

# What does the **Interaction Plateau** imply for Lifelong Learning Companions?

---

Kurt VanLehn

School of Computing and Informatics  
Arizona State University



# Outline

---

## ◆ Terminology

- Answer-based tutoring
- Step-based tutoring, with remediation via
  - » Hint sequences
  - » Dialogues

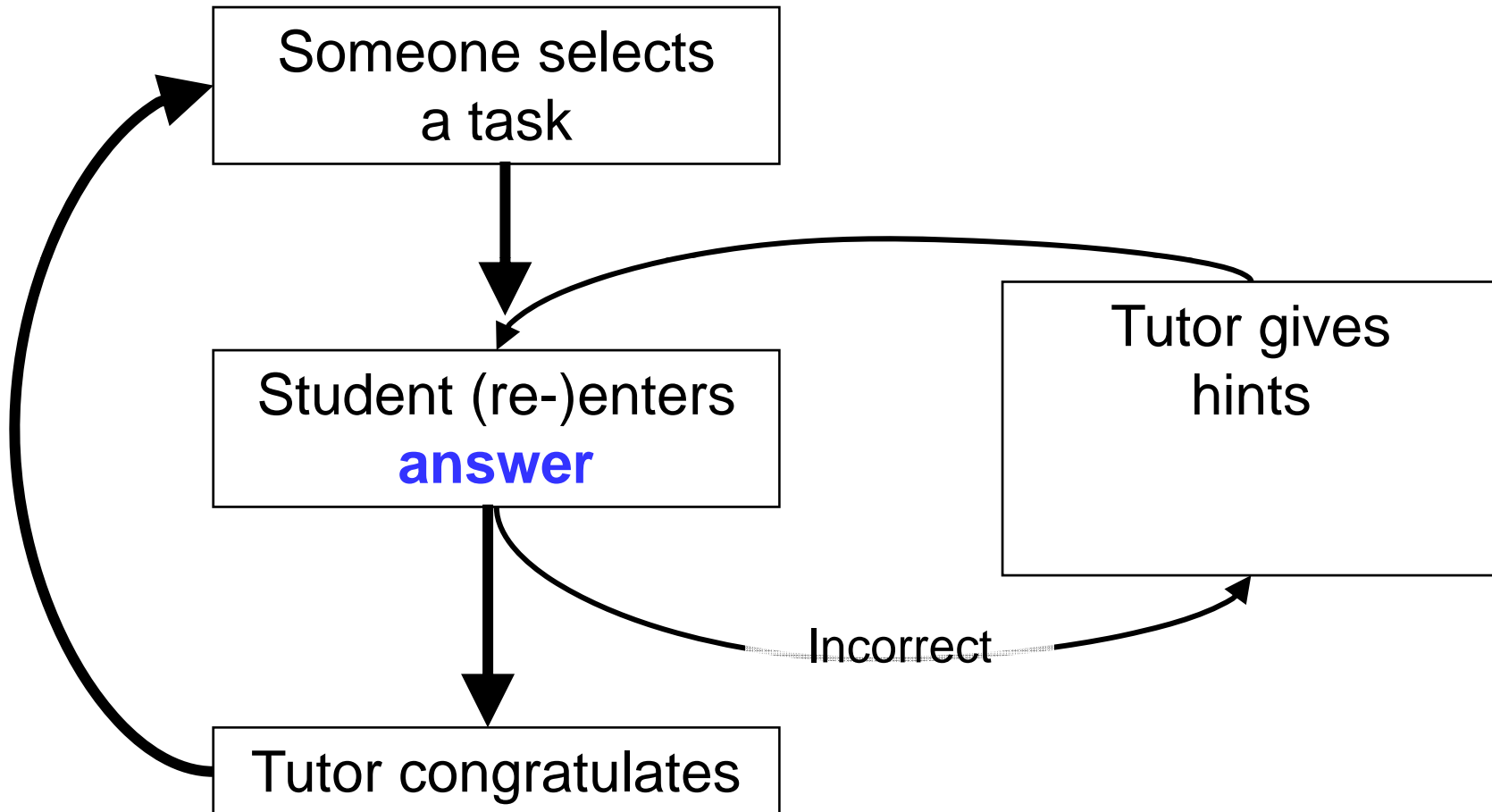
## ◆ Hypotheses

## ◆ Evidence

## ◆ Implications for Lifelong Learning Companions

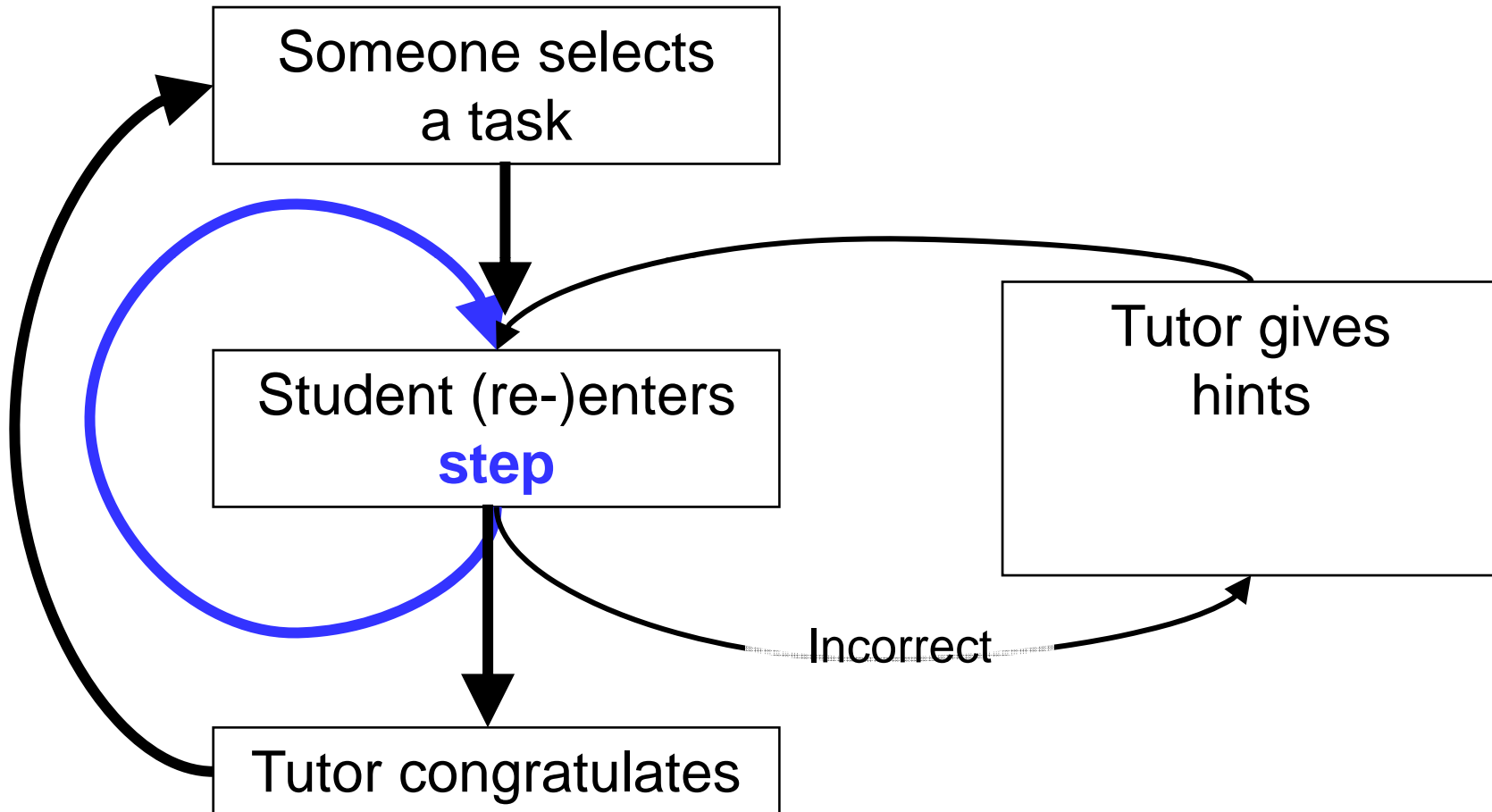
# Answer-based tutoring systems (= CAI, LMS) have a task loop

---



# Step-based tutoring systems (= ITS) also have a step loop

---



# Andes user interface

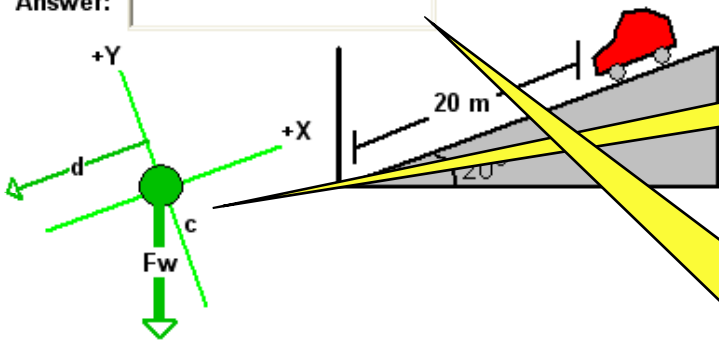
ANDES Physics Workbench - [dt5a.fbd]

File Edit Diagram Variable View Help

A 2000-kg car in neutral at the top of a 20.0 deg inclined driveway 20.0 m long slips its parking brake and rolls down.

If we ignore friction and drag, what would the magnitude of the velocity of the car be when it hits the garage door?

Answer:



Variables

Name	Definition	Dir	X-...	Y-Comp
T0	car starts rolling			
T1	car hits garage door			
x	axis		$\theta x = 20^\circ$	
mc	mass of car			
d	magnitude of the Displacement of car at time T0 to T1		$\theta d = 200^\circ$	d_x d_y
Fw	magnitude of the Weight Force on car at time T0 to T1 due to Earth		$\theta Fw = 270^\circ$	Fw_x Fw_y

1.  $mc = 2000 \text{ kg}$

2.  $d = 20.0 \text{ m}$

3.  $Fw_y = mc * g$

6.

7.

8.

9.

10.

T: Now that you have stated all of the given Information, you should start on the major principles. What quantity is the problem seeking?

S: The magnitude of the instantaneous Velocity of car at time T1

T: Yep. What is the first principle application that you would like to work on? Hint: this principle application will usually be one that mentions the sought quantity explicitly. Therefore it's equation may contain the sought quantity that the problem seeks.

270 degrees

00:10:30 SCORE: 39

Read a physics problem

Draw vectors

Type in equations

Type in answer

# Andes feedback and hints

Ask for a hint

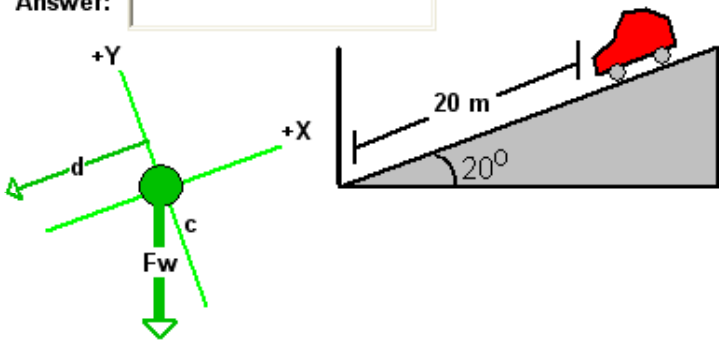
ANDES Physics Workbench - [dt5a.fbd]

File Edit Diagram Variable View Help

A 2000-kg car in neutral at the top of a 20.0 deg inclined driveway 20.0 m long slips its parking brake and rolls down.

If we ignore friction and drag, what would the magnitude of the velocity of the car be when it hits the garage door?

Answer:



T: Now that you have stated all of the given Information, you should start on the major principles. What quantity is the problem seeking?

S: The magnitude of the instantaneous Velocity of car at time T1

T: Yep. What is the first principle application that you would like to work on? Hint: this principle application will usually be one that mentions the sought quantity explicitly. Therefore it's equation may contain the sought quantity that the problem seeks.

Variables

Name	Definition	Dir	X-...	Y-Comp
T0	car starts rolling			
T1	car hits garage door			
x	axis	$\theta x=20^\circ$		
mc	mass of car			
d	magnitude of the Displacement of car at time T0 to T1	$\theta d=200^\circ$	d_x	d_y
Fw	magnitude of the Weight Force on car at time T0 to T1 due to Earth	$\theta Fw=270^\circ$	Fw_x	Fw_y

1. mc = 2000 kg

2. d = 20.0 m

3. Fw\_y = mc \* g

4.

5.

6.

7.

8.

9.

10.

Hints

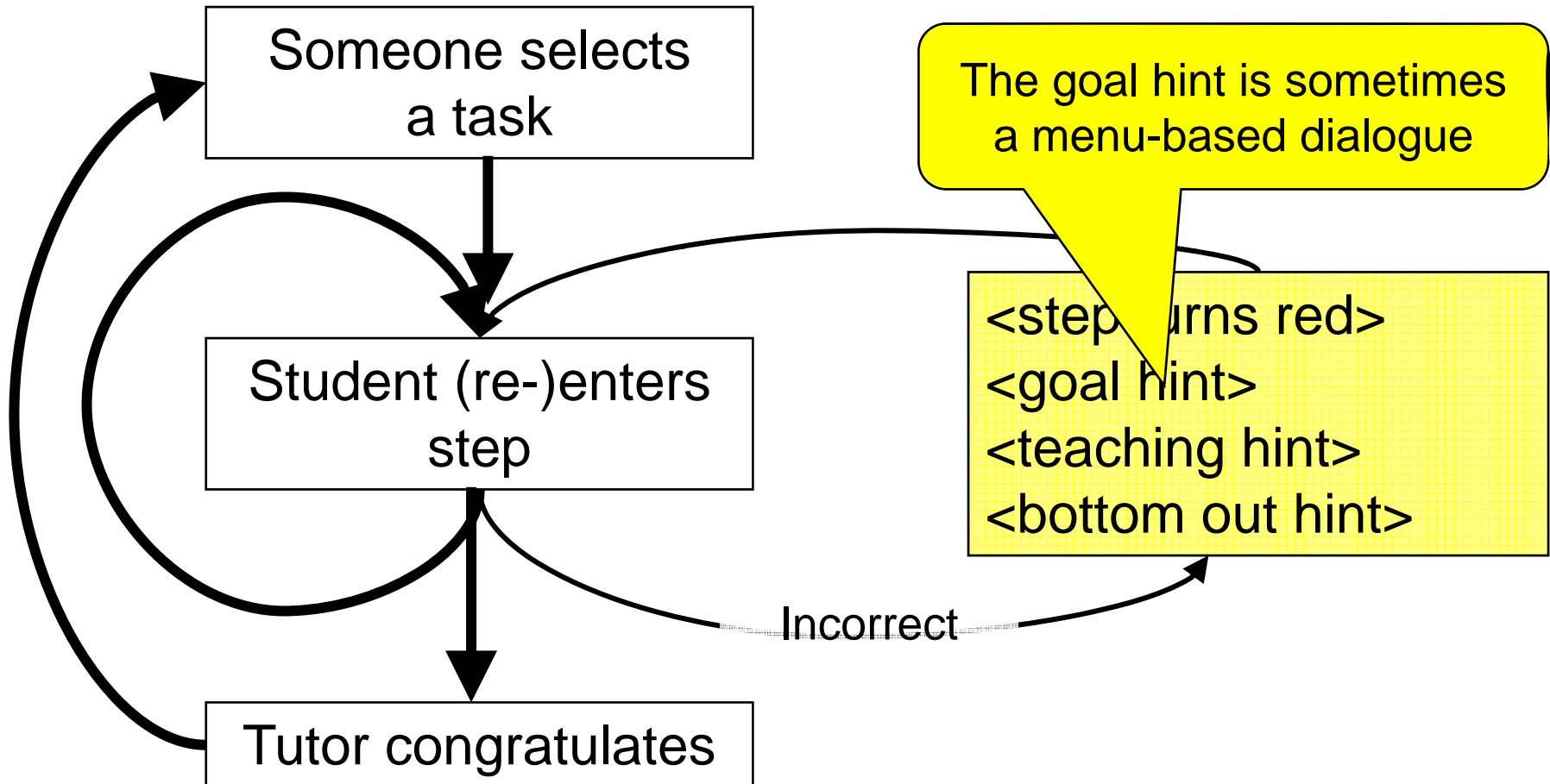
270 degrees

00:10:30 SCORE: 39

Green means correct  
Red means incorrect

# Andes remedies incorrect steps with hint sequences

---



# Cognitive Algebra I Tutor (Carnegie Learning)

**Problem**

A rock climber is currently on the side of a cliff 67 feet off the ground. She can climb on average about two and one-half feet per minute.

