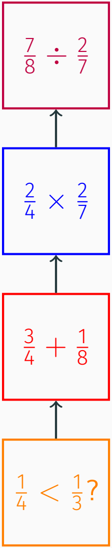
ROBUST EVALUATION MATRIX

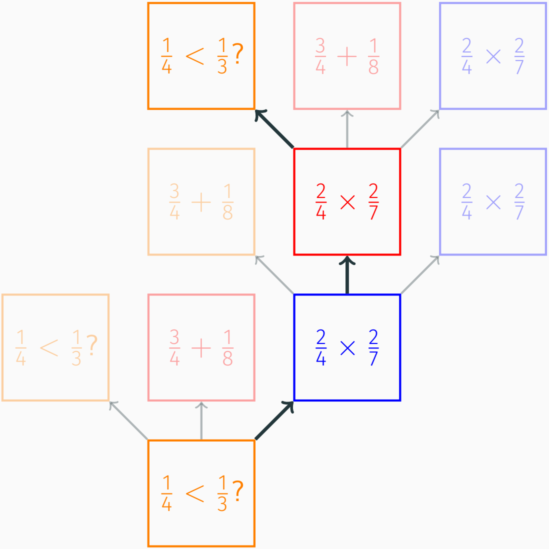
TOWARDS A MORE PRINCIPLED OFFLINE EXPLORATION OF INSTRUCTIONAL POLICIES

Shayan Doroudi, Vincent Aleven, Emma Brunskill

April 20, 2017

Carnegie Mellon University Stanford University


$\frac{7}{8} \div \frac{2}{7}$
$\frac{2}{4} \times \frac{2}{7}$
$\frac{3}{4} + \frac{1}{8}$
$\frac{1}{4} < \frac{1}{3}$?


$\frac{1}{4} < \frac{1}{3}$?
$\frac{3}{4} + \frac{1}{8}$
$\frac{2}{4} \times \frac{2}{7}$
$\frac{3}{4} + \frac{1}{8}$
$\frac{2}{4} \times \frac{2}{7}$
$\frac{2}{4} \times \frac{2}{7}$
$\frac{1}{4} < \frac{1}{3}$?
$\frac{3}{4} + \frac{1}{8}$
$\frac{2}{4} \times \frac{2}{7}$
$\frac{1}{4} < \frac{1}{3}$?

- Experiments take time and resources.

LIMITATIONS OF EXPERIMENTS

- Experiments take time and resources.
- Could be uninformative.

- Experiments take time and resources.
- Could be uninformative.
- Large space of possible interventions. How do we choose what to try?

Can we tell how well an intervention will do without running it?

How can we determine what intervention is best?

- Can use simulated students!

- Can use simulated students!
- Fit student model to past data.

- Can use simulated students!
- Fit student model to past data.
- Use student model to simulate an intervention.

Past Data

$\frac{3}{4} + \frac{1}{8}$

$\frac{1}{4} < \frac{1}{3}$?

