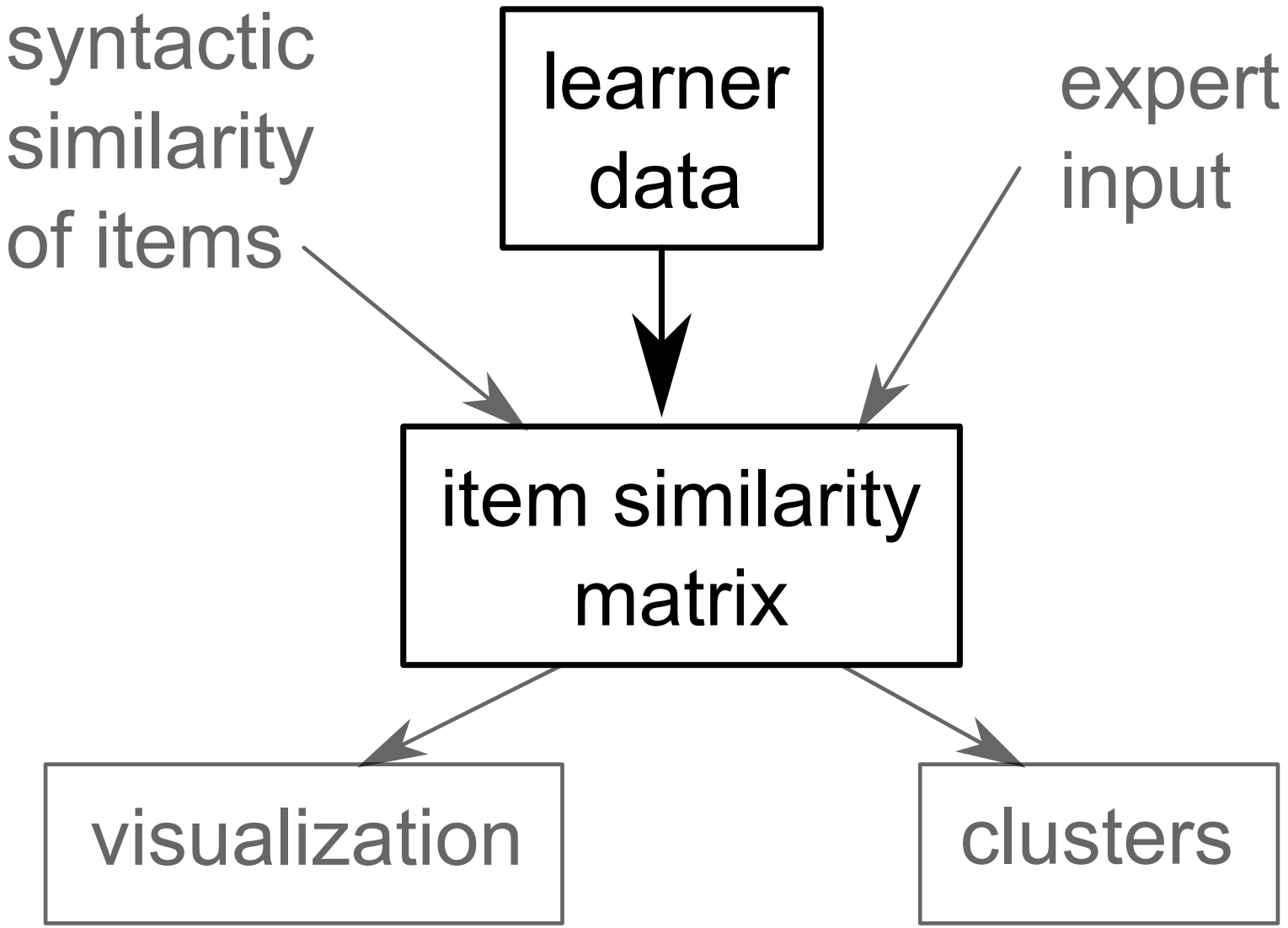
Large pool of items

- clustering
- visualization
- **outlier detection**
- ...



