# Effort-based Tutoring: an Empirical Approach to Intelligent Tutoring

Ivon Arroyo[1], Hasmik Mehranian[2], Beverly P. Woolf[1]

[1] *Department of Computer Science*

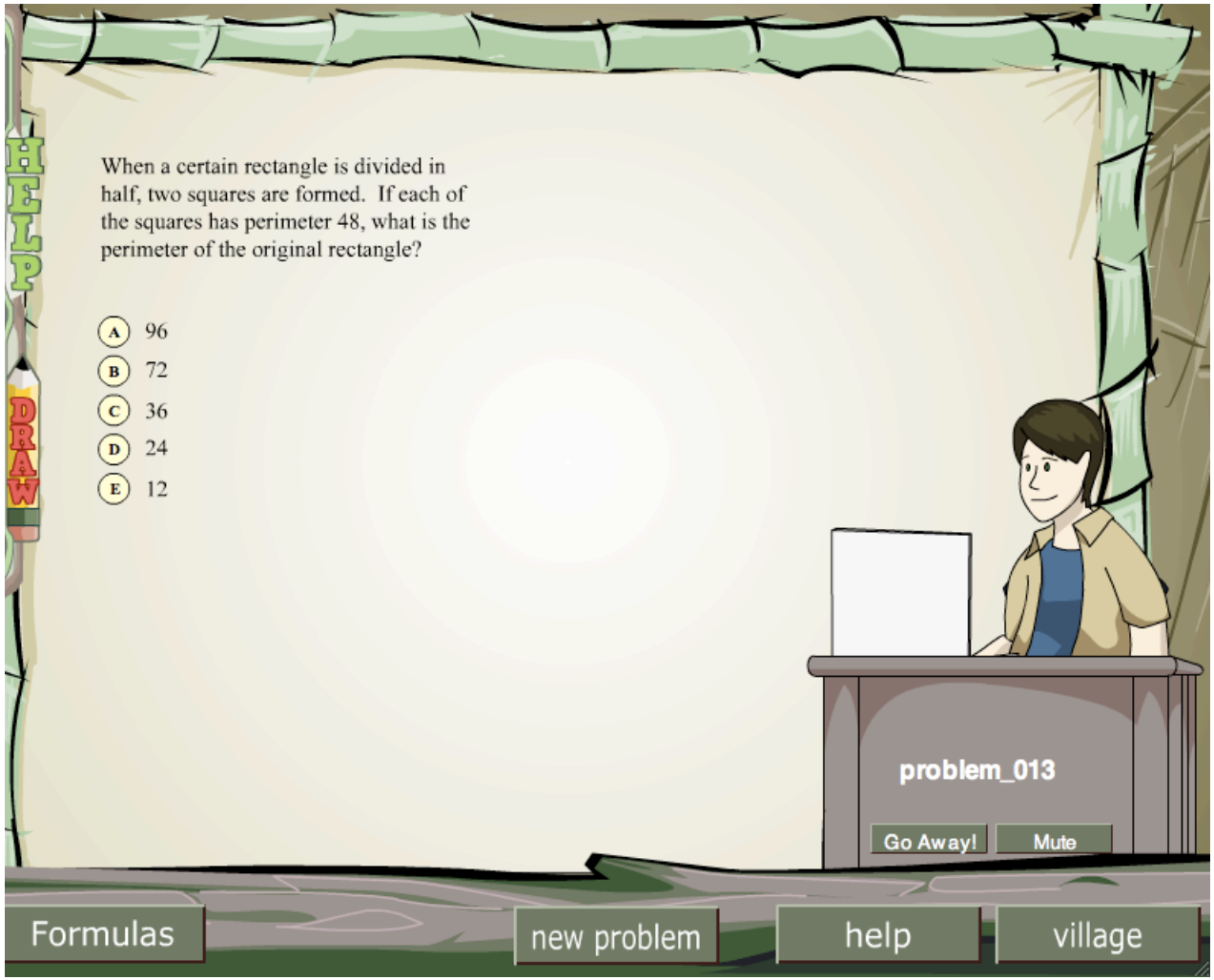
[2] *Dept. of Mechanical and Industrial Engineering*

*University of Massachusetts Amherst*

# Making Wayang Math Tutor Smarter

## From past student logs

*Multimedia Adaptive Tutoring System for Standardized-Tests Math Problems*



# What this paper is about

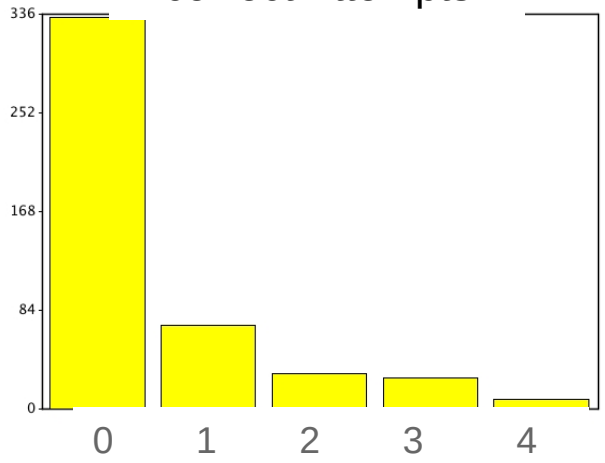
- **A description of how we made the Wayang Tutor Smarter**
  - From past student data
  - Given Content=Cases that students needed to know to solve
  - A wide range of students in public schools
- **A concrete procedure to add smartness to any ILE**
  - From past student data
  - Even for ill-defined domains
  - Regardless of amount of variety of content
  - High level of detail for replication
- **Unveiling parameters that regulate ITS functioning**
  - For optimization

# What do students do in Wayang?

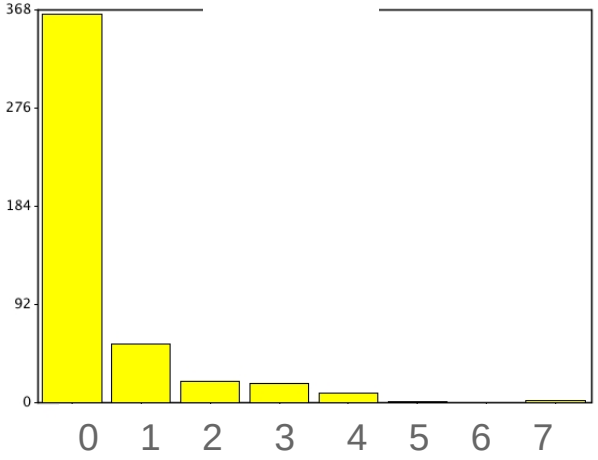
Three different ways to express their effort