$\frac{2}{4} \times \frac{2}{7}$

$\frac{7}{8} \div \frac{2}{7}$

Intervention

Intro to Calculus

Past Data

$\frac{3}{4}$

$\frac{2}{4}$

$\frac{7}{8} \div \frac{2}{7}$

Intervention

Calculus

Past Data

$\boxed{\frac{3}{4} + \frac{1}{8}}$

$\boxed{\frac{1}{4} < \frac{1}{3}?}$

$\boxed{\frac{2}{4} \times \frac{2}{7}}$

$\boxed{\frac{7}{8} \div \frac{2}{7}}$

Intervention

$\boxed{\frac{7}{8} \div \frac{2}{7}}$

$\uparrow$

$\boxed{\frac{2}{4} \times \frac{2}{7}}$

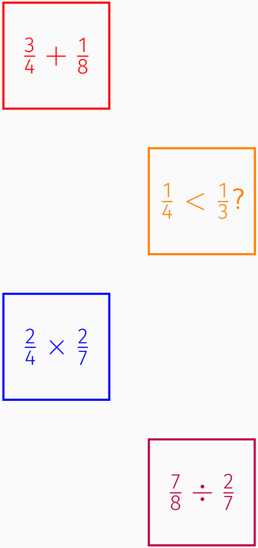
$\uparrow$

$\boxed{\frac{3}{4} + \frac{1}{8}}$

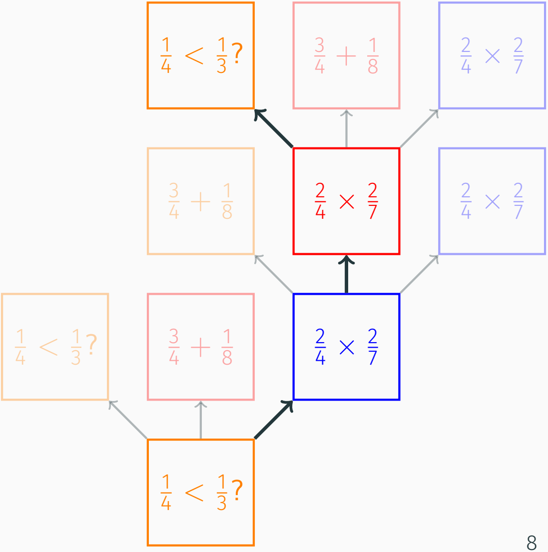
$\uparrow$

$\boxed{\frac{1}{4} < \frac{1}{3}?}$

Past Data



Intervention



Past Data

$\boxed{\frac{3}{4} + \frac{1}{8}}$

$\boxed{\frac{1}{4} < \frac{1}{3}?}$

$\boxed{\frac{2}{4} \times \frac{2}{7}}$

$\boxed{\frac{7}{8} \div \frac{2}{7}}$

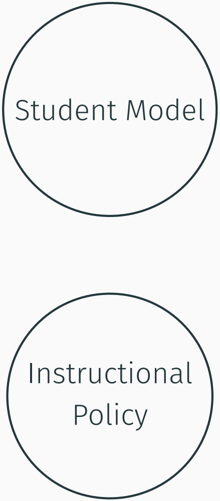
Intervention

Instructional Policy:
Method of sequencing activities, possibly adaptive with respect to student state.

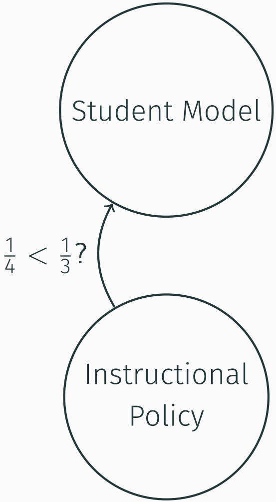
But what if the student model is inaccurate?

Simulate instructional policies on *several* student models!

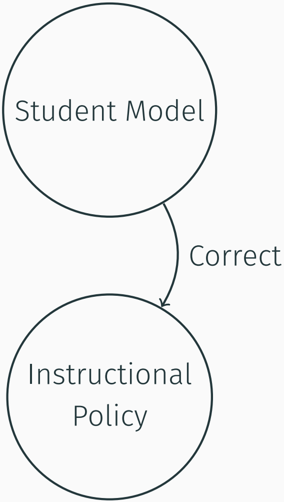
Simulate instructional policies on *several* student models!
Use many models we expect to be wrong, rather than using one model we hope to be right.