- 1 When will she be 92 feet off the ground?
- 2 In twenty minutes, how many feet above the ground will she be?
- 3 In 75 seconds, how far above the ground will she be?
- 4 Ten minutes ago, how far above the ground would she have been?

**Step: Label a column**

Quantity Name	CLIMBING TIME	HEIGHT ABOVE GROUND
Unit	MINUTES	FEET
Expression	$T$	$67 + 2.5T$
Question 1	10	92
Question 2	20	117
Question 3	1.25	70.125
Question 4	-10	42

**Step: Fill in a cell**

**Step: Define an axis**

**Step: Enter an equation**

Solve for T

$$25 = 2.5T$$

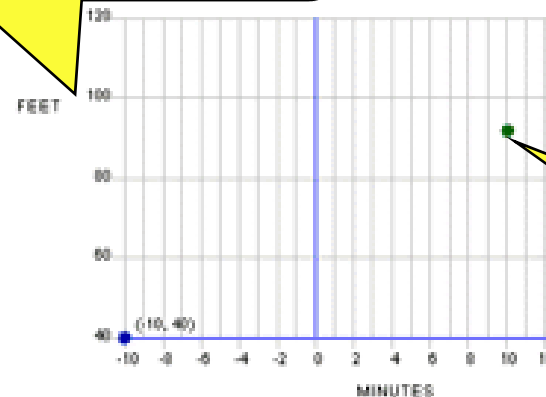
$$\frac{25}{2.5} = \frac{2.5T}{2.5}$$

$$10 = T$$

Divide both sides by

**Step: Divide both sides**

**Step: Plot a point**





# SLQ-Tutor (Addison Wesley)

## Problem

### Problem 30

List the titles and numbers of all movies that have won at least one Academy Award and have been made in or after 1988.

SELECT

title, number

FROM

movies

WHERE

aawon>1 and year>=1988

GROUP BY

HAVING

ORDER BY

Feedback Level

Hint

Submit Answer

Reset

Step

Step

Step

Submit!

Almost there - a few mistakes though. One of them is in the FROM clause. You can correct your query and press 'Submit' again, or try getting some more feedback.

Would you like to have another go?

Feedback

## Schema for the MOVIES Database

The general description of the database is available [here](#). Clicking on the name of a table brings up the table details. Keys in the attribute list are underlined, foreign keys are in *italics*.

The database that the problem refers to

Table Name

Attribute List

DIRECTOR

number lname fname born died

MOVIE

number title type aanom aawon year critics *director*

STAR

lname fname number born died city

CUSTOMER

lname fname number address rentals bonus jdate

TAPE

code *movie* pdate times *customer* hiredate

STARS\_IN

*movie* star role

# AutoTutor



The sun exerts a gravitational force on the earth as the earth moves in its orbit around the sun . Does the earth pull equally on the sun? Explain why.

Task

Tutor asks questions

## Log of previous turns

moves in its orbit around the sun . Does the earth pull equally on the sun? Explain why.

Student:

Tutor: Is there anything you can add to this?

Student:

Tutor: Kind of.

Tutor:

Tutor: How does Newton's law of motion apply to this situation?

Tutor:

Student:

Type your response here:

Dialogue history

Student types response

Submit

Settings...

# Steps are unordered and in natural language

---

◆ Why do most computers have a disk drive?  
Why can't they have only RAM?

- Steps {
1. RAM's content disappears when power quits, but disk content persist.
  2. RAM is usually holds less information than disk
  3. RAM takes battery power, so larger RAM takes more power
  4. Certain information, e.g., operating system and user files, must be stored permanently.

# (Human) tutor prompts for each missing step and hints until its correct

---

◆ Why do most computers have a disk drive? Why can't they have only RAM?

◆ S: RAM is too small. Only the disk is big enough.

Step 2  
done

◆ T: True. But if it had a lot of RAM, would that suffice?

◆ S: The battery would run out too fast.

Step 3  
done

◆ T: Excellent. What else?

◆ S: That's it.

◆ T: What if the battery dies?

◆ S: Oh. The RAM dies.

◆ T: Anything wrong with that?

◆ S: You lose your files.

Step 1  
done

◆ T: Besides the user's files, what else would be lost?

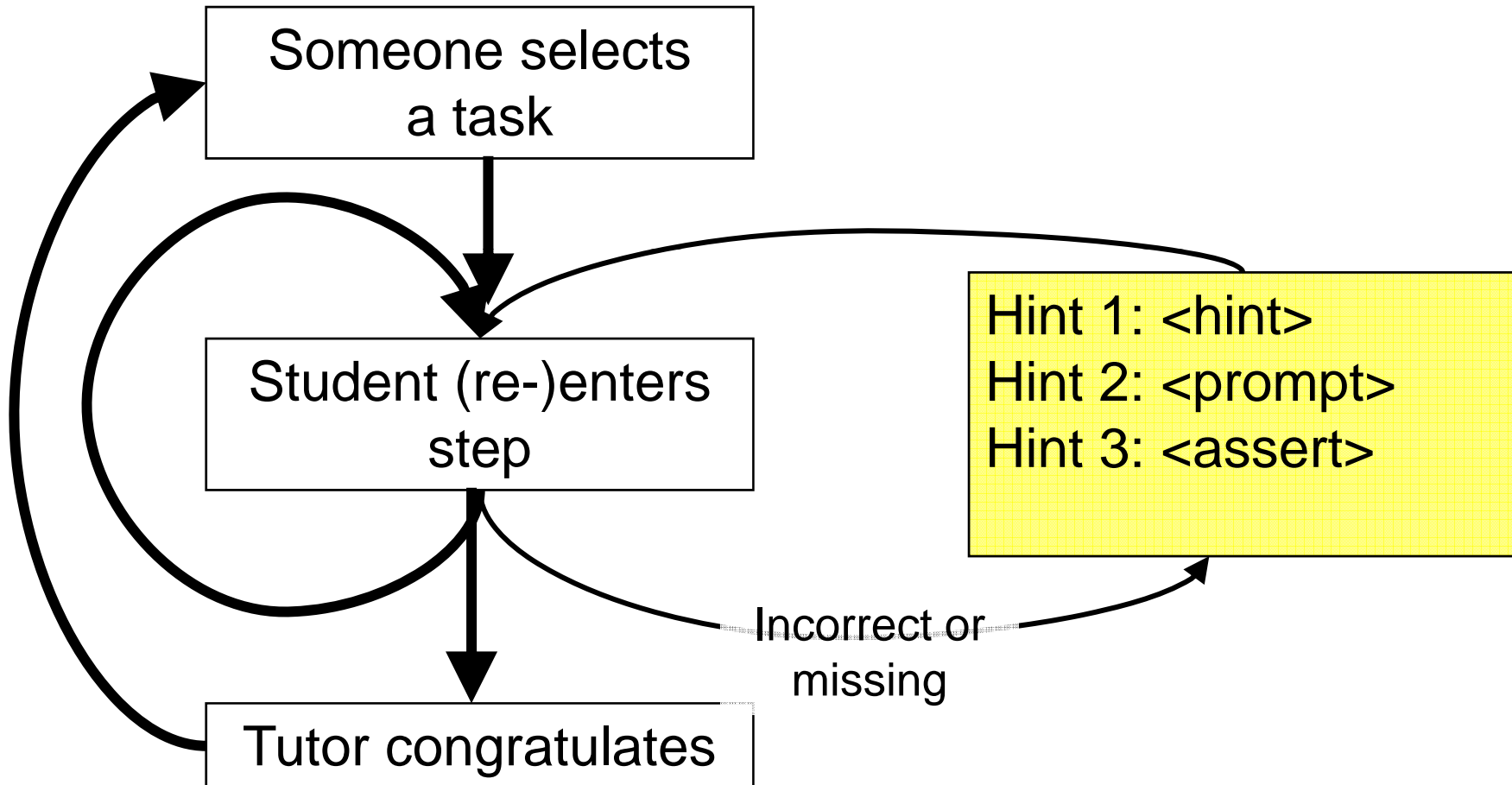
◆ S: Beats me.

◆ T: The operating system

Step 4  
done

# AutoTutor

---



# Student's screen for human tutoring and Why2/Atlas

**Problem**

4. A rock is thrown straight upward with a velocity  $v$ . What is its acceleration at the top of its path? What is its velocity when it comes back to the starting point? Why?

**Dialogue history**

Dialogue:

Tutor: What is the velocity just before it hits the ground?

Student: That will depend on the time that the ball is in the air/distance ball traveled.

Tutor: How will it compare with the velocity with which it was thrown up?

**Student's essay**

Enter your essay here:

At the very top of the path, where the velocity is equal to 0, the acceleration will be  $-9.8 \text{ m/(s}^2\text{s)}$ , the acceleration of gravity. It will no longer have any upward acceleration. When it comes back the its starting point (assuming that it is the ground) its velocity will be equal to 0 because its downward progress will be stopped by the ground.

**Student's turn in the dialogue**

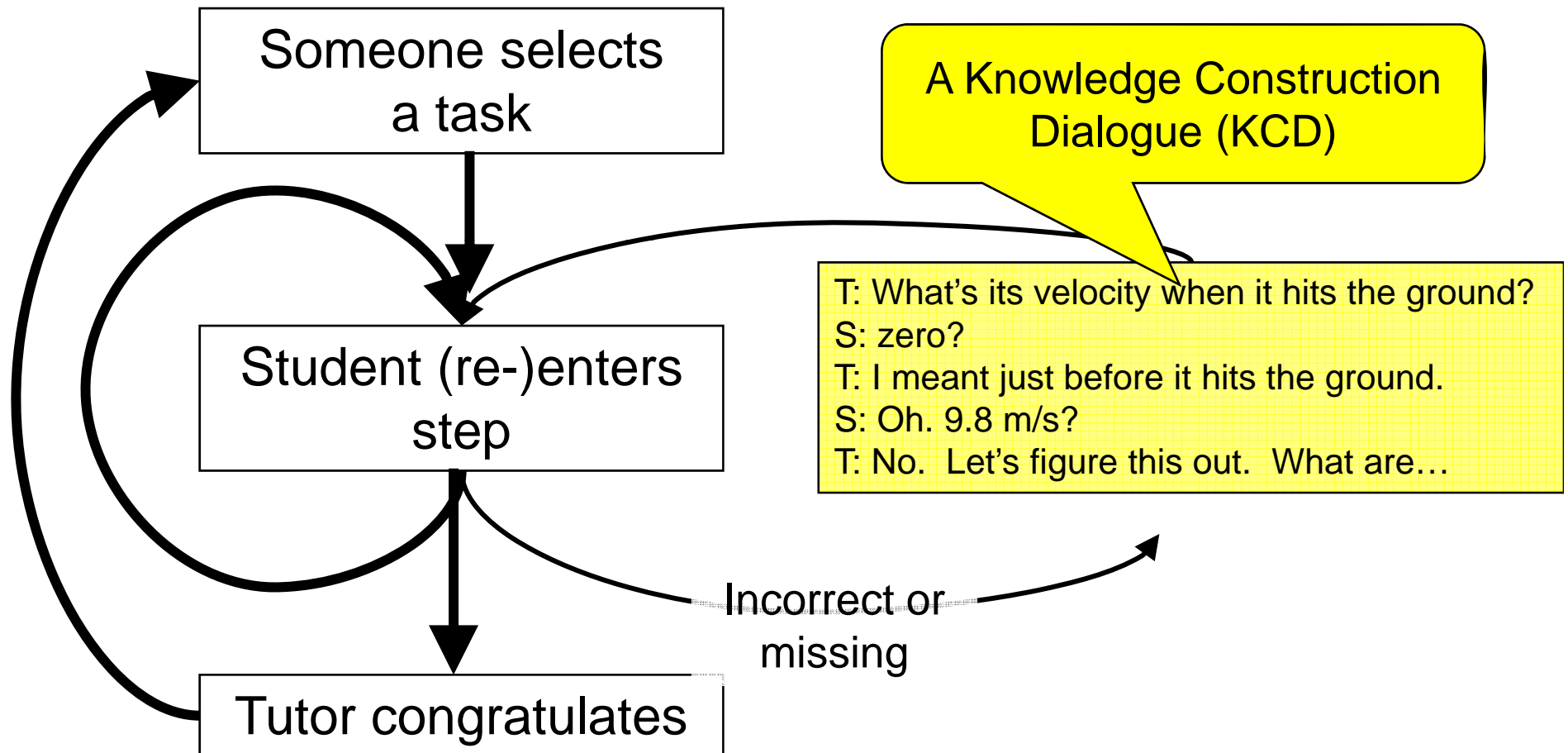
Tell Tutor:

They will be equal. The height the ball reaches will depend on the initial velocity.

Send Essay or Message

# Atlas, Circsim-Tutor, SE-Kermit

---



# Outline

---

## ◆ Terminology

- Answer-based tutoring
- Step-based tutoring, with remediation via
  - » Hint sequences
  - » Dialogues

## ◆ Hypotheses



Next

## ◆ Evidence

## ◆ Implications for lifelong learning companions



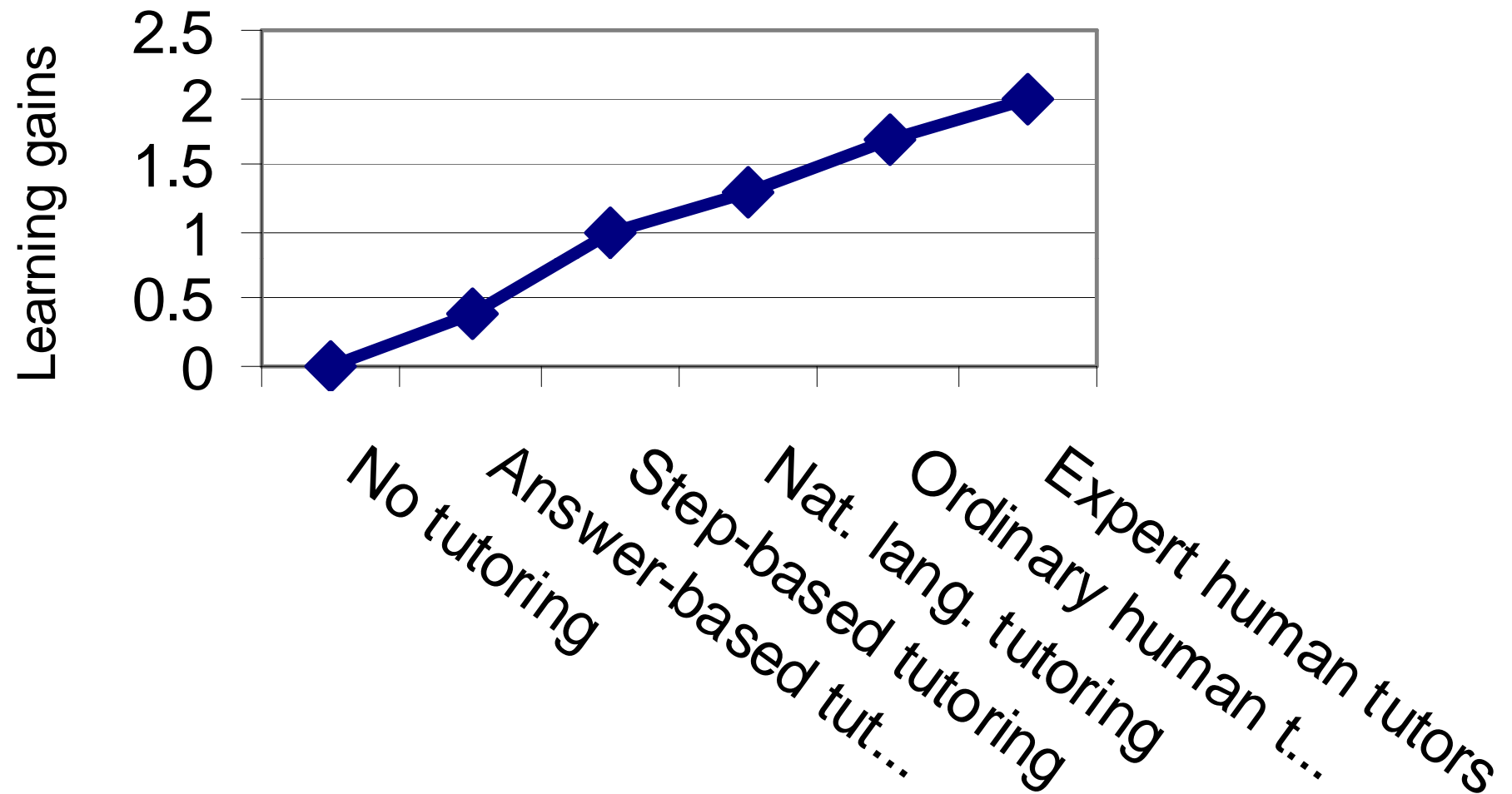
# Hypothesized ranking of tutoring, most effective first