Adaptive Learning

General approach

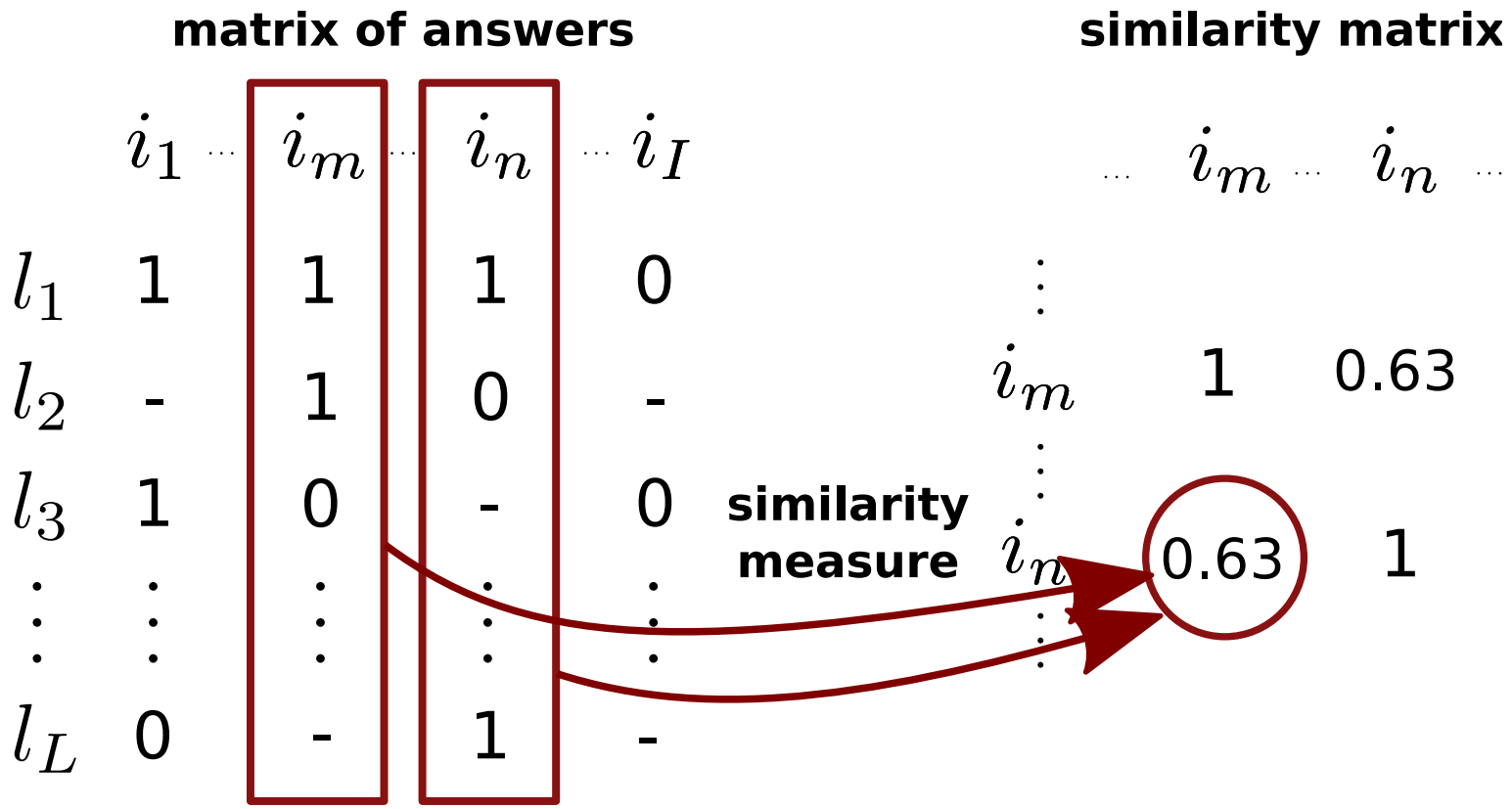


Research questions

- What similarity measures is suitable for EDM?
- How much data we need?
- How to combine more types of learner data?

Adaptive Learning

Similarity measures



Similarity measures

binary data

- 1 — correct
- 0 — incorrect
- input can be simplified:

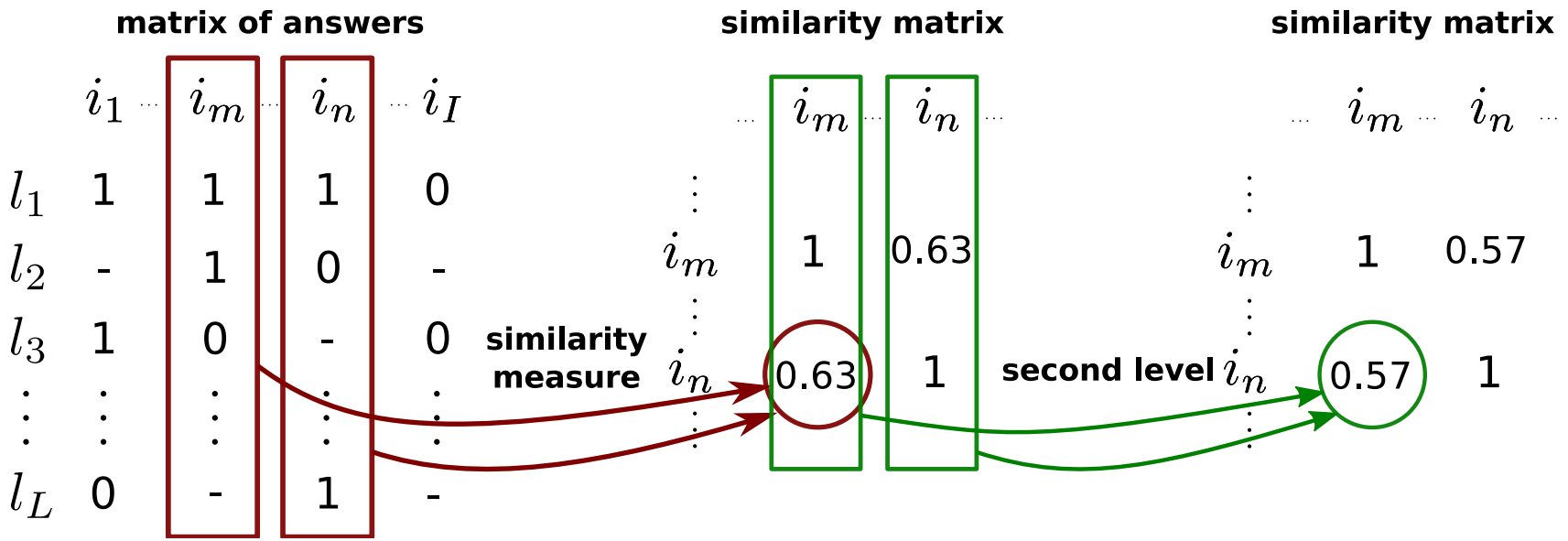
		item i	
		incorrect	correct
item j	incorrect	a	b
	correct	c	d

Adaptive Learning

Similarity measures

Yule	$S_y = (ad - bc)/(ad + bc)$
Pearson	$S_p = (ad - bc)/\sqrt{(a+b)(a+c)(b+d)(c+d)}$
Cohen	$S_c = (P_o - P_e)/(1 - P_e)$
	$P_o = (a + d)/n$
	$P_e = ((a+b)(a+c) + (b+d)(c+d))/n^2$
Sokal	$S_s = (a + d)/(a + b + c + d)$
Jaccard	$S_j = a/(a + b + c)$
Ochiai	$S_o = a/\sqrt{(a+b)(a+c)}$