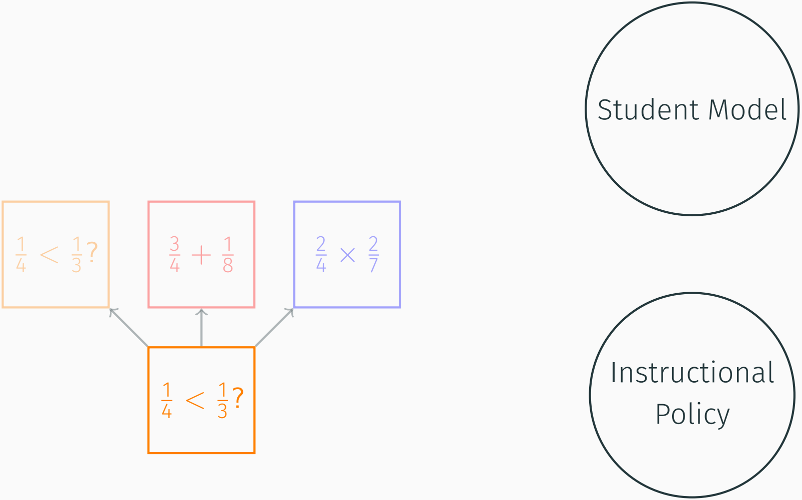

Student Model
Instructional Policy

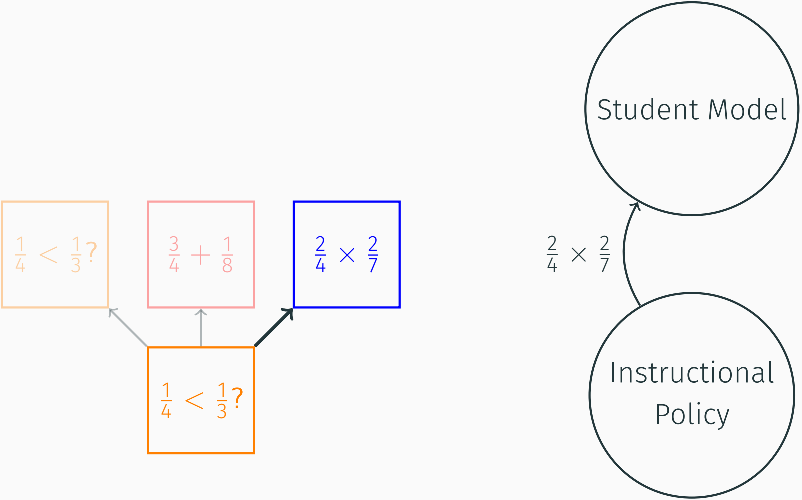
$\boxed{\frac{1}{4} < \frac{1}{3}?}$

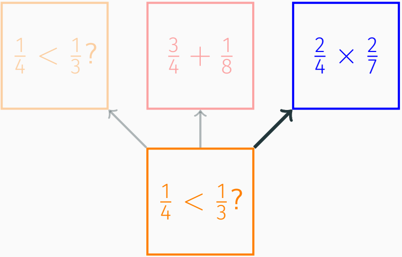


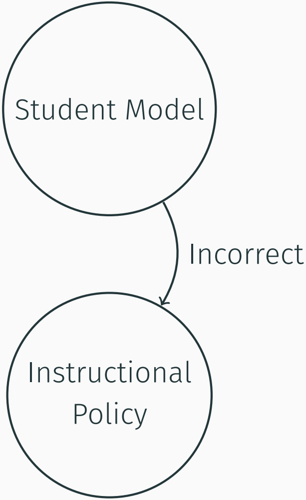
$\boxed{\frac{1}{4} < \frac{1}{3}?}$

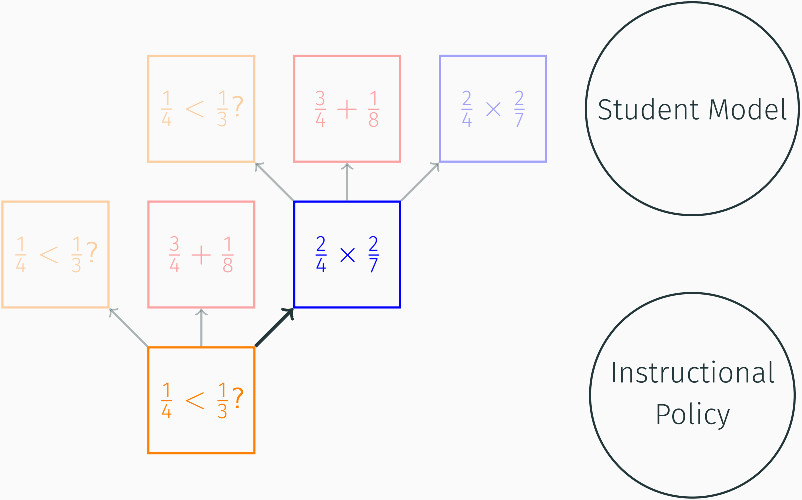


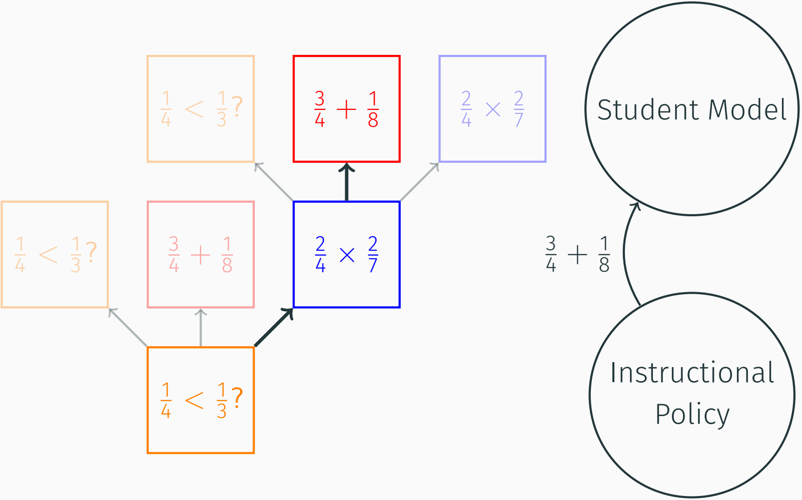

Student Model
