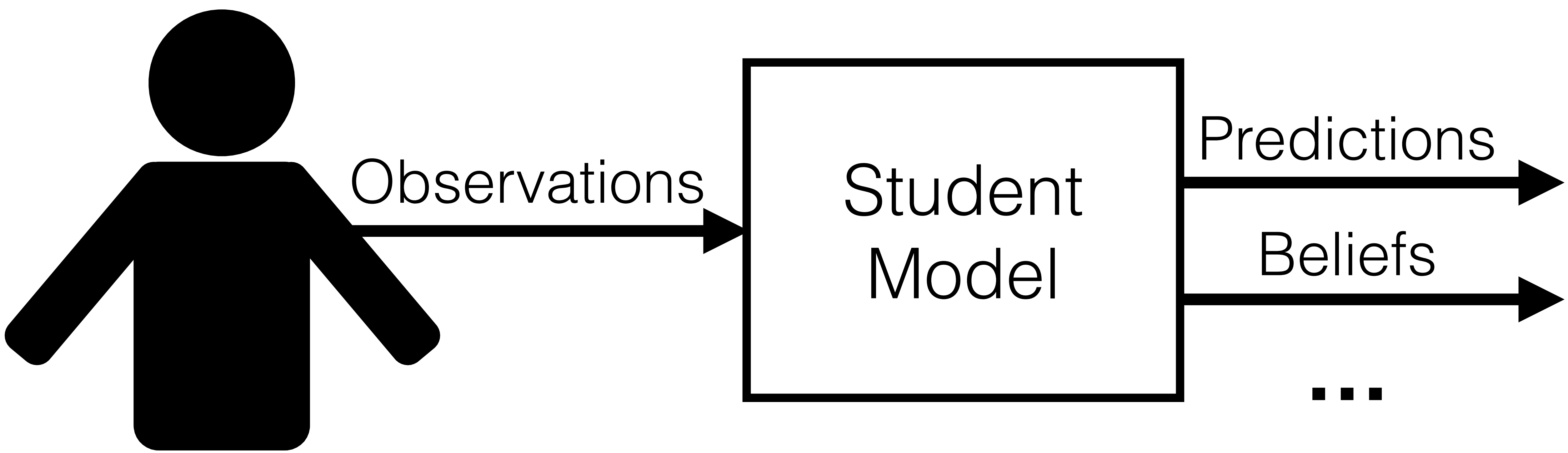
# From Predictive Models to Instructional Policies

Joseph Rollinson (jtrollinson@gmail.com)
Emma Brunskill (ebrun@cs.cmu.edu)

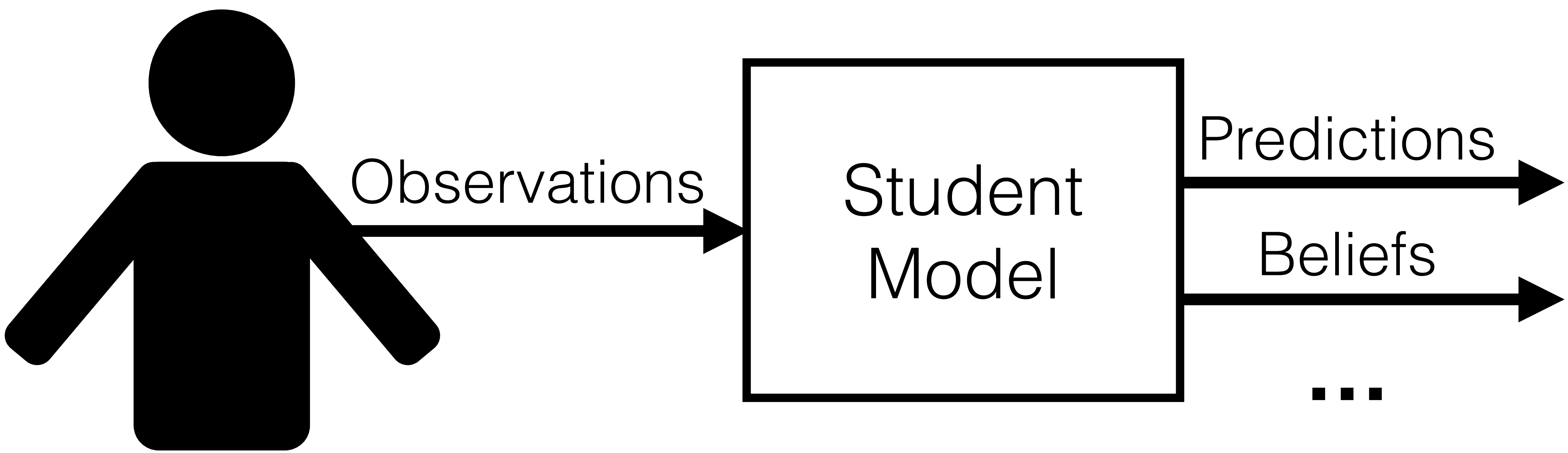
Carnegie Mellon

# Student models are a representation of the student




Corbett et al. 1994, Cen et al. 2006, Pavlik et al. 2009, Chi et al. 2011, Khajah 2014, Gong 2010, Pardos 2010, Falakmasir 2013

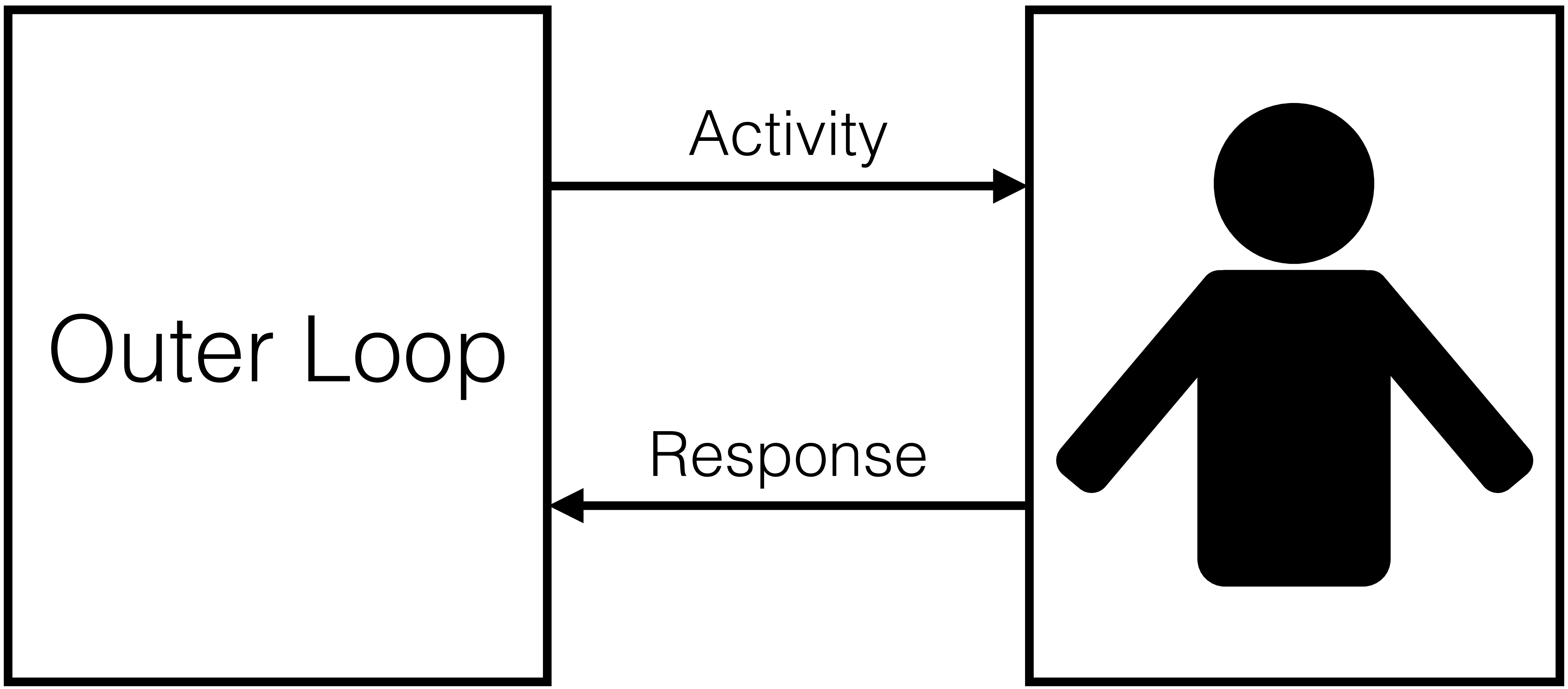
# Student models are a representation of the student