---

- A. Expert human tutors
- B. Ordinary human tutors
- C. Natural language tutoring systems  
(i.e., step-based tutoring systems with dialogue as remediation)
- D. Step-based tutoring systems  
with hint sequences as remediation
- E. Answer-based tutoring systems
- F. No tutoring

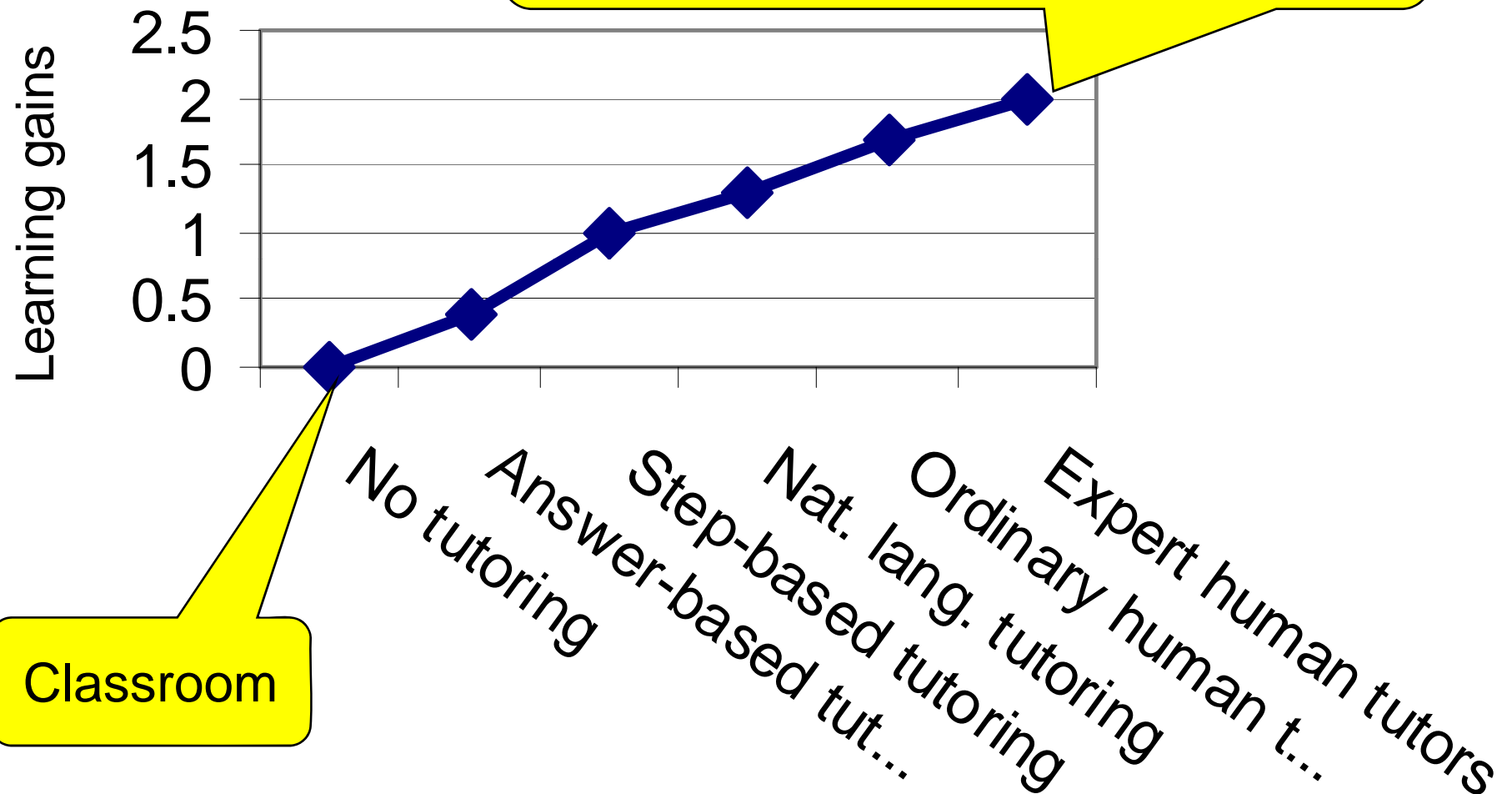
# Hypothesized effect sizes

---



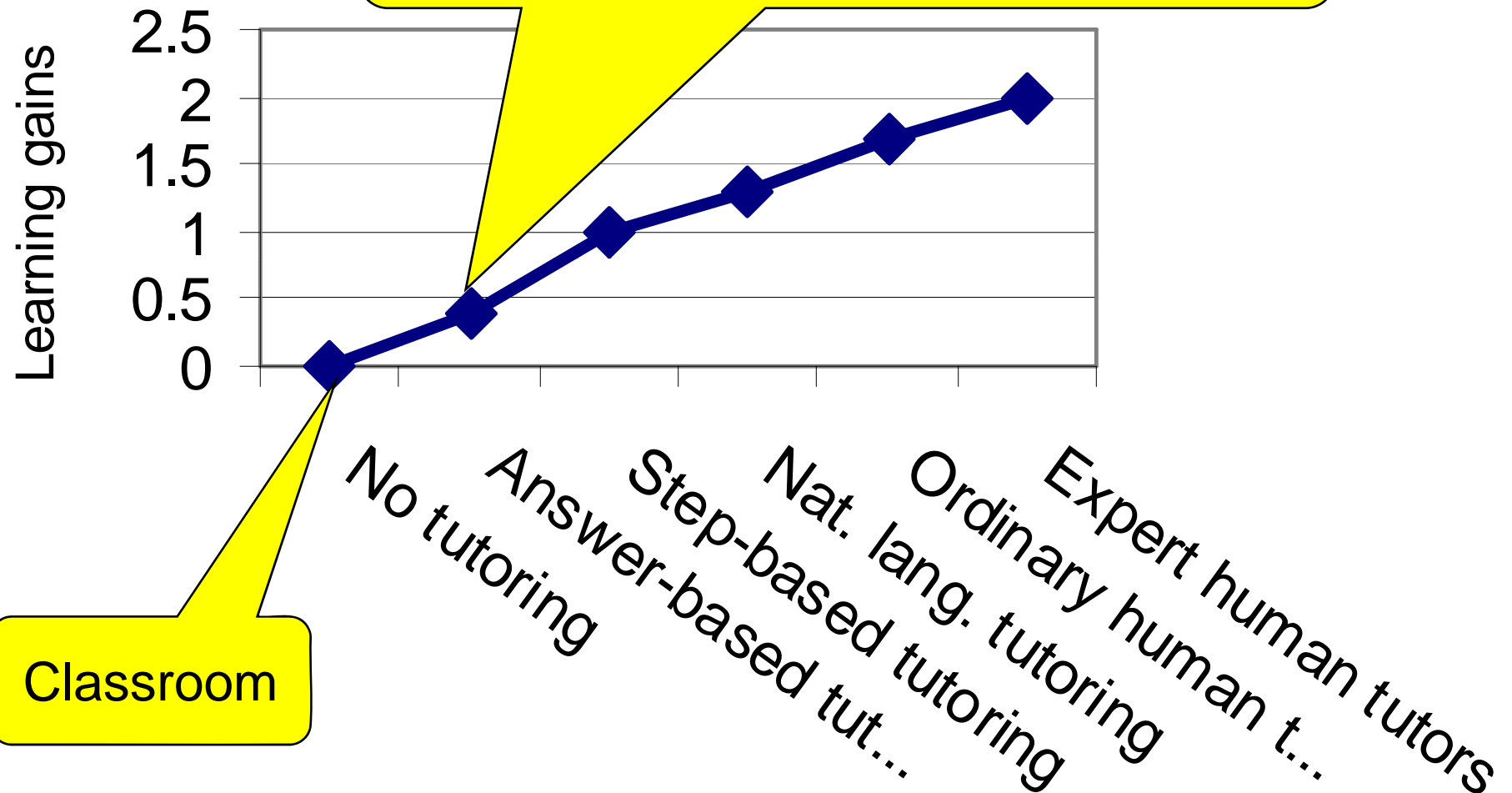
# Hypothesize

Bloom's (1984) 2-sigma: 4 weeks of human tutoring vs. classroom



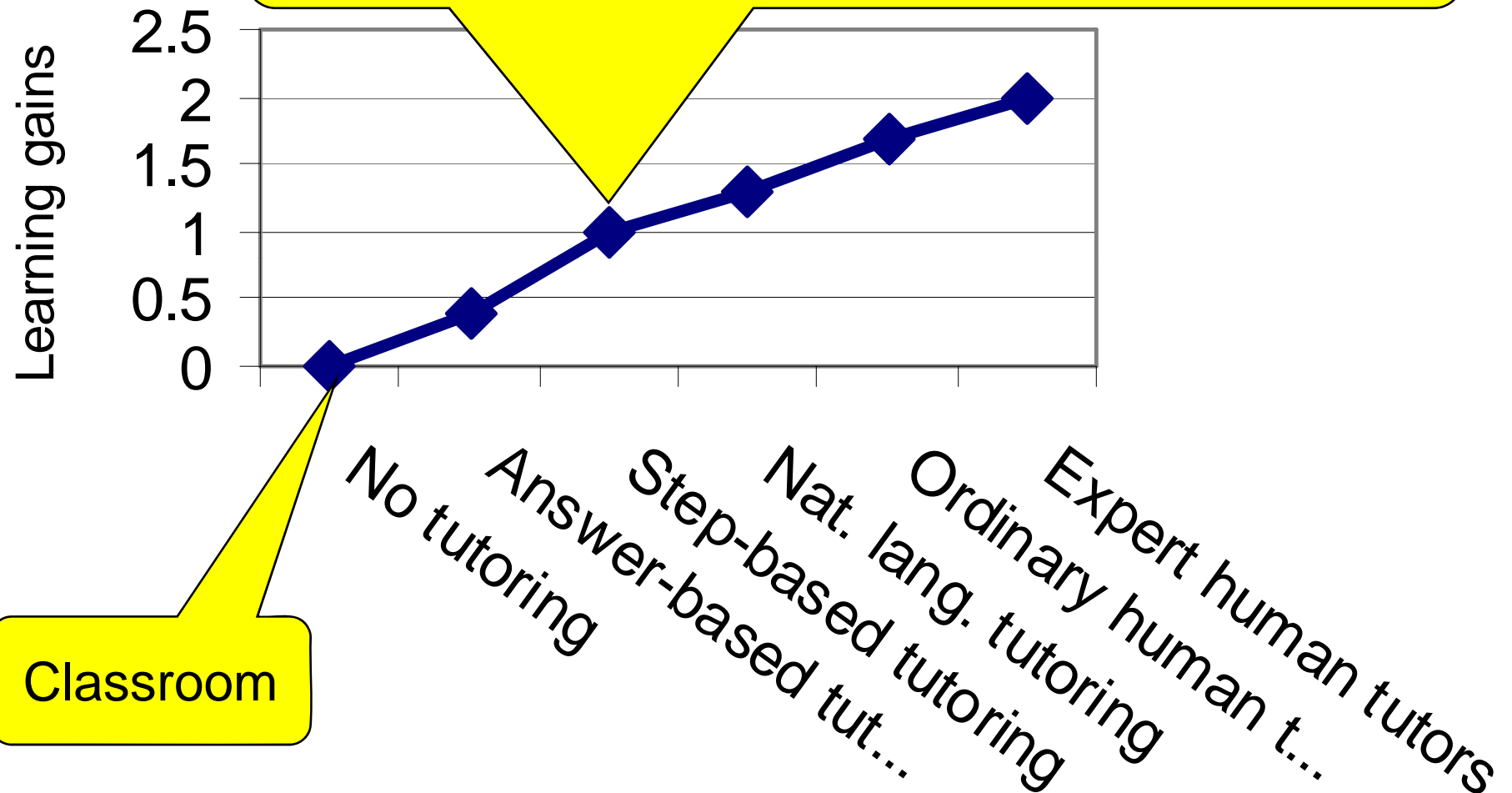
# Hypothesis

Kulik (1984) meta-analysis of CAI vs. classroom  $\rightarrow$  0.4 sigma



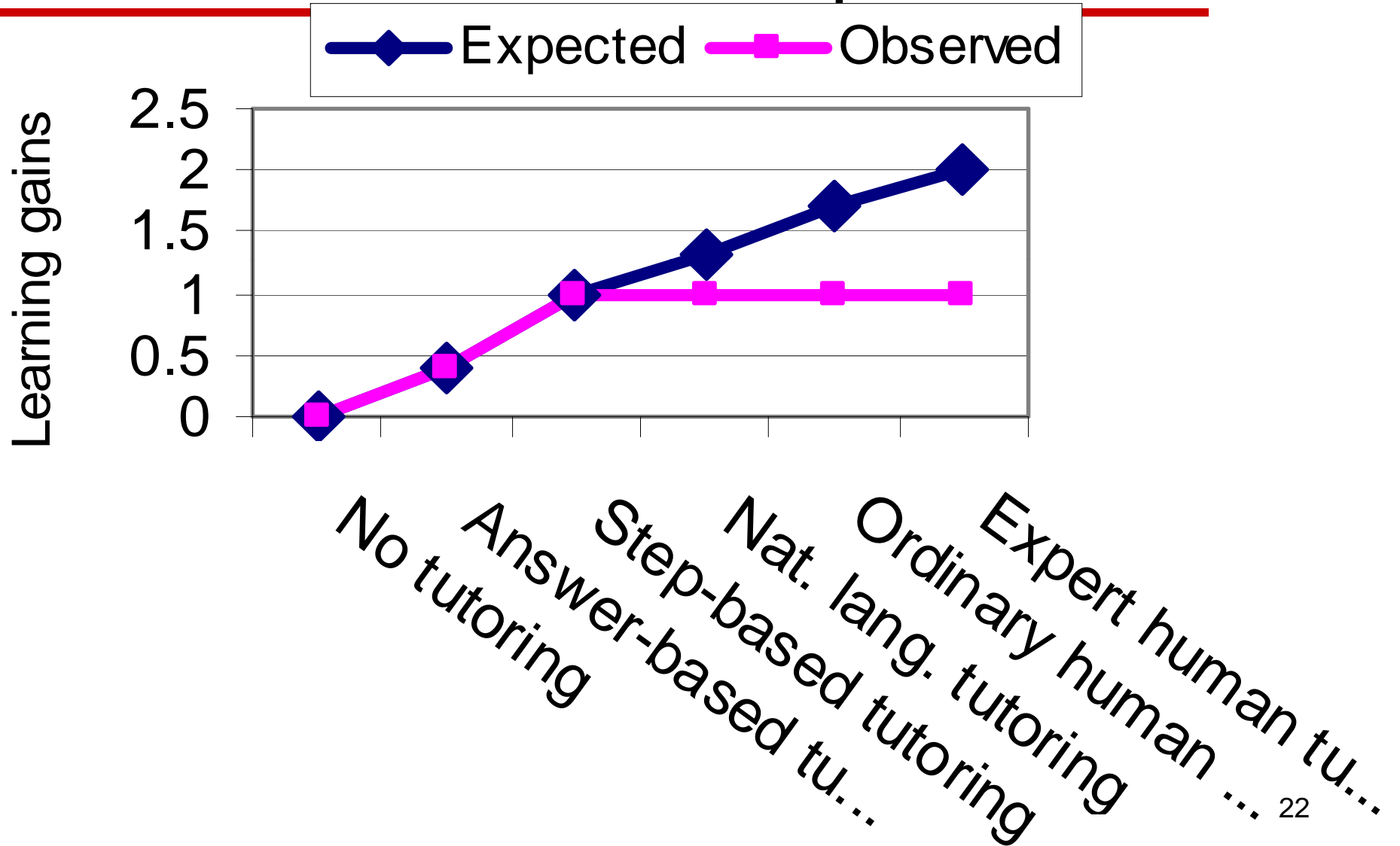
# Hypothesis

Many intelligent tutoring systems: e.g., Andes (VanLehn et al, 2005), Carnegie Learning's tutors...