Instructional Policy
$\frac{1}{4} < \frac{1}{3}?$
$\frac{3}{4} + \frac{1}{8}$
$\frac{2}{4} \times \frac{2}{7}$
$\frac{1}{4} < \frac{1}{3}?$

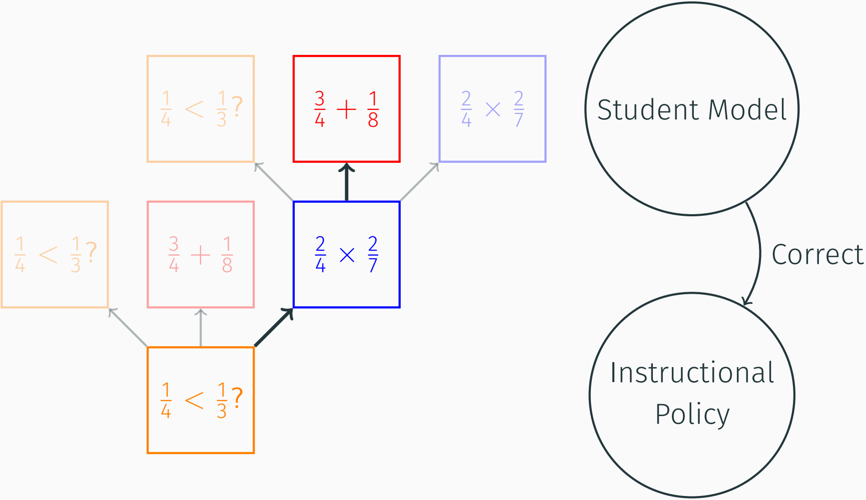

$\frac{1}{4} < \frac{1}{3}?$
$\frac{3}{4} + \frac{1}{8}$
$\frac{2}{4} \times \frac{2}{7}$
$\frac{1}{4} < \frac{1}{3}?$
Student Model
$\frac{2}{4} \times \frac{2}{7}$
Instructional Policy