Frequency of Behavior on one “easy” math problem in Wayang Outpost **484 cases**

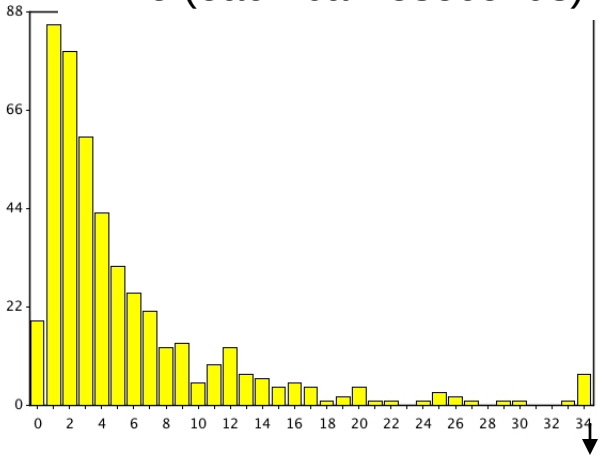
Incorrect Attempts



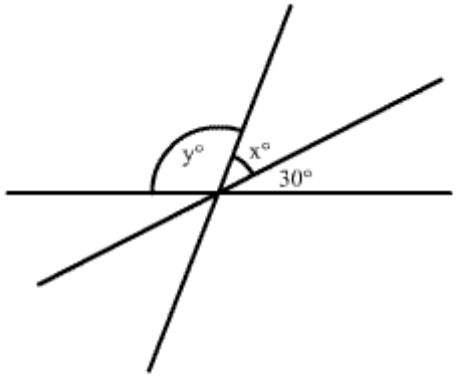

Hints




Time (each bar=5seconds)




>=2.8min




What is x + y?

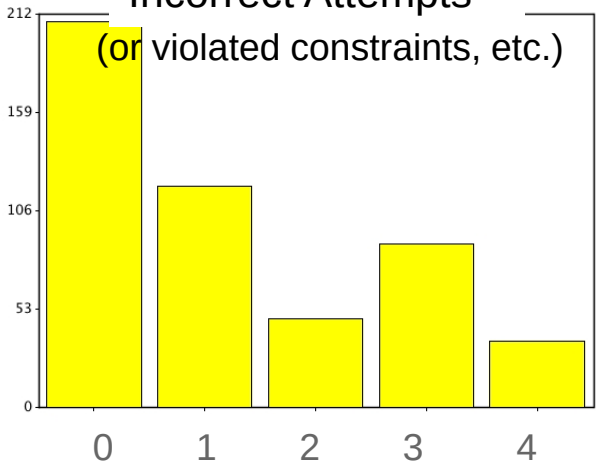
# What do students do in Wayang?

## Three different ways to express their effort

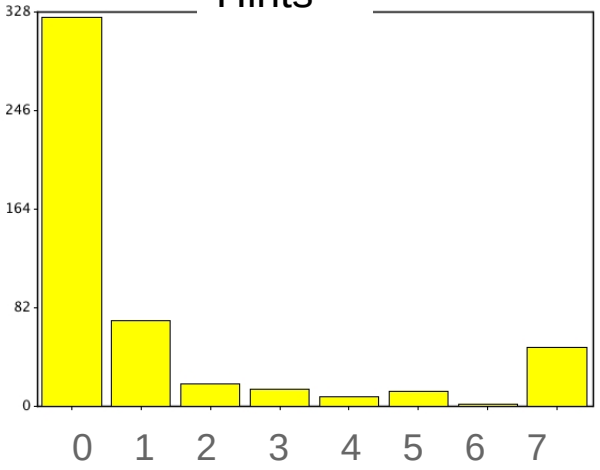
Frequency of Behavior on one “harder” math problem in Wayang Outpost

529 cases

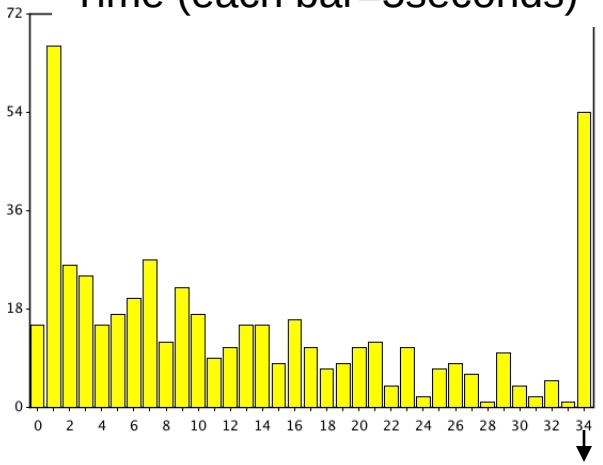
Incorrect Attempts
(or violated constraints, etc.)



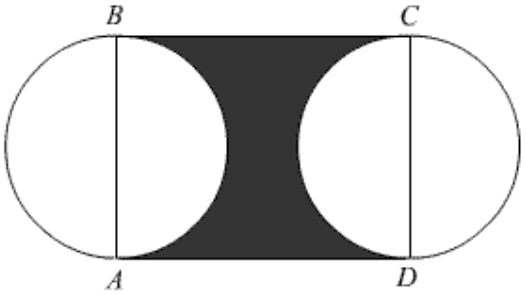

Hints




Time (each bar=5seconds)