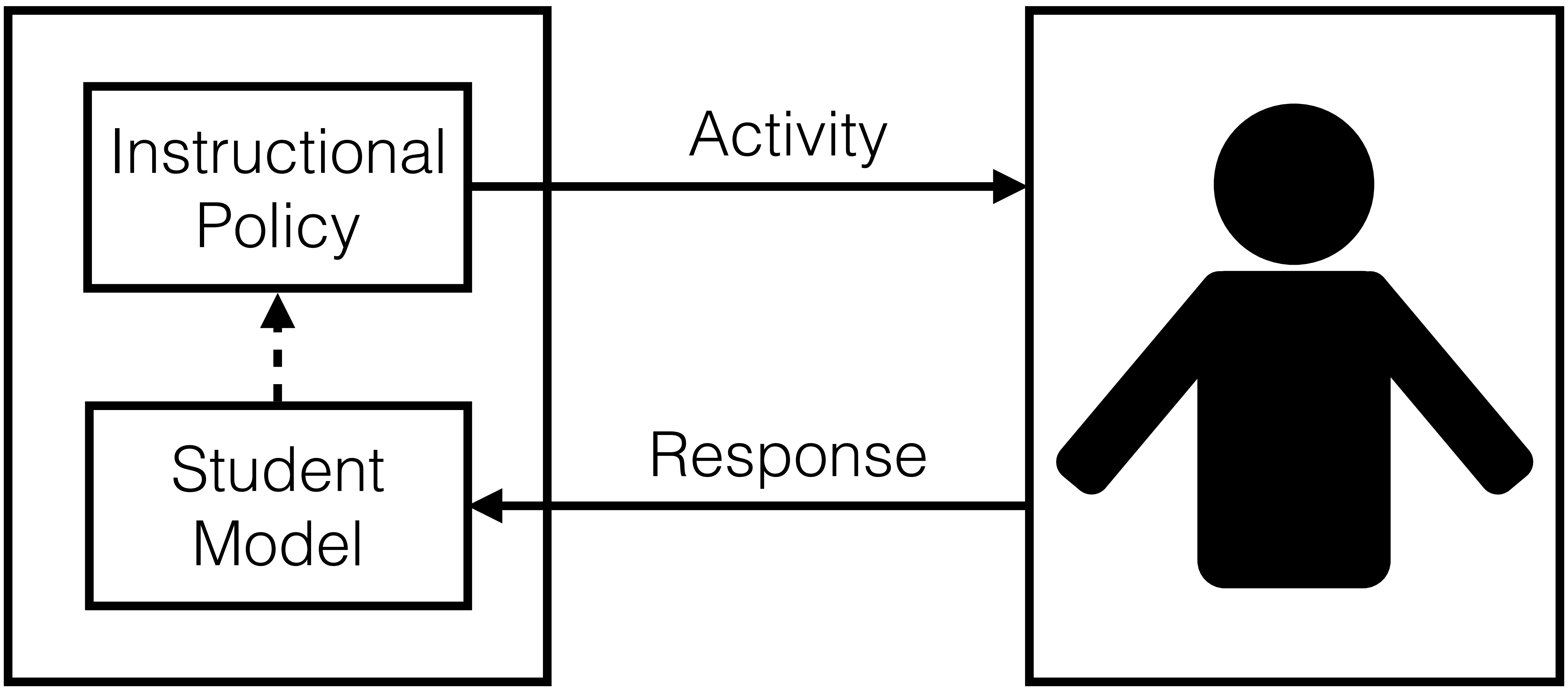

Much prior work building student models for predicting future student performance

Corbett et al. 1994, Cen et al. 2006, Pavlik et al. 2009, Chi et al. 2011, Khajah 2014, Gong 2010, Pardos 2010, Falakmasir 2013

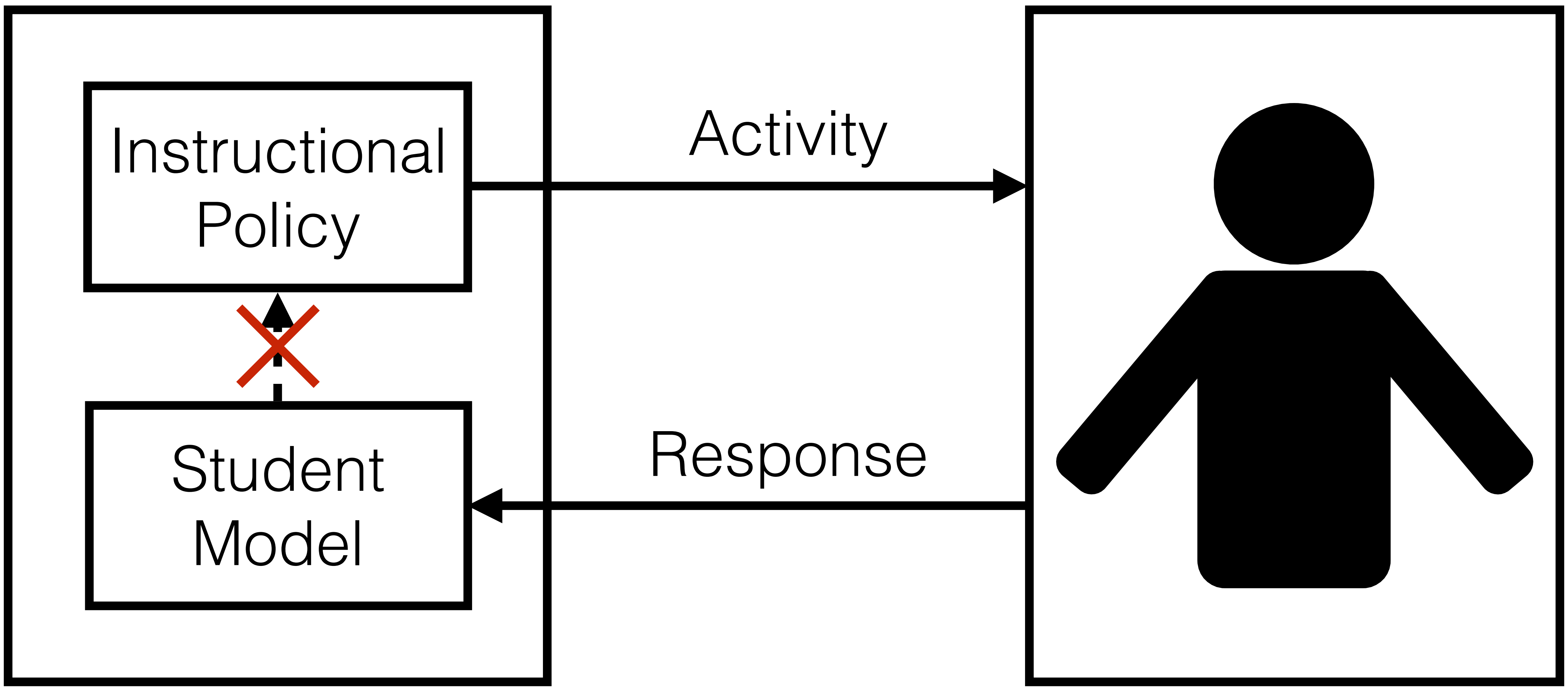
# Student models are also used with outer-loop instructional policies



# Student models are also used with outer-loop instructional policies



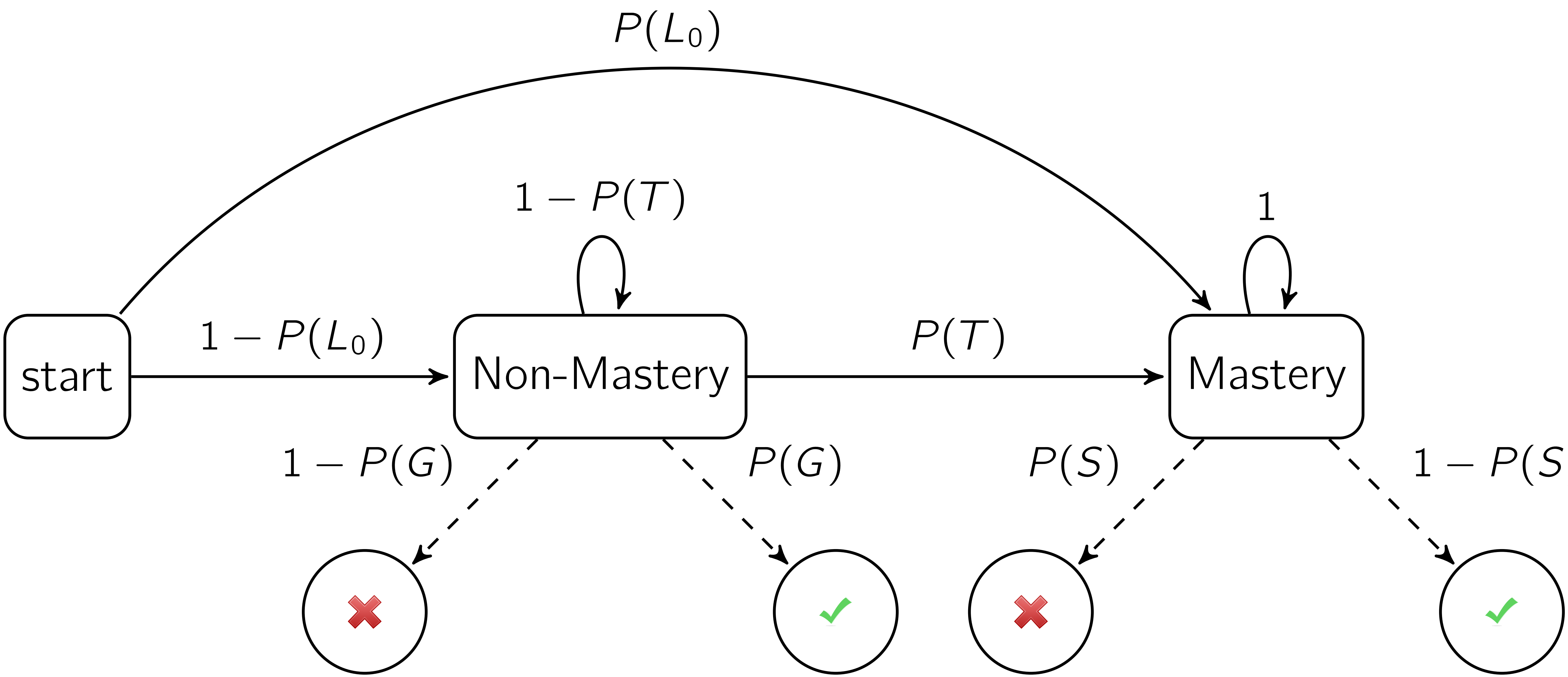
# Many predictive student models cannot be used with any existing instructional policy