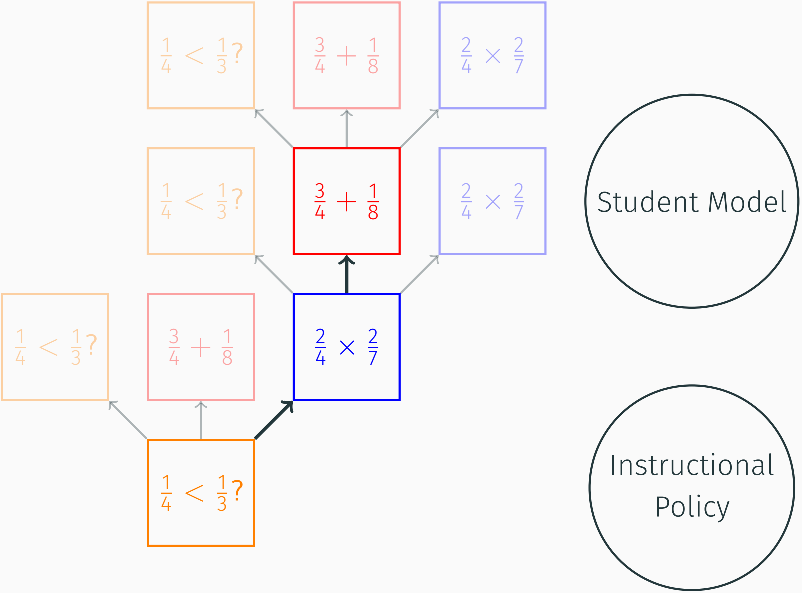

$\frac{1}{4} < \frac{1}{3}$?
$\frac{3}{4} + \frac{1}{8}$
$\frac{2}{4} \times \frac{2}{7}$
$\frac{1}{4} < \frac{1}{3}$?


Student Model
Incorrect
Instructional Policy

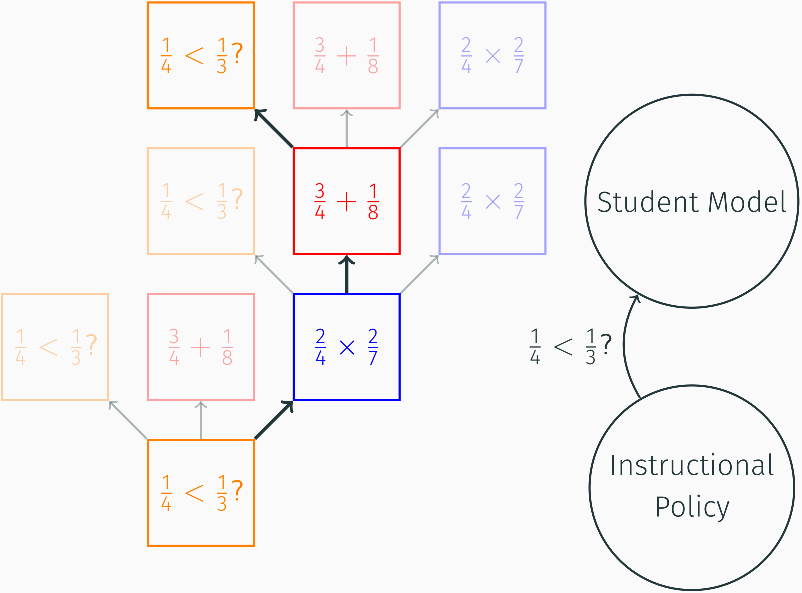

$\frac{1}{4} < \frac{1}{3}$?
$\frac{3}{4} + \frac{1}{8}$
$\frac{2}{4} \times \frac{2}{7}$
$\frac{1}{4} < \frac{1}{3}$?
$\frac{3}{4} + \frac{1}{8}$
$\frac{2}{4} \times \frac{2}{7}$
$\frac{1}{4} < \frac{1}{3}$?
Student Model
Instructional Policy