Second level of similarity

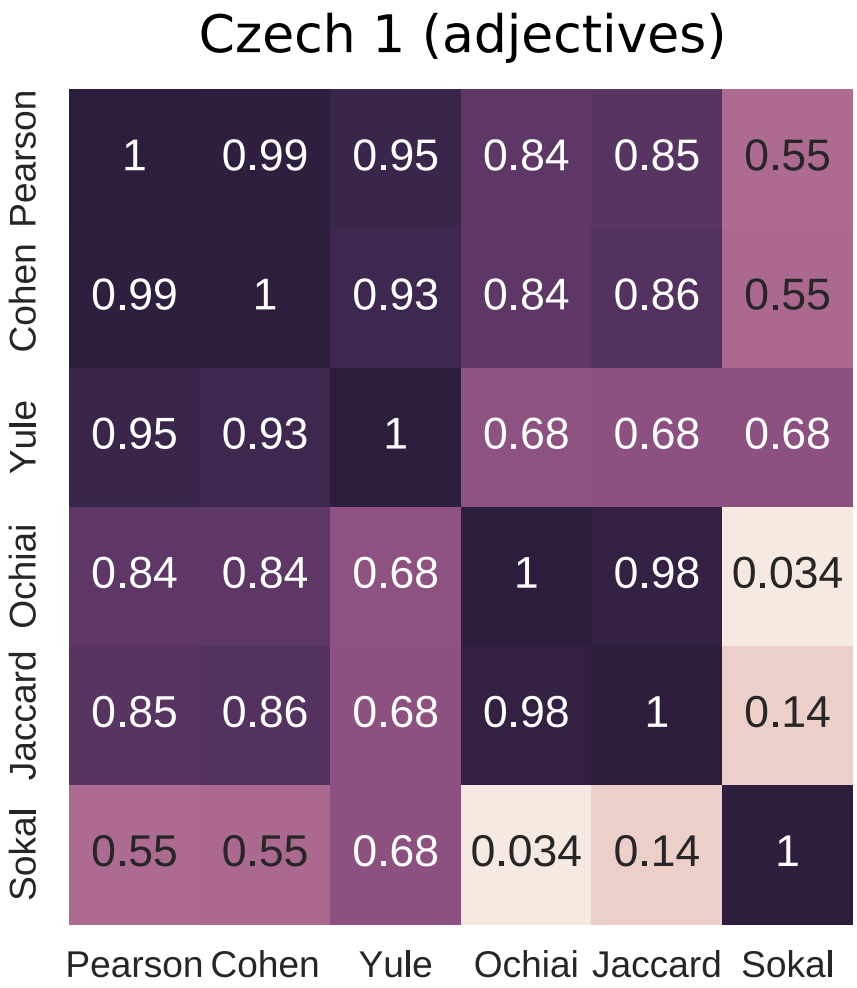


Second level of similarity

- 2 items are similar if they are *similarly* similar to other items
- more information used
- noise reduction
- necessary for some follow up algorithms