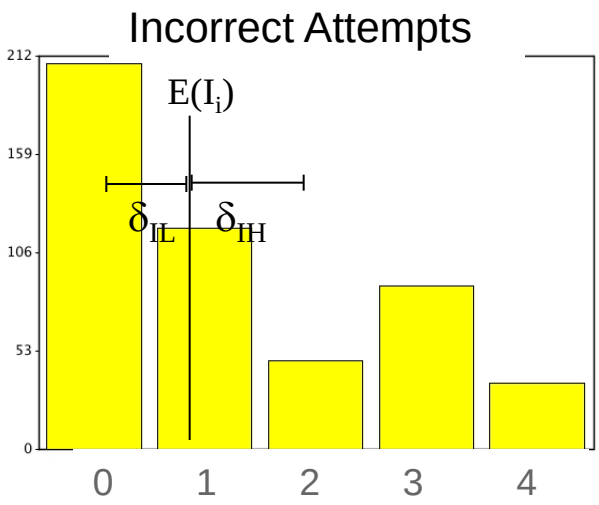

>=2.8min

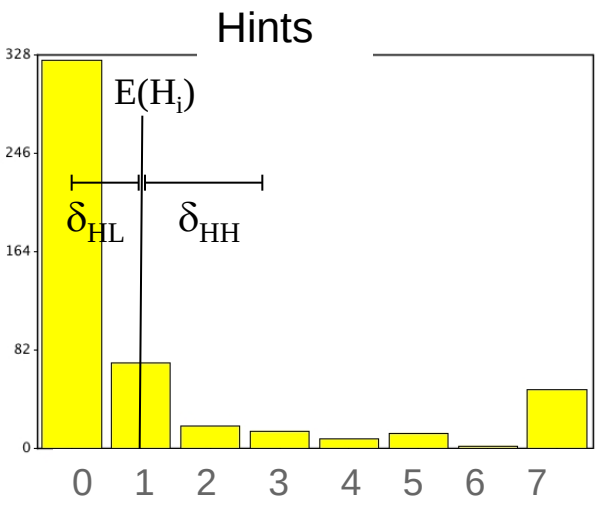


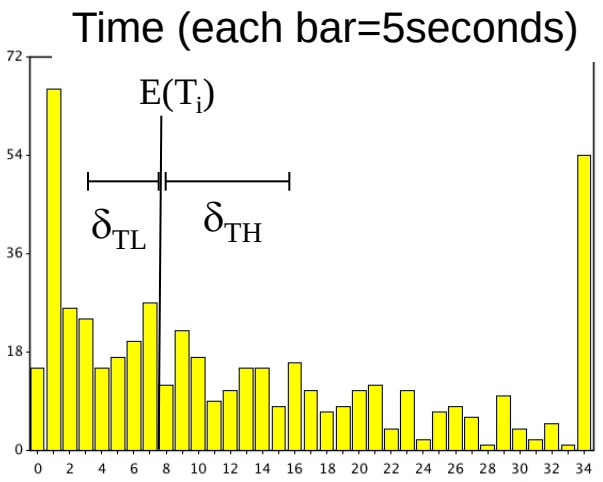

In rectangle $ABCD$, sides $AB$ and $CD$ pass through the centers of the two circles. If $AB=12$ and $AD=16$ what is the area of the shaded region?

# What is expected behavior?

In any problem $p_i$ i=1, .., N N=Total problems in system




$E(I_i)$ = mean (or median) Incorrect Attempts for problem $p_i$

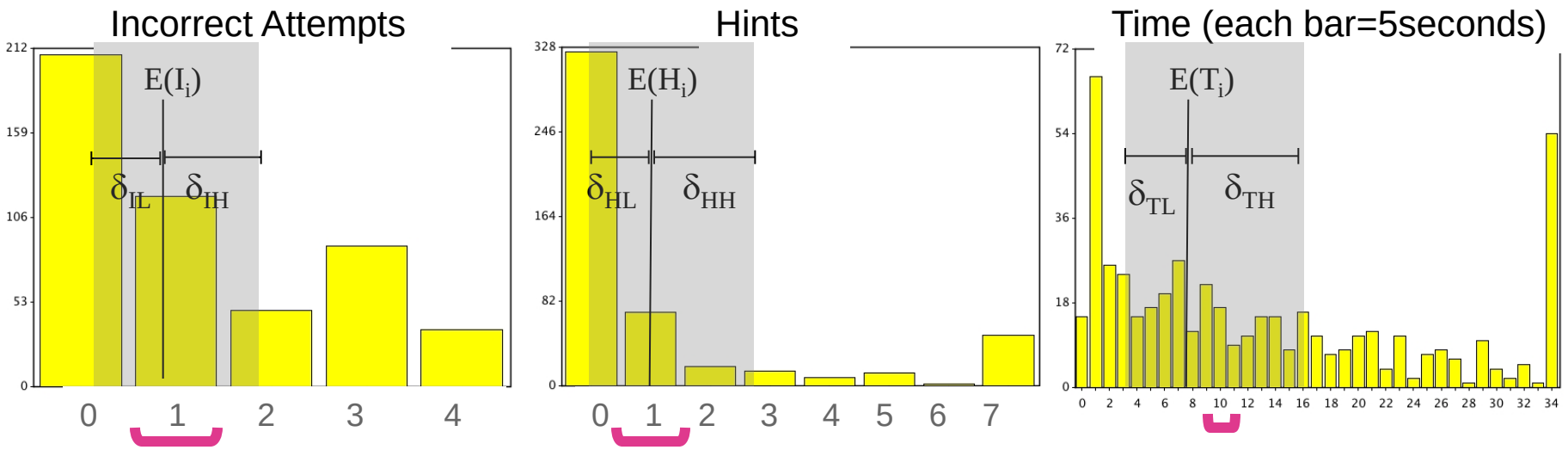
$\delta_{IH} = SD(I_i) * \theta_{HIGH}$ *e.g.* $\theta_{HIGH} = 0.5$

$\delta_{IL} = SD(I_i) * \theta_{LOW}$ *e.g.* $\theta_{LOW} = 0.25$

} Parameters

# What is expected behavior?

In any problem $p_i$ i=1, .., N N=Total problems in system

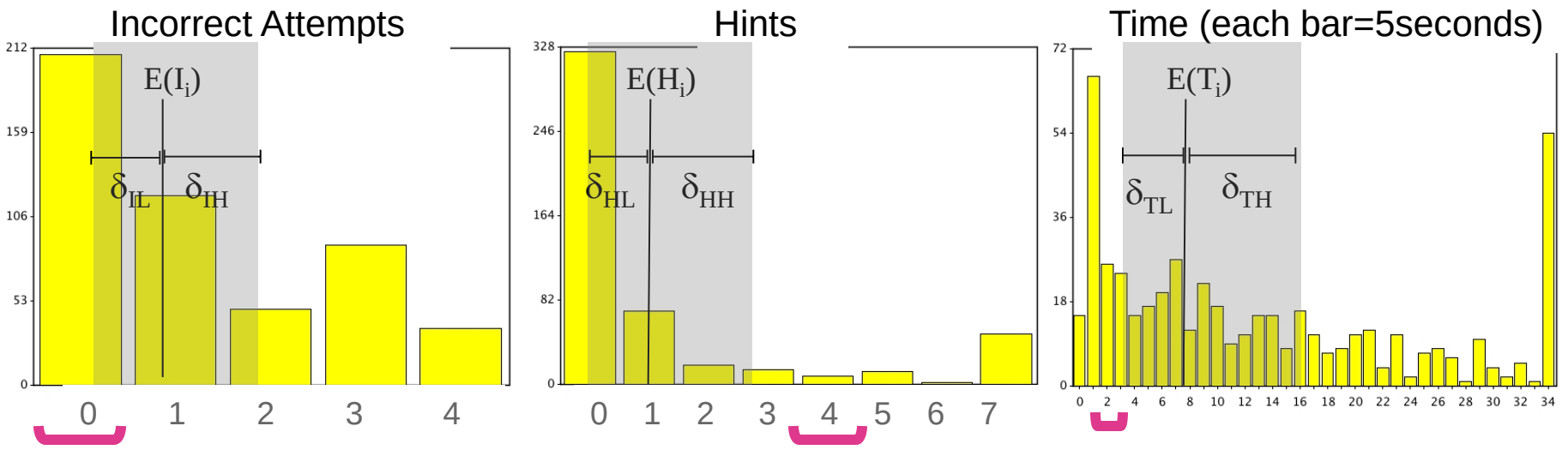



**Within expected behavior**

A new student encounters this problem…
Is their behavior within expectation, or atypical?