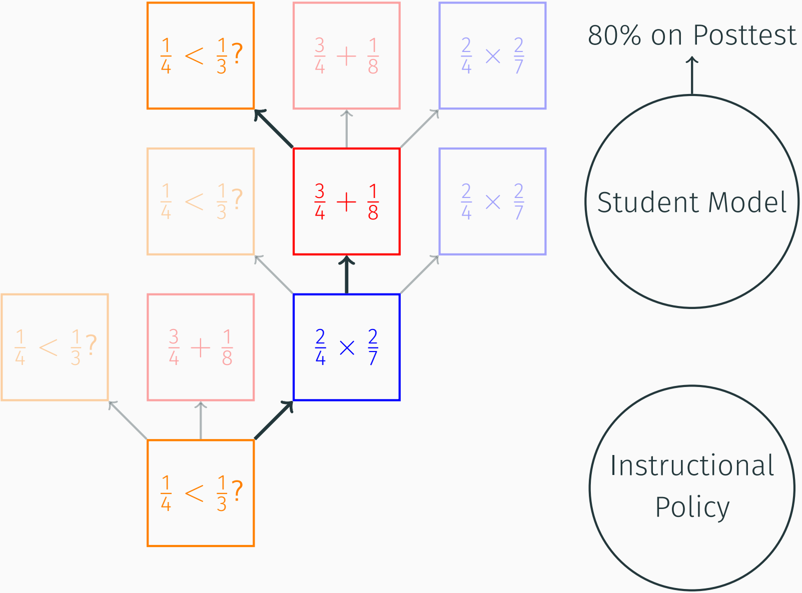

$\frac{1}{4} < \frac{1}{3}$?
$\frac{3}{4} + \frac{1}{8}$
$\frac{2}{4} \times \frac{2}{7}$
$\frac{1}{4} < \frac{1}{3}$?
$\frac{3}{4} + \frac{1}{8}$
$\frac{2}{4} \times \frac{2}{7}$
$\frac{1}{4} < \frac{1}{3}$?
Student Model
$\frac{3}{4} + \frac{1}{8}$
Instructional Policy

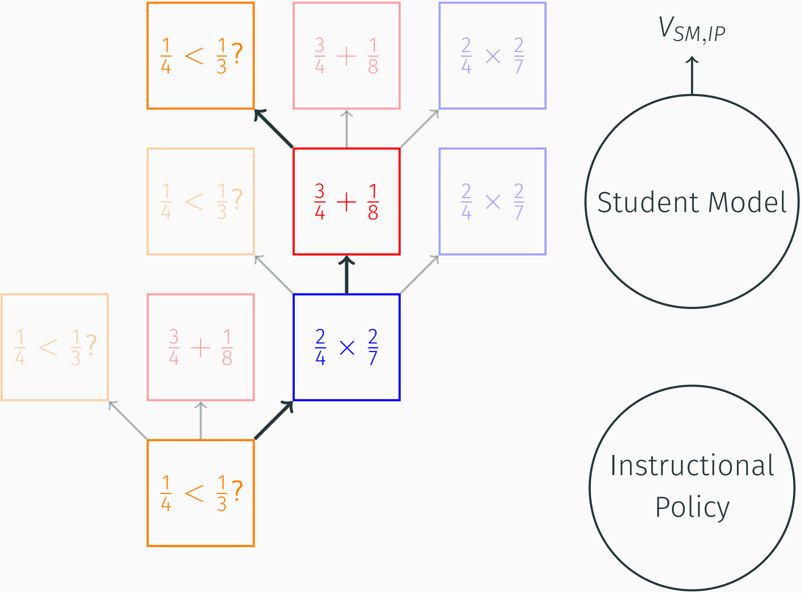

$\frac{1}{4} < \frac{1}{3}$?
$\frac{3}{4} + \frac{1}{8}$
$\frac{2}{4} \times \frac{2}{7}$
$\frac{1}{4} < \frac{1}{3}$?
$\frac{3}{4} + \frac{1}{8}$
$\frac{2}{4} \times \frac{2}{7}$
$\frac{1}{4} < \frac{1}{3}$?
Student Model
Correct
Instructional Policy



$\frac{1}{4} < \frac{1}{3}?$
$\frac{3}{4} + \frac{1}{8}$
$\frac{2}{4} \times \frac{2}{7}$
$\frac{1}{4} < \frac{1}{3}?$
$\frac{3}{4} + \frac{1}{8}$
$\frac{2}{4} \times \frac{2}{7}$
$\frac{1}{4} < \frac{1}{3}?$
$\frac{3}{4} + \frac{1}{8}$
$\frac{2}{4} \times \frac{2}{7}$
$\frac{1}{4} < \frac{1}{3}?$
Student Model
Instructional Policy

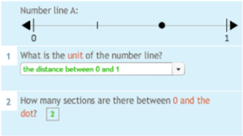

$\frac{1}{4} < \frac{1}{3}$?
$\frac{3}{4} + \frac{1}{8}$
$\frac{2}{4} \times \frac{2}{7}$
$\frac{1}{4} < \frac{1}{3}$?
$\frac{3}{4} + \frac{1}{8}$
$\frac{2}{4} \times \frac{2}{7}$
$\frac{1}{4} < \frac{1}{3}$?
$\frac{3}{4} + \frac{1}{8}$
$\frac{2}{4} \times \frac{2}{7}$
$\frac{1}{4} < \frac{1}{3}$?
Student Model
$\frac{1}{4} < \frac{1}{3}$?
Instructional Policy