Evaluation - correlation of measures

- Cohen - **Pearson**
- Ochiai - **Jaccard**
- Yule
- Sokal - the most different

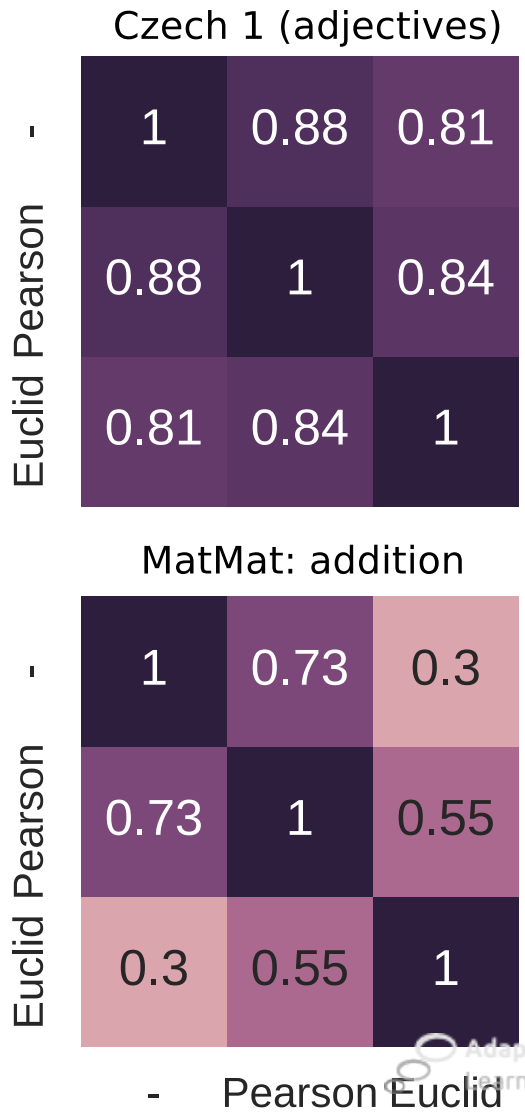


Evaluation - correlation of measures

- Cohen - **Pearson**
- Ochiai - **Jaccard**
- Yule
- Sokal - the most different

Second level of similarity

- brings change