# Contribution

Model agnostic instructional policy for the when-to-stop decision problem

# *Background*
# Bayesian Knowledge Tracing




Corbett, A. T., & Anderson, J. R. (1995). Knowledge tracing: modeling the acquisition of procedural knowledge. User Modeling and User-Adapted Interaction, 4, 253–278

## *Background*
## Performance Factors Model (PFM)

Logistic model for predicting student performance

**Features**

- Student (i)
- Skill (k)
- # Correct responses for skill (s)
- # Incorrect responses for skill (f)

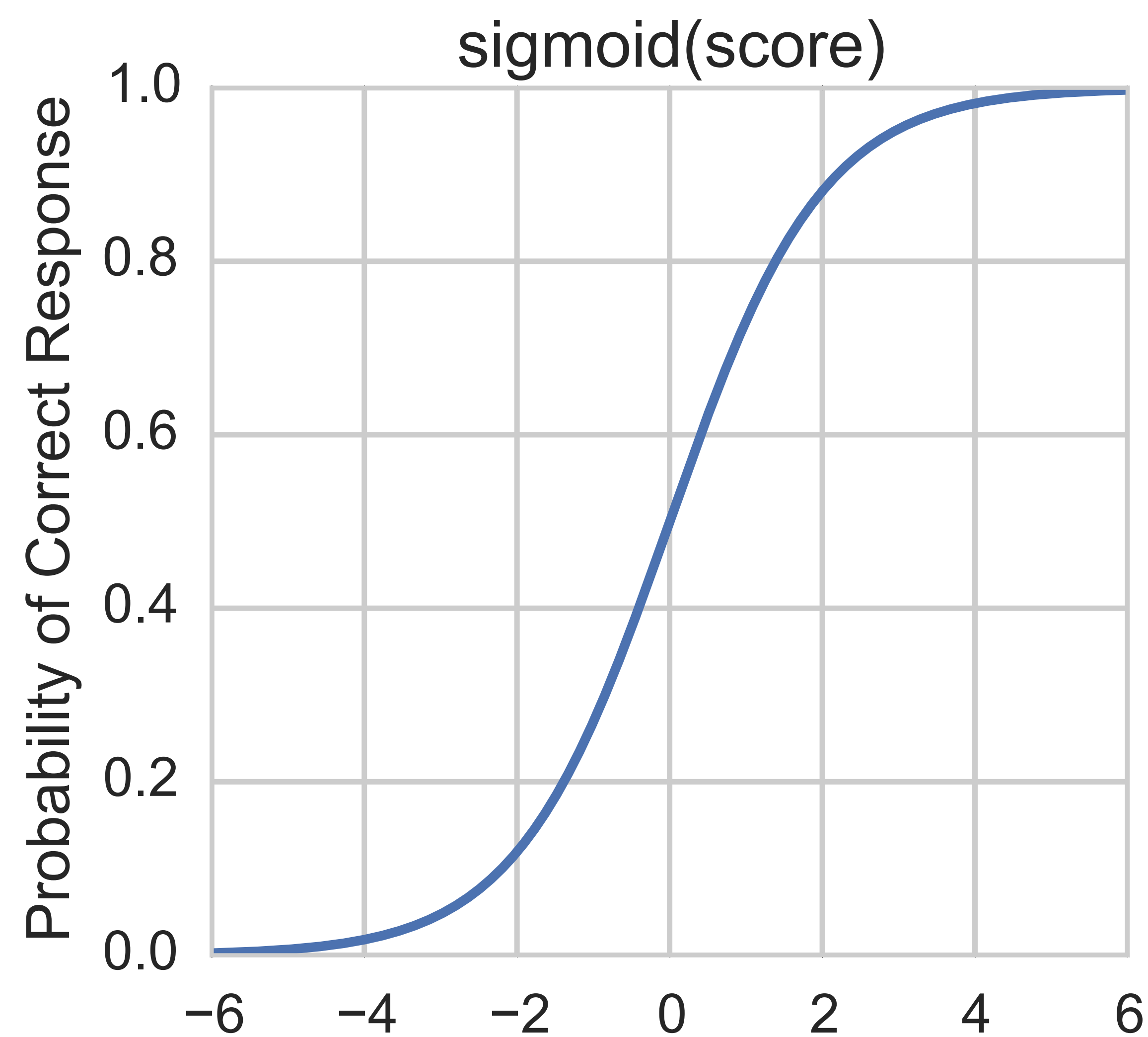
Cen et al. 2006, Pavlik et al. 2009, Chi et al. 2011

# *Background*
# Performance Factors Model (PFM)

Logistic model for predicting student performance

**Features**

- Student (i)
- Skill (k)
- # Correct responses for skill (s)
- # Incorrect responses for skill (f)




score $= \alpha_i + \beta_k + \mu_k s + \rho_k f$

Cen et al. 2006, Pavlik et al. 2009, Chi et al. 2011

# When-To-Stop Decision Problem

**Situation:** Teaching single skill with indistinguishable activities

**Observations:** Correctness of student responses

**Decision:** When to stop providing activities to student

# *Prior Work*
# Mastery Threshold Policy

Stop if we are confident that the student has mastered the skill

# *Prior Work*
# Mastery Threshold Policy

Stop if we are confident that the student has mastered the skill

$$P(M) > \Delta$$

# Issues with the Mastery Threshold Policy

1. Requires student model with concept of mastery

2. Will not stop if student cannot progress with given instruction (wheel-spinning)

Beck, Joseph E., and Yue Gong. "Wheel-spinning: Students who fail to master a skill." Artificial Intelligence in Education. Springer Berlin Heidelberg, 2013.

# *New Policy*
# Predictive Similarity Policy

Stop if we are confident that the student model's prediction of the student's performance will not change very much if the student is given another question

# *New Policy*
# Predictive Similarity Policy

Stop if we are confident that the student model's prediction of the student's performance will not change very much if the student is given another question

$$\Pr\left(\left|P_{t+1}(C) - P_t(C)\right| < \epsilon\right) > \delta$$

$$\Pr\left(\left|P_{t+1}(C) - P_t(C)\right| < \epsilon\right) > \delta$$

**3 Stopping Conditions:**

$$\Pr\left(\left|P_{t+1}(C) - P_t(C)\right| < \epsilon\right) > \delta$$

## 3 Stopping Conditions:

| | |
|---|---|
| $P_t(C) > \delta$ | Confident that student will respond correctly. |
| $\mid P_{t+1}(C) - P_t(C \mid C_t) \mid < \varepsilon$ | Prediction does not change much if student responds correctly. |

$$\Pr\left(\left|P_{t+1}(C) - P_t(C)\right| < \epsilon\right) > \delta$$

## 3 Stopping Conditions:

| | |
|---|---|
| $P_t(C) > \delta$<br><br>$\mid P_{t+1}(C) - P_t(C \mid C_t) \mid < \varepsilon$ | Confident that student will respond correctly.<br>Prediction does not change much if student responds correctly. |
| $P_t(\neg C) > \delta$<br><br>$\mid P_{t+1}(C) - P_t(C \mid \neg C_t) \mid < \varepsilon$ | Confident that student will respond incorrectly.<br>Prediction does not change much if student responds incorrectly. |

$$\Pr\left(\left|P_{t+1}(C) - P_t(C)\right| < \epsilon\right) > \delta$$

## 3 Stopping Conditions:

| | |
|---|---|
| $P_t(C) > \delta$<br><br>$\mid P_{t+1}(C) - P_t(C \mid C_t) \mid < \varepsilon$ | Confident that student will respond correctly.<br>Prediction does not change much if student responds correctly. |
| $P_t(\neg C) > \delta$<br><br>$\mid P_{t+1}(C) - P_t(C \mid \neg C_t) \mid < \varepsilon$ | Confident that student will respond incorrectly.<br>Prediction does not change much if student responds incorrectly. |
| $\mid P_{t+1}(C) - P_t(C \mid C_t) \mid < \varepsilon$<br><br>$\mid P_{t+1}(C) - P_t(C \mid \neg C_t) \mid < \varepsilon$ | Prediction does not change much no matter how the student's observation. |