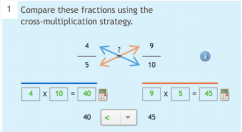

80% on Posttest
Student Model
Instructional Policy
$\frac{1}{4} < \frac{1}{3}$?
$\frac{3}{4} + \frac{1}{8}$
$\frac{2}{4} \times \frac{2}{7}$

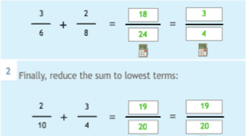

$\frac{1}{4} < \frac{1}{3}?$
$\frac{3}{4} + \frac{1}{8}$
$\frac{2}{4} \times \frac{2}{7}$
$\frac{1}{4} < \frac{1}{3}?$
$\frac{3}{4} + \frac{1}{8}$
$\frac{2}{4} \times \frac{2}{7}$
$\frac{1}{4} < \frac{1}{3}?$
$\frac{3}{4} + \frac{1}{8}$
$\frac{2}{4} \times \frac{2}{7}$
$\frac{1}{4} < \frac{1}{3}?$
$V_{SM,IP}$
Student Model
Instructional Policy

	Policy 1	Policy 2	Policy 3
Student Model 1	V_{SM_1,IP_1}	V_{SM_1,IP_2}	V_{SM_1,IP_3}
Student Model 2	V_{SM_2,IP_1}	V_{SM_2,IP_2}	V_{SM_2,IP_3}
Student Model 3	V_{SM_3,IP_1}	V_{SM_3,IP_2}	V_{SM_3,IP_3}
Student Model 4	V_{SM_4,IP_1}	V_{SM_4,IP_2}	V_{SM_4,IP_3}

CASE STUDY: FRACTIONS TUTOR







What goes wrong if we simulate on only a single student model?

- Used prior data to fit G-SCOPE Model (Hallak et al., 2015).

- Used prior data to fit G-SCOPE Model (Hallak et al., 2015).
- Used G-SCOPE Model to derive new Adaptive Policy.

- Used prior data to fit G-SCOPE Model (Hallak et al., 2015).
- Used G-SCOPE Model to derive new Adaptive Policy.
- Wanted to compare Adaptive Policy to a Baseline Policy (fixed, spiraling curriculum).

- Used prior data to fit G-SCOPE Model (Hallak et al., 2015).
- Used G-SCOPE Model to derive new Adaptive Policy.
- Wanted to compare Adaptive Policy to a Baseline Policy (fixed, spiraling curriculum).
- Simulated both policies on G-SCOPE Model to predict posttest scores (out of 16 points).

	Baseline	Adaptive Policy
Simulated Results	5.9 ± 0.9	9.1 ± 0.8

	Baseline	Adaptive Policy
Simulated Results	5.9 ± 0.9	9.1 ± 0.8
Experimental Results	5.5 ± 2.6	4.9 ± 1.8

- Used by Chi et al. (2011) and Rowe et al. (2014) in educational settings.

- Used by Chi et al. (2011) and Rowe et al. (2014) in educational settings.
- Rowe et al. (2014): New instructional policy estimated to be much better than random policy.

- Used by Chi et al. (2011) and Rowe et al. (2014) in educational settings.
- Rowe et al. (2014): New instructional policy estimated to be much better than random policy.
- But in experiment, no significant difference found (Rowe and Lester, 2015).

SINGLE MODEL SIMULATION

- Used by Chi et al. (2011) and Rowe et al. (2014) in educational settings.
- Rowe et al. (2014): New instructional policy estimated to be much better than random policy.
- But in experiment, no significant difference found (Rowe and Lester, 2015).
- If each policy is only simulated on the student model that was used to derive it, a sub-optimal policy might be predicted to be better than the optimal policy under the true student model (Mandel et al., 2014).

- Used by Chi et al. (2011) and Rowe et al. (2014) in educational settings.
- Rowe et al. (2014): New instructional policy estimated to be much better than random policy.
- But in experiment, no significant difference found (Rowe and Lester, 2015).
- If each policy is only simulated on the student model that was used to derive it, a sub-optimal policy might be predicted to be better than the optimal policy under the true student model (Mandel et al., 2014).
    - Even with an infinite amount of data!