- larger for smaller datasets



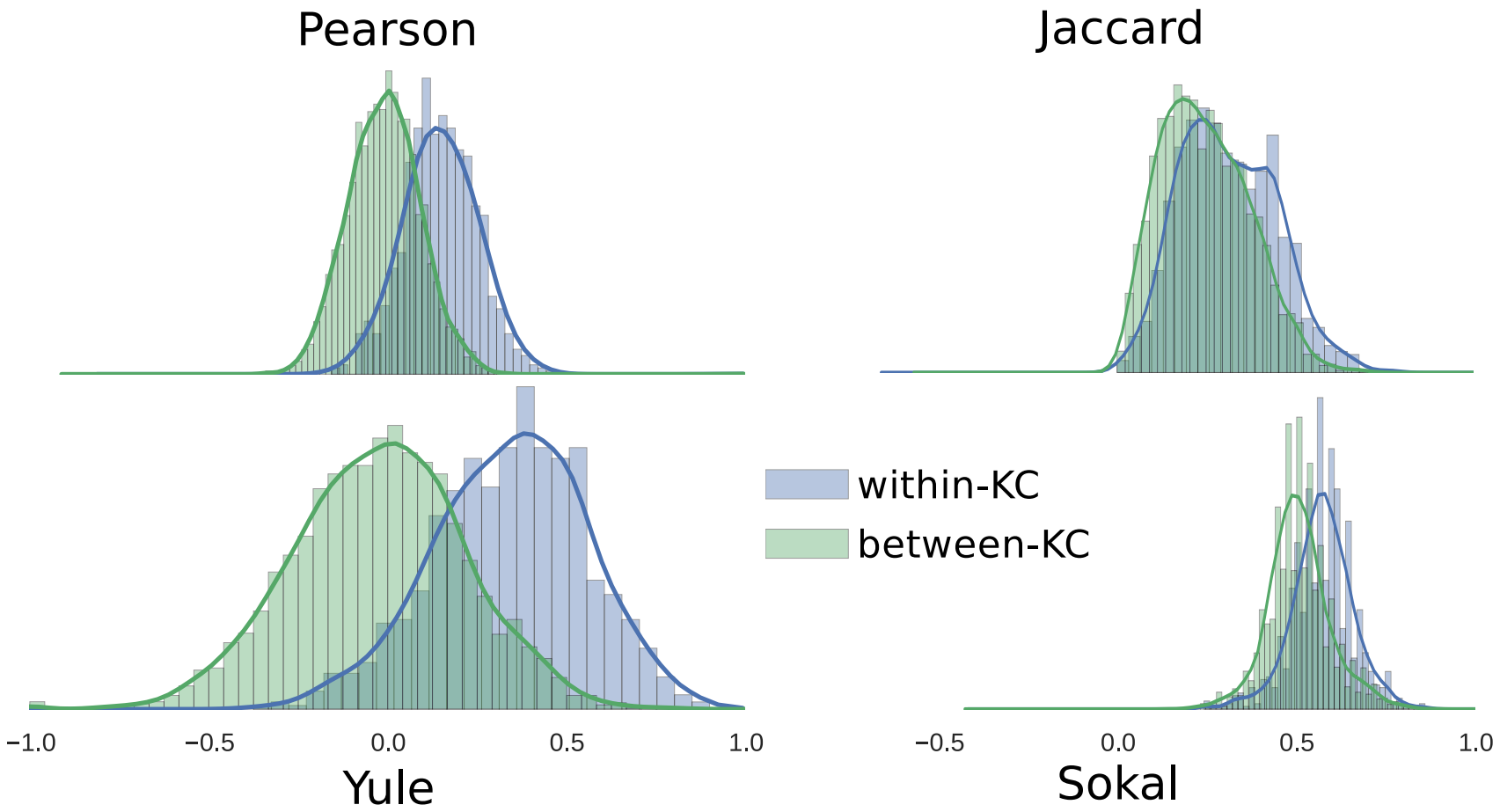
Simulated data

Simulated data

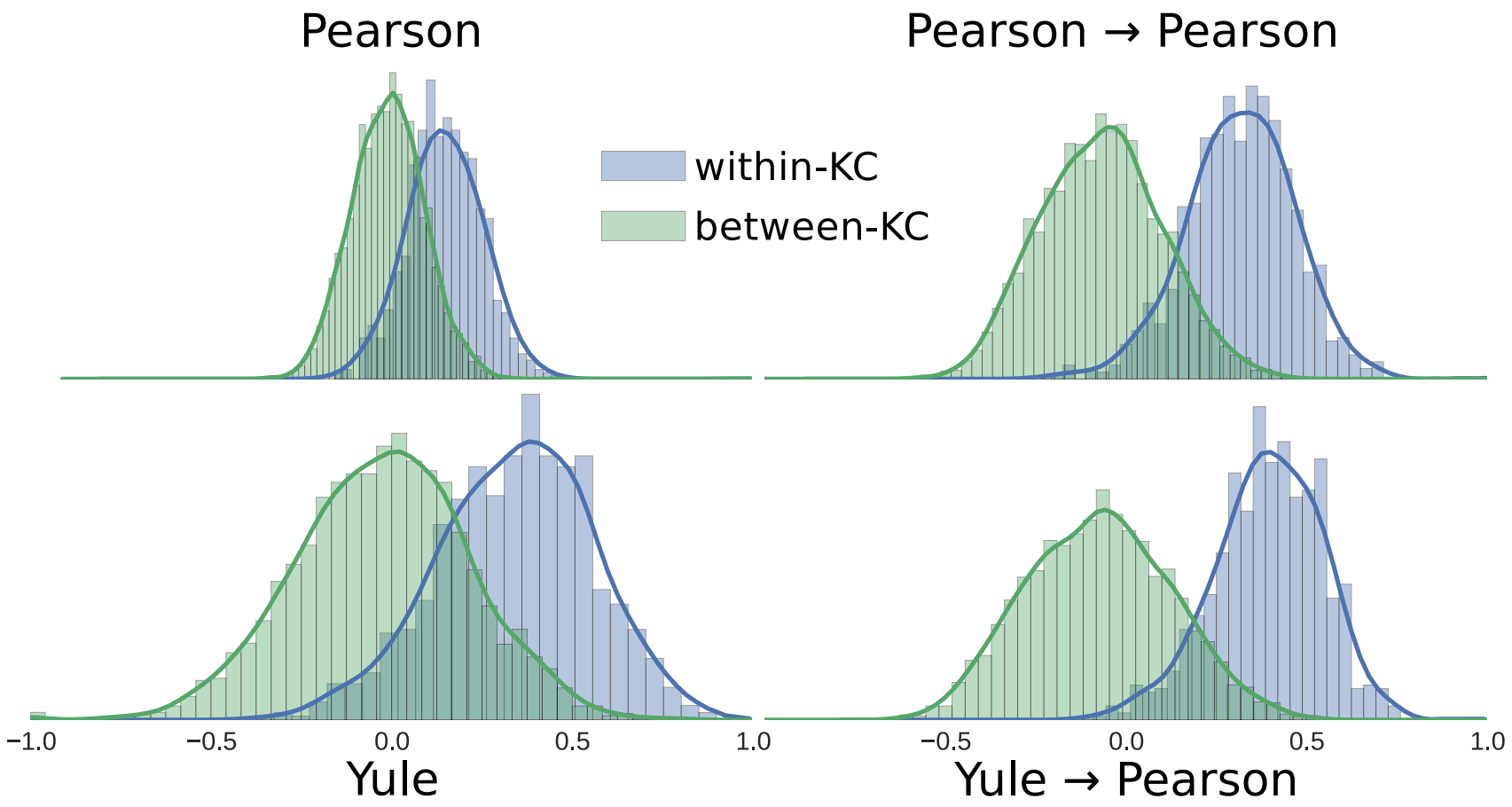
- we know *right answer*
- logistic model
  - learners have skills
  - items have difficulty
- typical setting
  - 100 learners
  - 5 knowledge components
  - 20 items per KC