My main claim:

There is an interaction plateau



# Outline

---

- ◆ Terminology
- ◆ Hypothesis
- ◆ Evidence
- ◆ Implications for Lifelong Learning Companions



Next

# Experiments had 6 conditions

(VanLehn, Graesser et al., 2007)

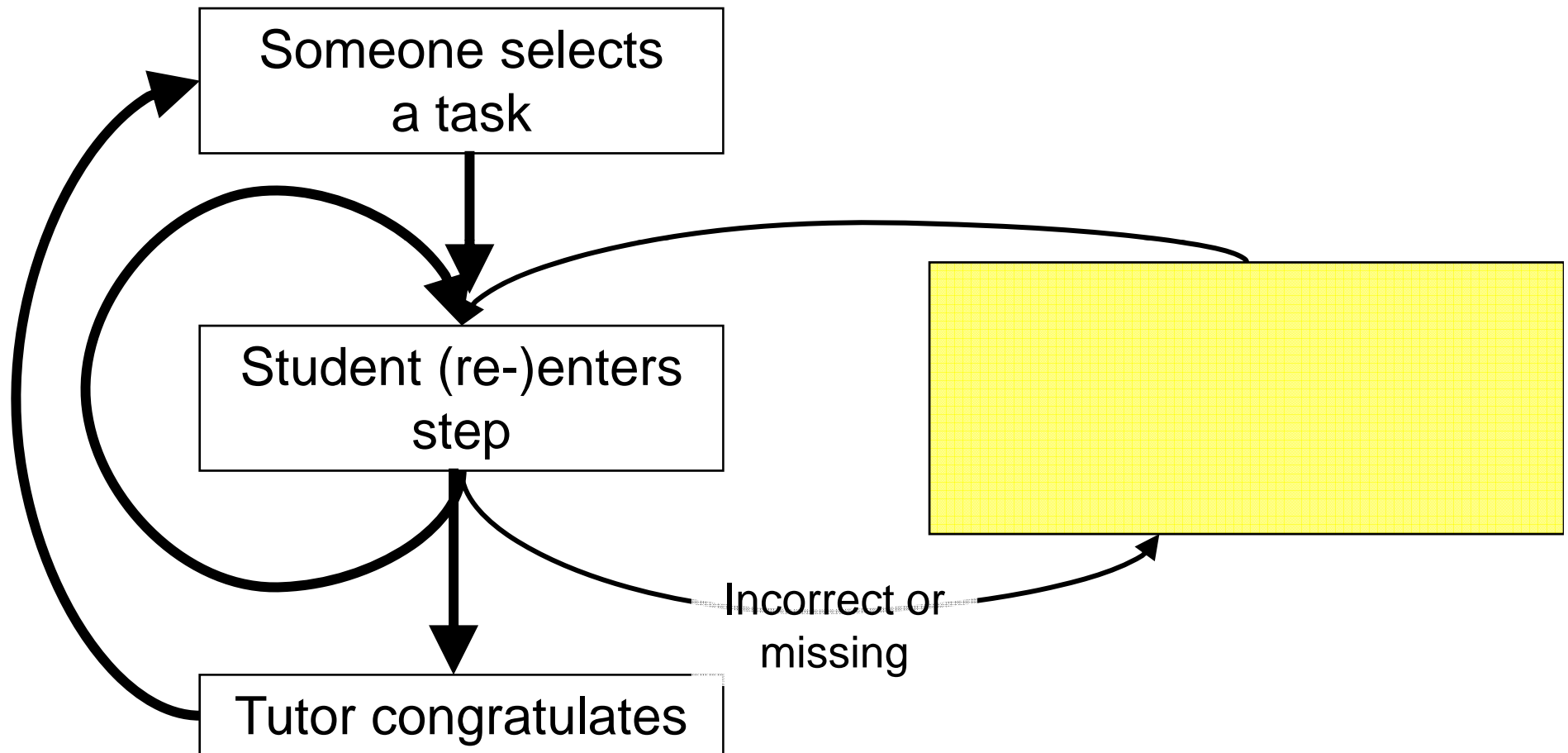
---

- ◆ Expert Human tutors
  - Typed
  - Spoken
- ◆ Natural language tutoring systems
  - Why2-AutoTutor (Graesser et al.)
  - Why2-Atlas (Jordan, Rosé, VanLehn et al.)
- ◆ Step-based tutoring system
  - Canned text remediation
- ◆ No tutoring
  - Textbook



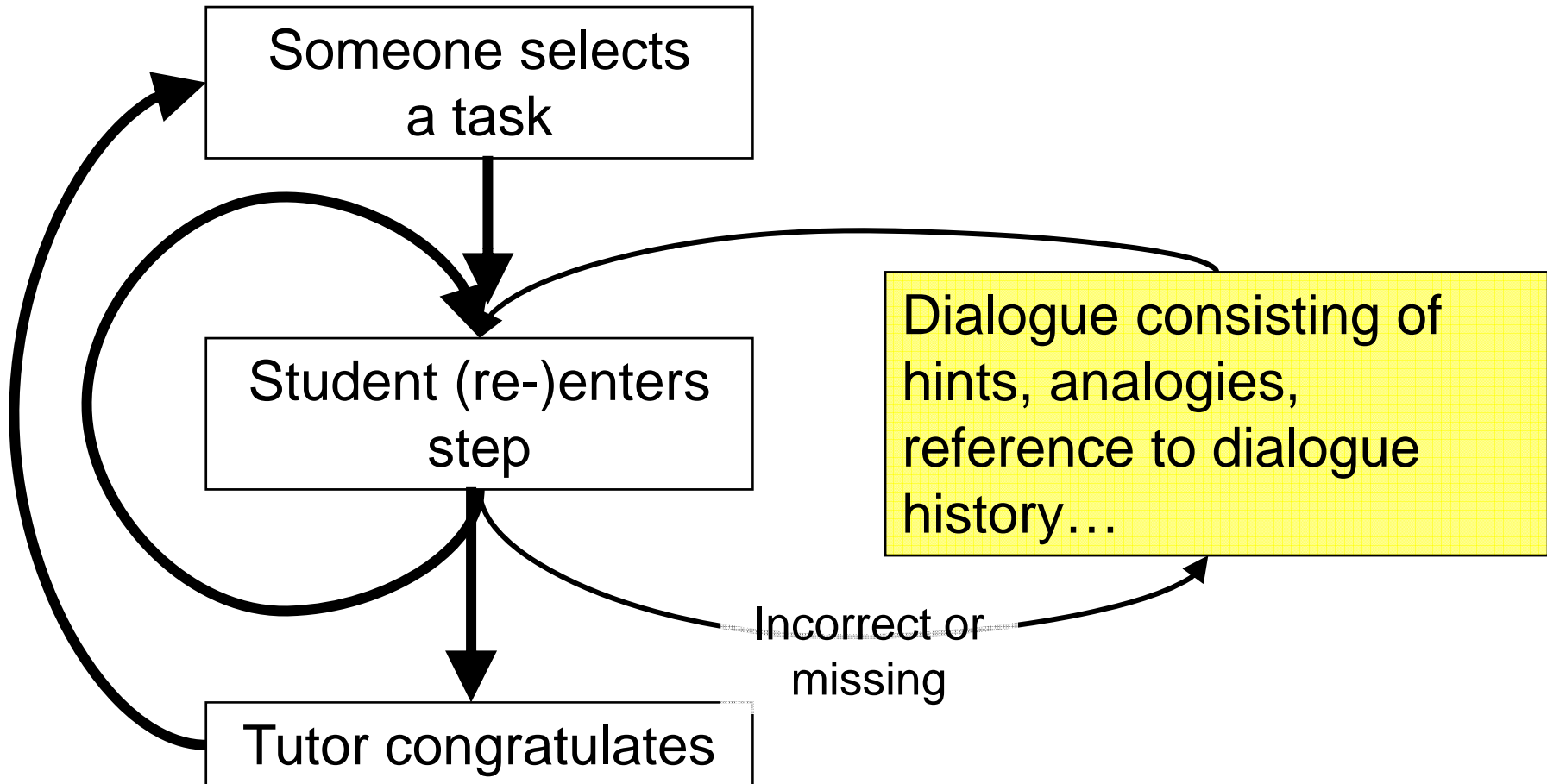
# Only difference between conditions was contents of yellow box