# What is odd behavior?

In any problem $p_i$ i=1, .., N N=Total problems in system




**Odd behavior**

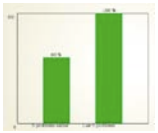
| Attempts < $E(I_i) - \delta_{IL}$ | Hints > $E(H_i) + \delta_{HH}$ | Time < $E(T_i) - \delta_{TL}$ |
|---|---|---|
| **Few Inc. Attempts** | **Lots of Hints** | **Little Time** |
| < | > | < |

# Pedagogical Moves in Wayang Outpost

| | Student Model<br>Estimate most likely scenario for student on $p_i$ | | | | Pedagogical Model Moves<br>Cognitive or Affective or Metacognitive | |
|---|---|---|---|---|---|---|
| | Mistakes | Hints | Time | Most Likely | Decision | Other Actions |
| 1 | < | < | < | Mastery without effort | Increase Problem Difficulty | Show learning progress |
| 2 | < | < | > | Toward Mastery | Maintain Problem Difficulty | Affective feedback: Praise Effort |
| 3 | < | > | < | Hint abuse, low effort | Reduce Problem Difficulty | Deemphasize importance of immediate success |
| 4 | < | > | > | Towards mastery, effort | Maintain Problem Difficulty | Praise effort |
| 5 | > | < | < | Quick guessing, low effort | Reduce Problem Difficulty | Deemphasize importance of immediate success |
| 6 | > | < | > | Hint avoidance and high effort | Reduce Problem Difficulty | Offer hints upon incorrect answer in the next problem |
| 7 | > | > | < | Quick guess and hint abuse | Reduce Problem Difficulty | Deemphasize importance of immediate success |
| 8 | > | > | > | Low mastery and High Effort | Reduce Problem Difficulty | Emphasize importance of effort and perseverance |
| 9 | Otherwise | | | Expected Behavior | Maintain Problem Difficulty | |

# Pedagogical Moves in Wayang Outpost

| | Student Model<br>Estimate most likely scenario for student on $p_i$ | | | | Pedagogical Model Moves<br>Cognitive or Affective or Metacognitive | |
|---|---|---|---|---|---|---|
| | Mistakes | Hints | Time | Most Likely | Decision | Other Actions |
| 1 | < | < | < | Mastery without effort | Increase Problem Difficulty | Show learning progress |
| 2 | < | < | > | Toward | Maintain Problem Difficulty | Affective feedback: Praise Effort |
| 3 | < | > | < | Hint abuse, low effort | Reduce Problem Difficulty | Deemphasize importance of immediate success |
| 4 | < | > | > | Towards mastery, effort | Maintain Problem Difficulty | Praise effort |
| 5 | > | < | < | Quick guessing, low effort | Reduce Problem Difficulty | Deemphasize importance of immediate success |
| 6 | > | < | > | Hint avoidance and high effort | Reduce Problem Difficulty | Offer hints upon incorrect answer in the next problem |
| 7 | > | > | < | Quick guess and hint abuse | Reduce Problem Difficulty | Deemphasize importance of immediate success |
| 8 | > | > | > | Low mastery and High Effort | Reduce Problem Difficulty | Emphasize importance of effort and perseverance |
| 9 | Otherwise | | | Expected Behavior | Maintain Problem Difficulty | |



# Pedagogical Moves in Wayang Outpost

| | Student Model Estimate most likely scenario for student on $p_i$ | | | | Pedagogical Model Moves Cognitive or Affective or Metacognitive | |
|---|---|---|---|---|---|---|
| | Mistakes | Hints | Time | Most Likely | Decision | Other Actions |
| 1 | < | < | < | Mastery without effort | Increase Problem Difficulty | Show learning progress |
| 2 | < | < | > | Toward | Maintain Problem Difficulty | Affective feedback: Praise Effort |
| 3 | < | > | < | Hint abuse, low effort | Reduce Problem Difficulty | Deemphasize importance of immediate success |
| 4 | < | > | > | Towards mastery, effort | Maintain Problem Difficulty | Praise effort |
| 5 | > | < | < | Quick guessing, low effort | Reduce Problem Difficulty | Deemphasize importance of immediate success |
| 6 | > | < | > | Hint avoidance and high effort | Reduce Problem Difficulty | Offer hints upon incorrect answer in the next problem |
| 7 | > | > | < | Quick guess and hint abuse | Reduce Problem Difficulty | Deemphasize importance of immediate success |
| 8 | > | > | > | Low mastery and High Effort | Reduce Problem Difficulty | Emphasize importance of effort and perseverance |
| 9 | Otherwise | | | Expected Behavior | Maintain Problem Difficulty | |

# Pedagogical Moves in Wayang Outpost

| | Student Model<br>Estimate most likely scenario for student on $p_i$ | | | | Pedagogical Model Moves<br>Cognitive or Affective or Metacognitive | |
|---|---|---|---|---|---|---|
| | Mistakes | Hints | Time | Most Likely | Decision | Other Actions |
| 1 | < | < | < | Mastery without effort | Increase Problem Difficulty | Show learning progress |
| 2 | < | < | > | Toward Mastery | Maintain Problem Difficulty | Affective feedback: Praise Effort |
| 3 | < | > | < | Hint abuse, low effort | Reduce Problem Difficulty | Deemphasize importance of immediate success |
| 4 | < | > | > | Towards mastery, effort | Maintain Problem Difficulty | Praise effort |
| 5 | > | < | < | Quick guessing, low effort | Reduce Problem Difficulty | Deemphasize importance of immediate success |
| 6 | > | < | > | Hint avoidance and high effort | Reduce Problem Difficulty | Offer hints upon incorrect answer in the next problem |
| 7 | > | > | < | Quick guess and hint abuse | Reduce Problem Difficulty | Deemphasize importance of immediate success |
| 8 | > | > | > | Low mastery and High Effort | Reduce Problem Difficulty | Emphasize importance of effort and perseverance |
| 9 | Otherwise | | | Expected Behavior | Maintain Problem Difficulty | |