# *Experiments*
# Methodology

1. Train student models on data set

2. Calculate expected amount of practice for each skill in dataset using instructional policy and student model

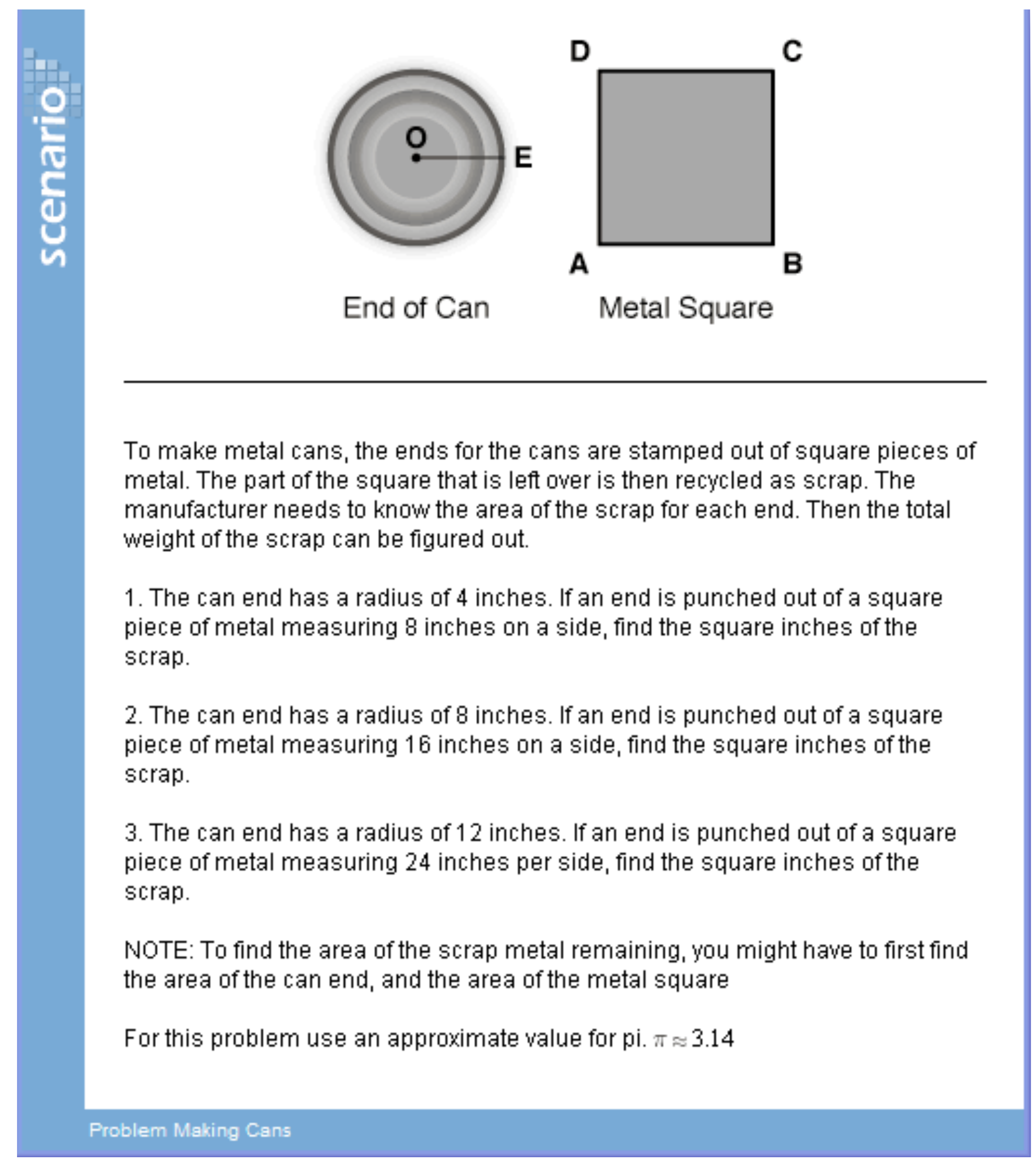
3. Compare expected amount of practice per skill

# *Dataset*
# KDD Cup Algebra I

> 3000 students

505 skills

BKT and PFM have similar predictive accuracy




J. Stamper, A. Niculescu-Mizil, S. Ritter, G. Gordon, and K. Koedinger. Algebra 1 2008-2009. challenge data set from kdd cup 2010 educational data mining challenge. find it at http://pslcdatashop.web.cmu.edu/kddcup/downloads.jsp.

# Expected Amount of Practice (ExpOps)

Metric of the number of questions given to students by a policy with a given student model.

J. I. Lee and E. Brunskill. The impact on individualizing student models on necessary practice opportunities. In EDM, 2012.

# Expected Amount of Practice (ExpOps)

Metric of the number of questions given to students by a policy with a given student model.

Comparison, **not** a measure of quality

J. I. Lee and E. Brunskill. The impact on individualizing student models on necessary practice opportunities. In EDM, 2012.

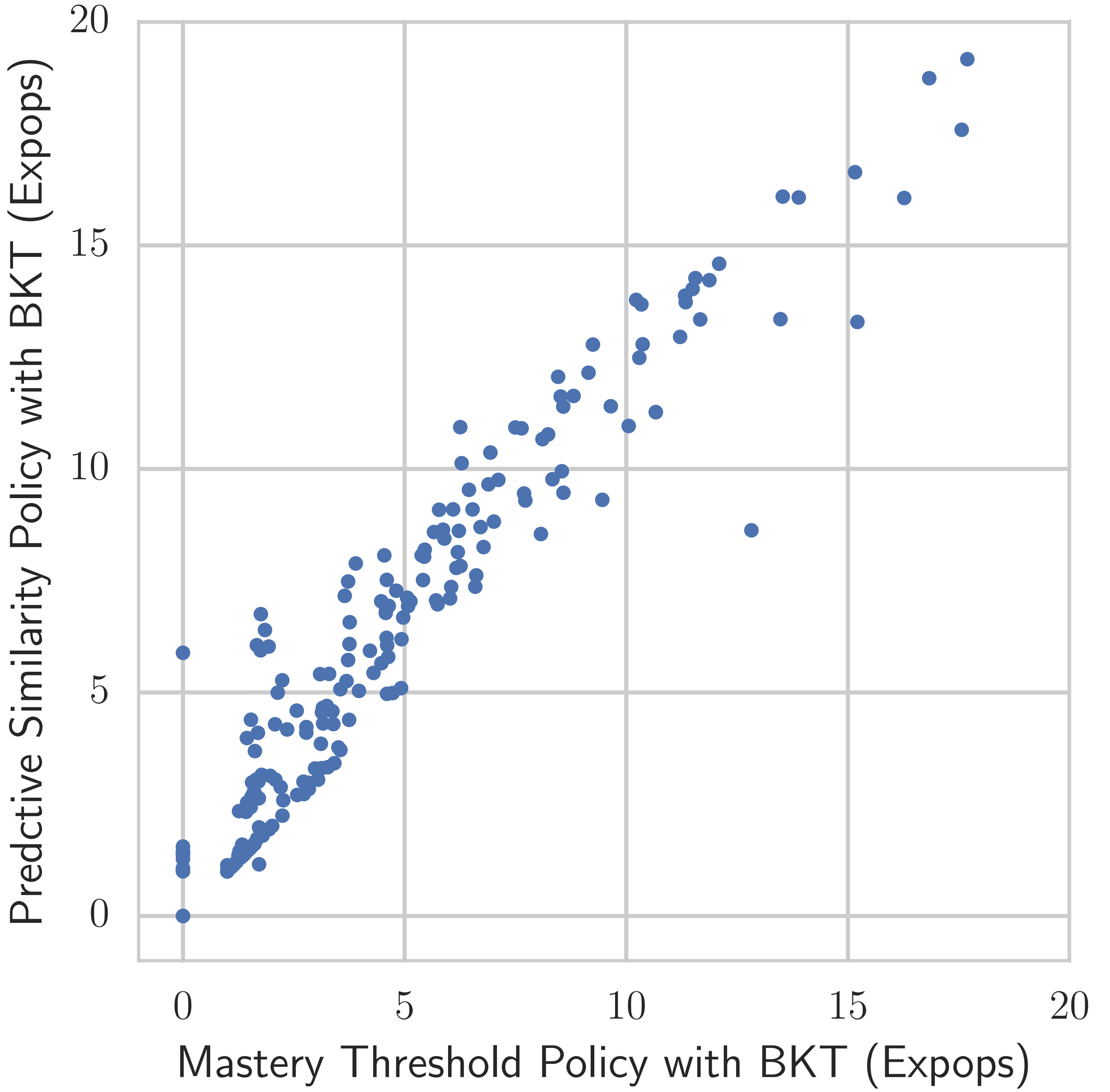
# *Experiment 1*
# Predictive Similarity vs. Mastery Threshold