---



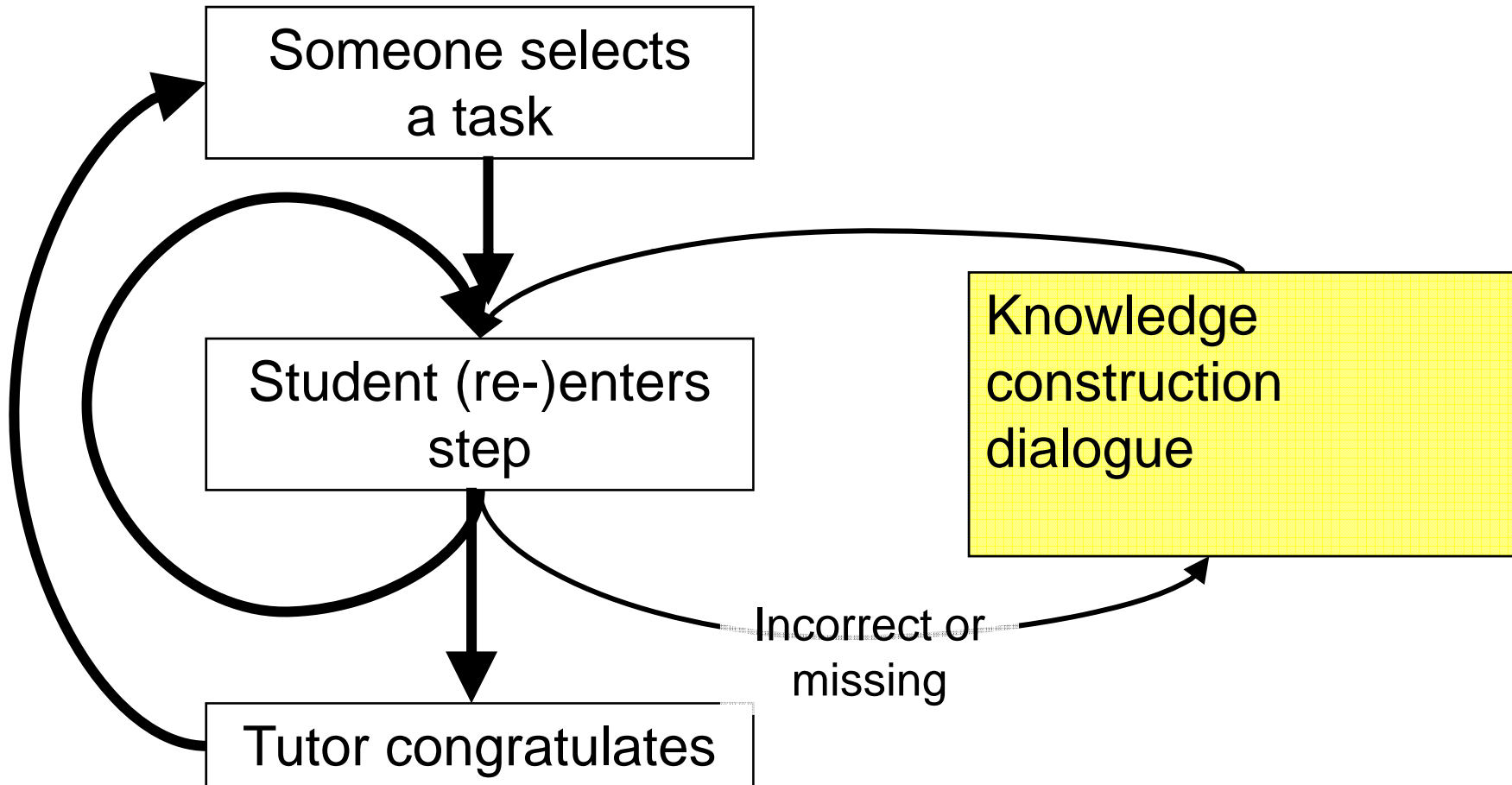
# Human tutors

---



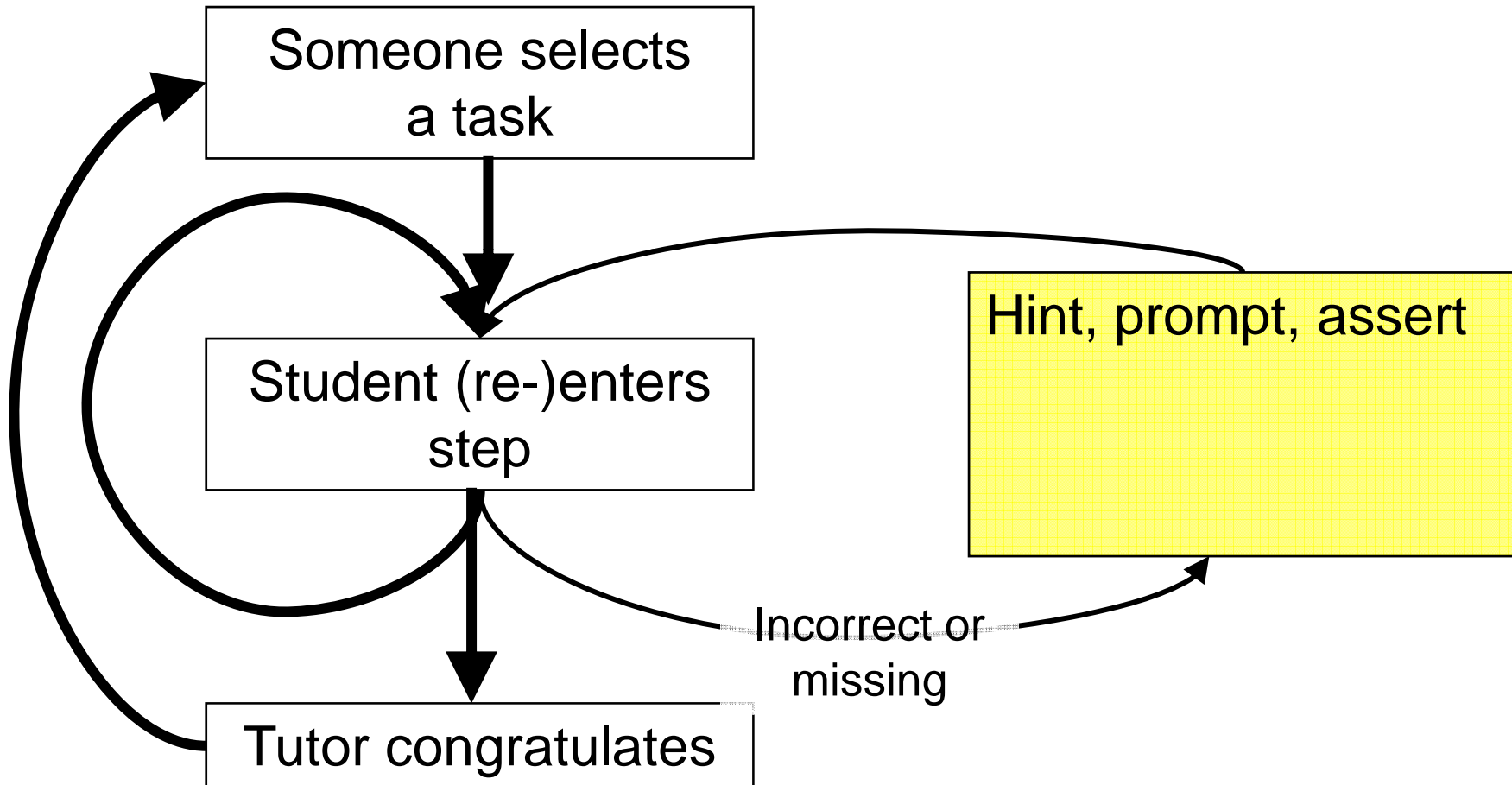
# Why2-Atlas

---



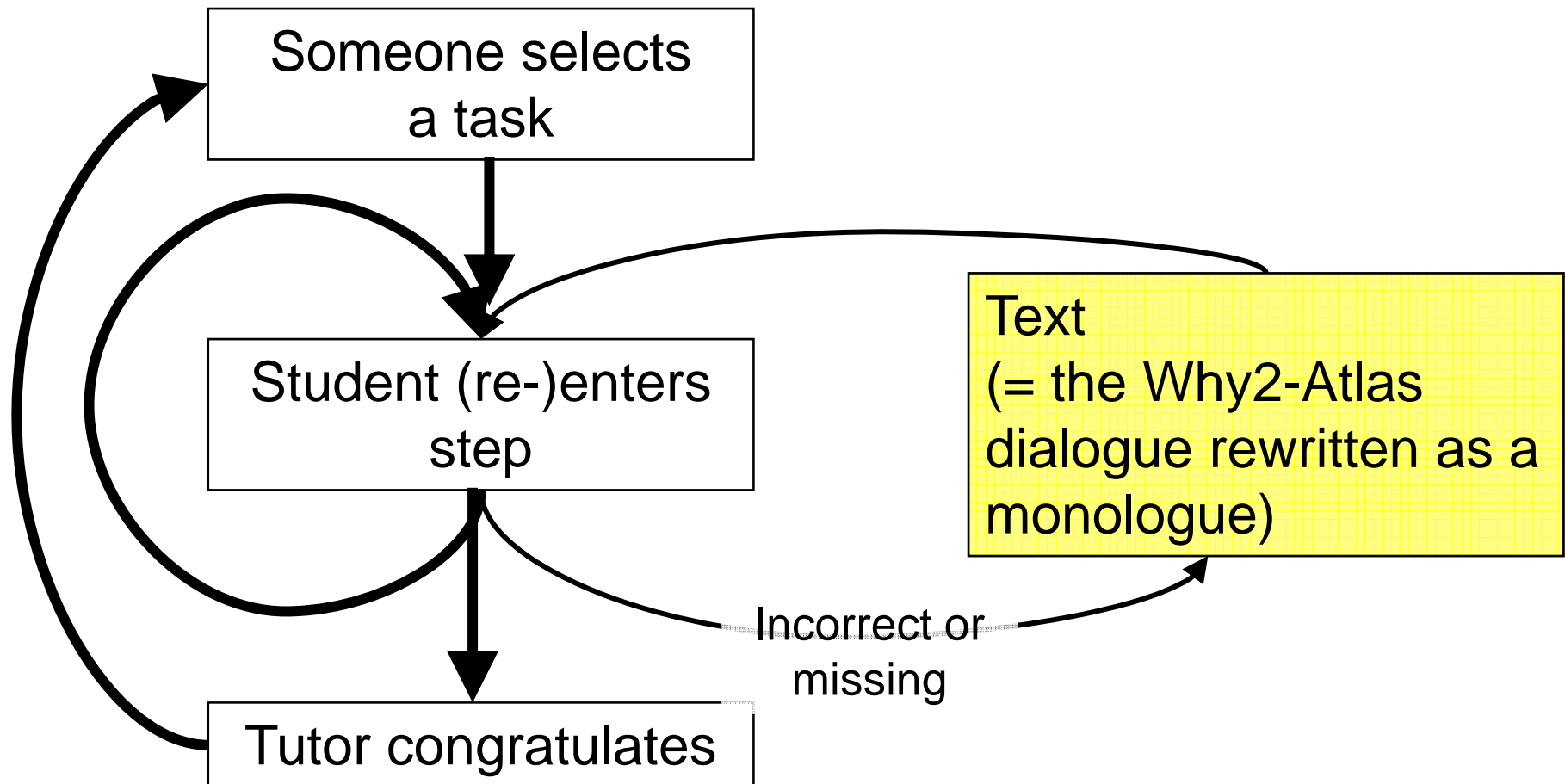
# Why2-AutoTutor

---



# Canned text

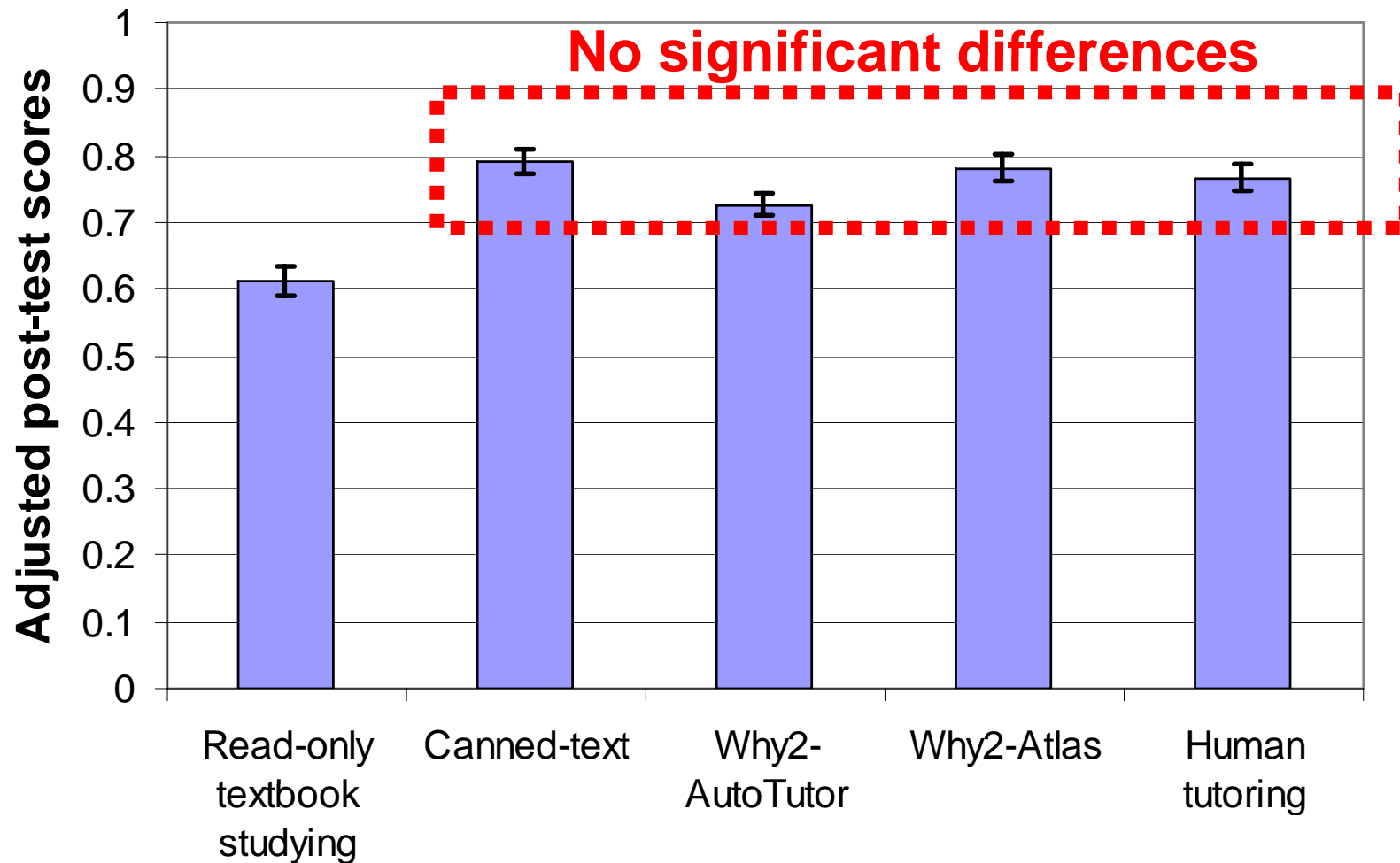
---



# Experiments 1 & 2

(VanLehn, Graesser et al., 2007)

---



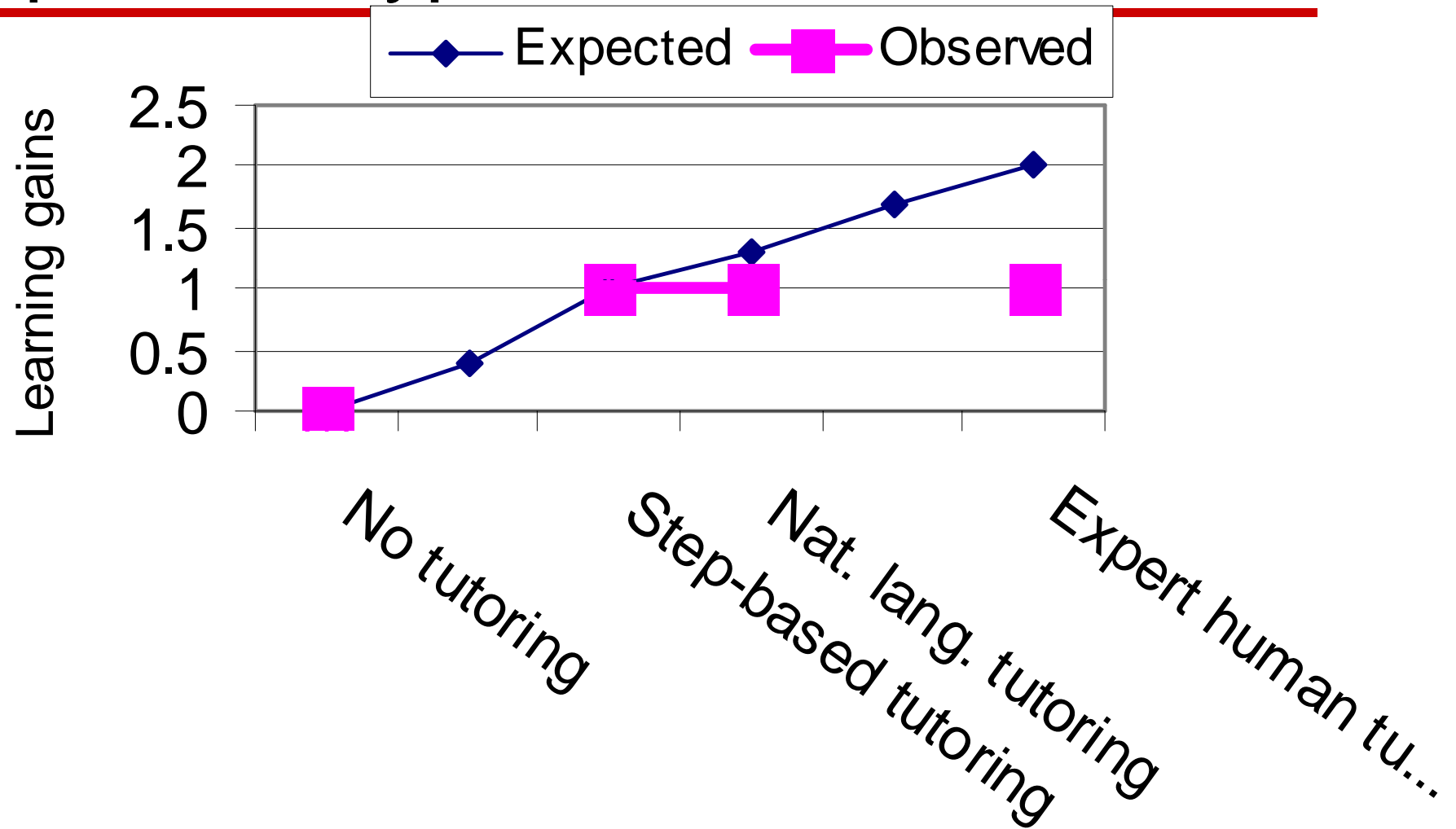
# Results from all 7 experiments

(VanLehn, Graesser et al., 2007)

---

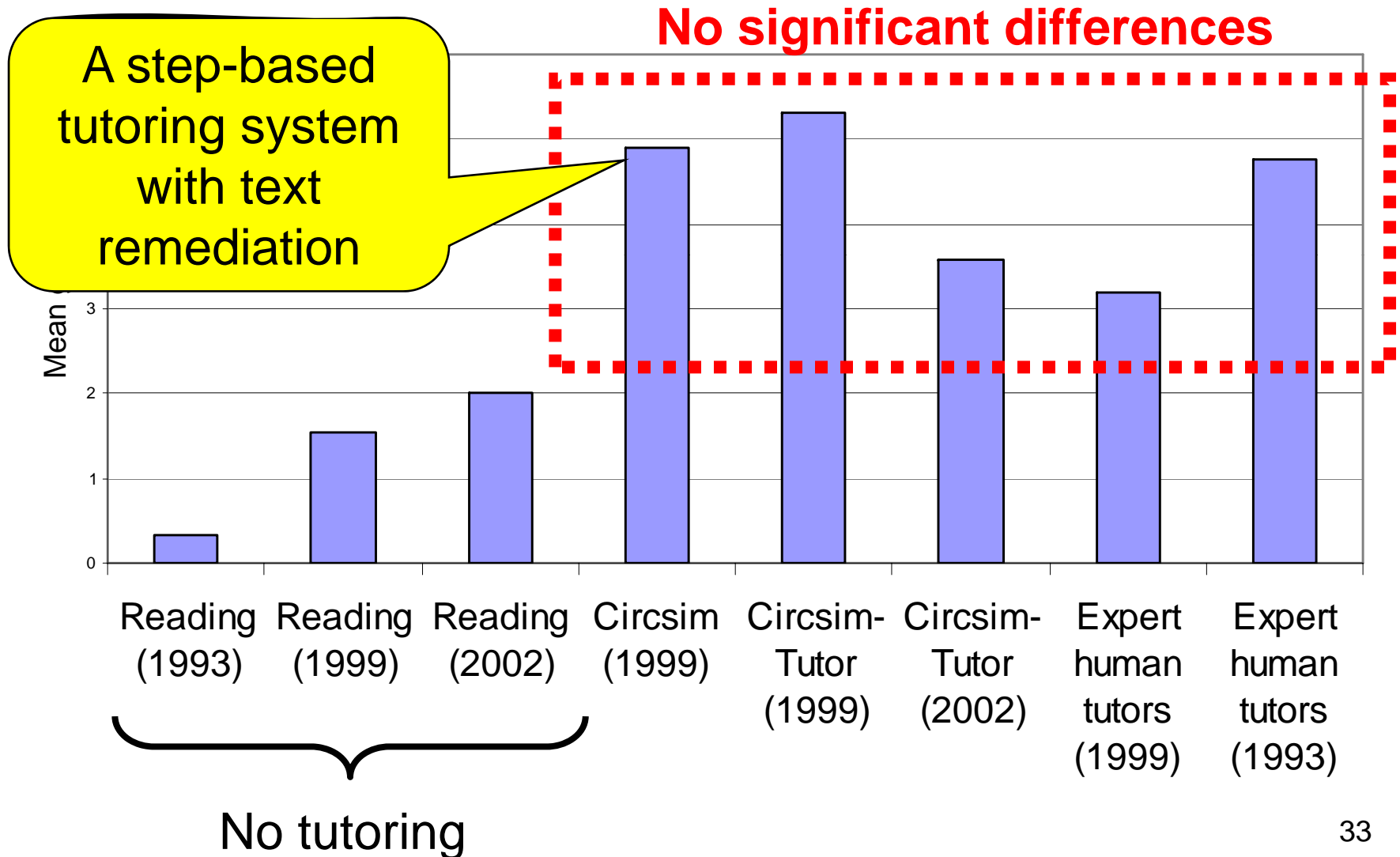
- ◆ Why2: Atlas = AutoTutor
- ◆ Why2 > Textbook
  - No essays
  - Content differences
- ◆ Human tutoring = Why2 = Canned text remediation
  - Exception: When pre-physics students worked with instruction authored for post-physics students, then Human tutoring > Canned text remediation