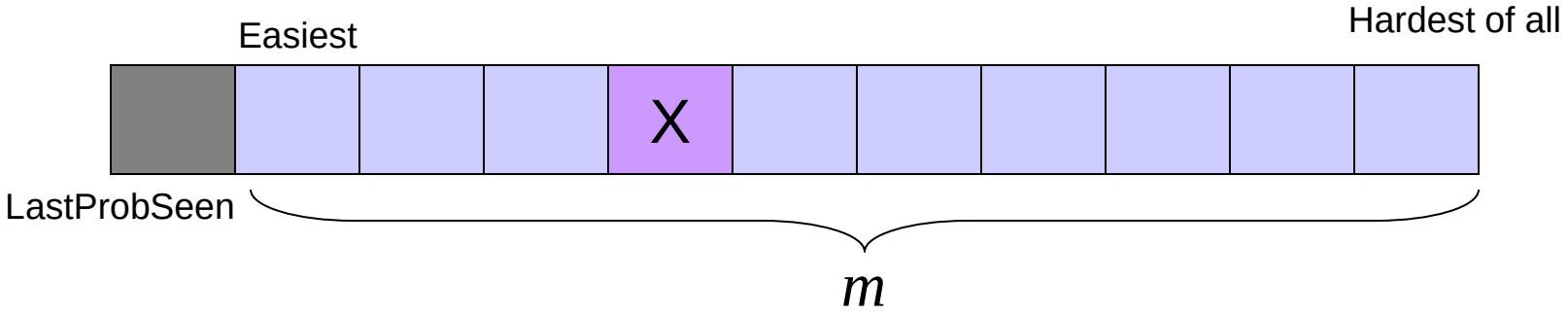
# Pedagogical Moves in Wayang Outpost

| | Student Model<br>Estimate most likely scenario for student on $p_i$ | | | | Pedagogical Model Moves<br>Cognitive or Affective or Metacognitive | |
|---|---|---|---|---|---|---|
| | Mistakes | Hints | Time | Most Likely | Decision | Other Actions |
| 1 | < | < | < | Mastery without effort | Increase Problem Difficulty | Show learning progress |
| 2 | < | < | > | Toward | Maintain Problem Difficulty | Affective feedback: Praise Effort |
| 3 | < | > | < | Hint abuse, low effort | Reduce Problem Difficulty | Deemphasize importance of immediate success |
| 4 | < | > | > | Towards mastery, effort | Maintain Problem Difficulty | Praise effort |
| 5 | > | < | < | Quick guessing, low effort | Reduce Problem Difficulty | Deemphasize importance of immediate success |
| 6 | > | < | > | Hint avoidance and high effort | Reduce Problem Difficulty | Offer hints upon incorrect answer in the next problem |
| 7 | > | > | < | Quick guess and hint abuse | Reduce Problem Difficulty | Deemphasize importance of immediate success |
| 8 | > | > | > | Low mastery and High Effort | Reduce Problem Difficulty | Emphasize importance of effort and perseverance |
| 9 | Otherwise | | | Expected Behavior | Maintain Problem Difficulty | |

# Increasing Problem Difficulty

At the next time step. Assume we know problem difficulty of items.

$H =$ Sorted list of harder math problems



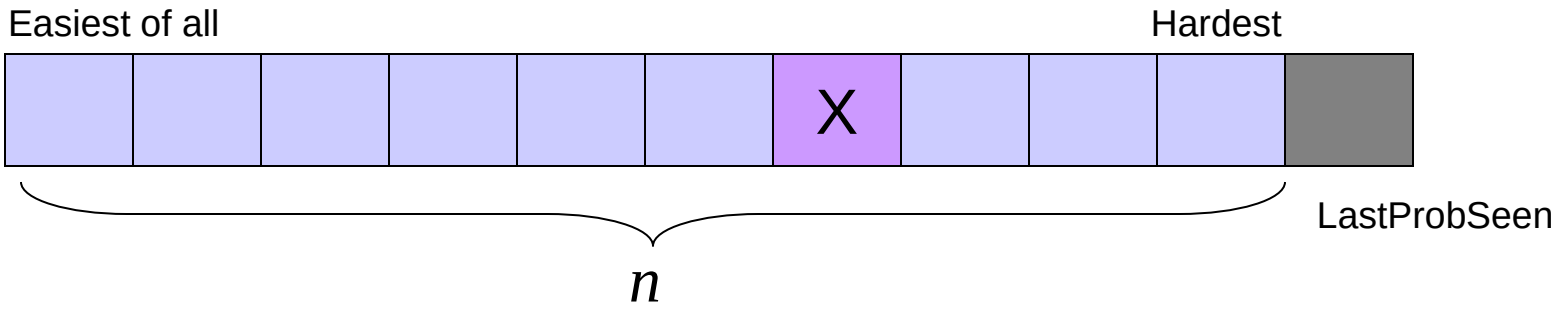

$$Harder(H[1..m],\gamma) = H\left\lfloor ceiling\left(\frac{m}{\gamma}\right)\right\rfloor$$

Parameter { $\gamma$ = 3 **--> Challenge rate**

# Decreasing Problem Difficulty

At the next time step. Assume we know problem difficulty of items.

$E =$ Sorted list of easier math problems




$$Easier(E[1..n],\gamma) = E\left\lfloor ceiling\left(n - \frac{n}{\gamma}\right)\right\rfloor$$

Parameter $\gamma$ = 3

# What is Problem Difficulty?

## Which problems are easier or harder?

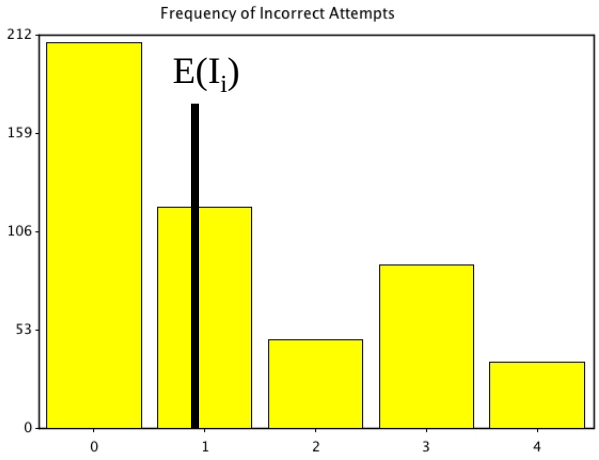
*Correctness effort factor* $dc_i = \frac{E(I_i)}{Max_{j=1}^{N}(E(I_j))}$ $dc_i \in [0,1]$

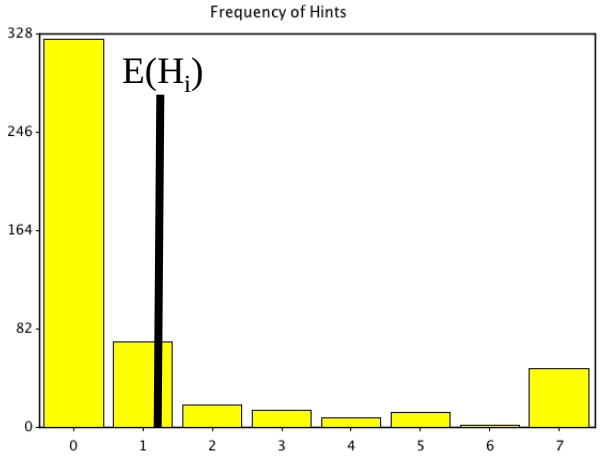
*Hint effort factor* $dh_i = \frac{E(H_i)}{Max_{j=1}^{N}(E(H_j))}$ $dh_i \in [0,1]$