1. Train *BKT* with EM for each skill in dataset

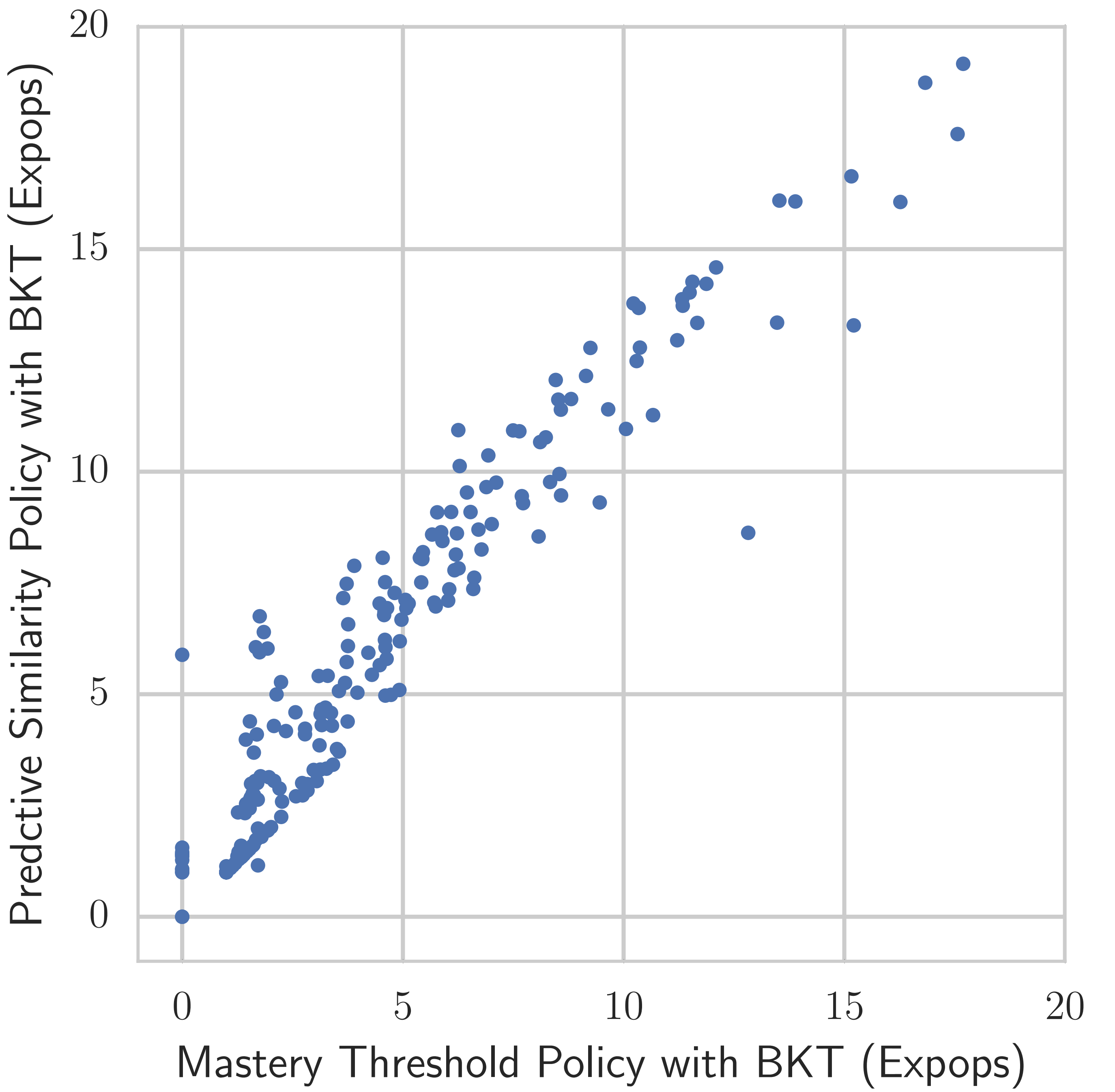
2. For each skill, calculate expected amount of practice using *Predictive Similarity* and *Mastery Threshold* policies with trained *BKTs*

3. Compare expected amount of practice on skills with non-degenerate BKTs

# *Experiment 1*
# Results



# *Experiment 1*
# Results




Predictive similarity policy makes similar decisions to mastery threshold policy (coef 0.95)

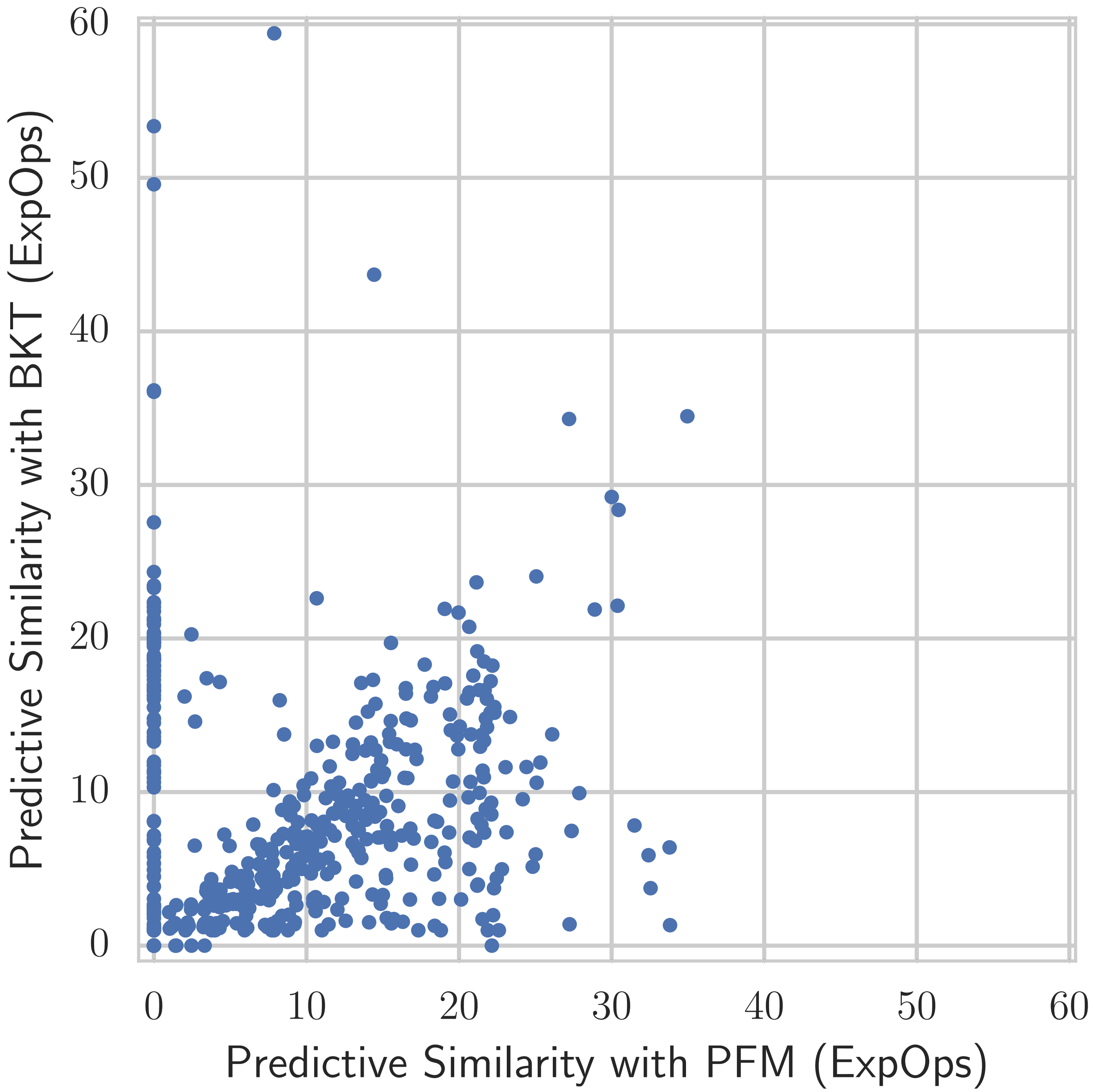
# *Experiment 2*
# BKT vs. PFM

1. Train *PFM* on KDD Cup dataset using logistic regression

2. Calculate expected amount of practice using *Predictive Similarity* policy with underlying *BKT* and *PFM* for each skill