# Why2 results support interaction plateau hypothesis



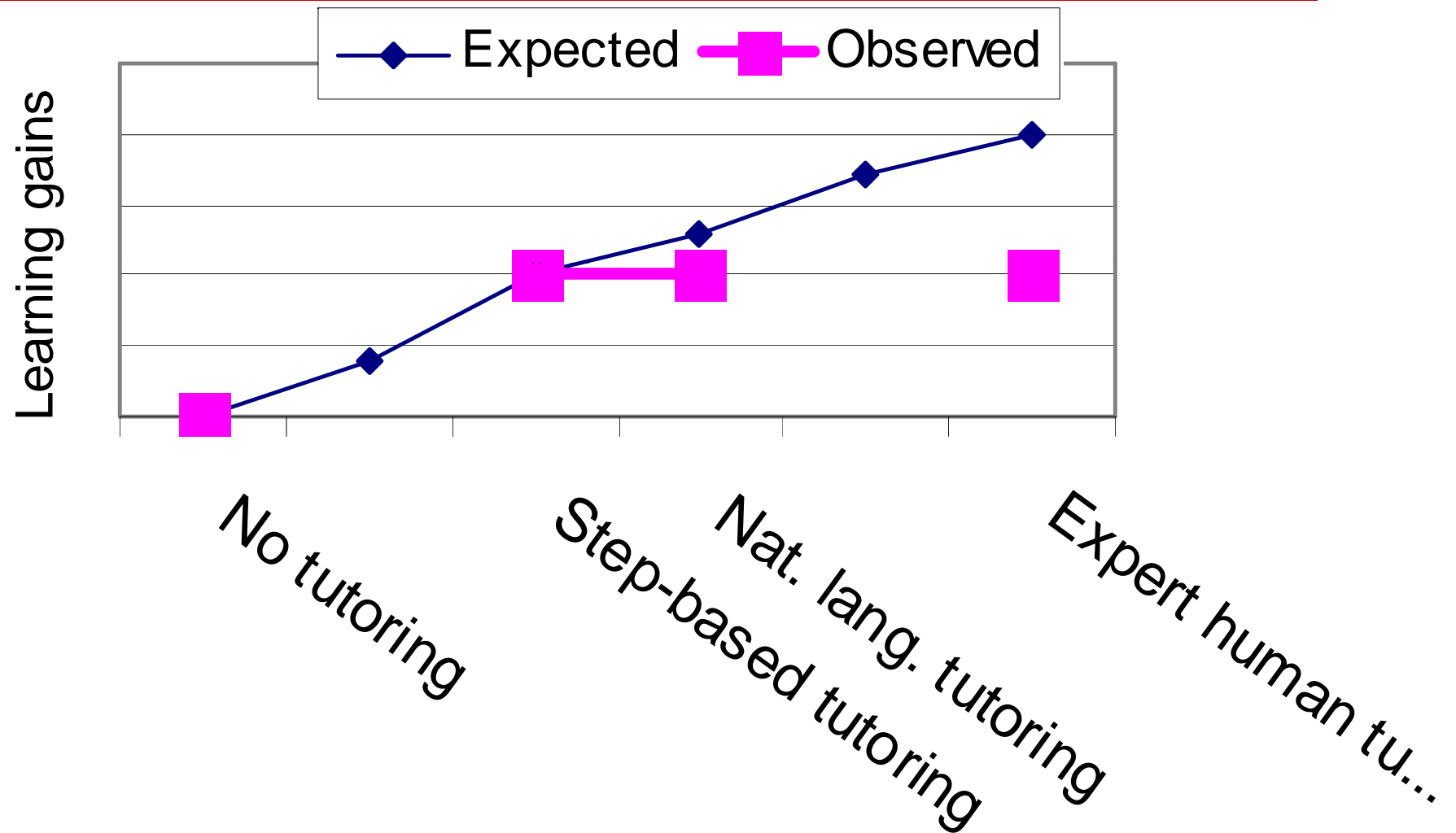


# Other evidence for the interaction plateau (Evens & Michael, 2006)



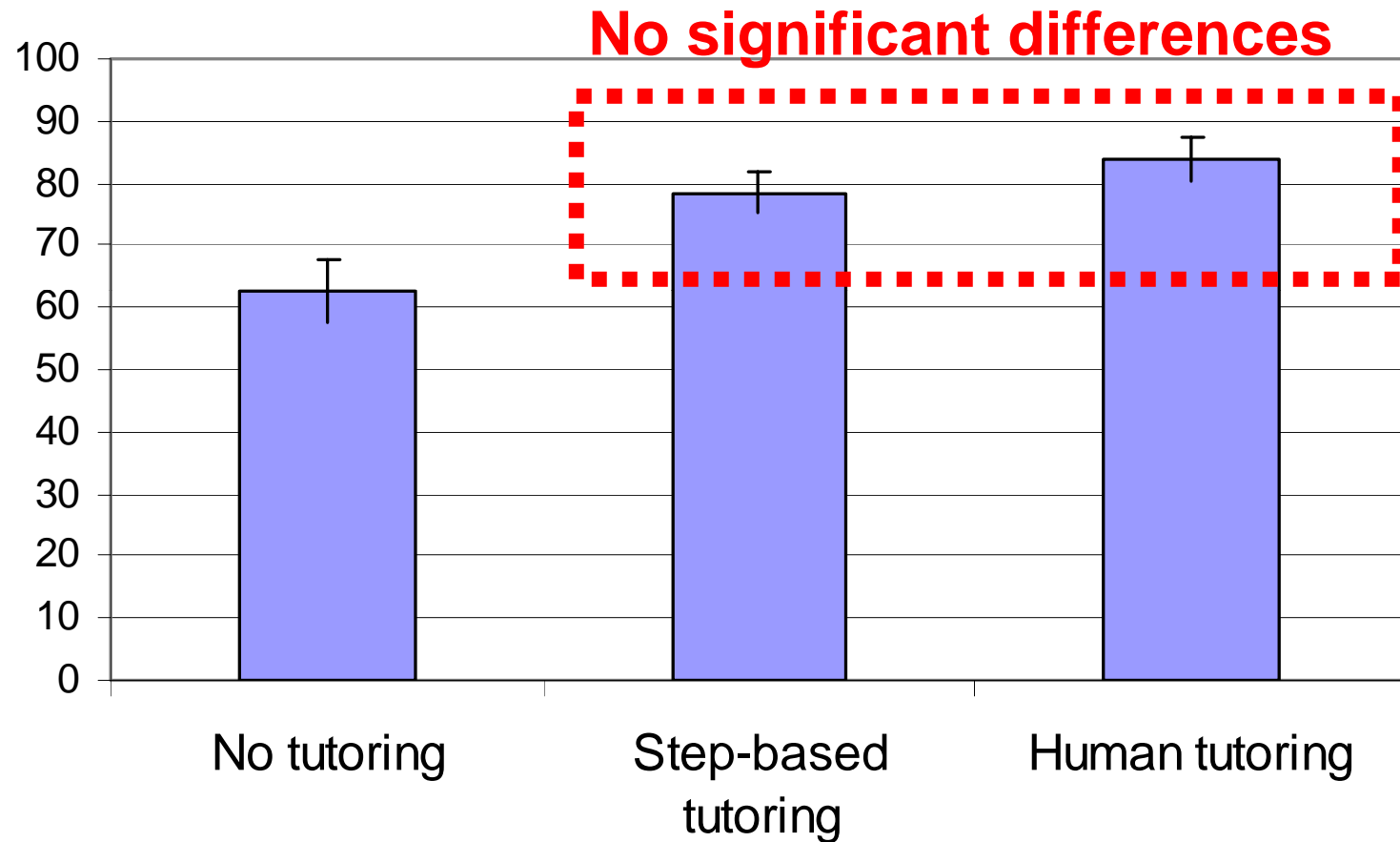
# Circsim results support interaction plateau hypothesis

---

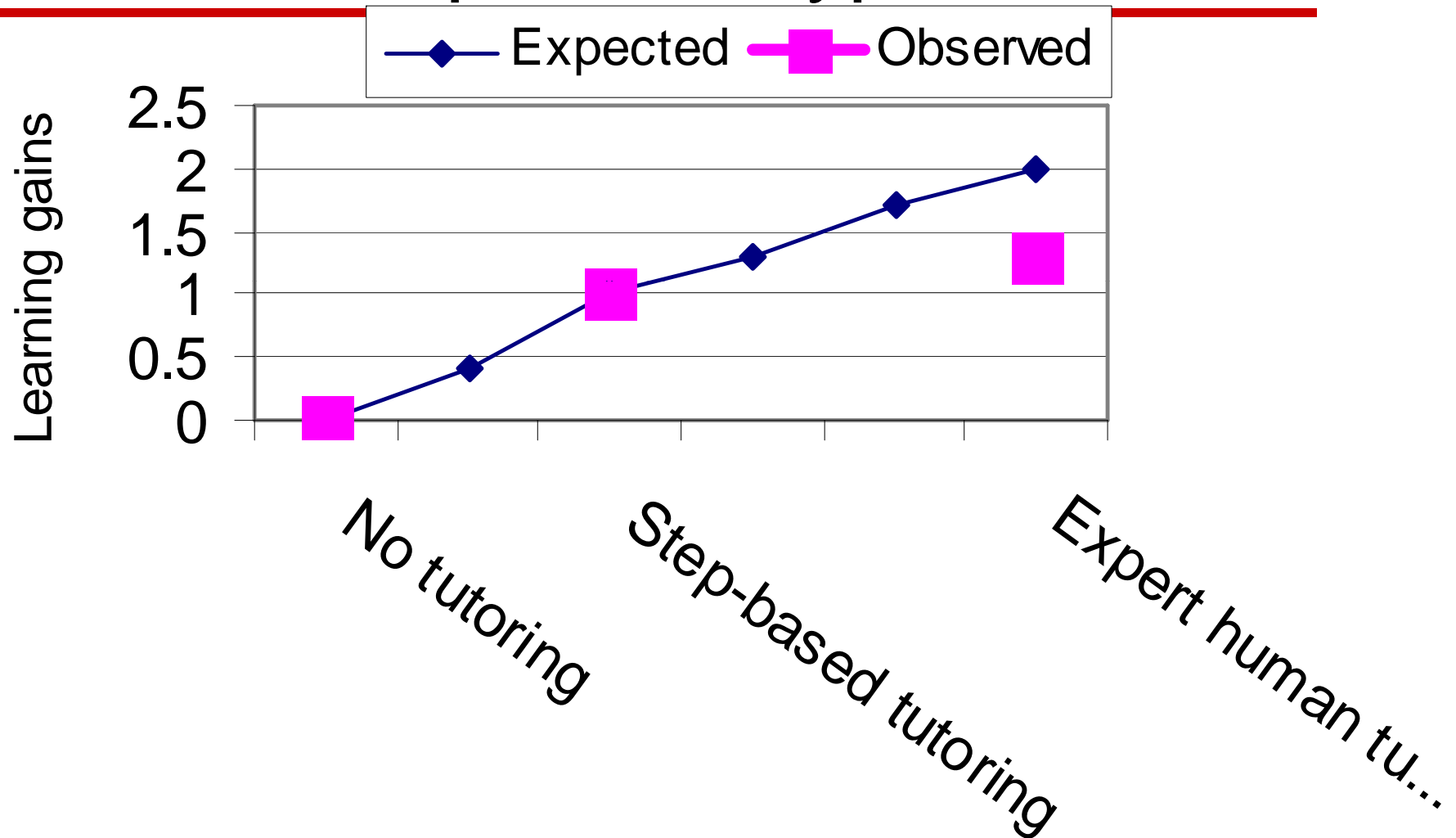


# Other evidence for the interaction plateau (Reif & Scott, 1999)

---



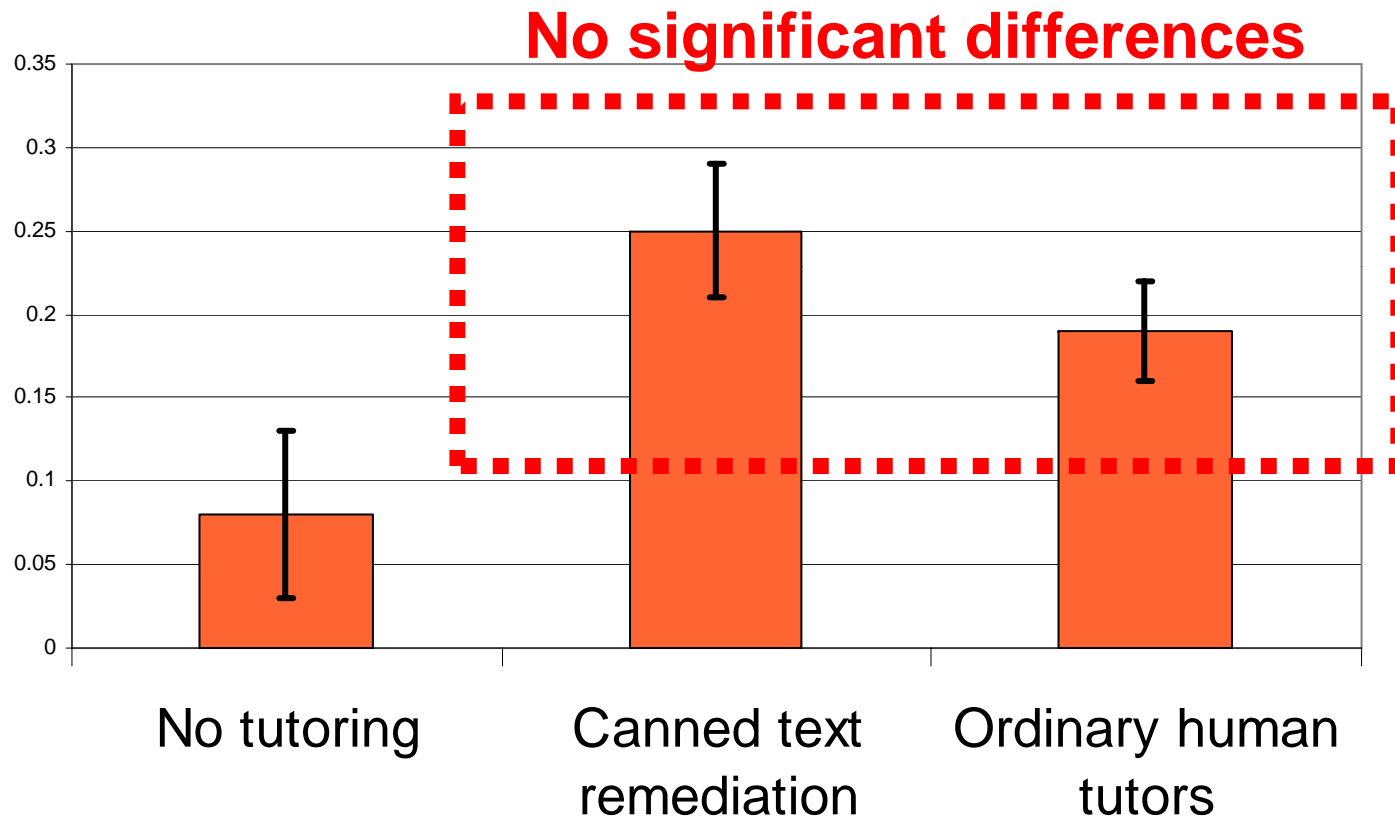
# Reif & Scott results support interaction plateau hypothesis