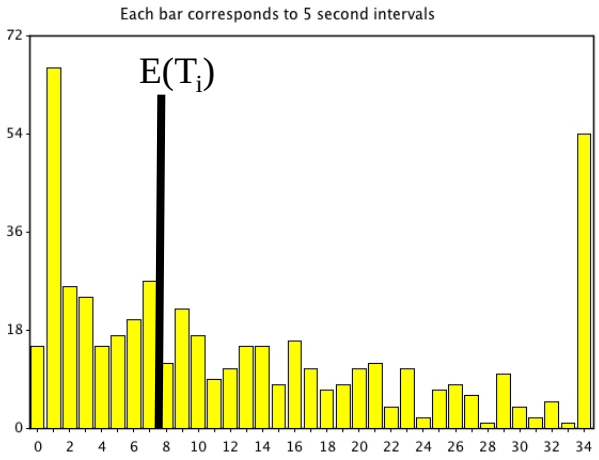
*Time effort factor* $dt_i = \frac{E(T_i)}{Max_{j=1}^{N}(E(T_j))}$ $dt_i \in [0,1]$

$$d_i = mean(dc_i, dt_i, dh_i)$$

$$d_i \in [0,1]$$




Reasonable… But is it really sound?

# Accuracy of Problem Difficulty?

Just too important to get them right

- They are too relevant to the Pedagogical Model moves

Just too likely they might be wrong

- They will be *biased* to the problem selector in place!
  - However, if you have a variety of past problem selectors…

Need criteria that can guarantee they are good-enough

- What kind of tests can we run?

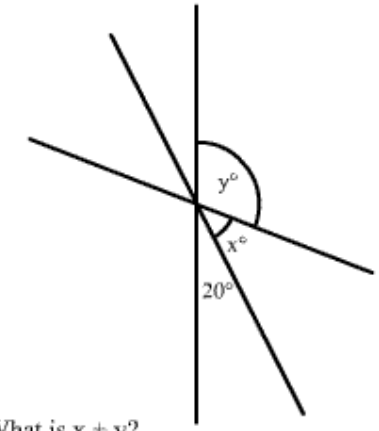
# Are difficulty estimations accurate?

## Criteria for evaluating the soundness of estimations

Axiom:

"Pairs of similar problems should have similar problem difficulty estimates"

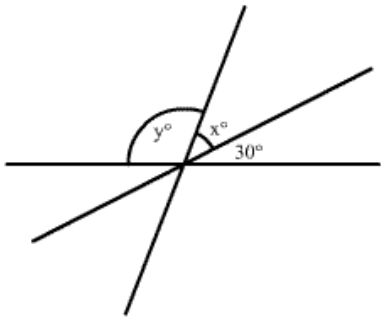
$p_{003}$




What is x + y?

- 20°
- 60°
- 155°
- 160°
- 180°

$p_{033}$




What is x + y?

- 50°
- 90°
- 150°
- 145°
- 60°

30 pairs of problems like this. It should happen that: $d_{p_i} \approx d_{p_{30+i}}$

# Are difficulty estimations accurate?

Criteria for evaluating the soundness of $d_i$ in relation to $d_{30+i}$

## Criteria 1: Similar to each other?

Pearson Correlation: $p<0.0001$, $R=.823$

# Are difficulty estimations accurate?

## Criteria for evaluating the soundness of $d_i$ in relation to $d_{30+i}$

Criteria 1: Similar to each other?

Criteria 2: Different from rest?

Pearson Correlation: p<0.0001, R=.823

Paired samples t(29)=7.35, p<.000

$$\left(d_{p_i} - d_{p_{30+i}}\right)^2 \leq \frac{\sum_{j=1, j\neq i}^{30} \left(d_{p_i} - d_{p_j}\right)^2}{30}$$

| | $d_{pi}$ | $d_{p30+i}$ | SE | MSE |
|---|---|---|---|---|
| 1 | 0.28 | 0.29 | 0.0001 | … |
| 2 | … | … | … | … |
| 3 | | | | |
| … | | | | |
| 30 | | | | |

# Are difficulty estimations accurate?

Criteria for evaluating the soundness of $d_i$ in relation to $d_{30+i}$

Criteria 1: Similar to each other?

Criteria 2: Different from rest?

Pearson Correlation: $p<0.0001$, $R=.823$

Paired samples $t(29)=7.35$, $p<.000$

# Are difficulty estimations accurate?

Criteria for evaluating the soundness of $d_i$ in relation to $d_{30+i}$

Criteria 1: Similar to each other?

Criteria 2: Different from rest?

Criteria 3: Different from each other?

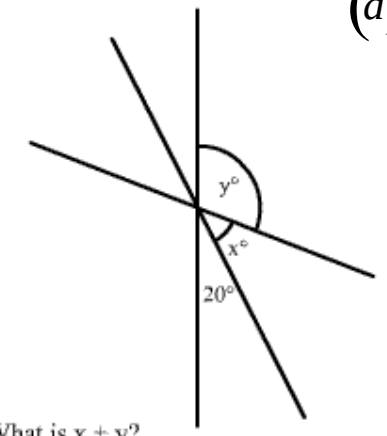
Pearson Correlation: $p<0.0001$, $R=.823$

Paired samples $t(29)=7.35$, $p<.000$

21 restrictions, $\chi^2(20)=5.25$, $p<0.05$

$$\left(d_{p_i} - d_{p_{30+i}}\right)^2 = \varepsilon_i$$

$p_{003}$

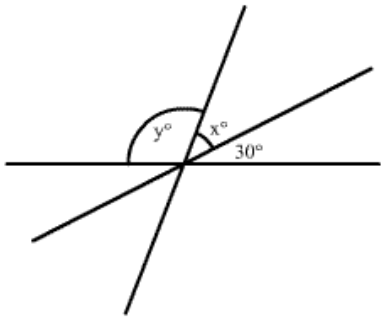



What is x + y?

- 20°
- 60°
- 155°
- 160°
- 180°

< > ?