Adaptive Learning

Evaluation


Pearson
Jaccard
within-KC
between-KC
−1.0
−0.5
0.0
0.5
1.0
Yule
−0.5
0.0
0.5
1.0
Sokal

Evaluation


Pearson
Pearson → Pearson
within-KC
between-KC
−1.0
−0.5
0.0
0.5
1.0
Yule
−0.5
0.0
0.5
1.0
Yule → Pearson

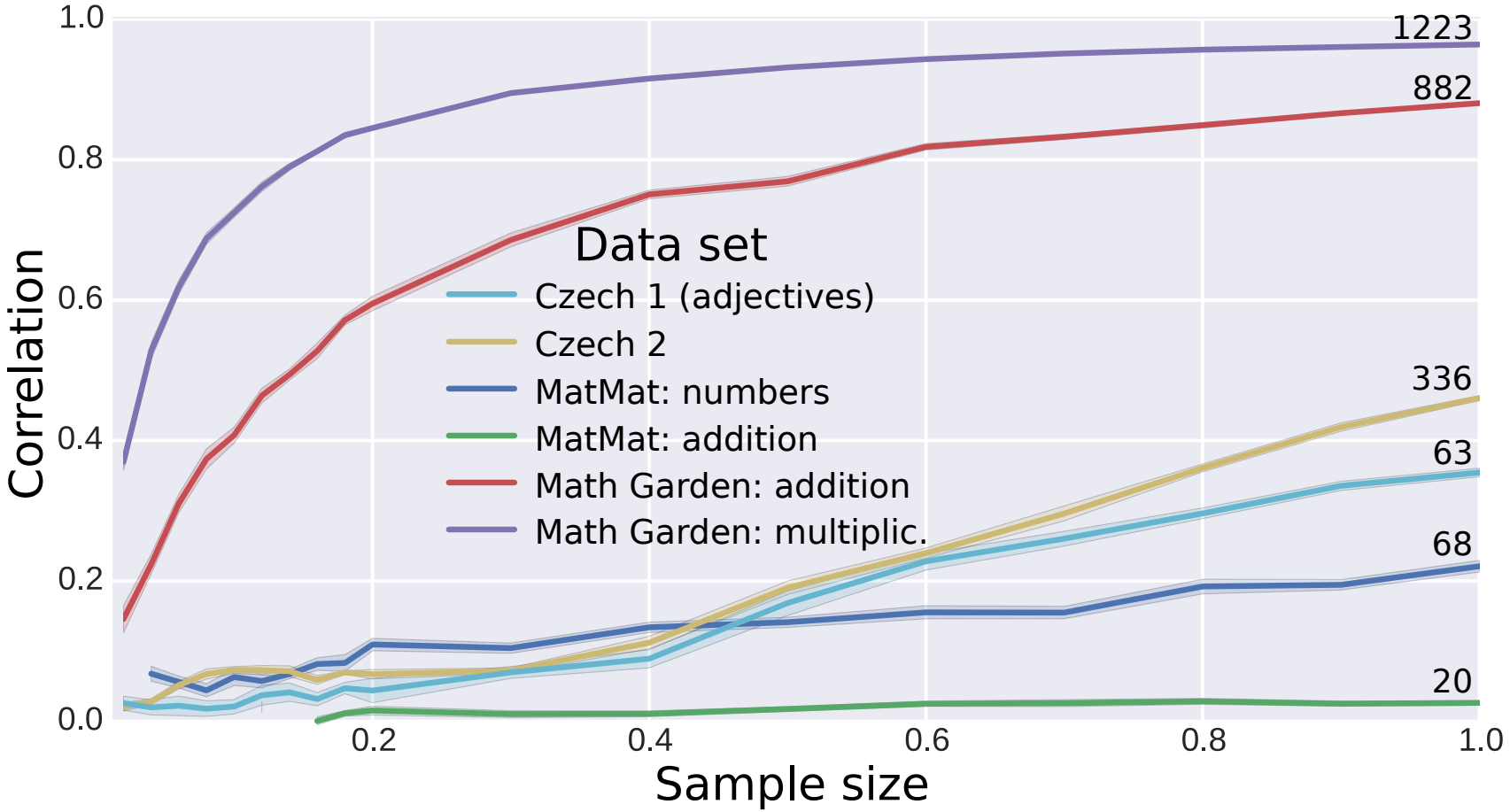
Evaluation - clustering

	Czech adjectives	100L 5KC	200L 5KC
Pearson	0.32 ± 0.02	0.48 ± 0.05	0.84 ± 0.05
Jaccard	0.31 ± 0.03	0.15 ± 0.04	0.29 ± 0.08
Yule	0.31 ± 0.03	0.43 ± 0.05	0.77 ± 0.07
Sokal	0.15 ± 0.06	0.18 ± 0.03	0.25 ± 0.05
Pearson $\rightarrow$ Euclid	$\mathbf{0.43} \pm 0.01$	$\mathbf{0.80} \pm 0.06$	$\mathbf{0.98} \pm 0.01$
Yule $\rightarrow$ Euclid	0.32 ± 0.02	0.65 ± 0.07	0.94 ± 0.04
Pearson $\rightarrow$ Pearson	0.41 ± 0.03	0.73 ± 0.06	0.96 ± 0.02
Yule $\rightarrow$ Pearson	0.32 ± 0.03	0.72 ± 0.06	0.97 ± 0.02

Do We Have Enough Data?

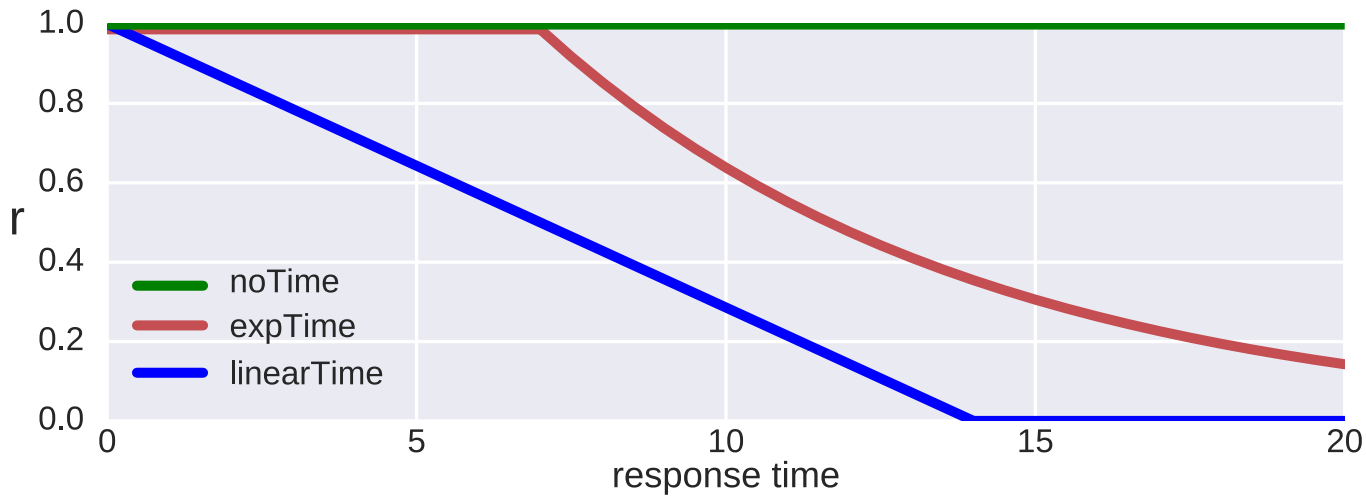
- stability of results
- split data to two halves
- how similarity measures correlate on these halves?