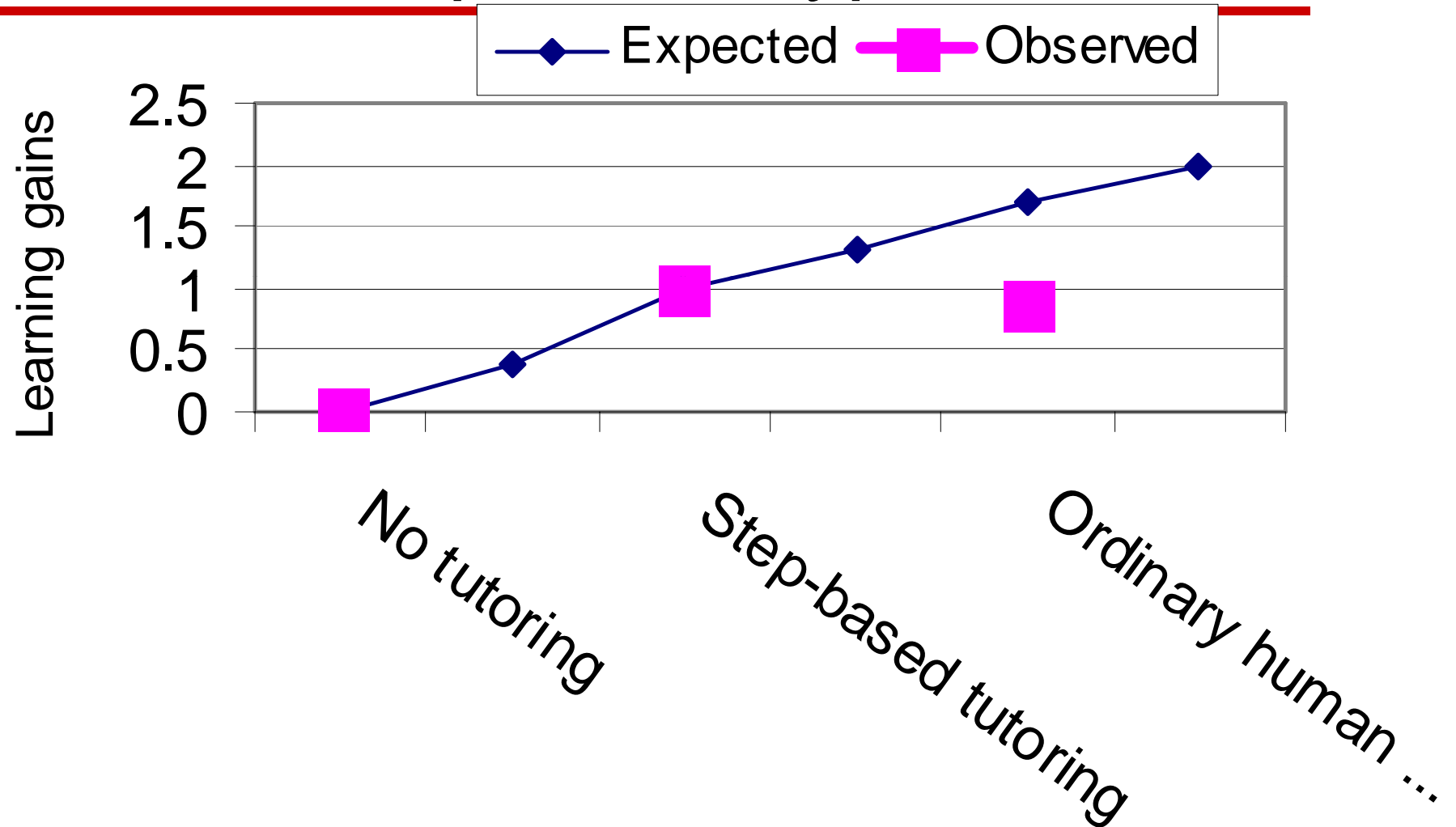
# Katz, Connelly & Allbritton (2003)

## post-practice reflection tutoring

---



# Katz et al. results support the interaction plateau hypothesis

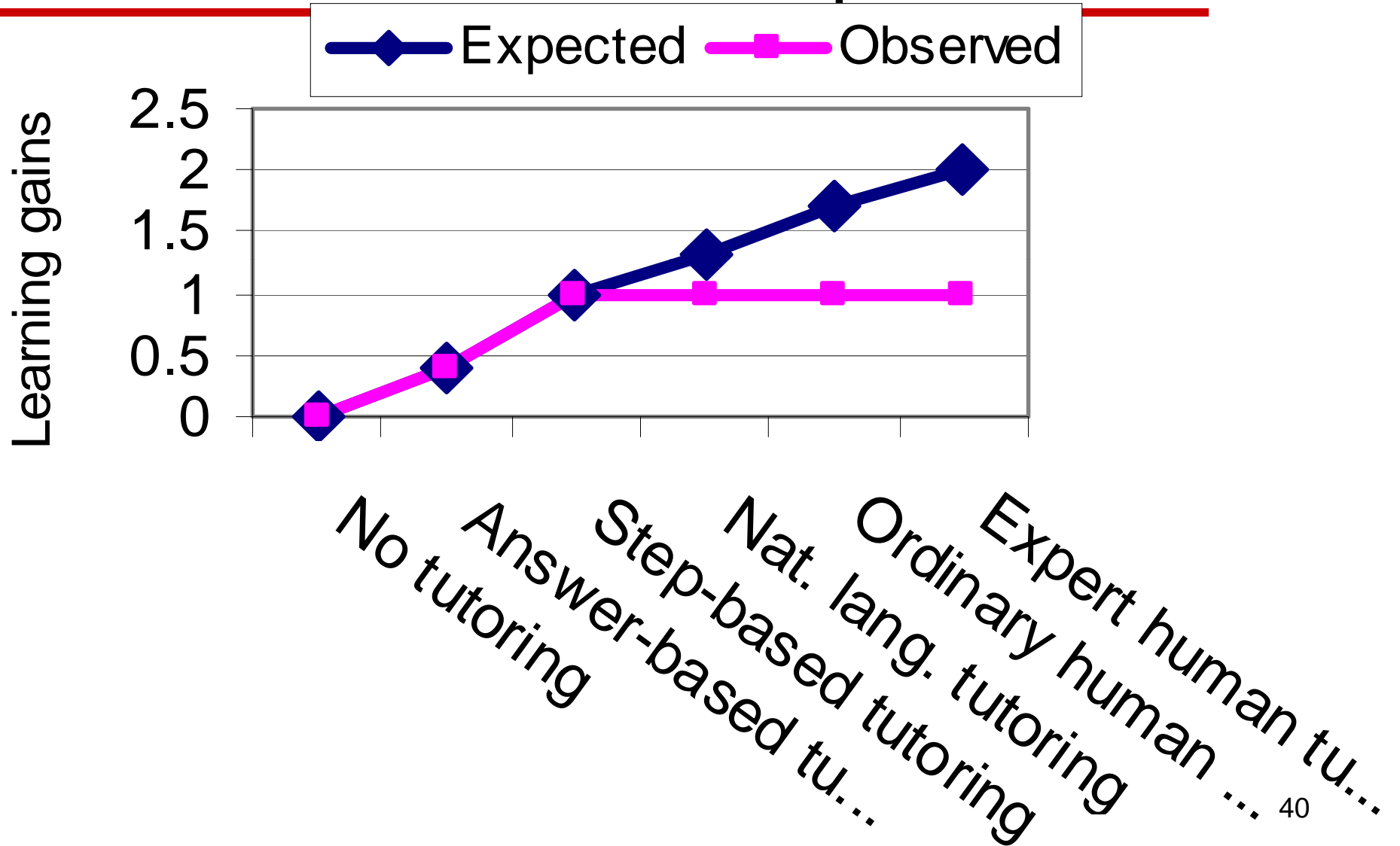


# Evidence: 43 comparisons (VanLehn, submitted)

Studies	No tutoring			Answer-based tutoring	Step-based tutoring			Human tutoring	
	Solve w/o feedback	Reading / studying	Copy example		Text remediation	Hint remediation	Dialog remediation	Ordinary / Low inter.	Expert / High inter.
1 Evens & Michaels, 2006									
2 Evens & Michaels, 2006									
3 Evens & Michaels, 2006		+							
4 VanLehn, Graesser et al., 2007									
5 VanLehn, Graesser et al., 2007									
6 VanLehn, Graesser et al., 2007		1.00							
7 Reif & Scott 1999									
8 Reif & Scott 1999		1.01							
9 Rose, Alevan et al., 2005						0.62			
10 Fossati et al., 2008									
11 Katz, Allbritton & Connelly, 2003									
12 Chae, Kim & Glass, 2005									
13 Johnson & Johnson, 1992									
14 Rose et al., 2001									
15 Litman et al., 2006									
16 Bloom, 1984		1.77, 2.06, 1.95, 1.58, 2.65, 2.11							
17 Merrill et al., 1995		2.65							
18 Azevedo et al., 2006		0.68							
19 Witherspoon et al., 2007		+							
20 Wood, Wood & Middleton, 1978						2.07			
21 Swanson, 1992						0.65			
22 Chi, Roy & Hausmann, 2008						0.66			
23 Chi, Roy & Hausmann, 2008									
24 Weerasinghe & Mitrovic, 2006									
25 Siler, Rose et al., 2002									
26 Person et al.		~0.50							
27 Graesser et al 2003		0.28, NS, 0.44							
28 Amott, Hastings & Allbritton, in press		0.46							
29 Craig et al, 2004; 2006		0.86, 0.71, NS, NS							
30 Craig et al., 2006 with questions		-0.69, -0.68							
31 Lane & VanLehn, 2005		0.96							
32 Heffernan et al., 2008						0.56			
33 Mendicino et al., in press									
34 Razzaq et al., 2008		0.54							
35 Moreno, Mayer et al., 2001		1.01, 0.98							
36 VanLehn et al., 2005		0.61, 0.25							
37 Corbett & Anderson, 2001		0.98							
38 Anderson et al, 1995		~1.00							
39 Koedinger et al., 1997		1.20, 0.30							
40 Roberts, Pioch & Ferguson, 2000									
41 Mitrovic & Ohlsson, 1999		0.75							
42 Suraweera & Mitrovic, 2004		0.63							
43 Charney, Reder & Kusbit, 1990; 1986			0.74						

Human tutoring  
=  
Step-based tutoring  
>  
Answer-based tutoring  
>  
No-tutoring

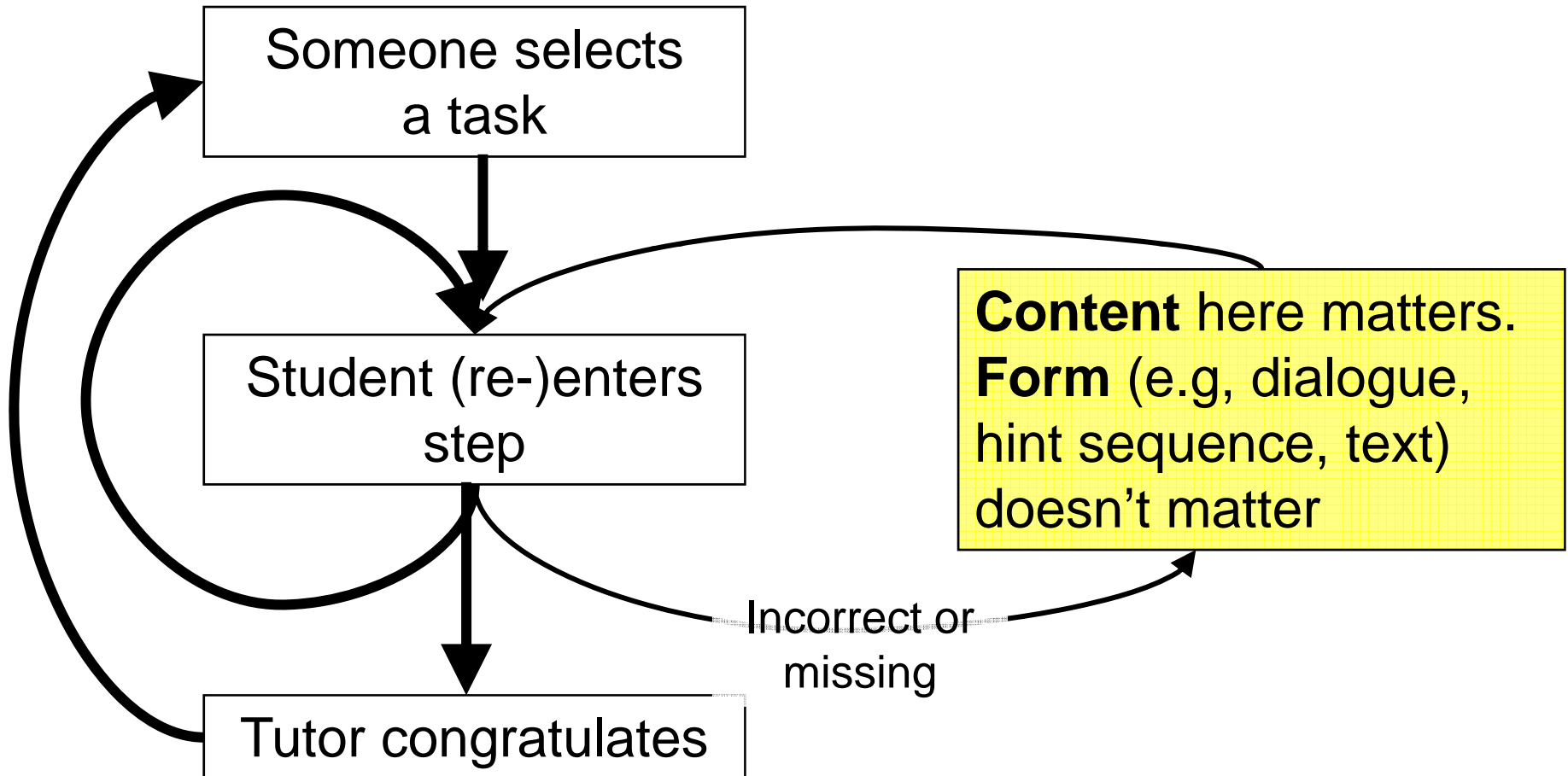
# My main claim (again): There is an interaction plateau





# Conclusion: How tutors present the remediation has no effect on learning gains

---



# Why is there an interaction plateau?

---

- ◆ Instructors design steps so that an ideal student can just bridge each one. Thus, simple remediation suffices if the student can't enact a step.
- ◆ When a student has failed to do a step, the student is motivated to learn from the remedial text, hint sequence or dialogue. Thus, more interactivity yields no added value.