$p_{033}$

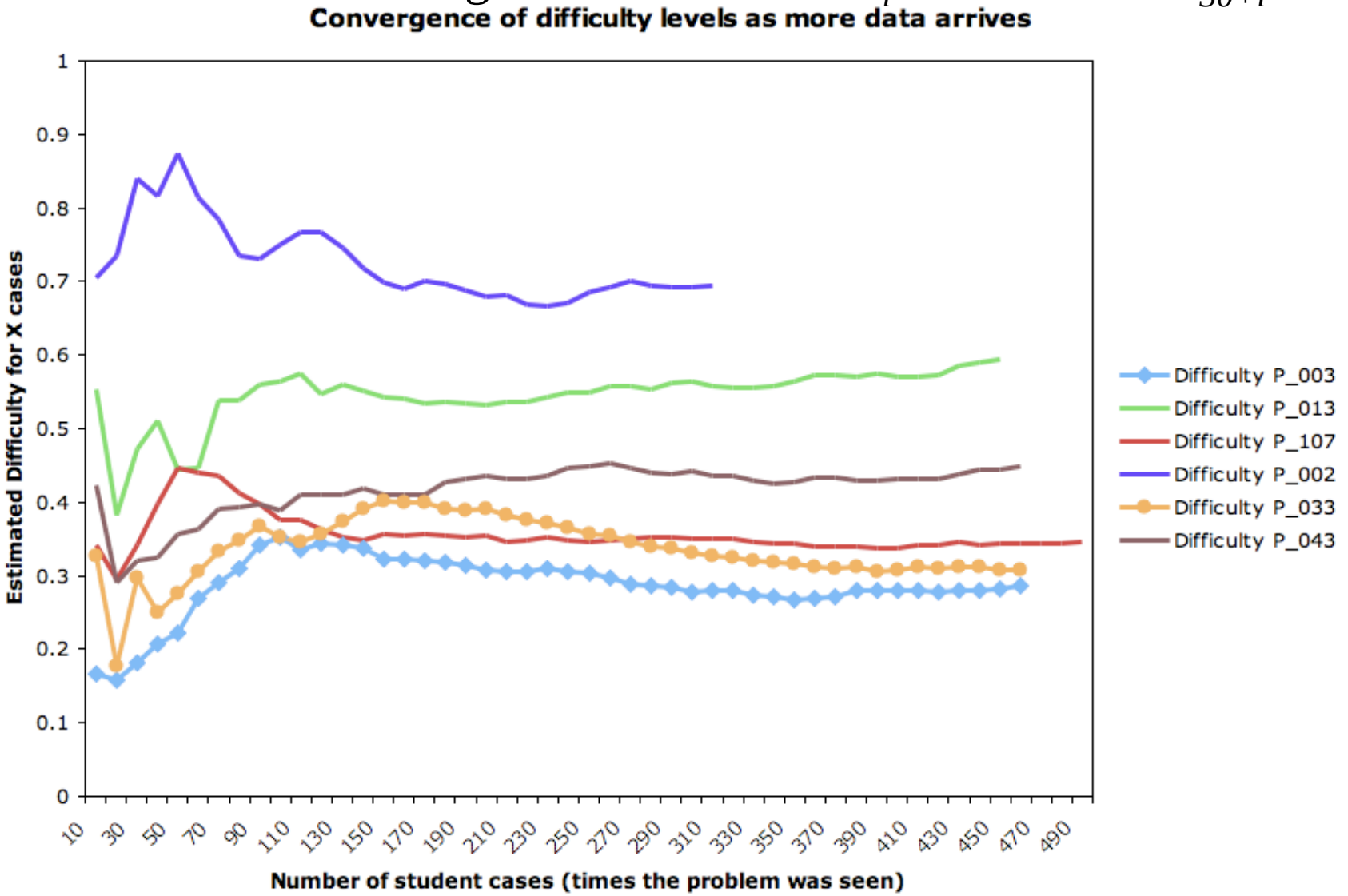



What is x + y?

- 50°
- 90°
- 150°
- 145°
- 60°

# Are difficulty estimations accurate?

Criteria for evaluating the soundness of $d_i$ in relation to $d_{30+i}$



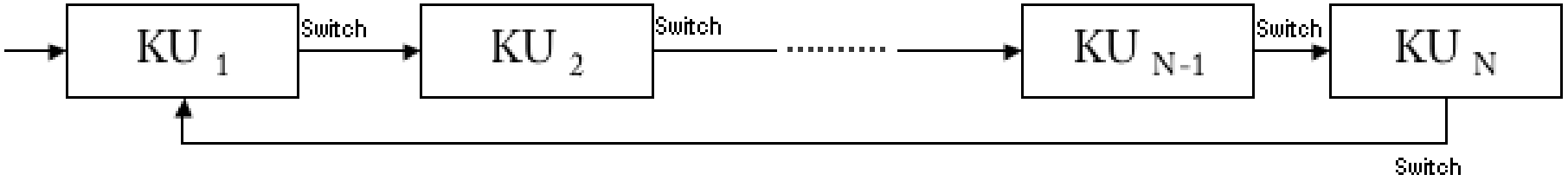
# Progressing Through Knowledge Units

- Intelligent/Adaptive Tutor =
  - Discarding some of the content
    - Some of it is inappropriate
    - Student is ready to move on
  - When has the student seen enough?
- Content Organization
  - Where should they move on to?
  - What are Knowledge Units?

# Switching through Knowledge Units

## Parameters to Switch Topics or Knowledge Units

Knowledge Units are Chunks of Problems that use similar skills




| Topic Switch Criterion | Reason | Parameter |
|---|---|---|
| 2.1 Topic Mastery was reached (e.g. enough ÒhardÓ problems answered correctly) | Cognitive | $M_{KU}$ |
| 2.2 Persistent failure to find a problem of desired difficulty | Content limitation | $F_{KU}$ |
| 2.3 Maximum time in Topic condition, or Maximum Number of Problems allowed | Classroom Implementation | $T_{KU}$ $N_{KU}$ |

# Does it improve learning?

Randomized Controlled Experiment (N=56) Spring 2004

56 high school students from different math classes in a public school
Pretest and Posttest, 9 (medium and hard) SAT-M problems

| | Within KU | Between KU |
|---|---|---|
| Experimental | $\gamma=2$<br>$\theta_{LOW}=0$<br>$\theta_{HIGH}=0$ | $N_{KU}$ = *fixed value* |
| Control | Random | $N_{KU}$ = *fixed value* |