	Baseline	Adaptive Policy
G-SCOPE Model	5.9 ± 0.9	9.1 ± 0.8

	Baseline	Adaptive Policy
G-SCOPE Model	5.9 ± 0.9	9.1 ± 0.8
Bayesian Knowledge Tracing Model	6.5 ± 0.8	7.0 ± 1.0

CASE STUDY: ROBUST EVALUATION MATRIX

	Baseline	Adaptive Policy
G-SCOPE Model	5.9 ± 0.9	9.1 ± 0.8
Bayesian Knowledge Tracing Model	6.5 ± 0.8	7.0 ± 1.0
Deep Knowledge Tracing Model	9.9 ± 1.5	8.6 ± 2.1

	Baseline	Adaptive Policy	*Awesome Policy*
G-SCOPE Model	5.9 ± 0.9	9.1 ± 0.8	*16*
Bayesian Knowledge Tracing Model	6.5 ± 0.8	7.0 ± 1.0	*16*
Deep Knowledge Tracing Model	9.9 ± 1.5	8.6 ± 2.1	*16*

See paper for another retrospective analysis of prior work (Rafferty et al., 2015) showing REM can be used:

See paper for another retrospective analysis of prior work (Rafferty et al., 2015) showing REM can be used:

- to find good policies, robust to the choice of the model

See paper for another retrospective analysis of prior work (Rafferty et al., 2015) showing REM can be used:

- to find good policies, robust to the choice of the model
- to spot bad policies that single model simulation could not catch

- Simulating instructional policies on various student models can help determine what policy is best (if any) in advance of running an experiment.

- Simulating instructional policies on various student models can help determine what policy is best (if any) in advance of running an experiment.
- Showed retrospective analysis of how REM could have been used to inform experiment.

- Simulating instructional policies on various student models can help determine what policy is best (if any) in advance of running an experiment.
- Showed retrospective analysis of how REM could have been used to inform experiment.
- Next Step: Close the loop.

Use many models we expect to be wrong, rather than using one model we hope to be right.

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grants R305A130215 and R305B150008 to Carnegie Mellon University. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Dept. of Education.

REFERENCES

Chi, M., VanLehn, K., Litman, D., and Jordan, P. (2011). Empirically evaluating the application of reinforcement learning to the induction of effective and adaptive pedagogical strategies. *User Modeling and User-Adapted Interaction*, 21(1-2):137--180.

Hallak, A., Schnitzler, C. F., Mann, T., and Mannor, S. (2015). Off-policy model-based learning under unknown factored dynamics. In *Proceedings of the 32nd International Conference on Machine Learning (ICML-15)*, pages 711--719.

Mandel, T., Liu, Y.-E., Levine, S., Brunskill, E., and Popovic, Z. (2014). Offline policy evaluation across representations with applications to educational games. In *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems*, pages 1077--1084. International Foundation for Autonomous Agents and Multiagent Systems.

Rafferty, A. N., Brunskill, E., Griffiths, T. L., and Shafto, P. (2015). Faster teaching via pomdp planning. *Cognitive Science*.

Rowe, J. P. and Lester, J. C. (2015). Improving student problem solving in narrative-centered learning environments: A modular reinforcement learning framework. In *International Conference on Artificial Intelligence in Education*, pages 419--428. Springer.

Rowe, J. P., Mott, B. W., and Lester, J. C. (2014). Optimizing player experience in interactive narrative planning: A modular reinforcement learning approach. In *AIIDE*.

- Estimator that gives unbiased and consistent estimates for a policy!

- Estimator that gives unbiased and consistent estimates for a policy!
- Can have very high variance when policy is different from prior data.

- Estimator that gives unbiased and consistent estimates for a policy!
- Can have very high variance when policy is different from prior data.
- Example: Worked example or problem-solving?

- Estimator that gives unbiased and consistent estimates for a policy!
- Can have very high variance when policy is different from prior data.
- Example: Worked example or problem-solving?
    - 20 sequential decisions $\Rightarrow$ need over 2^{20} students

- Estimator that gives unbiased and consistent estimates for a policy!
- Can have very high variance when policy is different from prior data.
- Example: Worked example or problem-solving?
    - 20 sequential decisions $\Rightarrow$ need over 2^{20} students
    - 50 sequential decisions $\Rightarrow$ need over 2^{50} students!