# Step-based ITS should be **more effective** than expert human tutors

---

## ◆ Inner step loop

- Although ITS's remediations are simpler (e.g., mere hint sequences), they are just as effective
- ITS makes fewer mistakes interpreting steps

## ◆ Outer task loop

- ITS can do more accurate, deeper assessments
- ITS can accurately index a larger library of tasks

# Outline

---

- ◆ Terminology
- ◆ Hypothesis
- ◆ Evidence
- ◆ Implications for Lifelong Learning Companions



Next

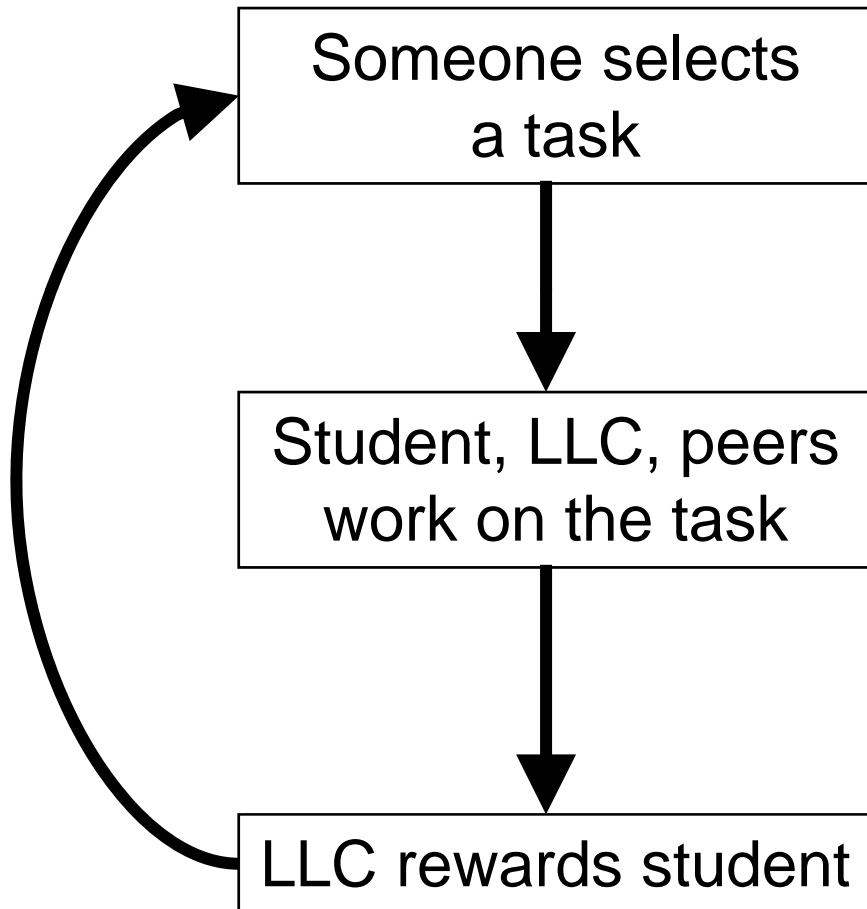
# If we duck the NL problem, then...

---

- ◆ E.g., avoid tasks that require NL

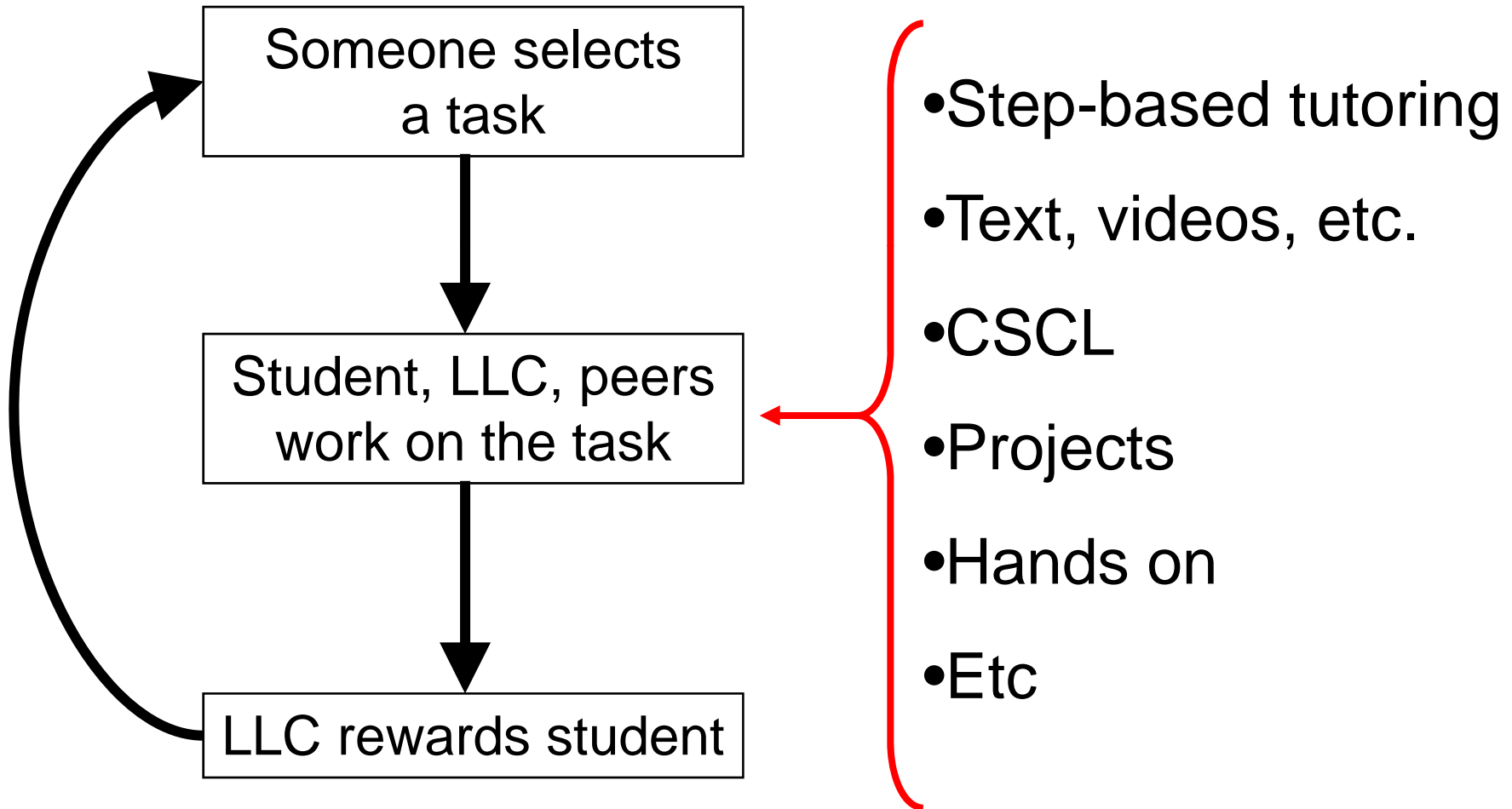
# A lifelong learning companion (LLC) could be just a fancy LMS!

---



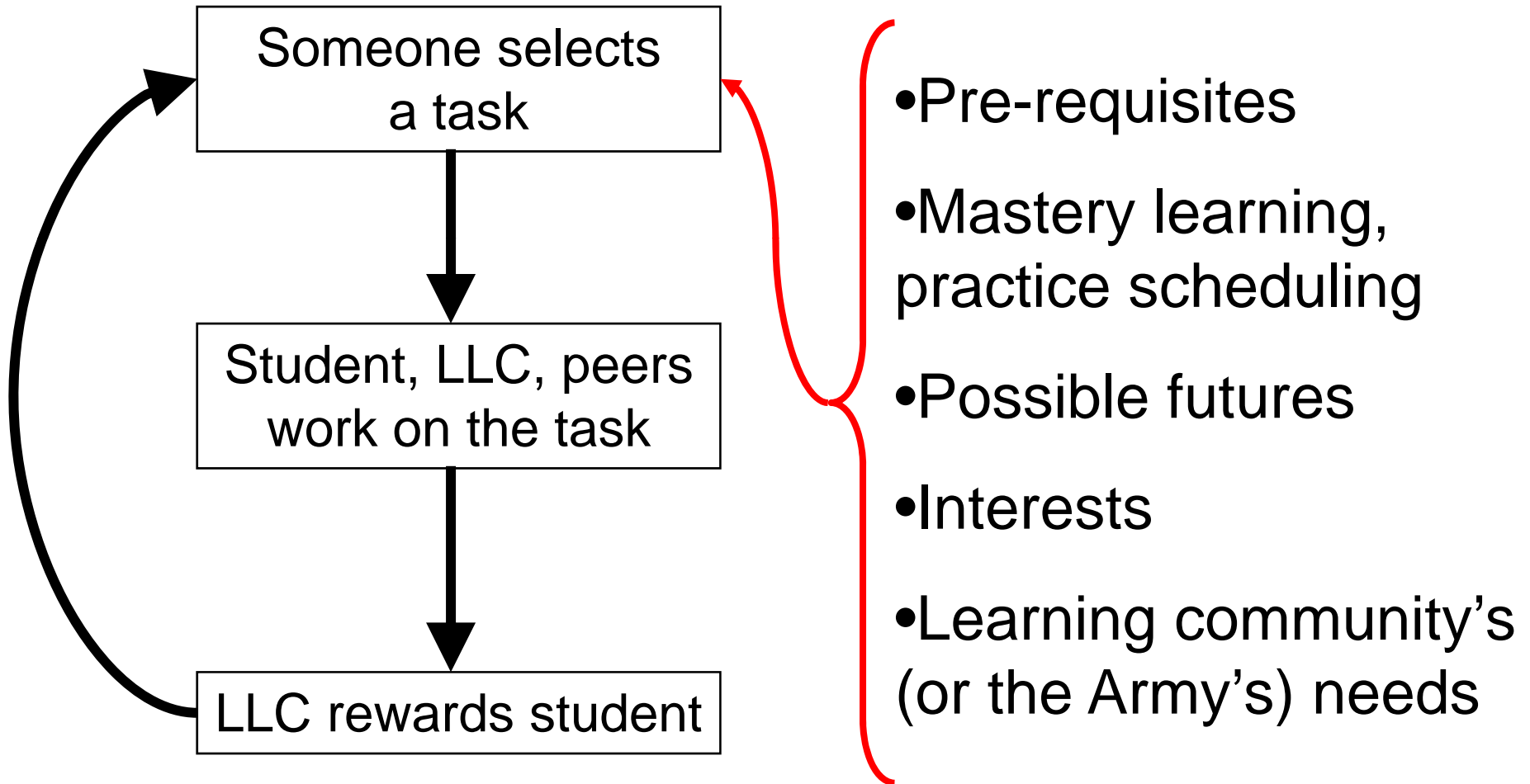
# LLC supports many task types (feasible, sort of)

---



# Task selection gives learner more control (feasible)

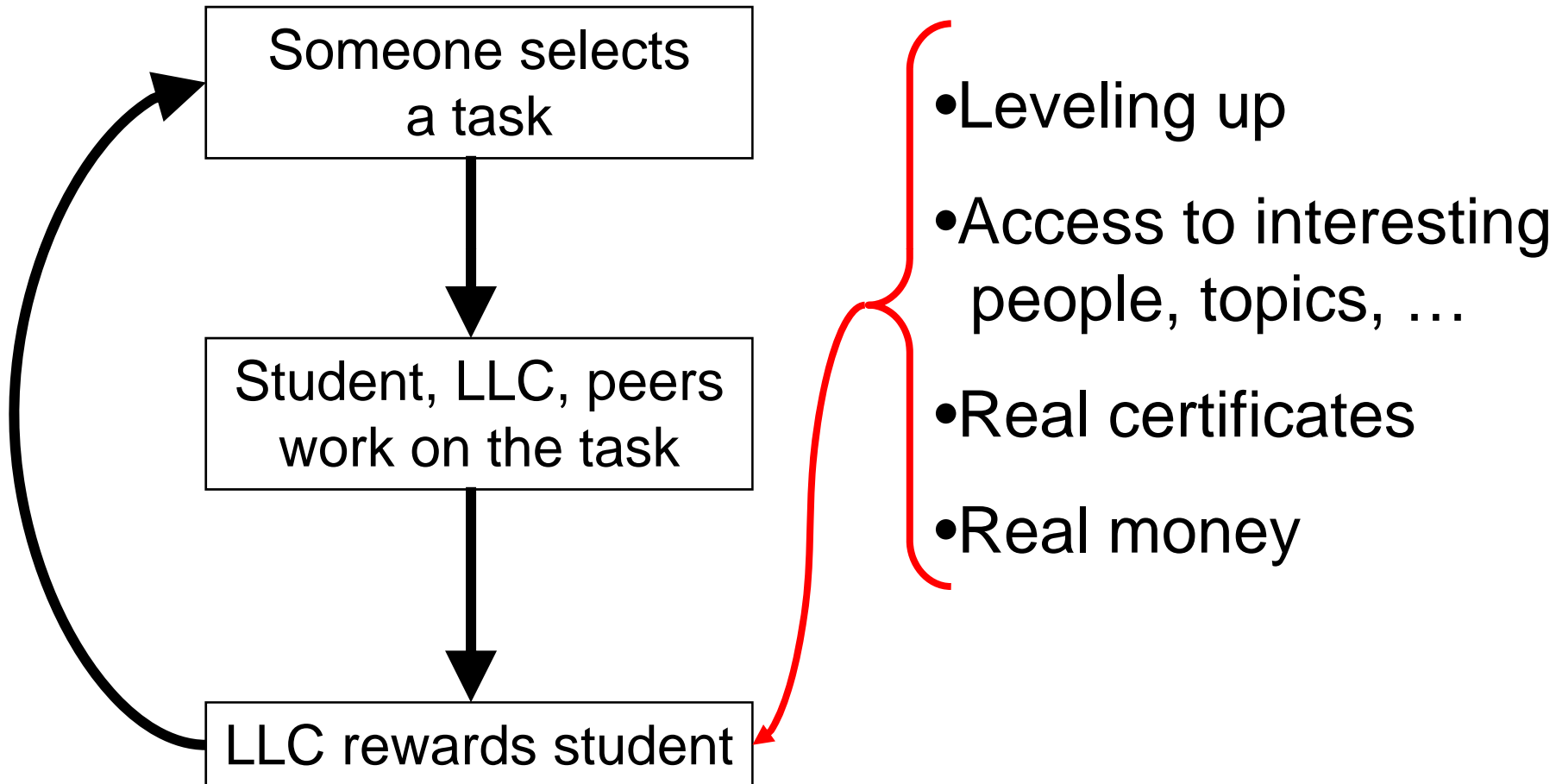
---





# LLC gives larger rewards (feasible?)

---



# BUT: LLC must understand tasks well enough to select, support, reward

---

## ◆ LMS solution

- Meta data for each task

## ◆ ITS solution

- Solution graph (like Collagen's HTN) for each task
- Perhaps generated by a problem solver/planner

## ◆ How to understand other types of tasks?

- CSCL tasks
- Projects
- Explorations...

# Biggest problem is knowledge “engineering” bottleneck

---

- ◆ Lifelong = 80 years = 30K days @ 4 hrs/day = 120K hours = 200K tasks
- ◆ Andes has ~500 tasks, covers 2 semesters
- ◆ Open ended & growing library
- ◆ Must automate the task analysis
  - LSA assigns metadata?
  - Social computing?

# Summary

---

- ◆ 200K analyzed tasks (hard)
  - ◆ Novel task selection, reward (feasible)
  - + ◆ NL for tasks that need them (optional)
- 
- ◆ A lifelong learning companion

---

◆ Back up slides are next

# When should natural language be used in step-based tutoring systems?

---

- ◆ Only when there is no alternative
  - Spoken commands of COVE (Roberts, 2001)
  - Spoken commands of SCoT (Pon-Barry et al., 2006)
  - Tactical Iraqi (Johnson et al.)
  - Predator & Aegis team training
  - Essays of Why2 ?
  - Many others...

# Why did Bloom (1984) observe 2.0 effect size?

---

## ◆ Three conditions

- Human tutors with mastery learning with a 90% threshold → 2.0 sigma
- Classroom with mastery learning with an 80% threshold → 1.0 sigma
- Classroom

## ◆ My interpretation:

- It's the mastery learning thresholds, not the human tutoring

# Why does the public think expert human tutors are gods?

---

- ◆ Human tutors > step-based tutoring system when the material is so far over the student's head that they rarely understand the text/hints used for remediation (VanLehn et al., 2007)
- ◆ **Step-based tutoring systems are not yet common**