Do We Have Enough Data?



Response times

- additional information
- correctness and response time to one measure of success
- response: $0/1 \rightarrow [0, 1]$

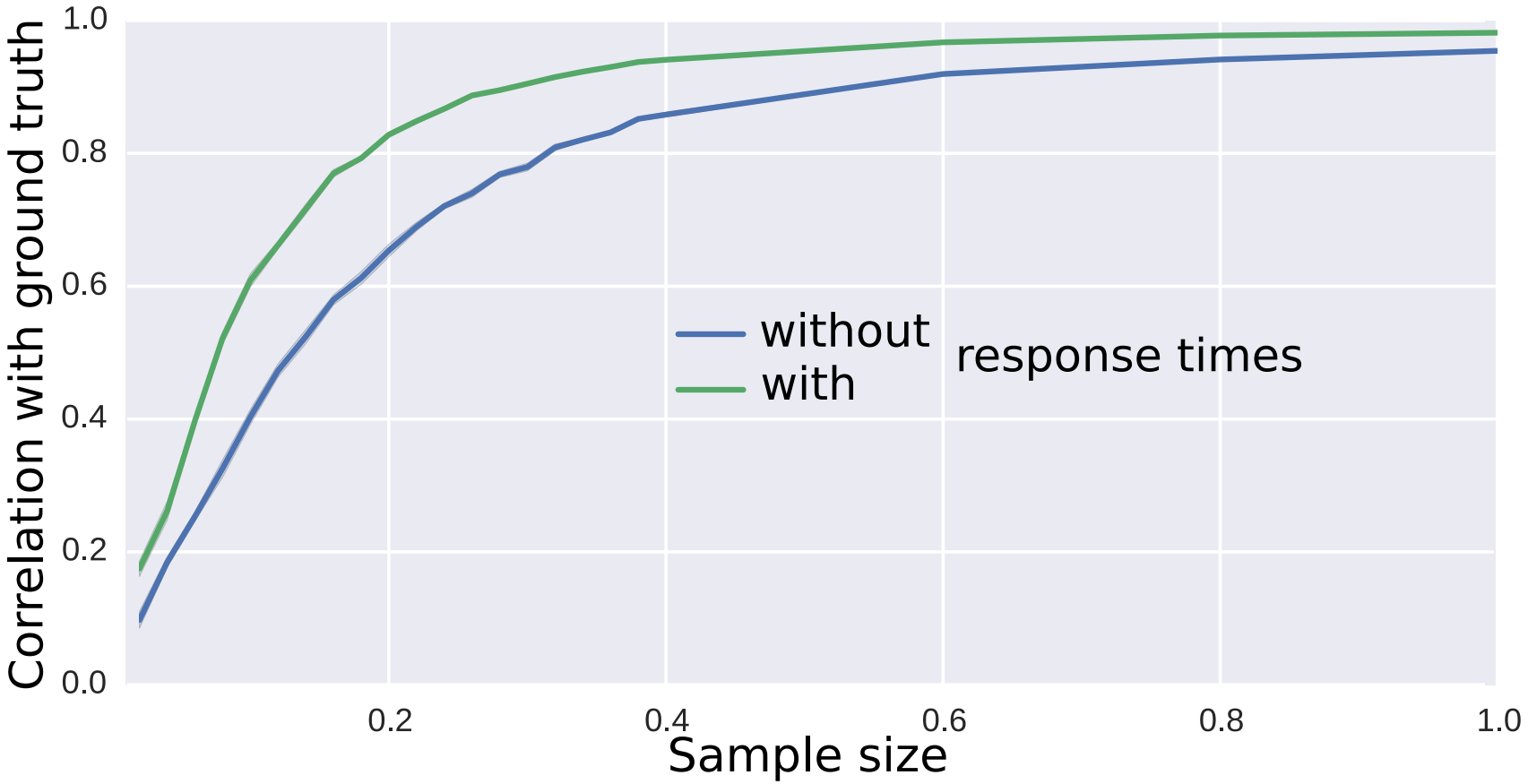


Respose times - Math Garden

- Math Garden - large datasets: $\sim 1M$ of answer on 30 items
- small impact of time information - correlation > 0.9
- but what with smaller datasets?

Adaptive Learning

Respose times - MathGarden



Conclusion

- **Pearson metric is a good default**
- Pearson, Cohen and Yule are good
- second level improve results
- we should check that we have sufficient data
- response time can help with small datasets

Adaptive Learning