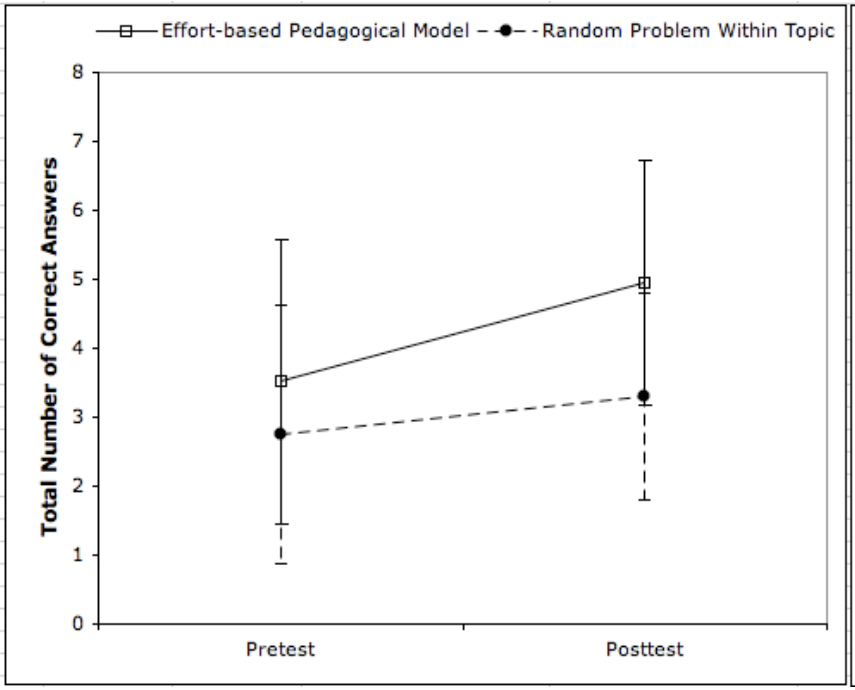
**Only problem difficulty moves, no "other actions"**

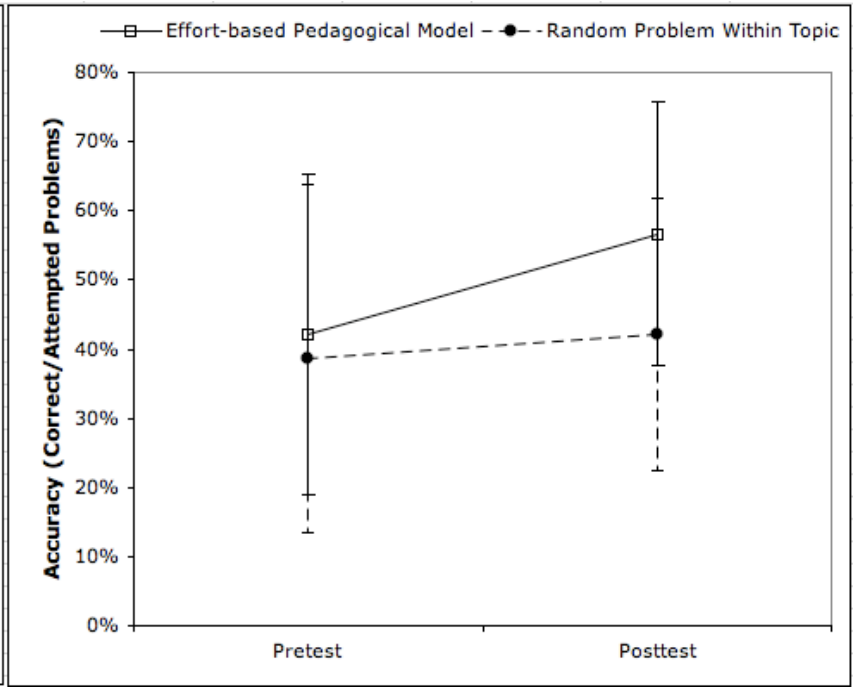
# Does it improve learning?

Randomized Controlled Experiment (N=56) Spring 2004

56 high school students from different math classes in a public school

Pretest and Posttest, 9 (medium and hard) SAT-M problems




ANCOVA for Posttest Score $F(55,1)=8.4$, $p=.006$

# Conclusions

- We can discern expected from odd behavior
- We can unveil parameters that regulate ITS functioning
- A procedure to make Learning Environments Smarter
- Tests to evaluate accuracy of problem difficulty
- Evidence that adaptive problem sequencing works
  - Students learned more with adaptive problem selection
  - No learning companions in experiment!

# Future Work

- A variety of new behaviors
  - Drawing tools
  - Example viewing
  - Emotion Self-reports
- Combinations of possible behaviors are too many
  - Partial combinations
  - Analyzing clusters of behaviors
- Sequential behavior
  - Motifs and Time-based Patterns
    - Identifying High-Level Student Behavior Using time-based Motif Discovery (Shanabrook et al., EDM 2010)

# Effort-based Tutoring: an Empirical Approach to Intelligent Tutoring

Ivon Arroyo[1], Hasmik Mehranian[2], Beverly P. Woolf[1]

[1] *Department of Computer Science*

[2] *Dept. of Mechanical and Industrial Engineering*

*University of Massachusetts Amherst*

# What ITS research needs

- Sophisticated Integrated Student Models (SM)
  - *Discern* between Knowledge, Engagement, Affect, Meta-cognition
- Smarter Pedagogical Models (PM)
  - Juggle a variety of student models
  - Effective pedagogical components: in what situations do they work?
- Evidence of Success at Optimizing Learning
  - From the combination of SM and PM
    - Evaluating PM assuming a perfect SM (Beck, 2000)
- What parameters regulate ITS functioning?
  - For optimization
  - For replication: to add smartness to an ILE?

# Why an Empirical Approach to ITS?
## Some reasons

- How students use ILEs is not known at design time
  - We don't know how students misuse
  - We don't know how much students misuse
- Because it can help us to see an integral view
  - Misuse can throw off our estimates of student knowledge
    - E.g. incorrect attempts may appear as unknowing
  - Specific behaviors may be attributed to different reasons
- To understand Domain better
  - *Objective* Problem Difficulty < > *Subjective* Problem Difficulty
- We could learn to make the ILE smarter

# An Integral Wholesome View

Factors that regulate/mediate the functioning of an ITS



Student Behavior
(Engagement, Use/Misuse)

Pedagogical Moves

Knowledge Tracing

Content Availability

Affect Models
Emotions/Motivation

Content Quality
Help quality

Metacognition Models

Item Difficulty

Interface (HCI)
Answer format?

***Student Models***

# The Problem of Intelligent Tutoring

What are we trying to solve?

- Optimizing Learning
  - And Motivation, and Meta-cognition…
- What might help us to get us there?
  - Modeling Student Knowledge and Others (SM)
  - Good Pedagogical Material (Content)
  - Good Pedagogical Decisions (PM)

# Problems to Implementation of ITS

- Gaming throws off the Knowledge Estimates
- Parameters to Knowledge Tracing
  - Didn’t know exactly what to set
- Ill-defined:
  - Content does not break down nicely into skills
- Step back: Take an Wholesome, Integrated View
  - Can we write a “recipe” to follow to add smartness to an ILE?