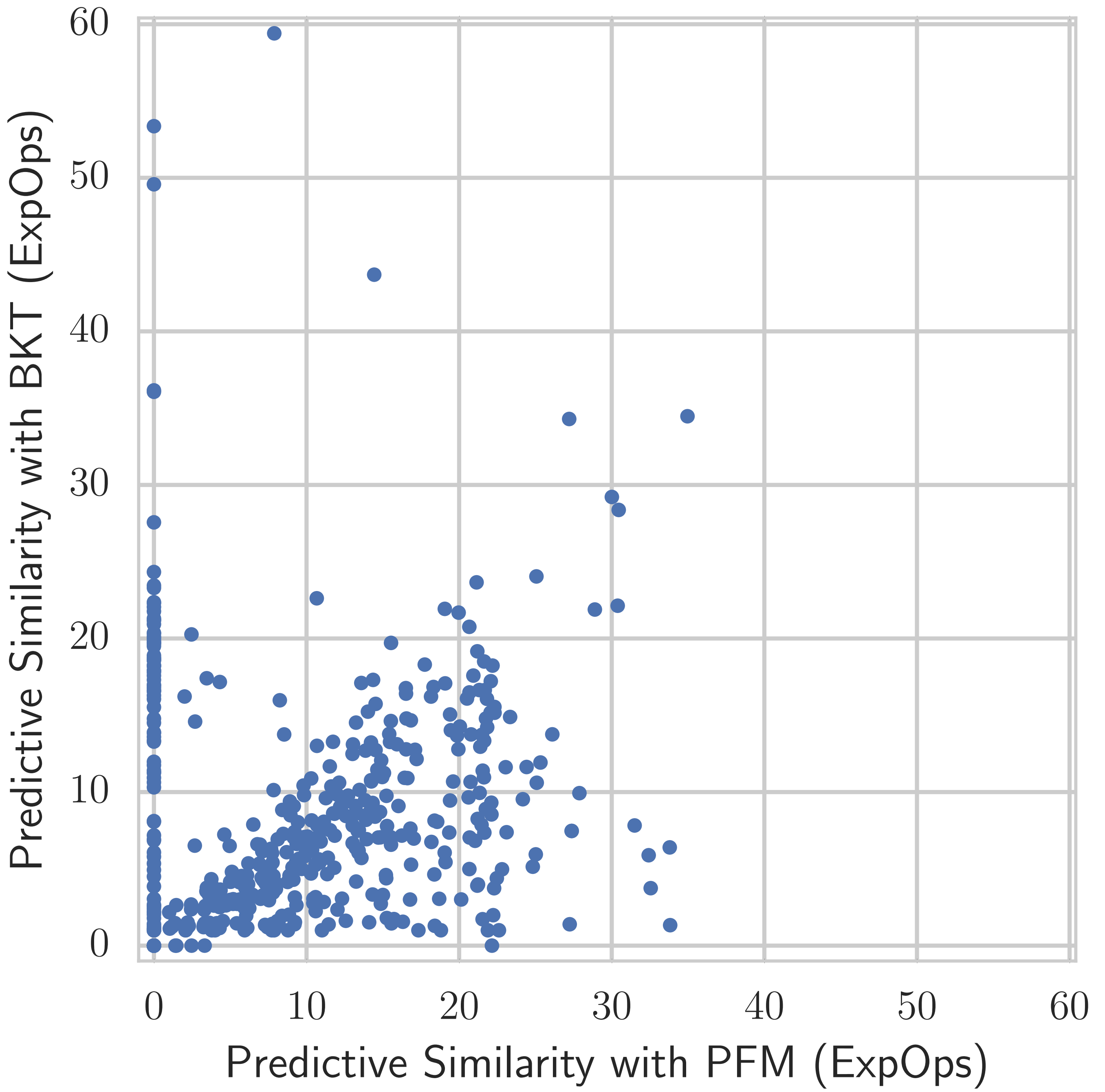
3. Compare expected amount of practice values

# PFM vs. BKT



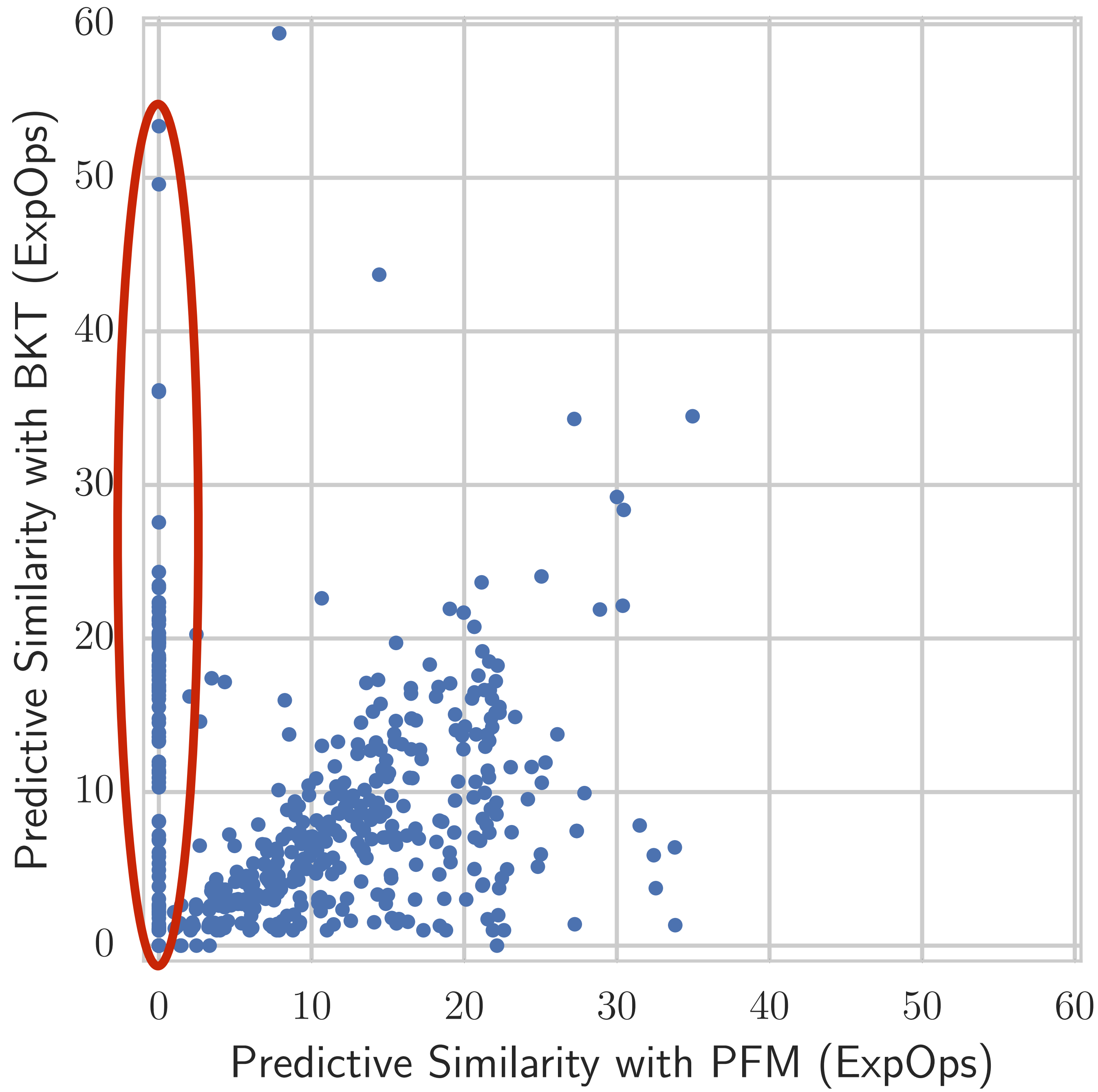
60
50
40
30
20
10
0
Predictive Similarity with BKT (ExpOps)
0
10
20
30
40
50
60
Predictive Similarity with PFM (ExpOps)

# PFM vs. BKT




PFM based policy either:

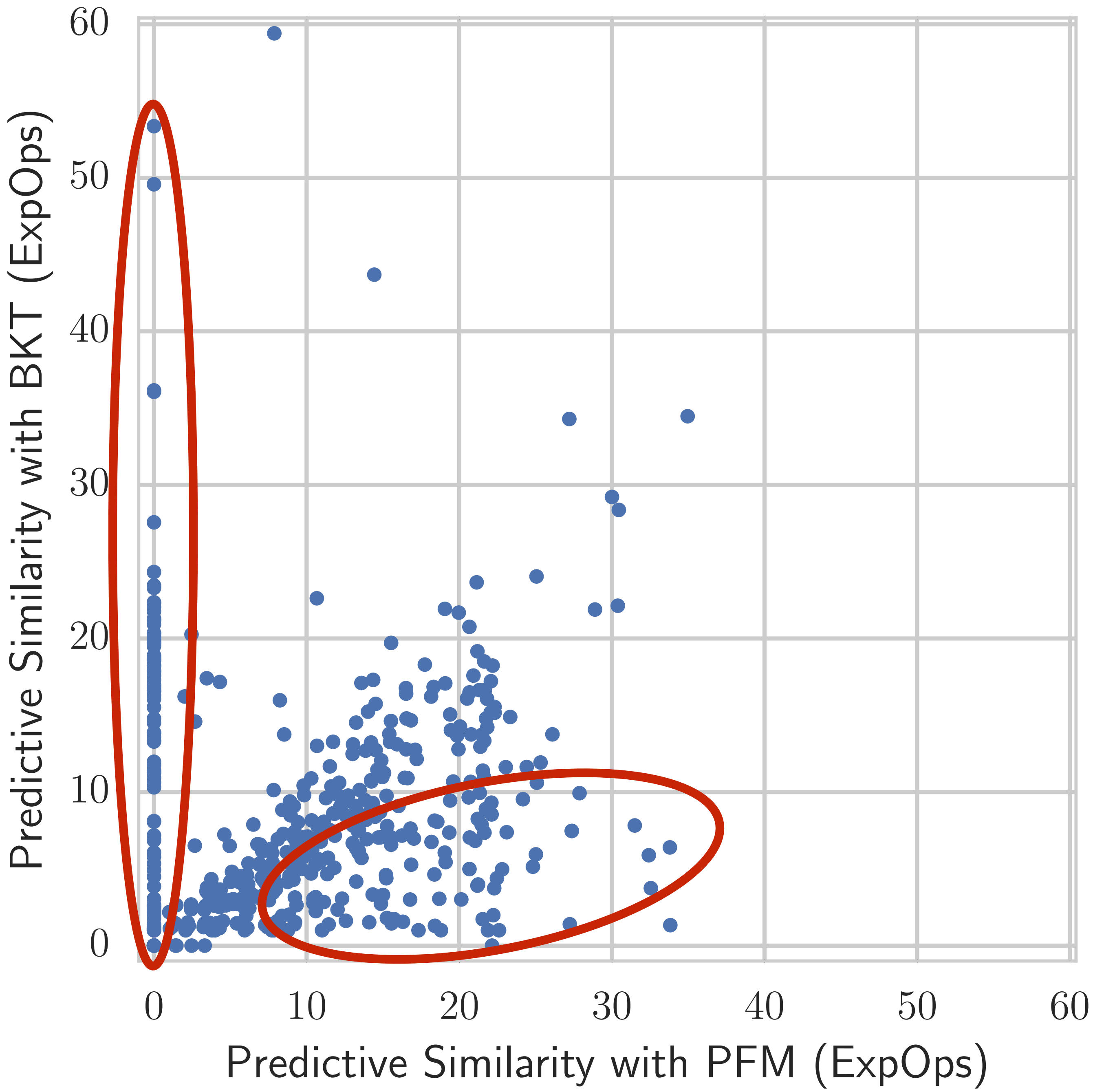
# PFM vs. BKT




PFM based policy either:

- Stops immediately

# PFM vs. BKT




PFM based policy either:

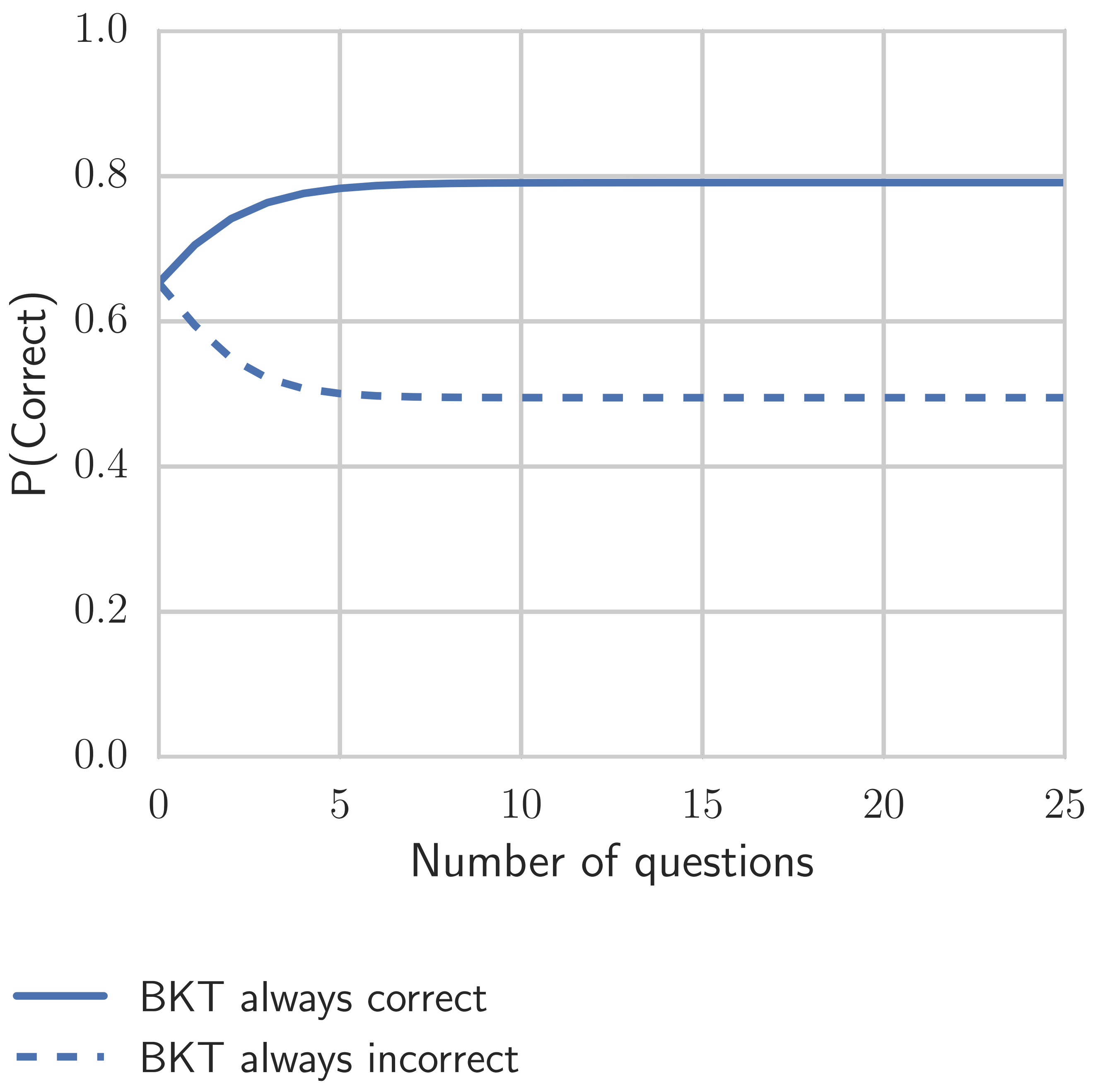
- Stops immediately
- Longer than BKT based policy

# *Diving In*
# Comparing BKT and PFM by skill

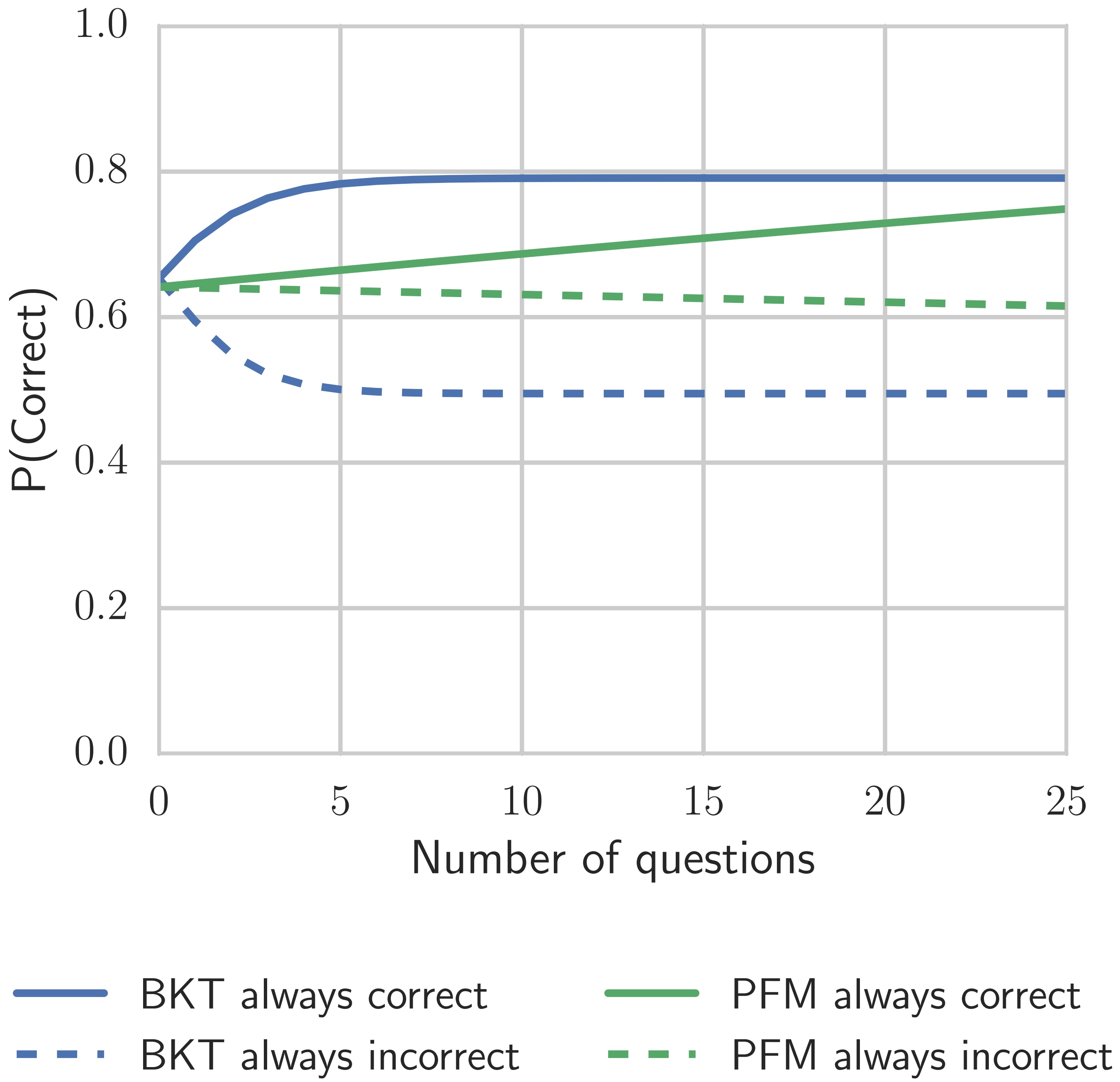
Calculate student model predictions for skill if:

- simulated student always responds *correctly*
- simulated student always responds *incorrectly*

# Skill: PFM Immediately stops

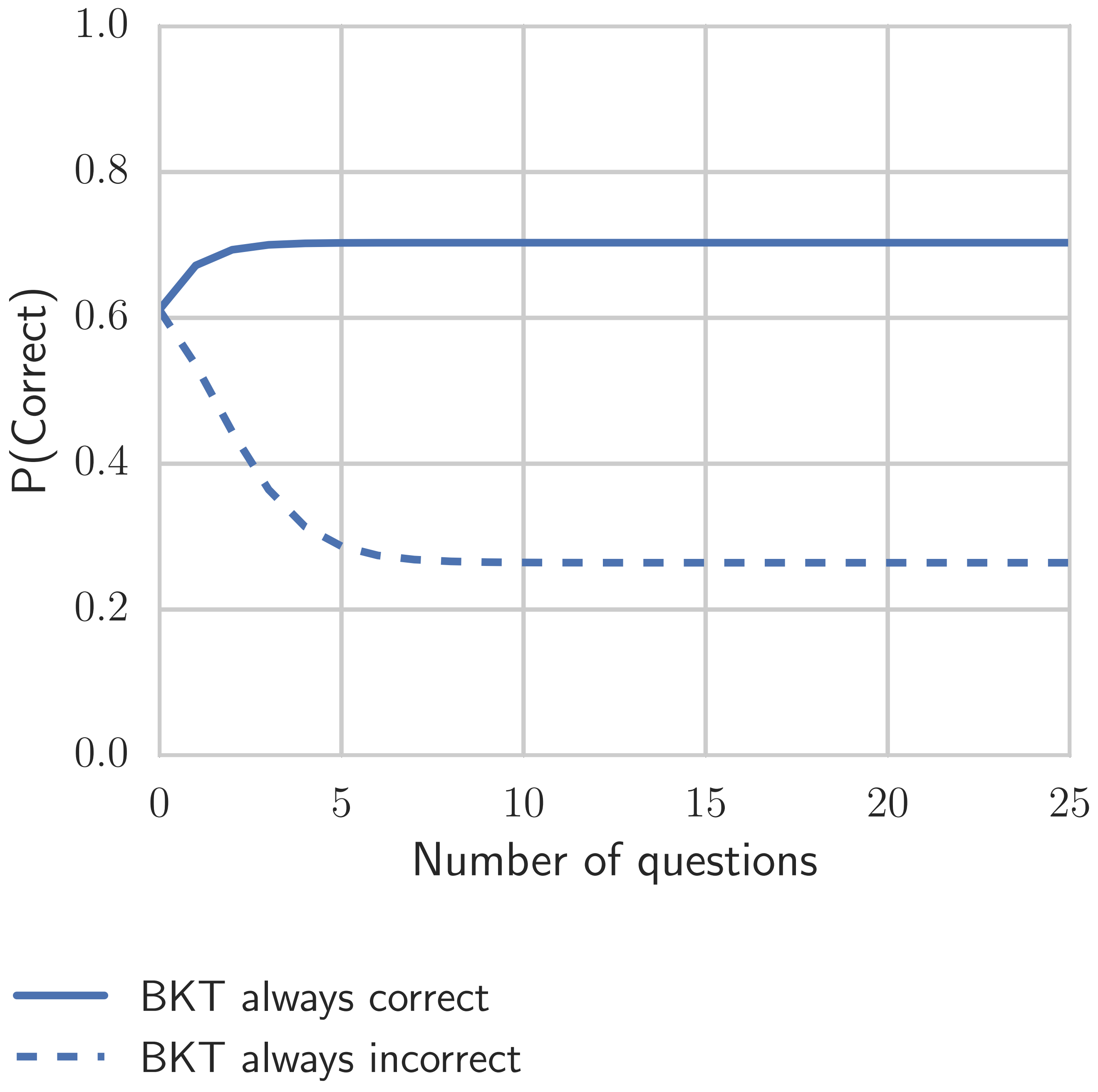


# Skill: PFM Immediately stops

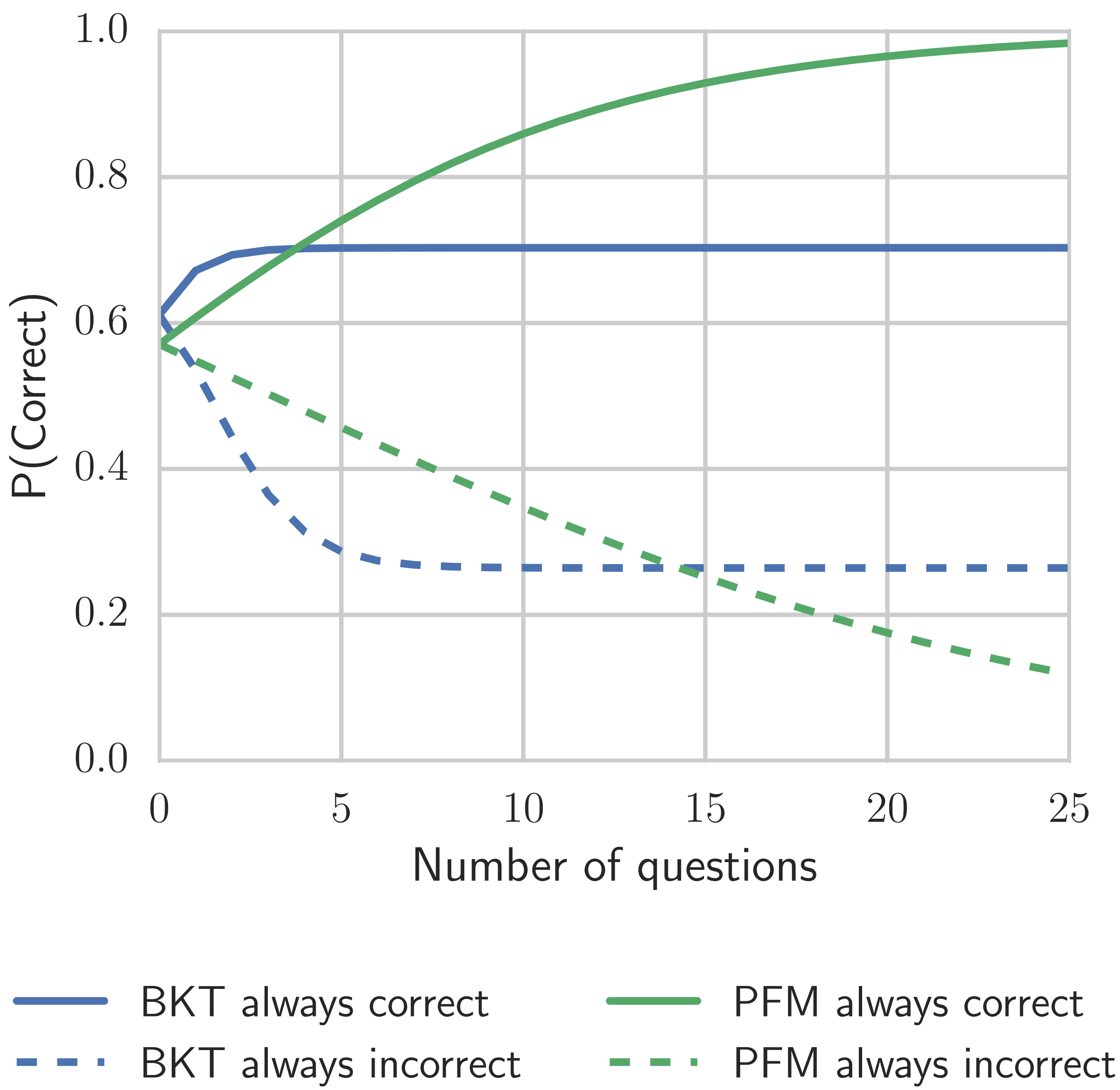



PFM predictions change very slowly.

# Skill: PFM longer than BKT



# Skill: PFM longer than BKT




PFM predictions asymptote much later than BKT predictions

# Discussion / Summary

- *Contribution*: a model-agnostic when-to-stop instructional policy called predictive similarity
- Predictive similarity policy acts like the mastery threshold policy when used with a BKT
- Models with similar predictive accuracies may lead to very different instructional behavior

# Future Work

- Perform experiments on another dataset
- Incorporating other observations into the predictive similarity policy
- Expanding predictive similarity policy to longer horizons
- Model agnostic instructional policies for more complicated instructional decisions (e.g. multiple skills)
- Method for evaluating policies

# Questions?